Why Pashmina Goat Produces Long Hair-fiber and Barbari doesn't: A Differential Gene Expression Study

Rashid Saif [1, 2], Tania Mahmood[2], Aniqa Ejaz[2], Saeeda Zia[3]

[1] Institute of Biotechnology, Gulab Devi Educational Complex, Lahore, Pakistan
[2] Decode Genomics, Punjab University Employees Housing Scheme (II), Lahore, Pakistan
[3] Department of Sciences and Humanities, National University of Computer and Emerging Sciences, Lahore, Pakistan

Corresponding author: rashid.saif37@gmail.com

Abstract: The Pashmina and Barbari are two famous goat breeds found in the wide areas of the Indo-Pak region. Pashmina is famous for its long hair-fiber (Cashmere) production while Barbari is not-selected for this trait. So, the mRNA expression profiling in the skin samples of both breeds would be an attractive and judicious approach for detecting putative genes involved in this valued trait. Here, we performed differential gene expression analysis on publicly available RNA-Seq data from both breeds. Out of 44,617,994 filtered reads of Pashmina and 55,995,999 of Barbari which are 76.48% and 73.69% mapped to the ARS1 reference transcriptome assembly respectively. A Pairwise comparison of both breeds resulted in 47,159 normalized expressed transcripts while 8,414 transcripts are differentially expressed above the significant threshold. Among these, 4,788 are upregulated in Pashmina while 3,626 transcripts are upregulated in Barbari. Fifty-nine transcripts harbor 57 genes including 32 LOC genes and 24 are annotated genes which were selected on the basis of TMM counts > 500. Genes with ectopic expressions other than uncharacterized and LOC symbol genes are Keratins *(KRT)* and Keratin Associated Proteins *(KRTAPs)*, *CystatinA&6*, *TCHH, SPRR4, PPIA, SLC25A4, S100A11, DMKN, LOR, ANXA2, PRR9* and *SFN.* All of these genes are likely to be involved in keratinocyte differentiation, sulfur matrix proteins, dermal papilla cells, hair follicles proliferation, hair curvature, wool fiber diameter, hair transition, hair shaft differentiation and its keratinization. These differentially expressed reported genes are critically valuable for enhancing the quality and quantity of the pashmina fiber and overall breed improvement. This study will also provide important information on hair follicle differentiation for further enrichment analyses and introducing this valued trait to other goat breeds as well.

Key words: Transcriptome analysis, *Capra hircus*, Differential gene expression, Pashmina goat, Barbari goat, RNA-seq

1. Introduction

Hair follicles (HFs) are composed of integumental tissues which consist of outer dermal and inner epidermal layer cells. Their interaction initiates the hair cycle composed of anagen (proliferation), telogen (resting), catagen (regression), exogen (shedding) and kinogen (new growth) phases that ultimately results in keratinized end products including hairs, skin, horns etc. [1]. It is well recognized that the contribution of animal's hair in thermal insulation and visual display affect the human commerce for which they harvest the fiber [2]. Commercially valuable fiber produced are of goats that usually consist of double coat, the outer long hairs made up of primary hair follicles (PHFs) while the secondary hair follicles (SHFs) make up the inner wool coat (the textile fiber cashmere). The under wool coat is < 19 µm in diameter that is obtained mainly from Pashmina goat breed [3]. The fitness of pashmina fiber depends on number and small diameter of SHFs. It is of prime significance to the country's economy as it is globally valued for its warmth, fine, soft and thick fiber. In the past for centuries, this Pashmina goat have roamed in western zones of Tibetan plateau and now it habitats the high altitude of Himalayas in Pakistan, Nepal and Northern India [4]. As a functional mini-organ, the molecular mechanisms controlling HFs in indigenous livestock are now addressed widely. RNA-Seq technology due to its cost-effectiveness provides such platform to elucidate the novel transcripts, transcripts with low or high abundance and differential gene expression among various breeds. This including other high throughput sequencing technologies can also provide a detailed and accurate molecular insight of the organism under consideration for which the genomic resources are limited.

Herein, we focused to analyze the publically available datasets of Pashmina and Barbari goat's transcriptome skin sample. In contrast to Pashmina goat breed, Barbari is small, short-haired, meat and milk yielding breed and is indigenous to Pakistan, Nepal, India, Mauritius and Vietnam [5]. We compared the digital transcript expression profiles from both breeds to look for highly expressed genes related to HFs.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2021 doi:10.20944/preprints202103.0187.v1

2. Materials and Methods

2.1. Sample source and quality analysis

Whole transcriptome skin sample of Pashmina and Barbari goats were selected to analyze the differential expression of wool/fiber producing genes. Both breeds are shown in Figure 1.

Figure 1. Pashmina goat (left) [6] vs. Barbari (right) [7].

Sequencing reads were taken from ENA source under the experiment accession: SRX3414248 and SRX6655258 respectively. Prior to mapping and quantification of paired end fastq sequences, files were first subjected to quality checks using FastQC (v0.11.8) software [8]. Only the clean reads were then retained by filtering low quality bases by Trimmomatic (v0.36) using sliding window approach [9].

2.2. Normalization of quantified expression levels

We calculated number of mapped reads to each transcript directly from filtered reads by Salmon tool [10] in mapping based mode providing *Capra hircus* transcriptome reference fasta file to perform mapping as an intermediate step. The normalization of raw counts using Trinity's tool script [11], abundance_extimates_to_matrix.pl reported TMM (Trimmed Mean of M-values) and TPM (Transcripts Per Million) values which are effective to eliminate sequencing and gene length biases.

2.3. Gene expression profiling and visualization

The differential expression in transcripts of Pashmina and Barbari goat's skin sample was analyzed to figure out the expression of key genes related to wool production. The number of expressed transcripts in TPM normalized matrix was counted using count_matrix_features_given_MIN_TPM_threshold.pl script of Trinity [12] and then used plot function on R for visual inference. Pairwise comparison between the two samples using counts matrix identified DE transcripts by running script run_DE_analysis.pl. The parameter method was set to edgeR and dispersion value 0.1. This generated MA and volcano plots. From TMM matrix, differentially expressed transcripts (DETs) were extracted and clustered using analyze_diff_exp.pl script. Transcripts were considered significant that have p value at most 0.001 and log2 fold-change (log2FC) of 2. Heatmap and correlation graph were also generated [13].

3. Results

3.1. Quality checks and mapping

Trimming of 47,830,768 reads of Pashmina breed resulted in survival of 44,617,994 (93.28%) pair of reads while from a total of 55,995,999 reads of Barbari 55,502,832 (99.12%) pair of reads retained after cleaning of low quality bases. Overall GC% of both breeds is 52%. Quality scores graphs is shown in (Figure 2). In addition, 76.48% (34126799 reads) of Pashmina breed and 73.69% (40902778 reads) of Barbari mapped successfully to the reference ARS1transcriptome.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2021 doi:10.20944/preprints202103.0187.v1

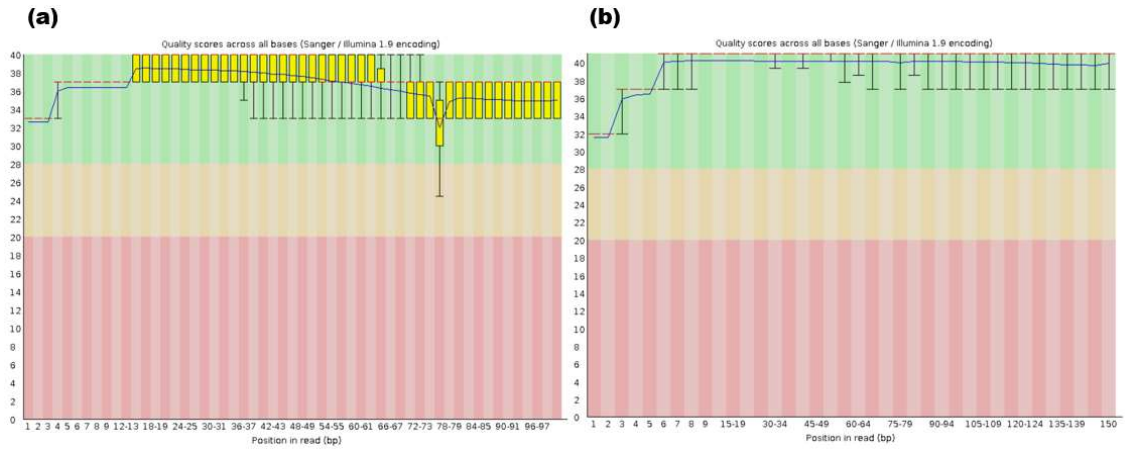

Figure 2. Graphical output of FastQC report showing the quality scores across all bases. (a) Pashmina breed. (b) Barbari breed. The yellow bar exhibits interquartile range, red line is for median value while blue line signifies mean quality of bases.

3.2. Identification and screening of expressed transcripts

To detect the key genes related to cashmere production in Pashmina and Barbari, the pairwise comparison of both breeds was carried out. On the basis of TPM matrix maximally, 47,159 transcripts are found to be expressed. (Figure 3) displays the distribution of these expressed transcripts as MA and Volcano plots.

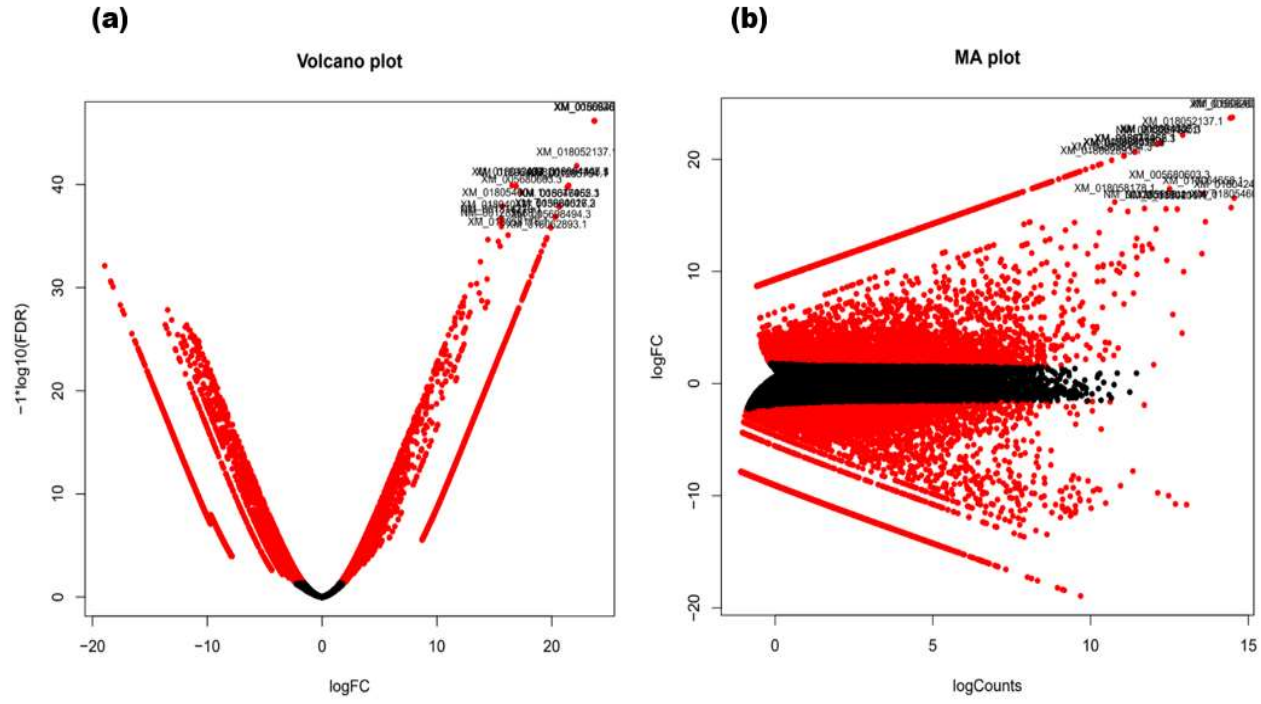

Figure 3. Distribution of transcripts is displayed. (a) Volcano plot (b) MA plot. Non-significant transcripts are represented with black dots. The red dots on right side in volcano plot and upper one in MA plot are upregulated transcripts. Top most genes are annotated with transcript IDs.

While calculating the DETs at a significant threshold (p-value cutoff for FDR < 0.001 and log2FC > 2), we found 8,414 transcripts that appeared to be differentially expressed between Pashmina and Barbari (Table S1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2021 doi:10.20944/preprints202103.0187.v1

Among the DETs, 4,788 are upregulated in Pashmina (Table S2) while 3,626 transcripts are upregulated in Barbari breed (Table S3).

3.3. High differential expression of HF genes

Two-dimensional hierarchical clustering of differentially expressed transcripts was performed to group the similar ones based on the expression patterns which classified the differentially expressed features on a heatmap as shown in (Figure 4).

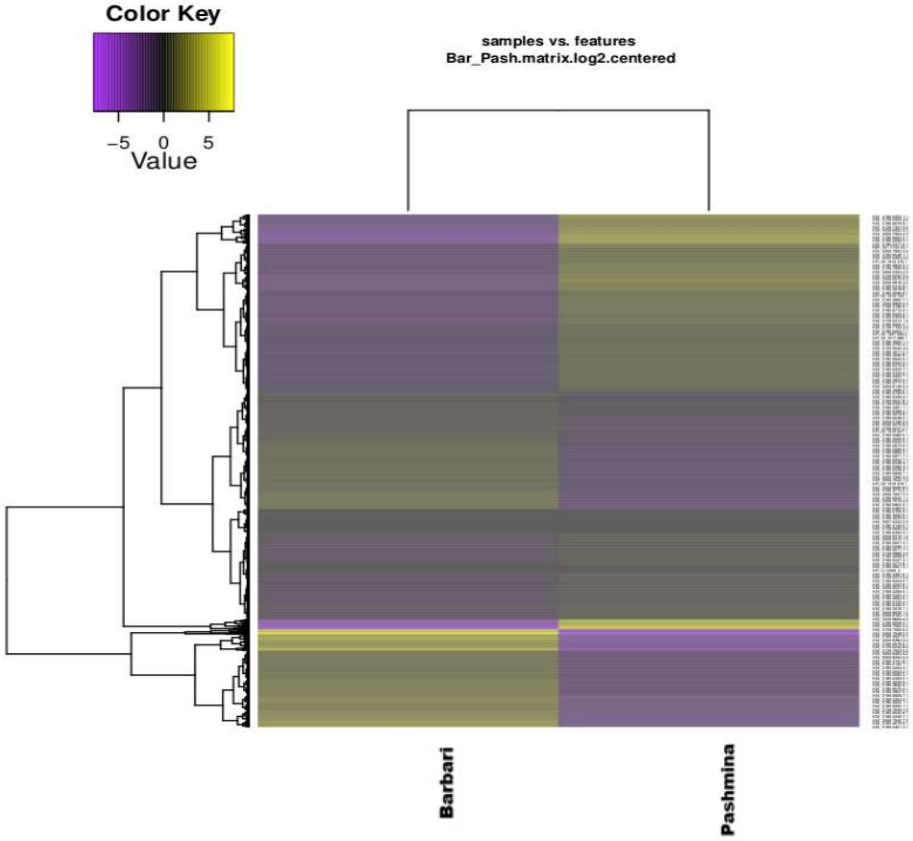

Figure 4. Hierarchical clustering of DETs in Pashmina and Barbari goat based on TMM expressions on heatmap. The upregulated expression of transcripts are shown in yellow while down-regulated transcripts are represented in purple color.

The transcripts with the highest expression in Pashmina and Barbari were separately analyzed to look for the genes involved in wool production. A high expression level having the average normalized counts > 500 was selected to single out the most expressed genes related to HFs that divulged 59 transcripts consisting of 57 genes including 32 LOC symbol genes and 24 known genes. Table 1 detailed the known genes that have role in keratinization pathway, hair development or morphogenesis, fiber diameter etc.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2021 doi:10.20944/preprints202103.0187.v1

Table 1. Fold Changes of DEGs responsible for HFs in both breeds

Transcript ID	Genes	Gene names	LogFC	Pashmina TMM counts	Barbari TMM counts
XM_005692310.3	DMKN	Dermokine	-18.20295068	944.451	0
XM_013969000.2	KAP8	Keratin 8	-13.45105267	3176.066	0.29
NM_001285774.1	KRTAP3-1	Keratin associated protein 3-1	-11.60625424	5065.256	1.697
XM_013971229.2	KRTDAP	Keratinocyte differentiation associated protein	-13.34343458	2123.409	0.216
NM_001285767.1	KRTAP11-1	Keratin associated protein 11-1	-11.39093149	2758.346	0.988
XM_018047719.1	KRT6A	keratin 6A	-11.10724961	1619.46	0.633
XM_018047718.1	KRT5	Keratin 5	-10.78919982	12527.875	6.214
XM_005693818.3	KRT14	Keratin 14	-10.73352246	13575.964	7.116
XM_013977322.1	SPRR4	Small proline rick protein 4	-12.11735104	838.589	0.178
XM_018046116.1	TCHH	Trichohyalin	-10.43801428	518.08	0.321
XM_018063392.1	KRTDAP	Keratinocyte differentiation associated protein	-11.3513932	2312.868	1.045
XM_005677520.2	PRR9	Proline rich 9	-10.27471289	2590.013	1.928
XM_018047715.1	KRT1	Keratin 1	-9.990481309	8496.427	7.334
XM_005676807.3	SFN	Stratifin	-10.15748235	961.992	0.765
XM_018065044.1	KRT10	Keratin 10	-9.733969775	7700.574	7.99
XM_018047714.1	KRT77	Keratin 77	-9.751593306	799.402	0.802
NM_001285719.1	KRT27	Keratin 27	-9.574450764	2504.223	2.952
NM_001285766.1	KRT25	Keratin 25	-9.463088883	2720.404	3.451
XM_005675022.3	CSTA	Cystatin-A	-10.93663814	715.197	0.397
XM_018046124.1	LOR	Loricin	-9.608742566	641.14	0.747
XM_005675653.3	DMKN	Dermokine	-8.890178133	695.94	1.296
XM_005692309.3	CST6	Cystatin-6	-7.796853214	18161.91	84.967
XM_018043270.1	ANXA2	Annexin A2	-3.618405089	1732.492	127.817
XM_018054288.1	S100A11	S100 calcium binding protein A11	-3.280579422	1092.421	118.984
XM_005677583.3	PPIA	peptidylprolyl isomerase A	-2.607281603	1246.828	201.647
XM_018042040.1	SLC25A4	solute carrier family 25 member 4	6.799276849	27.689	2777.476

Table 3 detailed the identified LOC symbol genes that are differentially expressed. These are involved in keratinization pathway, keratinocyte differentiation, follicle transition from catagen to telogen phase while rest of LOC symbol genes are uncharacterized and thus are not mentioned and counted here.

Table 2. Comprehensive list of differentially expressed LOC genes in both breed.

Transcript ID	Genes	Gene names	LogFC	Pashmina TMM counts	Barbari TMM counts
XM_018065087.1	LOC108638296	keratin, high-sulfur matrix protein, B2D-like	-13.09954778	1197.751	0.136
XM_005693797.3	LOC102170546	keratin-associated protein 3-3	-13.65637106	1377.916	0.097
XM_018047728.1	LOC102183211	keratin, type II cuticular Hb1-like	-11.79556929	1148.294	0.271
XM_005674695.3	LOC100861181	keratin-associated protein 7-1	-11.93109531	3225.366	0.826
XM_005674699.3	LOC100860930	keratin-associated protein 6-1	-13.53441427	1517.974	0.121
XM_005693813.3	LOC102176457	keratin, type I microfibrillar, 47.6 kDa-like	-11.50435984	1664.06	0.511
XM_018065058.1	LOC102176726	Keratin, type I microfibrillar 48 kDa, 8C-1	-11.33365608	1668.517	0.576
XM_005677560.3	LOC102183026	cornifin-A-like	-11.3070629	3227.395	1.24
XM_018046520.1	LOC102187909	flaggirin-2	-11.4889734	608.103	0.192
XM_018065059.1	LOC102179881	keratin, type I microfibrillar, 47.6 kDa	-11.15974894	1331.023	0.518
XM_018065052.1	LOC102179515	keratin, type I cytoskeletal 15	-11.09359986	1096.238	0.448
XM_005674706.2	LOC102182256	keratin-associated protein 26-1	-12.2732842	779.321	0.144
XM_018065090.1	LOC102177561	keratin, high-sulfur matrix protein, IIIA3-like	-11.81026409	1090.782	0.268
NM_001285717.1	LOC100861381	Hair acidic keratin 1	-10.90665807	1287.425	0.602
XM_018065055.1	LOC102177275	keratin, type I cytoskeletal 17	-10.65089629	3005.338	1.669
XM_018053338.1	LOC108636735	keratin-associated protein 6-1-like	-12.79359006	993.626	0.135
XM_018065089.1	LOC108638297	keratin, high-sulfur matrix protein, B2C	-11.21869324	1432.433	0.554
XM_018065091.1	LOC108638298	keratin, high-sulfur matrix protein, IIIA3-like	-12.73755675	764.901	0.132
XM_018047726.1	LOC102183766	keratin, type II cuticular Hb1	-10.78463873	855.636	0.433
NM_001285770.1	LOC100861175	keratin associated protein 13.1	-11.3240876	1700.814	0.692
XM_013962365.2	LOC102183026	cornifin-A-like	-15.9313294	577.511	0
XM_018065093.1	LOC108638300	keratin, high-sulfur matrix protein, IIIA3-like	-11.42757254	865.616	0.308
XM_005674703.3	LOC102181431	keratin-associated protein 13-1	-11.66937062	516.926	0.146
XM_018065079.1	LOC108638291	keratin, high-sulfur matrix protein, B2D-like	-11.0776741	599.149	0.287
XM_018047727.1	LOC102185436	keratin, type II microfibrillar, component 7C	-10.0443469	1290.057	1.09

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2021 doi:10.20944/preprints202103.0187.v1

XM_005693796.3	LOC102170264	keratin, high sulfur matrix protein, IIIB3	-10.21592398	1079.87	0.905
XM_005688097.3	LOC102175311	calmodulin	-9.90827306	1082.197	1.086
XM_018047717.1	LOC102177231	keratin, type II cytoskeletal 71	-9.108103679	2887.431	4.589
XM_005679937.3	LOC102184223	keratin, type II microfibrillar, component 5	-8.432753572	539.167	1.367
XM_005692259.2	LOC102169125	major allergen I polypeptide chain 2	-8.304047277	1587.193	5.515
XM_018062299.1	LOC106503120	major allergen I polypeptide chain 1	-7.835300928	880.146	4.487
XM_005675654.3	LOC102186806	cystatin-B	-3.575714584	1002.589	87.747

* [14] refers to the gene names included in this table.

4. Discussion

To investigate the genes that are responsible for wool/cashmere production, we performed the differential expression analysis on the transcriptome profile of skin samples from two goat breeds, Pashmina and Barbari. Pashmina goat provides protection against the cold harsh environment via its luxurious fiber while Barbari goat breed inhabits the hot arid environment and are valued for meat and milk yield [4, 5]. The top most DETs were then functionally annotated to get molecular insight of its corresponding genes.

Gene ontology terms of highly expressed HF transcripts in Pashmina and Barbari goat breed revealed that most of the genes are related to keratin differentiation and are involved in keratinization pathways. Several scientific evidence is available and many studies report the genes concurrent to our results. It is previously suggested that keratins (KRT) and Keratin associated proteins (KRTAPs) are the main structural component of hair fibers and sheath [15]. The interaction of fibrous KRTs and matrix KRTAPs are the foundation of cornified skin appendages such as hairs, horns. An analysis on Chinese Tan sheep's skin transcriptome revealed the potential genes for curly fleece and hair/wool formation including *KRT27, PRR9* and *CST6* that ranged from 3,353 to 7,813 FPKM reads [16]. Another study reported the high expression of *SFN* gene in Changthangi goat [15]. This Stratifin gene regulates the epithelial keratinization by directly regulating the cell cycle. It has been best selected for human dermal papilla cells (DPCs) as well. Similarly, they analyzed *CSTA* and *KRT77* gene having fold change values of 4.63 and 3.46 respectively that helps in keratinocyte differentiation. A recent analysis on differential expression in Canine anagen, telogen HFs and interfollicular epidermis (IFE) [13] presented that hair follicle associated structural genes (*TCHH, KRT25*) have high expression levels 3,997.42 and 10,456.27 during anagen phase. Similarly, *KRT1, KRT5, KRT10, KRT14* and *LOR* are epidermal structure-associated genes and have high TPM counts (21,712.07, 12,733.06, 33,847.68, 29,472.32 and 7229.10) in IFE samples. RNA-seq analyses on *Ovis aries* revealed *S10011A* gene association with hair/wool development in them [17]. Small proline-rich protein 4 that has TMM count > 500 in our Pashmina breed is a class of cornified envelope and are found around 300 kb of epidermal differentiation complex. An analysis on sheep found its expression in HFs, epidermis and capillaries [18].

Importantly, the LOC symbol genes were also targeted in this study with known functions. A little is known about them but they have been characterized and are previously reported. A study elucidated the role of these LOC symbol genes in transition of HFs from catagen to telogen phase/telogen induction and their differential expression in either phase with FPKM, log2(FC) and P-values [19]. Other uncharacterized proteins such as LOC108635855 encoded by XR_001917982.1, LOC108635656 (XR_001917911.1) and many others were also divulged in Pashmina breed with high expression counts but they are not included here.

In conclusion, Pashmina and Barbari breed's skin transcriptome sample were subjected to differential expression analysis to search out potential genes related to HFs. Differential expression patterns above a significant normalized count in the two breeds under consideration suggested the involvement of those genes in keratinization pathways, keratinocyte differentiation, sulfur matrix proteins, dermal papilla cells, HFs proliferation, hair curvature, wool fiber diameter, hair transition, hair shaft differentiation and hair follicle keratinization. All DETs identified in this study can be biologically evaluated in future for pashmina/cashmere fiber evolvement in animals and provides some possible clues for further investigations on gene enrichment analysis of hair development genes.

References

1. Fan Y, Wu R, Qiao X, Zhang Y, Wang R, Su R, et al. Hair follicle transcriptome profiles during the transition from anagen to catagen in Cashmere goat (Capra hircus). Genet Mol Res 2015; 14 (4): 17904-15.
2. Galbraith H. Fundamental hair follicle biology and fine fiber production in animals. animal 2010; 4 (9): 1490-509.
3. Ryder ML. The use of goat hair: an introductory historical review: Anthropozoologica; 1993.
4. Malik I, Wani F. Pashmina goat farming in cold arid desert of Ladakh: A geographical study of Changthang region. International Journal of Advanced Scientific Research and Management 2017; 2: 1-10.
5. Porter V, Alderson L, Hall SJ, Sponenberg DP. Mason's World Encyclopedia of Livestock Breeds and Breeding, 2 Volume Pack: Cabi; 2016.
6. shutterstock-Pashmina goat images. Available from: https://www.shutterstock.com/search/pashmina+goat.
7. Domestic Animal Breeds. Available from: https://domesticanimalbreeds.com/barbari-goat-breed-everything-you-need-to-know/.
8. Brown J, Pirrung M, McCue LA. FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. Bioinformatics 2017; 33 (19): 3137-9.
9. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30 (15): 2114-20.
10. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nature methods 2017; 14 (4): 417-9.
11. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols 2013; 8 (8): 1494-512.
12. Henschel R, Lieber M, Wu L-S, Nista PM, Haas BJ, LeDuc RD, editors. Trinity RNA-Seq assembler performance optimization. Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment: Bridging from the eXtreme to the campus and beyond; 2012.
13. Wiener DJ, Groch KR, Brunner MA, Leeb T, Jagannathan V, Welle MM. Transcriptome Profiling and Differential Gene Expression in Canine Microdissected Anagen and Telogen Hair Follicles and Interfollicular Epidermis. Genes 2020; 11 (8): 884.
14. National Library of Medicine-Center for Biotechnology Information Search database. Available from: https://www.ncbi.nlm.nih.gov/search/.
15. Ahlawat S, Arora R, Sharma R, Sharma U, Kaur M, Kumar A, et al. Skin transcriptome profiling of Changthangi goats highlights the relevance of genes involved in Pashmina production. Scientific reports 2020; 10 (1): 1-10.
16. Kang X, Liu G, Liu Y, Xu Q, Zhang M, Fang M. Transcriptome profile at different physiological stages reveals potential mode for curly fleece in Chinese tan sheep. PloS one 2013; 8 (8): e71763.
17. Zlobin AS, Volkova NA, Borodin PM, Aksenovich TI, Tsepilov YA. Recent advances in understanding genetic variants associated with growth, carcass and meat productivity traits in sheep (Ovis aries): an update. Archives animal breeding 2019; 62 (2): 579-83.
18. Jinshan Z, Huaiyuan Q, Jingjing X, Liu N, Rongwei H, Perez-Campo F, et al. Discovery of genes and proteins possibly regulating mean wool fiber diameter using cDNA microarray and proteomic approaches. Scientific Reports (Nature Publisher Group) 2020; 10 (1).

 doi:10.20944/preprints202103.0187.v1

19. Follicle transition from Anagen to Catagen / Catagen Induction. Available from: http://lms.snu.edu.in/pidb/pashmina_cycling/ana_cat.php.