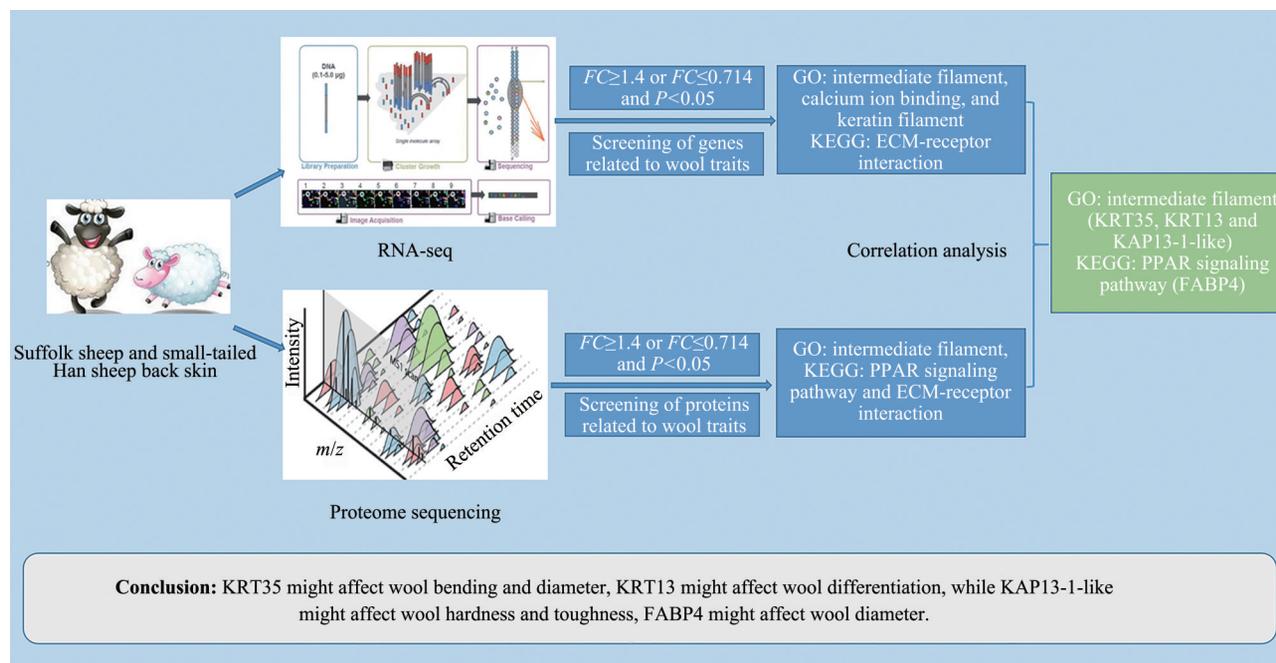




Revealing The Key Genes Affecting Wool Traits on The Basis of Skin Tissue RNA-seq and Proteome Sequencing*

WU Jin-Qiang, HAO Xiao-Jing, ZHAO Hong-Xia, DONG Ya-Jie, WANG Rong,
 ZHANG Peng-Xiang, WANG Hai-Dong, HE Xiao-Yan**
 (College of Animal Medicine, Shanxi Agricultural University, Taigu 030801, China)

Graphical abstract



Abstract Objective Wool is a high-grade raw textile material, and the physical properties of wool are directly related to the quality of the wool. The present study was to search for the genes affecting wool traits and to explore the complex molecular mechanism affecting wool traits. **Methods** This study selected 3 Suffolk sheep and 3 small-tailed Han sheep and took samples of their back skin tissue using RNA sequencing (RNA-seq) and proteome sequencing analysis of genes, proteins, and related signalling pathways that cause differences in wool traits. **Results** RNA-seq showed that after sequencing, 230 406 674 raw data points and 222 049 370 clean data points were obtained, of which the percentage of Q20 bases was over 99.9%, and the percentage of Q30

* This study was supported by grants from The National Natural Science Foundation of China (31772690) and the Natural Science Foundation of Shanxi Province (201701D121106).

** Corresponding author.

Tel: 86-13753418018, E-mail: sxndhxy@163.com

Received: May 31, 2022 Accepted: July 28, 2022

bases was over 98%. With a fold change (FC) ≥ 1.4 or $FC \leq 0.714$ and $P < 0.05$ as the standard, 1 213 differentially expressed genes (DEGs) were screened out, among which there were 644 upregulated genes and 569 downregulated genes in Suffolk sheep in comparison with small-tailed Han sheep. The gene ontology (GO) enrichment found that intermediate filament, calcium ion binding, and keratin filament were significantly enriched, indicating that they might be related to wool traits. The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment found that the signalling pathway affecting wool traits might be the ECM-receptor interaction. Proteome sequencing showed that with $FC \geq 1.4$ or $FC \leq 0.714$ and $P < 0.05$ as the standard, 99 differentially expressed proteins (DEPs) were screened out, among which there were 47 upregulated proteins and 52 downregulated proteins in Suffolk sheep in comparison with small-tailed Han sheep. The GO enrichment found that intermediate filament was significantly enriched, indicating that it might be related to wool traits. The KEGG enrichment found that the signalling pathways affecting wool traits might be the peroxisome proliferator-activated receptor (PPAR) and the ECM-receptor interaction. A combined analysis of RNA-seq and proteome sequencing found that a total of 15 significantly different genes were detected in both RNA-seq and proteome sequencing, of which 13 were positively correlated and 2 were negatively correlated. Intermediate filament was significantly enriched, KRT35, KRT13 and KAP13-1-like genes might be the key candidate genes to affect wool traits. The PPAR signalling pathway was significantly enriched and might be a key candidate pathway to affect wool traits. The FABP4 gene might be a key candidate gene to affect wool traits. **Conclusion** Among them, KRT35 might affect wool bending and diameter, KRT13 might affect wool differentiation, while KAP13-1-like might affect wool hardness and toughness, FABP4 might affect wool diameter. These results will expand our understanding of the complex molecular mechanisms affecting sheep wool traits and provide a basis for subsequent studies.

Key words RNA-sequencing, proteome sequencing, Suffolk sheep, small-tailed Han sheep, wool traits

DOI: 10.16476/j.pibb.2022.0257

Wool is an important raw material in the textile industry^[1]. It has the advantages of good elasticity, strong hygroscopicity, and good warmth retention. Wool textiles are famous for their luxurious, elegant, and comfortable natural style; cashmere in particular is referred to as “soft gold”^[2]. Wool fineness, length, strength, and flexure are all important economic traits that affect wool quality^[3]. China is a major wool-producing country, and it is also the country with the largest amount of wool processing and consumption. The output of high-quality wool, especially superfine wool, cannot meet the needs of the growing domestic processing enterprises. Domestic wool textile processing enterprises require the importing of a large amount of wool from Australia, New Zealand, and other countries for processing every year. Therefore, it is particularly important to improve wool properties and wool quality. Suffolk sheep are characterized by early maturity, and rapid growth and development, with a hair length of 7–8 cm, a fineness of 50–58, and a net hair rate of about 60%. Their coat is white, but a small number of colored fibers can occasionally be found. Small-tailed Han sheep have white and heterogeneous coats, with a small amount of dry and dead hair, and a few individuals have colored spots on their heads (Figure 1). In comparison with Suffolk sheep wool, the amount of shearing is low, the quality

of wool is poor, and all of the wool is heterogeneous, meaning that it does not meet the requirements of light textile materials. Although these two kinds of sheep are both meat sheep, their wool quality is quite different. Therefore, a comparative study of the reasons for the differences in wool traits between Suffolk sheep and small-tailed Han sheep is of great significance for exploring the complex molecular mechanisms that affect wool traits.



Fig. 1 Back hair of Suffolk and small-tailed Han sheep

RNA sequencing (RNA-seq) is a next-generation high-throughput sequencing technology^[4]. It was first proposed and applied in yeast in 2008^[5]. RNA-seq technology is a high-throughput sequencing platform capable of detecting low-abundance transcripts, identifying novel transcript units, and revealing their differential expression across samples^[6]. The proteome was first proposed by Marc Wilkins, a

scientist at Macquarie University in Australia, at the Siena two-dimensional gel electrophoresis conference in 1994^[7]. “Proteome” refers to the entire collection of proteins encoded by the genome. The proteome looks for differences between proteins through precise quantification and identification. It is a holistic and dynamic study at the protein level, one that can be used for validation and as an important complement to genomics research^[8]. In this study, we referred to others to conduct a sequencing study on wool traits by taking the back skin of sheep as the research object^[9-10]. The back skin of Suffolk sheep and small-tailed Han sheep was selected as the test material, and the gene and protein expression differences in the skin tissue of Suffolk sheep and small-tailed Han sheep were compared by RNA-seq and proteome sequencing. The differential genes were verified by quantitative real-time PCR (qRT-PCR), and the differential proteins were verified by Western blot. The influencing factors of wool traits and the correlation between RNA-seq and proteome sequencing are discussed in this study in attempt to reveal the complex molecular mechanism affecting wool traits.

1 Materials and methods

1.1 Materials

The research samples came from Jiexiu Sheep Farm. There were 3 female Suffolk sheep and 3 female small-tailed Han sheep aged 40 months under the same feeding and management conditions, which were marked as S01, S02, and S03 and as B01, B02, and B03. For each sheep, we used procaine (1.5 ml, 3%) for local anesthesia to reduce animal pain. After hair shearing and alcohol disinfection, 3 pieces of back skin tissue were taken from each sheep, and each tissue was separated by 2–3 cm. Biopsy punch (Miltex, Japan) was used for skin extraction with a diameter of 8 mm and a thickness of approximately 5–6 mm. Yunnan Baiyao powder (Yunnan Baiyao Group Co., Ltd., China) was applied immediately to stop the bleeding. Then the samples were quickly put into the liquid nitrogen until RNA and total protein extraction. In this study, skin samples were collected in accordance with the International Guiding Principles for Biomedical Research Involving Animals and were reviewed and approved by the Ethics Committee of Animal Experiments of Shanxi Agricultural

University (Taigu, China). In our study, these activities did not require a specific permit and the animals did not involve endangered or protected species.

1.2 RNA-Seq of sheep back skin

1.2.1 Total RNA extraction

Total RNA was isolated and purified using TRIzol reagent (Invitrogen, Carlsbad, CA, USA), following the manufacturer's procedure. The RNA amount and purity of each sample were quantified using a NanoDrop ND-1000 (NanoDrop, Wilmington, DE, USA). The RNA integrity was assessed by a Bioanalyzer 2100 (Agilent, CA, USA) with an RNA integrity number (*RIN*) >7.0 and confirmed by electrophoresis with denaturing agarose gel. We used a concentration >50 mg/L, *RIN*>7.0, A_{260}/A_{280} >1.8, and total RNA >1 µg to meet the requirements of the downstream experiments.

1.2.2 RNA library construction

Poly(A) RNA was purified from 1 µg total RNA by two rounds of purification using a Dynabeads Oligo (dT) (Thermo Fisher, cat. 25-61005, USA). Then, the poly(A) RNA was fragmented into small pieces using a Magnesium RNA Fragmentation Module (NEB, cat. e6150, USA) at 94°C for 5–7 min. Following this, the cleaved RNA fragments were reverse-transcribed to create the cDNA by SuperScript™ II Reverse Transcriptase (Invitrogen, cat. 1896649, USA), which was next used to synthesise U-labeled second-stranded DNAs with *E. coli* DNA polymerase I (NEB, cat. m0209, USA), RNase H (NEB, cat. m0297, USA), and dUTP solution (Thermo Fisher, cat. R0133, USA). An A-base was then added to the blunt ends of each strand, preparing them for ligation to the indexed adapters. Each adapter contained a T-base overhang for ligating the adapter to the A-tailed fragmented DNA. Single index or dual index adapters were ligated to the fragments, and size selection was performed with AMPure XP beads. After the heat-labile UDG enzyme (NEB, cat. m0280, USA) treatment of the U-labeled second-stranded DNAs, the ligated products were amplified with PCR under the following conditions: initial denaturation at 95°C for 3 min; 8 cycles of denaturation at 98°C for 15 s, annealing at 60°C for 15 s, and extension at 72°C for 30 s; and then final extension at 72°C for 5 min. The average insert size for the final cDNA library was

(300±50) bp. Finally, we performed the 2×150 bp paired-end sequencing (PE150) on an Illumina Novaseq™ 6000 (LC-Bio Technology Co., Ltd., Hangzhou, China) following the vendor's recommended protocol.

1.2.3 Bioinformatics analysis

Cutadapt software was used to remove the reads that contained adaptor contamination. After we removed the low-quality bases and undetermined bases, we used HISAT2 software to map reads to the genome. The mapped reads of each sample were assembled using StringTie with default parameters. Then, all transcriptomes from all samples were merged to reconstruct a comprehensive transcriptome using gffcompare software. After the final transcriptome was generated, StringTie and ballgown were used to estimate the expression levels of all

transcripts and to perform the expression level for mRNAs by calculating *FPKM* ($FPKM = [\text{total exon fragments}/\text{mapped reads (millions)} \times \text{exon length (kB)}]$). The differentially expressed mRNAs were selected with $FC \geq 1.4$ or $FC \leq 0.714$ and $P < 0.05$ by R package edgeR^[11] or DESeq2^[12], and then, GO enrichment and KEGG enrichment analyses were conducted on the differentially expressed mRNAs.

1.2.4 Verification of RNA-seq results by qRT-PCR

We used qRT-PCR to validate RNA-seq data and gene expression levels. The extracted RNA was reverse-transcribed to synthesize cDNA using Vazyme's HiScript III RT SuperMix for qPCR (+gDNA wiper), and 6 genes were randomly selected for qRT-PCR detection using ChamQ Universal SYBR qPCR Master Mix. The primer sequences are shown in Table 1.

Table 1 Primer sequences for qRT-PCR

Gene	Sequence (5'→3')	Products size/bp	Anneal temperature/°C
<i>CDO1</i>	F: ACACGGCAGCAGTATCCACG R: TGGTTTTCCCTCAGGATTCTTTCAG	156	56
<i>CST3</i>	F: CAGCGAGTTCAACAAGCG R: GTAAACCTGGAAGGAGCACA	214	56
<i>FABP4</i>	F: TCAGTGTAATGGGGATGTGG R: GGTAGCAGTGACACCGTTCAT	251	58
<i>KRT10</i>	F: GTCCCAACTAGCCCTAAAACA R: ATTCTGGCACTCGGTCTCC	148	57
<i>KRT35</i>	F: GGTTCAACACCCAGACAGAGG R: TCCAGGGCATTGACCGTAC	112	58
<i>SI00A4</i>	F: TAGGGAAAAGGACGGATGA R: CAGGACAGGAAGACGCAGTA	106	56
<i>18S</i>	F: GAAGGGCACCACCAGGAGT R: CAGACAAATCACTCCACCAA	198	56–58

1.2.5 qRT-PCR data analysis

The qRT-PCR results were analyzed by the $2^{-\Delta\Delta Ct}$ method^[13].

1.3 Proteome sequencing of sheep back skin

1.3.1 Protein extraction

The samples were frozen in liquid nitrogen and ground with a pestle and mortar. A fivefold volume of TCA/acetone (1 : 9) was added to the powder, and the mixture was mixed by vortex. The mixture was placed at -20°C for 4 h and then centrifuged at 6 000×g for 40 min at 4°C. The supernatant was discarded. The pre-cooling acetone was added, and the mixture was washed 3 times. The precipitation was air dried. A

30-fold volume of SDT buffer (4% SDS, 100 mmol/L Tris-HCl, pH7.6) was added to 20–30 mg powder; the mixture was then mixed together, followed by boiling for 5 min. The lysate was sonicated and then boiled for 15 min. After being centrifuged at 14 000×g for 40 min, the supernatant was filtered with 0.22 μm filters. The filtrate was quantified with the BCA Protein Assay Kit (P0012, Beyotime). The sample was stored at -80 °C.

1.3.2 SDS-PAGE separation

A total of 20 μg of proteins of each sample was mixed with 6×loading buffer and then boiled for 5 min. The proteins were separated on 12%

SDS-PAGE gel. Protein bands were visualized by Coomassie Blue R-250 staining.

1.3.3 Filter-aided sample preparation digestion

A total of 200 μg of proteins of each sample was reduced with 50 mmol/L DTT for 30 min at 56°C. Then, the detergent, DTT, and other low-molecular-mass components were removed using UA buffer (8 mol/L urea, 150 mmol/L Tris-HCl, pH 8.5) by repeated ultrafiltration (Sartorius, 30 ku). Following this, 100 μl iodoacetamide (100 mmol/L iodoacetamide (IAA) in urea (UA) buffer) was added to block the reduced cysteine residues, and then, the samples were incubated for 30 min in darkness. The filters were washed with 100 μl UA buffer three times and, then, with 100 μl of the 25 mmol/L NH_4HCO_3 buffer twice. Finally, the protein suspensions were digested with 4 μg trypsin (Promega) in 40 μl of 25 mmol/L NH_4HCO_3 buffer overnight at 37°C, and the resulting peptides were collected as a filtrate^[14].

1.3.4 Mass spectrometry (MS) analysis

The peptide of each sample was desalted on C18 Cartridges and then concentrated by vacuum centrifugation and reconstituted in 40 μl of 0.1% (v/v) formic acid. The peptide content was estimated by UV light spectral density at 280 nm using an extinction coefficient of 1.1 of 0.1% (w/v) solution that was calculated on the basis of the frequency of tryptophan and tyrosine in vertebrate proteins. LC-MS/MS analysis was performed on a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) that was coupled to Easy nLC (Thermo Fisher Scientific). A total of 2 μg of peptide was loaded onto the C18-reversed-phase analytical column (Thermo Fisher Scientific, Acclaim PepMap RSLC 50 $\mu\text{m}\times 15\text{ cm}$, nano viper, P/N164943) in buffer A (0.1% formic acid) and separated with a linear gradient of buffer B (80% acetonitrile and 0.1% formic acid) at a flow rate of 300 nL/min. The linear gradient was as follows: 5% buffer B for 5 min, 5%–28% buffer B for 90 min, 28%–38% buffer B for 15 min, 38%–100% buffer B for 5 min, and held in 100% buffer B for 5 min. MS data were acquired using a data-dependent top10 method that dynamically chose the most abundant precursor ions from the survey scan (350–1 800 m/z) for HCD fragmentation. MS1 scans were acquired at a resolution of 70 000 at m/z 200 with an AGC target of 3×10^6 and a maxIT of 50 ms. MS2 scans were acquired at a resolution of 17 500 at m/z 200 with an

AGC target of 2×10^5 and a maxIT of 45 ms, and the isolation width was 2 m/z . Only ions with a charge state between 2 and 6 and a minimum intensity of 2×10^3 were selected for fragmentation. Dynamic exclusion for the selected ions was 30 s. Normalized collision energy was 27 eV.

1.3.5 Protein quantitative analysis

This project used MaxQuant software (version 1.6.14.0) for the database search and the label-free quantitation (LFQ)^[15] algorithm for quantitative analysis. The global false discovery rate cutoff for peptide and protein identification was set at 0.01. Protein abundance was calculated on the basis of normalized protein spectral intensities. Proteins with an $FC\geq 1.4$ or $FC\leq 0.714$ and $P<0.05$ were considered differentially expressed proteins (DEPs).

1.3.6 Bioinformatics analysis

First, all protein sequences were aligned to the Ovis aries database, which was downloaded from the ensembl; only the sequences in the top 10 and with E-values $\leq 1\times 10^{-3}$ were kept. Second, we selected the GO term of the sequence with top Bit-Score by Blast2GO. Then, we completed the annotation from GO terms to proteins using the Blast2GO Command Line. After the elementary annotation, InterProScan^[16] was used to search the EBI database on the basis of motifs, and then, we added the functional information of the motifs to proteins in order to improve annotation. Then, we carried out further improvements in annotation and connection between GO terms by ANNEX. Fisher's exact test was used to enrich GO terms through the comparison of the number of differentially expressed proteins and total proteins correlated to GO terms. A pathway analysis was performed using the KEGG database. Fisher's exact test was used to identify the significantly enriched pathways through the comparison of the number of differentially expressed proteins and total proteins correlated to pathways.

1.3.7 Validation of proteome sequencing results by Western blot

We used Western blot to verify the proteome sequencing results. The grouped samples were extracted by the SDT cleavage method, and four proteins were randomly selected for verification. The loading amount of each sample was 400 μg . Electrophoretic separation was performed on 10% SDS-PAGE in order to obtain clear electrophoretic

bands, which were transferred to NC membranes and exposed in a dark room using an ECL kit. The signal bands were scanned, and the final film was analyzed by ImageJ for NDUFS3 (Sangon Biotech, D122746), CORO1A (Sangon Biotech, D163499), FABP4 (Sangon Biotech, D120618), RPL10 (Sangon Biotech, D151841), and β -actin (CW BIO, CW0103S) gray band values.

1.4 Correlation analysis between RNA-seq and proteome sequencing of back skin

As the translation product of mRNA, proteins performed specific functions. Transcriptome data and proteome data could be correlated simply by mRNA-protein translation relationships, and the genes that really play a role could be locked from the many directions of transcriptome sequencing. The specific steps were as follows: the transcriptome data were

correlated with gene data, and gene data were correlated with protein data. The correlation matrix was obtained by taking genes as a bridge, and the data were screened by using significant difference data and regulatory relationship. Finally, a GO and KEGG enrichment analysis was carried out on the data.

2 Results and analysis

2.1 Quality control detection of sequencing data

2.1.1 Quality control detection of RNA-seq data

After sequencing, a total of 230 406 674 original data and 222 049 370 clean data were obtained, of which the Q20 base percentage was over 99.9% and the Q30 base percentage was over 98% (Table 2). This showed that the sequencing data were of good quality and could be used for subsequent test analysis.

Table 2 Data output for each sample

Sample	Raw data read	Raw data base	Valid data read	Valid data base	Valid ratio (reads)	Q20/%	Q30/%	GC content/%
S_01	38 759 774	5.81 G	37 222 630	5.58 G	96.03	99.97	98.02	51
S_02	35 773 546	5.37 G	34 521 524	5.18 G	96.50	99.97	98.08	51
S_03	42 211 192	6.33 G	40 481 506	6.07 G	95.90	99.98	98.35	50
B_01	39 494 238	5.92 G	38 172 048	5.73 G	96.65	99.98	98.06	51
B_02	35 896 008	5.38 G	34 687 818	5.20 G	96.63	99.98	98.02	50.50
B_03	38 271 916	5.74 G	36 963 844	5.54 G	96.58	99.98	98.04	51

2.1.2 Quality control detection of proteome sequencing data

The mass spectrometry data in this experiment were collected from a Q Exactive Plus high-resolution mass spectrometer, which was able to obtain high-quality MS1 and MS2 spectra. The MS spectrum data were analyzed using Andromeda, resulting in a final score for each MS2 spectrum. It was seen that the MS2 Andromeda score was ideal, wherein about 70%

of the peptides scored above 60 points. Each set of label-free data used peptide false discovery rate (FDR) ≤ 0.01 and protein $FDR \leq 0.01$ as screening criteria in the qualitative analysis. When we combined the data with the obtained distribution of excellent peptide scores, we found that the quality of MS experimental data was high (Figure 2).

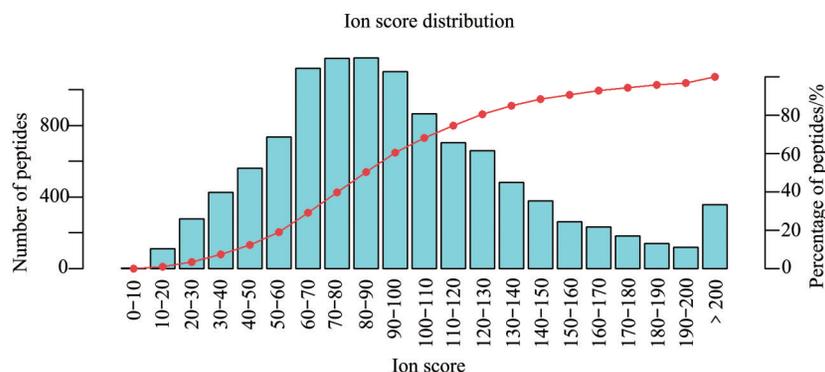


Fig. 2 Peptide ion score distribution

2.2 Analysis results of DEGs and DEPs

2.2.1 Analysis of DEGs

The purpose of the differential expression analysis was to find out the differentially expressed genes (DEGs) between the two groups of samples. Setting the difference multiple $FC \geq 1.4$ or $FC \leq 0.714$ and $P < 0.05$ as the standard, we screened out a total of 1 213 DEGs, among which there were 644 upregulated genes and 569 downregulated genes in Suffolk sheep in comparison with small-tailed Han sheep. We clustered and analyzed the genes in accordance with the similarity of the gene expression profiles of the samples (Figure 3). The results showed that, in terms of sequencing, Suffolk sheep and small-tailed Han sheep had the same population quality, small intra-group differences, and good biological repeatability.

2.2.2 GO and KEGG enrichment analysis of DEGs

A total of 3 588 gene functions were involved in GO enrichment. The top 20 functions with the most significant differences were the extracellular region (GO: 0005576), extracellular space (GO: 0005615), endopeptidase inhibitor activity (GO: 0004866), intermediate filament (GO: 0005882), negative regulation of endopeptidase activity (GO: 0010951), acute-phase response (GO: 0006953), cysteine-type endopeptidase inhibitor activity (GO: 0004869), structural constituent of the epidermis (GO: 0030280), structural molecule activity (GO: 0005198), MHC class II protein complex (GO: 0042613), oligopeptide transmembrane transporter activity (GO: 0035673), calcium ion binding (GO: 0005509), keratin filament (GO: 0045095), negative regulation of dendritic spine development (GO: 0061000), high-density lipoprotein particle (GO: 0034364), antigen processing and presentation of peptide or polysaccharide antigen *via* MHC class II (GO: 0002504), cornified envelope (GO: 0001533), oligopeptide transmembrane transport (GO: 0035672), chromosome passenger complex (GO: 0032133), odontogenesis (GO: 0042476) (Figure 4a). Among them, the gene functions that play a role in wool traits might be the intermediate filament, calcium ion binding, and keratin filament, among others.

All DEGs involved a total of 285 KEGG biological pathways, of which the top 20 pathways with the most significant differences were staphylococcus aureus infection (ko05150), cell

adhesion molecules (CAMs) (ko04514), antigen processing and presentation (ko04612), complement and coagulation cascades (ko04610), glycosaminoglycan biosynthesis-keratan sulfate (ko00533), hematopoietic cell lineage (ko04640), alpha-linolenic acid metabolism (ko00592), arachidonic acid metabolism (ko00590), estrogen signaling pathway (ko04915), extracellular matrix (ECM)-receptor interaction (ko04512), phagosome (ko04145), platinum drug resistance (ko01524), glutathione metabolism (ko00480), drug metabolism-cytochrome P450 (ko00982), basal cell carcinoma (ko05217), linoleic acid metabolism (ko00591), tuberculosis (ko05152), herpes simplex virus 1 infection (ko05168), glycosylphosphatidylinositol (GPI)-anchor biosynthesis (ko00563), asthma (ko05310). Among them, the ECM-receptor interaction might play an important role in wool traits (Figure 4b). Further analysis found that ten differential genes, including ENSOARG00000005796, HSPG2, and ITGA7, were involved in the ECM-receptor interaction pathway.

2.2.3 qRT-PCR to verify the sequencing results

The results showed that the expression trends of these six genes in qRT-PCR were basically the same as those in RNA-seq (Figure 5), indicating that the sequencing data were of good quality and could be used for subsequent analysis.

2.2.4 Analysis of DEPs

Setting the difference multiple $FC \geq 1.4$ or $FC \leq 0.714$ and $P < 0.05$ as the standard, we screened out 99 DEPs, among which there were 47 upregulated proteins and 52 downregulated proteins in Suffolk sheep in comparison with small-tailed Han sheep.

2.2.5 GO and KEGG enrichment analysis of DEPs

A total of 592 protein functions were involved in GO enrichment. The top 20 functions with the most significant differences were the extracellular exosome (GO: 0070062), intermediate filament (GO: 0005882), structural molecule activity (GO: 0005198), immunoglobulin binding (GO: 0019865), high-density lipoprotein particle receptor binding (GO: 0070653), ADP biosynthetic process (GO: 0006172), phagolysosome assembly (GO: 0001845), elastic fiber assembly (GO: 0048251), chylomicron (GO: 0042627), nucleotide phosphorylation (GO: 0046939), ribosomal large subunit binding (GO: 0043023), AMP metabolic process (GO: 0046033), T cell mediated cytotoxicity

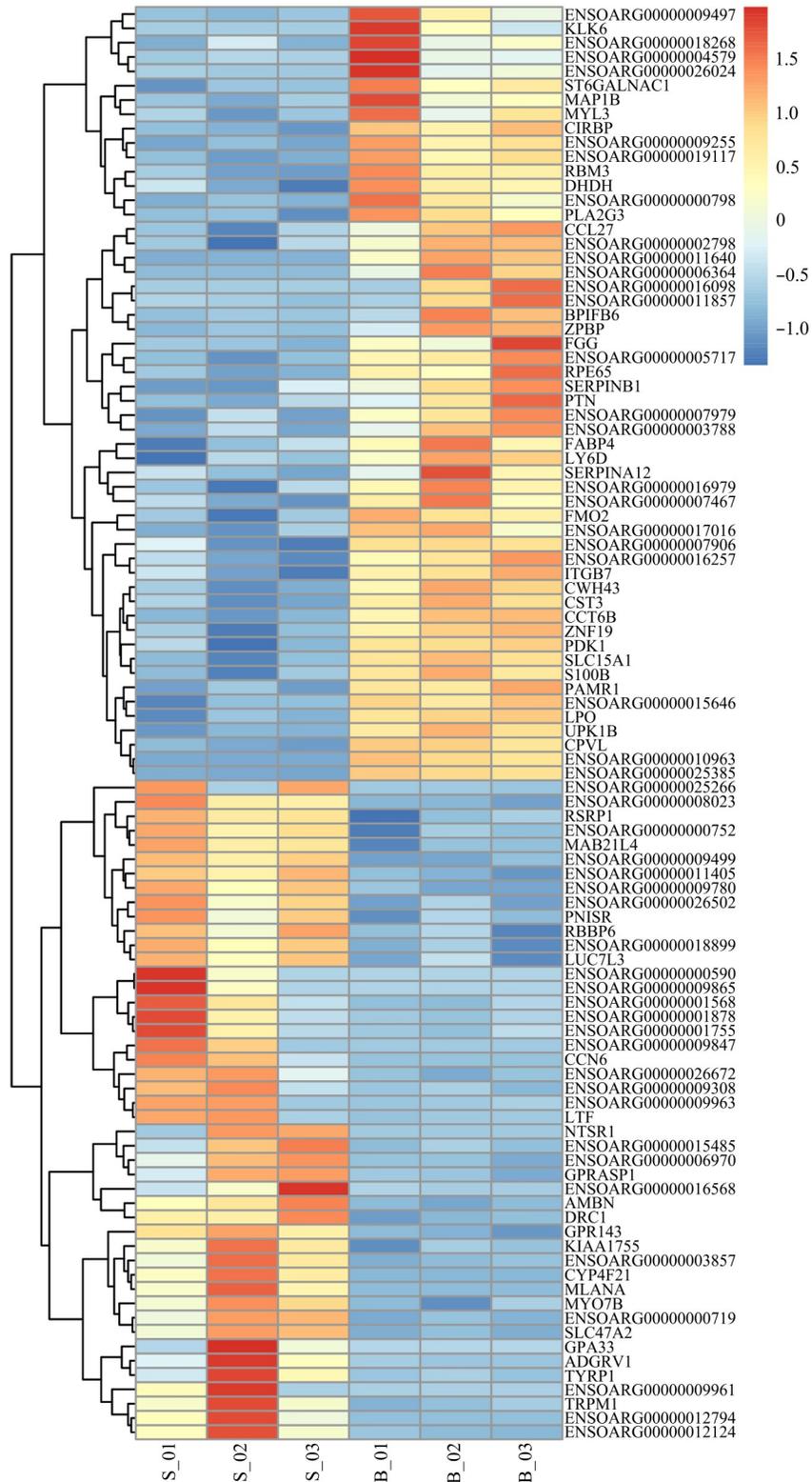


Fig. 3 Cluster analysis of DEGs

In order to better intuitively reflect the clustering expression patterns, we used $\lg(FPKM+1)$ to display them. For biological repetition, Z-values ($Z_{\text{sample}_i} = (FPKM_{\text{sample}_i} - \text{Mean}_{FPKM}) / \text{Standard deviation}_{FPKM}$ of all samples) were used to display gene expression. The abscissa is the sample, and the ordinate is the screened DEGs. Different colors represent different gene expression levels. The colors from blue to white to red represent the expression level from low to high-red represents highly expressed genes and blue represents lowly expressed genes.

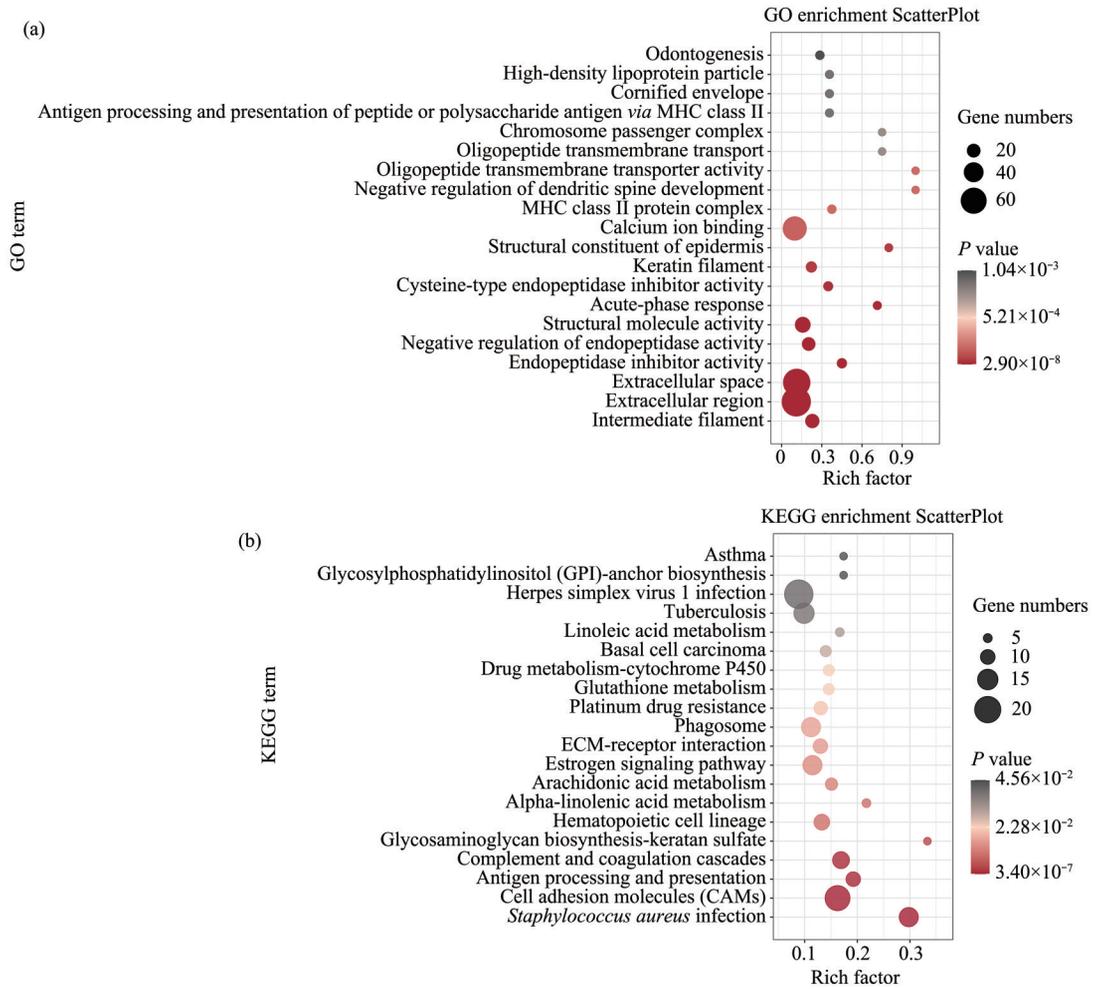


Fig. 4 GO and KEGG enrichment analysis results of DEGs

(a) GO enrichment analysis results of DEGs. (b) KEGG enrichment analysis results of DEGs.

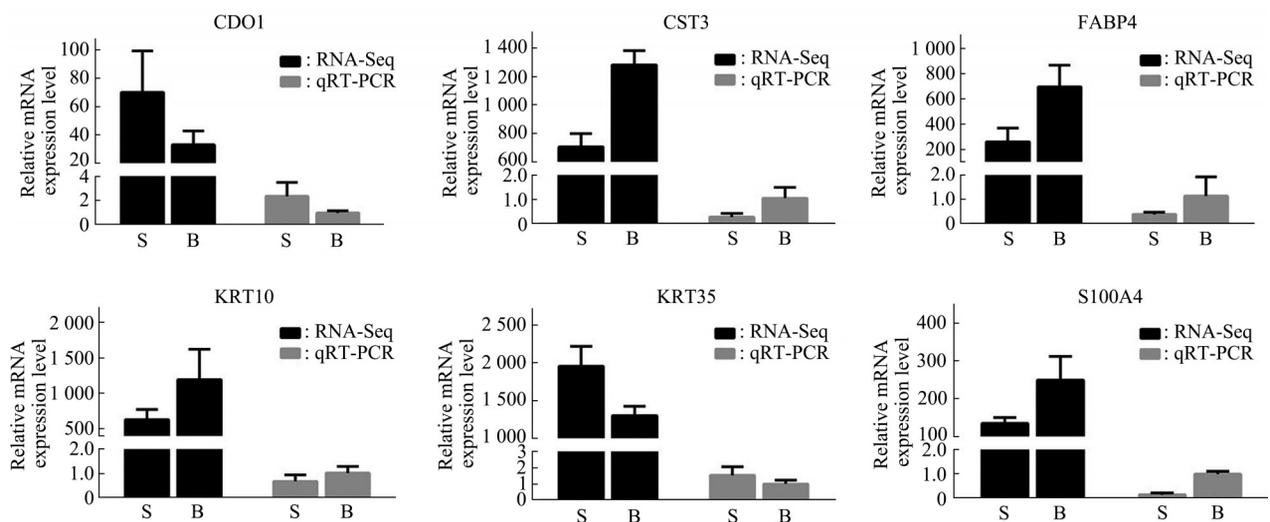


Fig. 5 Relative expression levels of some genes in the RNA-seq and qRT-PCR verification results

S: suffolk sheep; B: small-tailed Han sheep.

(GO:0001913), transport (GO:0006810), triglyceride catabolic process (GO:0019433), proteolysis involved in cellular protein catabolic process (GO:0051603), adenylate kinase activity (GO:0004017), aminopeptidase activity (GO:0004177), cellular response to thyroid hormone stimulus (GO:0097067), cysteine-type peptidase activity (GO:0008234) (Figure 6a). The key function affecting wool traits might be intermediate filament, and this result was consistent with RNA-seq.

The DEPs involved a total of 232 KEGG biological pathways, of which the top 20 pathways with the most significant differences were the estrogen signaling pathway (ko04915), prion diseases (ko05020), PPAR signaling pathway (ko03320), focal adhesion (ko04510), ECM-receptor interaction (ko04512), African trypanosomiasis (ko05143), human papillomavirus infection (ko05165), vibrio

cholerae infection (ko05110), apoptosis (ko04210), amoebiasis (ko05146), vasopressin-regulated water reabsorption (ko04962), cholesterol metabolism (ko04979), ribosome (ko03010), relaxin signaling pathway (ko04926), platelet activation (ko04611), vitamin B6 metabolism (ko00750), regulation of lipolysis in adipocytes (ko04923), oocyte meiosis (ko04114), drug metabolism-cytochrome P450 (ko00982), Parkinson disease (ko05012) (Figure 6b). Among them, the signaling pathways that affect wool traits might be the PPAR signaling pathway and the ECM-receptor interaction. Further analysis found that four proteins, including FABP9, FABP4, APOA1 and APOC3, were involved in the PPAR signaling pathway. Four proteins, including TNXB, COL6A5, VWF and COL1A1, were involved in the ECM-receptor interaction pathway.

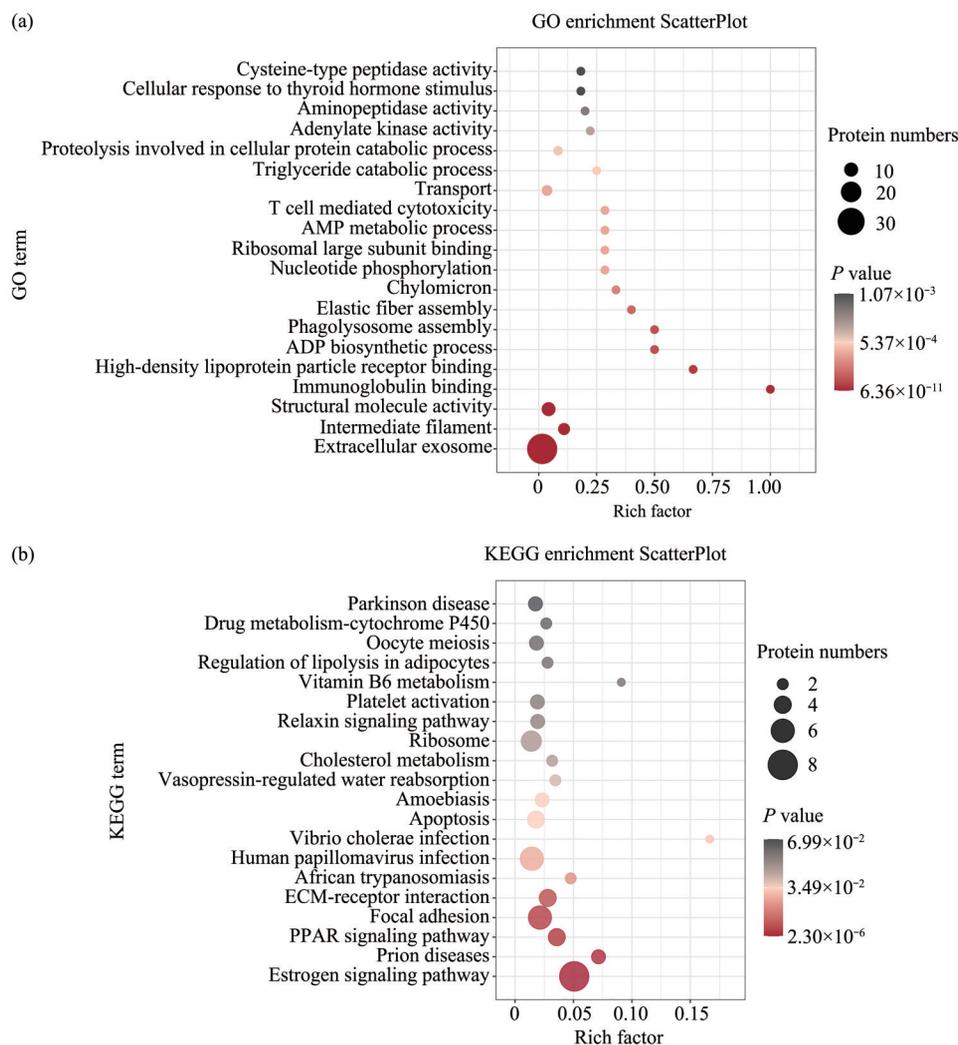


Fig. 6 GO and KEGG enrichment analysis results of DEPs

(a) GO enrichment analysis results of DEPs. (b) KEGG enrichment analysis results of DEPs.

2.2.6 Western blot verification of sequencing results

The Western blot detection results showed that the expression trends of the 4 selected proteins were

essentially the same as the proteome sequencing results (Figure 7), indicating that the sequencing data were of good quality and could be used for subsequent analysis.

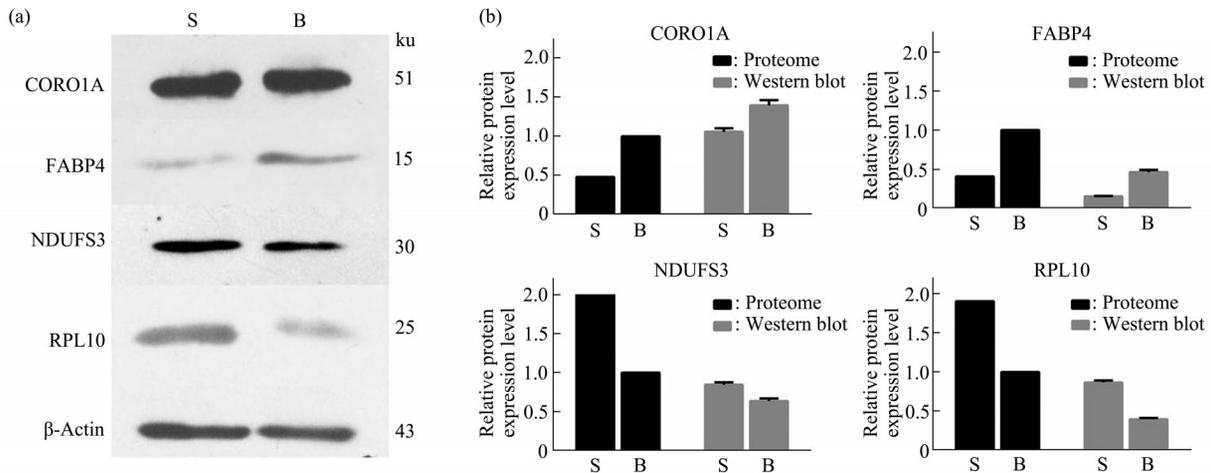


Fig. 7 The relative expression levels of some proteins in the proteome sequencing and the results of the Western blot verification

(a) Western blot results. (b) Comparison of proteome sequencing results and Western blot results. S: suffolk sheep; B: small-tailed Han sheep.

2.3 Association analysis of RNA-seq and proteome sequencing

The significantly different genes detected in both RNA-seq and proteome sequencing were screened by the method of intersecting Venn diagrams. Among them, 15 genes with significant differences were screened (Figure 8). The genes were LOC101106046-keratin-associated protein 13-1-like, ENSOARG0000002492, KRTDAP, ENSOARG0000006366, FABP4, AHCY, RBM3, FMO2, ENSOARG00000013504, LOC101104808-aldehyde oxidase 4-like, KRT35, KRT13, AMDHD2, SELENBP1 and SYNE2. We found that only

ENSOARG0000006366 and FMO2 were negatively correlated and that the rest were positively correlated (Table 3).

The correlation between RNA-seq and proteome sequencing found that a total of 81 functions were enriched, of which the top 20 functions with the most significant differences were: intermediate filament (GO: 0005882), flavin adenine dinucleotide binding (GO:0050660), selenium binding (GO:0008430), cold acclimation (GO:0009631), oxygen metabolic process (GO: 0072592), N-acetylglucosamine-6-phosphate deacetylase activity (GO: 0008448), N-acetylglucosamine catabolic process (GO:0006046), nuclear lumen (GO: 0031981), NADPH oxidation (GO: 0070995), carbon-nitrogen ligase activity, with glutamine as amido-N-donor (GO: 0016884), nuclear migration along microfilament (GO: 0031022), organic acid metabolic process (GO:0006082), toxin metabolic process (GO:0009404), cellular response to cold (GO:0070417), adenosylhomocysteinase activity (GO: 0004013), S-adenosylmethionine cycle (GO: 0033353), N-acetylneuraminate catabolic process (GO: 0019262), meiotic nuclear membrane microtubule tethering complex (GO:0034993), N,N-dimethylaniline monooxygenase activity (GO: 0004499), ribosomal large subunit binding (GO: 0043023) (Figure 9a). Among them, the function that

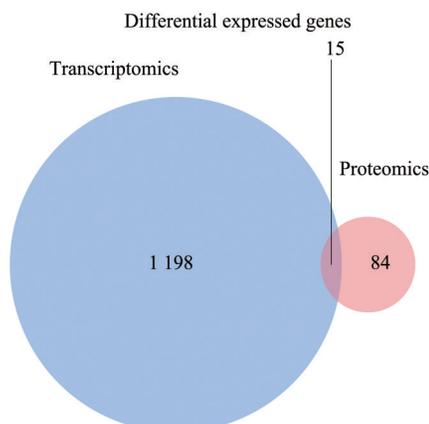


Fig. 8 The significantly different genes detected in both RNA-seq and proteome sequencing

Table 3 Screened significantly different genes

Proteins ID	Genes ID	Gene name	Proteome <i>FC</i>	Proteome <i>P</i> value	RNA-seq <i>FC</i>	RNA-seq <i>P</i> value
ENSOARP0000000837.1	ENSOARG0000000815	LOC1011060	3.35	0.01	2.00	0.01
ENSOARP0000002644.1	ENSOARG0000002492	—	2.24	0.01	1.65	0.05
ENSOARP00000005155.1	ENSOARG00000004804	KRTDAP	0.37	0.01	0.42	0.01
ENSOARP00000006824.1	ENSOARG00000006366	—	0.33	0.01	1.65	0.01
ENSOARP00000010025.1	ENSOARG00000009344	FABP4	0.41	0.01	0.28	0.01
ENSOARP00000010038.1	ENSOARG00000009354	AHCY	1.96	0.01	1.46	0.03
ENSOARP00000013129.1	ENSOARG00000012250	RBM3	0.47	0.01	0.00	0.01
ENSOARP00000013327.1	ENSOARG00000012434	FMO2	1.90	0.02	0.43	0.05
ENSOARP00000014473.1	ENSOARG00000013504	—	0.48	0.01	0.38	0.01
ENSOARP00000017455.1	ENSOARG00000016231	LOC101104808	0.59	0.05	0.40	0.01
ENSOARP00000018025.1	ENSOARG00000016784	KRT35	1.73	0.05	1.50	0.01
ENSOARP00000018205.1	ENSOARG00000016951	KRT13	0.29	0.01	0.56	0.01
ENSOARP00000019966.1	ENSOARG00000018598	AMDHD2	0.45	0.01	0.45	0.01
ENSOARP00000022560.1	ENSOARG00000020993	SELENBP1	1.82	0.03	1.88	0.01
ENSOARP00000022729.1	ENSOARG00000021147	SYNE2	1.74	0.04	1.45	0.03

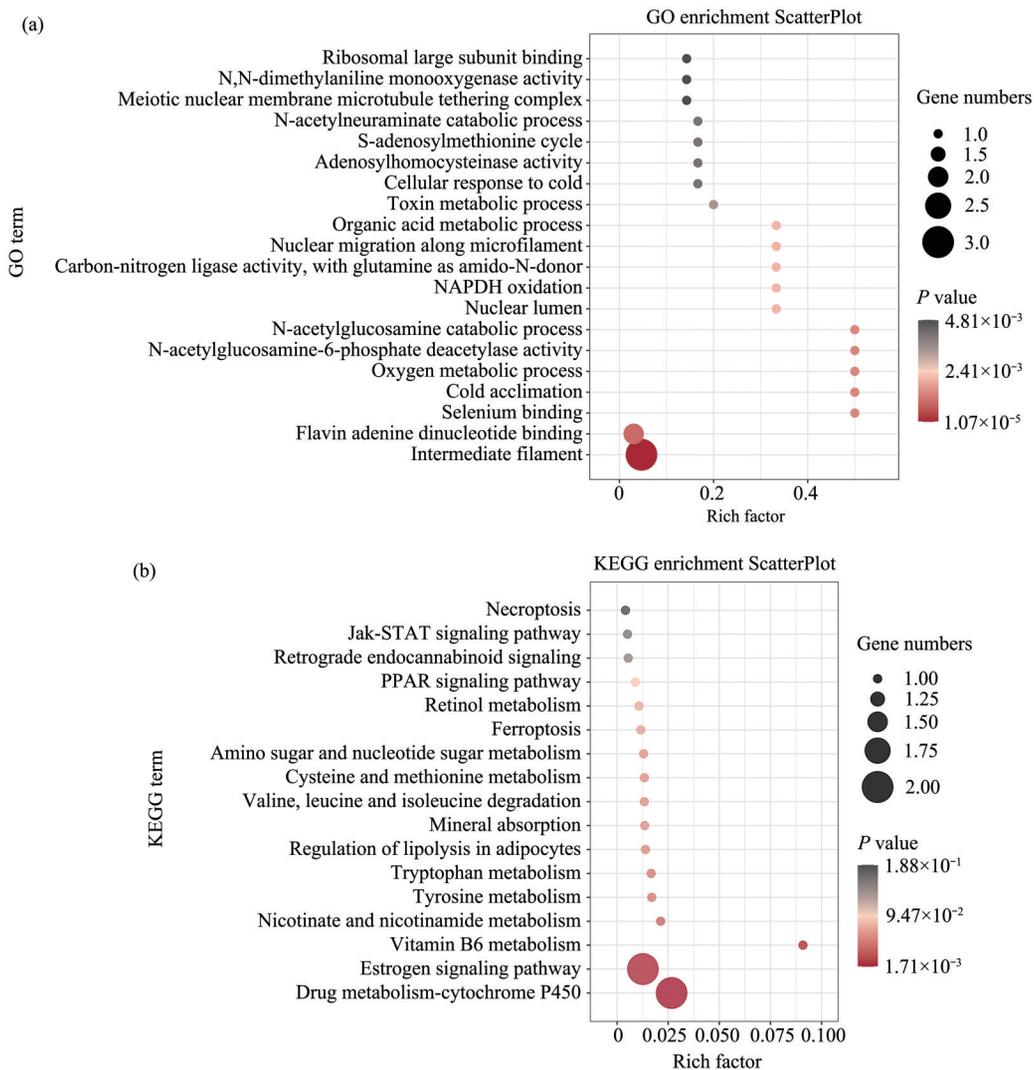


Fig. 9 The correlation analysis between RNA-seq and proteome sequencing analysis enrichment results

played a role in wool traits might be intermediate filament. The involved KAP13-1-like, KRT35 and KRT13 might be a key candidate gene that affects wool traits.

The correlation between RNA-seq and proteome sequencing showed that there were 17 KEGG biological pathways involved in the difference. They were drug metabolism-cytochrome P450 (ko00982), estrogen signaling pathway (ko04915), vitamin B6 metabolism (ko00750), nicotinate and nicotinamide metabolism (ko00760), tyrosine metabolism (ko00350), tryptophan metabolism (ko00380), regulation of lipolysis in adipocytes (ko04923), mineral absorption (ko04978), valine, leucine and isoleucine degradation (ko00280), cysteine and methionine metabolism (ko00270), amino sugar and nucleotide sugar metabolism (ko00520), ferroptosis (ko04216), retinol metabolism (ko00830), PPAR signaling pathway (ko03320), retrograde endocannabinoid signaling (ko04723), Jak-STAT signaling pathway (ko04630), necroptosis (ko04127) (Figure 9b). The KEGG pathway that played a role in wool traits might be the PPAR signaling pathway. The FABP4 involved was found to be a key candidate gene that affects wool traits.

3 Discussion

The quality of wool is closely related to wool fineness, density, and curvature. With the development of high-throughput sequencing technology, genes affecting wool traits have been continuously discovered. It plays a positive role on the research of molecular regulation mechanism of hair traits.

The results of RNA-seq showed that 1 213 DEGs were screened out with $FC \geq 1.4$ or $FC \leq 0.714$ and $P < 0.05$ as the standard. In comparison with small-tail Han sheep, Suffolk sheep had 644 upregulated genes and 569 downregulated genes. Differential gene affected wool traits by affecting intermediate filament, calcium ion binding, and keratin filament formation. Kang^[17] performed transcriptome sequencing analysis on Tan sheep skin tissue at 1 month and 48 months. It was found that keratin family members KRT25, KRT5, KRT71, KRT14, and among others were found to be related to hair structure formation. The high expression of keratin was closely related to the regulation of calcium levels. Calcium levels were a

key factor in controlling terminal differentiation of keratinocytes^[18]. Zhao^[19] found that the expression levels of KRT1, KRT10, and KRT24 related to cashmere fibers were upregulated in the transcriptome analysis of skin tissues of Liaoning cashmere goats and Ziwuling black goats. Intermediate filament skeleton genes, such as KRTAPs, KRTAPs-like, and a few KRTs, as well as the differentially expressed KRTAPs and KRTAPs-like genes obtained by sequencing, showed downregulated expression patterns in skin tissues of Liaoning cashmere goats. In addition, GO enrichment found that intermediate filament and intermediate filament cytoskeleton genes were significantly enriched. Transcriptomic studies on the skin of long-haired and short-haired Inner Mongolia cashmere goats found that DEGs were enriched in intermediate filaments and in the intermediate filament cytoskeleton^[20]. Therefore, intermediate filament, keratin filament, and calcium ion binding might have effects on wool traits. The KEGG pathway that affected wool traits was the ECM-receptor interaction pathway. The ECM consists of a complex mixture of macromolecules, playing an important role in tissue and organ morphogenesis and the maintenance of cellular and tissue structure and function. Zhu *et al.*^[21] found that the ECM-receptor interaction and the Wnt/ β -catenin/Lef1 signaling pathway were closely related to hair follicle morphogenesis. An analysis of the DEGs related to coat color of Tarim red deer found that the most significant enrichment of the KEGG signaling pathway was the ECM-receptor interaction^[22]. In addition, the transcriptomic analysis of the early skin hair follicle development of northern Shanbei white cashmere goats also found that the KEGG pathway was mainly enriched in the cancer pathway and the ECM-receptor interaction signaling pathway^[23]. This study also found that the ECM-receptor interaction was significantly enriched in the KEGG signaling pathway, indicating that it might have effects on wool traits.

Proteome sequencing is a validating and important complement of RNA-seq research^[8]. Proteome sequencing showed that 99 DEPs were screened out, among which there were 47 upregulated proteins and 52 downregulated proteins in Suffolk sheep in comparison with small-tailed Han sheep. GO enrichment showed that the gene affecting wool traits was intermediate filament. Intermediate filament

protein was a heterodimeric complex that plays a pivotal role in the hair shaft for its mechanical strength, hair shape, *etc*^[24]. Thus, intermediate filaments were ubiquitous in biological structures including hair^[25]. In addition, keratin also belongs to the intermediate filament family^[26]. Coincidentally, a recent proteomic study that analyzed fiber diversity in sheep and goats also found significant enrichment in keratin or keratin-related proteins, proteins associated with hair growth, and fatty acid synthesis (FABP4 and FABP5)^[27]. KEGG enrichment found that PPAR signaling pathway and ECM-receptor interaction were significantly enriched. PPARs belong to the nuclear hormone receptor superfamily, which were an effective key regulators of epidermal development^[28]. PPARs was critical for maintaining skin barrier permeability, inhibiting keratinocyte growth, promoting keratinocyte terminal differentiation, and regulating skin inflammation. Meanwhile, they may also have protective effects on human hair follicle epithelial stem cells^[29]. In addition, Yue *et al.*^[30] also found that the PPAR pathway may play an important role in the initiation of secondary hair follicles. Studies such as those of Di-Poï and Schmutz showed that PPAR plays a potential role in hair follicle growth and melanocyte differentiation^[31-32]. The study of Li *et al.*^[33] on DEGs in Rex rabbit skin with different coat densities found that the PPAR signaling pathway may be involved in the development of coats. In addition, Dr. Fan's research^[20] on the skin transcriptome of long-haired and short-haired Inner Mongolian cashmere goats found that the KEGG pathway was mainly enriched in the biosynthesis of unsaturated fatty acids, fatty acid metabolism, alpha-linolenic acid metabolism, and the PPAR signaling pathway. Therefore, PPAR signaling pathway might have effects on wool traits.

There were 15 significantly different genes detected in both RNA-seq and proteome sequencing, of which 13 were positively correlated and 2 were negatively correlated. The generation of negative correlation is a normal result because transcription and translation are two different biological processes. After the transcription of RNA, functional protein can be obtained through a series of processes, such as the turnover of RNA and protein, post-translation modification, protein conformational changes, and protein hydrolysis. Correlation analysis found that the functions of intermediate filament was significantly

enriched, and the related genes were KAP13-1-like, KRT35 and KRT13. The PPAR signaling pathway was significantly enriched, and the involved gene was FABP4. KRT is closely related to hair as a natural fibrous structural protein^[34]. Through the analysis of the lncRNAs of cashmere fineness, Zheng *et al.*^[35] found that KRT35 was related to the cashmere fineness in the anagen phase. Asymmetrical expression of KRT35 gene in secondary follicles was associated with bulb deflection and follicle curvature, suggesting a role in the determination of follicle and fibre morphology^[36]. KRT13, a 54 ku type 1 acidic intermediate filament protein often paired with KRT4, is expressed in suprabasal layers of non-cornified stratified squamous epithelia^[37]. Microdissection of hair follicles revealed that KRT13 was mainly located in the main body of the hair follicle and participated in the differentiation of the hair follicle^[38-39]. KAP13-1-like is similar to KAP gene and may play the role of KAP gene. KAPs were structural components of wool fibers and variation in the genes encoding the KAPs can affect wool traits^[40]. KAPs determine some of the most important physical attributes of the fully-keratinised hair fibre, including hardness, toughness and pliability, and in linking intermediate filaments to one another, either directly or indirectly, with a resultant increase in durability and resistance to degradation by microorganisms^[41]. FABP4 is an adipocyte-type fatty acid binding protein that is mainly expressed in adipocytes and macrophages. It is a chaperone in the process of intracellular free fatty acid transport. FABP4 can combine with free fatty acid and transport it into cells. At the same time, FABP4 is involved in regulating gene expression, cell proliferation, and cell differentiation^[42-44]. Fat tissue plays an important role in all aspects of skin physiology, from regulating hair follicle morphogenesis to wound healing^[45]. Skin fat cells develop from dermis cells, independently of subcutaneous fat tissue, and are characterized by limited early expression of FABP4. This study strengthens the link between hair follicle and skin fat cell biology^[46]. A genome-wide association study found that the FABP4 gene is an important factor of resistance to wool rot in Merino sheep^[47]. Li *et al.*^[27] found significant differences of FABP4 protein when they compared the wool and cashmere of small-tailed Han sheep. Likewise, when wool and cashmere from goats and sheep were compared, FABP4 was also

found to be differentially expressed in wool and cashmere, and the expression of FABP4 was higher in wool samples than in cashmere, which show the correlation of this protein with hair diameter. The correlation between FABP4 and wool traits has been further demonstrated.

4 Conclusion

In this study, RNA-seq on the back skin tissue of Suffolk sheep and small-tailed Han sheep showed that intermediate filament, calcium ion binding and keratin filament were significantly enriched, and the signaling pathway affecting wool traits might be the ECM-receptor interaction. Proteome sequencing showed that intermediate filament was significantly enriched. The signaling pathways affecting wool traits might be the PPAR signaling pathway and the ECM-receptor interaction. The combined analysis of RNA-seq and proteome sequencing showed that Intermediate filament was significantly enriched, KRT35, KRT13 and KAP13-1-like genes might be the key candidate genes to affect wool traits. The PPAR signaling pathway was significantly enriched and might be a key candidate pathway to affect wool traits. The FABP4 gene might be a key candidate gene to affect wool traits. These results will expand our understanding of the complex molecular mechanisms affecting sheep wool traits and provide a basis for subsequent studies.

References

- [1] Liu G, Liu R, Tang X, *et al.* Expression profiling reveals genes involved in the regulation of wool follicle bulb regression and regeneration in sheep. *Int J Mol Sci*, 2015, **16**(5): 9152-9166
- [2] Ge W, Zhang W, Zhang Y, *et al.* A single-cell transcriptome atlas of cashmere goat hair follicle morphogenesis. *Genomics Proteomics Bioinformatics*, 2021, **19**(3): 437-451
- [3] Zhao H, Hu R, Li F, *et al.* Five SNPs within the FGF5 gene significantly affect both wool traits and growth performance in fine-wool sheep (*Ovis aries*). *Front Genet*, 2021, **12**: 732097
- [4] Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet*, 2019, **20**(11): 631-656
- [5] Nagalakshmi U, Wang Z, Waern K, *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 2008, **320**(5881): 1344-1349
- [6] Pareek C S, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet*, 2011, **52**(4): 413-435
- [7] Dunn M J. 2D electrophoresis: from protein maps to genomes. proceedings of the international meeting, Siena, Italy, September 5-7, 1994. *Electrophoresis*, 1995, **16**(7): 1077-322
- [8] 褚光辉. 基于 iTRAQ 技术的结直肠癌蛋白标志物鉴定[D]. 上海: 海军军医大学, 2018
- Chu G H. Identification of Colorectal Cancer Protein Biomarker Based on iTRAQ[D]. Shanghai: Naval Medical University, 2013
- [9] 陈文硕. 粗毛羊和细毛羊初级毛囊基板发育期皮肤的转录组分析[D]. 武汉: 华中农业大学, 2018
- Chen W S. Comparative Study on Sheep Skin Transcriptome to Reveal Primary Wool Follicle Development Between Coarse and Fine Wool Sheep[D]. Wuhan: Huazhong Agricultural University, 2018
- [10] Ma G W, Chu Y K, Zhang W J, *et al.* Polymorphisms of FST gene and their association with wool quality traits in Chinese Merino sheep. *PLoS One*, 2017, **12**(4): e0174868
- [11] Robinson M D, McCarthy D J, Smyth G K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 2010, **26**(1): 139-140
- [12] Perteu M, Perteu G M, Antonescu C M, *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*, 2015, **33**(3): 290-295
- [13] Livak K J, Schmittgen T D L. Analysis of relative gene expression data using real-time quantitative PCR and the 2-DDCt method. *Methods*, 2001, **25**(4): 402-408
- [14] Wiśniewski J R, Zougman A, Nagaraj N, *et al.* Universal sample preparation method for proteome analysis. *Nat Methods*, 2009, **6**(5): 359-362
- [15] Cox J, Hein M Y, Lubner C A, *et al.* Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*, 2014, **13**(9): 2513-2526
- [16] Quevillon E, Silventoinen V, Pillai S, *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res*, 2005, **33**(Web Server issue): W116-W120
- [17] 康晓龙. 基于转录组学滩羊卷曲被毛形成的分子机制研究[D]. 北京: 中国农业大学, 2013
- Kang X L. Transcriptome Study on the Molecular Mechanism of Curly Coat Phenotype in Chinese Tan Sheep[D]. Beijing: China Agricultural University, 2013
- [18] Collin C, Gautier B, Gaillard O, *et al.* Protective effects of taurine on human hair follicle grown *in vitro*. *Int J Cosmet Sci*, 2006, **28**(4): 289-298
- [19] 赵孟丽. 绒山羊皮肤组织转录组分析及KRTAP基因对绒品质的影响[D]. 兰州: 甘肃农业大学, 2021
- Zhao M L. Analysis of Skin Transcriptome Profiling and Effect of KRTAP Genes on Cashmere Quality in Goats[D]. Lanzhou: Gansu Agricultural University, 2021
- [20] 范一星. 不同毛被类型内蒙古绒山羊皮肤毛囊周期性生长及相关分子机理研究[D]. 呼和浩特: 内蒙古农业大学, 2019
- Fan Y X. Study on Different Hair Types of Inner Mongolia Cashmere Goat in Periodic Changes of Skin and Hair Follicles and Related Molecular Mechanisms[D]. Huhhot: Inner Mongolia Agricultural University, 2019

- [21] Zhu B, Xu T, Yuan J, *et al.* Transcriptome sequencing reveals differences between primary and secondary hair follicle-derived dermal papilla cells of the Cashmere goat (*Capra hircus*). PLoS One, 2013, **8**(9): e76282
- [22] Eli S, Mamat T, Ababaikeri B, *et al.* Screening and preliminary analysis of Tarim Red Deer (*Cervus elaphus yarkandensis*) hair color differentially expressed genes based on RNA-Seq. Acta Veterinaria et Zootechnica Sinica, 2020, **51**(10): 2413-2424
- [23] 高晔. 陕北白绒山羊皮肤毛囊发育早期的转录组学分析及VDR基因的功能研究[D]. 西安:西北农林科技大学, 2017
Gao Y. Comparative Transcriptome Analysis of Fetal Skin Reveals Key Genes Related to Hair Follicle Morphogenesis in Shanbei White Cashmere Goats and Functional Research of VDR[D]. Xiayang: Northwest A&F University, 2017
- [24] Ishii D, Abe R, Watanabe S, *et al.* Stepwise characterization of the thermodynamics of trichocyte intermediate filament protein supramolecular assembly. J Mol Biol, 2011, **408**(5): 832-838
- [25] Murthy N S, Wang W, Kamath Y. Structure of intermediate filament assembly in hair deduced from hydration studies using small-angle neutron scattering. J Struct Biol, 2019, **206**(3): 295-304
- [26] Chou C C, Buehler M J. Structure and mechanical properties of human trichocyte keratin intermediate filament protein. Biomacromolecules, 2012, **13**(11): 3522-3532
- [27] Li Y, Zhou G, Zhang R, *et al.* Comparative proteomic analyses using iTRAQ-labeling provides insights into fiber diversity in sheep and goats. J Proteomics, 2018, **172**: 82-88
- [28] Billoni N, Buan B, Gautier B, *et al.* Expression of peroxisome proliferator activated receptors (PPARs) in human hair follicles and PPAR alpha involvement in hair growth. Acta Derm Venereol, 2000, **80**(5): 329-334
- [29] Ramot Y, Mastrofrancesco A, Camera E, *et al.* The role of PPAR γ -mediated signalling in skin biology and pathology: new targets and opportunities for clinical dermatology. Exp Dermatol, 2015, **24**(4): 245-251
- [30] Yue Y, Guo T, Yuan C, *et al.* Integrated analysis of the roles of long noncoding RNA and coding RNA expression in sheep (*Ovis aries*) skin during initiation of secondary hair follicle. PLoS One, 2016, **11**(6): e0156890
- [31] Di-Poi N, Michalik L, Desvergne B, *et al.* Functions of peroxisome proliferator-activated receptors (PPAR) in skin homeostasis. Lipids, 2004, **39**(11): 1093-1099
- [32] Schmuth M, Haqq C M, Cairns W J, *et al.* Peroxisome proliferator-activated receptor (PPAR) -beta/delta stimulates differentiation and lipid accumulation in keratinocytes. J Invest Dermatol, 2004, **122**(4): 971-983
- [33] Li B, Li C F, Zhu Y L, *et al.* Detection of differentially expressed genes in Rex rabbit skin with different wool density. Acta Veterinaria et Zootechnica Sinica, 2014, **45**(10): 1631-1639
- [34] Ledford B, Barron C, Van Dyke M, *et al.* Keratose hydrogel for tissue regeneration and drug delivery. Semin Cell Dev Biol, 2022, **128**: 145-153
- [35] Zheng Y Y, Sheng S D, Hui T Y, *et al.* An integrated analysis of cashmere fineness lncRNAs in cashmere goats. Genes (Basel), 2019, **10**(4): 266
- [36] Yu Z, Gordon S W, Nixon A J, *et al.* Expression patterns of keratin intermediate filament and keratin associated protein genes in wool follicles. Differentiation, 2009, **77**(3): 307-316
- [37] Li Q, Yin L, Jones L W, *et al.* Keratin 13 expression reprograms bone and brain metastases of human prostate cancer cells. Oncotarget, 2016, **7**(51): 84645-84657
- [38] Fukuyama M, Tsukashima A, Kimishima M, *et al.* Human iPS cell-derived cell aggregates exhibited dermal papilla cell properties in *in vitro* three-dimensional assemblage mimicking hair follicle structures. Front Cell Dev Biol, 2021, **9**: 590333
- [39] Liu Y M, Liu W, Jia J S, *et al.* Abnormalities of hair structure and skin histology derived from CRISPR/Cas9-based knockout of phospholipase C-delta 1 in mice. J Transl Med, 2018, **16**(1): 141
- [40] Ullah F, Jamal S M, Ekegbu U J, *et al.* Polymorphism in the ovine keratin-associated protein gene KRTAP7-1 and its association with wool characteristics. J Anim Sci, 2020, **98**(1): skz381
- [41] Fraser R D B, Parry D A D. Trichocyte keratin-associated proteins (KAPs). Adv Exp Med Biol, 2018, **1054**: 71-86
- [42] Garin-Shkolnik T, Rudich A, Hotamisligil G S, *et al.* FABP4 attenuates PPAR γ and adipogenesis and is inversely correlated with PPAR γ in adipose tissues. Diabetes, 2014, **63**(3): 900-911
- [43] Cataltepe S, Arıkan M C, Liang X, *et al.* Fatty acid binding protein 4 expression in cerebral vascular malformations: implications for vascular remodelling. Neuropathol Appl Neurobiol, 2015, **41**(5): 646-656
- [44] Fuseya T, Furuhashi M, Matsumoto M, *et al.* Ectopic fatty acid-binding protein 4 expression in the vascular endothelium is involved in neointima formation after vascular injury. J Am Heart Assoc, 2017, **6**(9): e006377
- [45] Islam N, Garza L A. Adipose and hair function: an aPPARent connection. J Invest Dermatol, 2018, **138**(3): 480-482
- [46] Wojciechowicz K, Gledhill K, Ambler C A, *et al.* Development of the mouse dermal adipose layer occurs independently of subcutaneous adipose tissue and is marked by restricted early expression of FABP4. PLoS One, 2013, **8**(3): e59811
- [47] Smith W J, Li Y, Ingham A, *et al.* A genomics-informed, SNP association study reveals FBLN1 and FABP4 as contributing to resistance to fleece rot in Australian Merino sheep. BMC Vet Res, 2010, **6**: 27

基于皮肤组织转录组学和蛋白质组学测序揭示影响羊毛性状的关键基因*

吴晋强 郝晓静 赵红霞 董亚洁 王 荣 张鹏翔 王海东 赫晓燕**

(山西农业大学动物医学学院, 太谷 030801)

摘要 目的 羊毛是高档纺织原料,羊毛的物理性质直接关系到羊毛品质。本研究旨在挖掘影响羊毛性状的基因,探索影响羊毛性状的复杂分子机制。**方法** 本研究选取萨福克羊和小尾寒羊各3只作为试验样本,取背部皮肤组织,采用转录组学(RNA-seq)和蛋白质组学测序分析造成羊毛性状差异的基因、蛋白质及相关信号通路。**结果** 转录组测序表明:测序完成后,共得到230 406 674个原始数据和222 049 370个干净数据,其中Q20碱基百分比为99.9%以上,Q30碱基百分比为98%以上。以差异倍数 $FC \geq 1.4$ 或 $FC \leq 0.714$ 且 $P < 0.05$ 作为标准,由此筛选出1 213个差异表达基因(DEGs),其中萨福克羊与小尾寒羊相比,上调基因有644个,下调基因有569个。GO富集发现中间丝、钙离子结合、角蛋白丝显著富集,表明其可能与羊毛性状相关。KEGG富集发现,影响羊毛性状的信号通路可能是ECM-受体相互作用。蛋白质组测序表明:以差异倍数 $FC \geq 1.4$ 或 $FC \leq 0.714$ 且 $P < 0.05$ 作为标准,由此筛选出99个差异表达蛋白(DEPs),其中萨福克羊与小尾寒羊相比,上调蛋白有47个,下调蛋白有52个。GO富集发现中间丝等显著富集,表明其可能与羊毛性状相关。KEGG富集发现,影响羊毛性状的信号通路可能是过氧化物酶体增殖物激活受体(PPAR)信号通路和ECM-受体相互作用。转录组与蛋白质组联合分析发现:转录组和蛋白质组均检测到的显著差异基因共有15个,其中13个为正相关,2个为负相关。中间丝显著富集,其中涉及的KRT35、KRT13、KAP13-1-like基因可能是影响羊毛性状的关键候选基因,PPAR信号通路显著富集,可能是影响羊毛性状的关键候选通路,FABP4基因可能是影响羊毛性状的关键候选基因。**结论** KRT35可能影响羊毛直径和弯曲;KRT13可能影响羊毛分化;KAP13-1-like可能影响羊毛硬度和韧性;FABP4可能影响羊毛直径。这些结果将扩展对影响绵羊毛性状的复杂分子机制的理解,并为后续的研究提供基础。

关键词 转录组测序,蛋白质组测序,萨福克羊,小尾寒羊,羊毛性状

中图分类号 R318, Q38

DOI: 10.16476/j.pibb.2022.0257

* 国家自然科学基金(31772690)和山西省自然科学基金(201701D121106)资助项目。

** 通讯联系人。

Tel: 13753418018, E-mail: sxndhxy@163.com

收稿日期: 2022-05-31, 接受日期: 2022-07-28