



Changthangi Pashmina Goat Genome: Sequencing, Assembly, and Annotation

Basharat Bhat^{1*}, Nazir A. Ganai¹, Ashutosh Singh², Rakeeb Mir³, Syed Mudasir Ahmad¹, Sajad Majeed Zargar⁴ and Firdose Malik⁵

¹ Division of Animal Biotechnology, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar, India, ² Department of Life Science, Shiv Nadar University, Greater Noida, India, ³ Department of Biotechnology, Baba Ghulam Shah Badshah University, Rajouri, India, ⁴ Division of Plant Biotechnology, Sher-e-Kashmir University of Agricultural Sciences Technology of Kashmir, Srinagar, India, ⁵ Division of Temperate Sericulture, Sher-e-Kashmir University of Agricultural Sciences and Technology of Kashmir, Srinagar, India

OPEN ACCESS

Edited by:

Zhe Zhang,
South China Agricultural University,
China

Reviewed by:

Lei Zhou,
China Agricultural University, China
Priscila S. N. de Oliveira,
Federal University of São Carlos,
Brazil

*Correspondence:

Basharat Bhat
bb284@snu.edu.in

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 14 April 2021

Accepted: 22 June 2021

Published: 20 July 2021

Citation:

Bhat B, Ganai NA, Singh A, Mir F,
Ahmad SM, Majeed Zargar S and
Malik F (2021) Changthangi Pashmina
Goat Genome: Sequencing,
Assembly, and Annotation.
Front. Genet. 12:695178.
doi: 10.3389/fgene.2021.695178

Pashmina goats produce the world's finest and the most costly animal fiber (Pashmina) with an average fineness of 11–13 microns and have more evolved mechanisms than any known goat breed around the globe. Despite the repute of Pashmina goat for producing the finest and most sought-after animal fiber, meager information is available in the public domain about Pashmina genomics and transcriptomics. Here we present a 2.94 GB genome sequence from a male Changthangi white Pashmina goat. We generated 294.8 GB (>100X coverage) of the whole-genome sequence using the Illumina HiSeq 2500 sequencer. All cleaned reads were mapped to the goat reference genome (2,922,813,246 bp) which covers 97.84% of the genome. The Unaligned reads were used for *de novo* assembly resulting in a total of 882 MB non-reference contigs. *De novo* assembly analysis presented in this study provides important insight into the adaptation of Pashmina goats to cold stress and helps enhance our understanding of this complex phenomenon. A comparison of the Pashmina goat genome with a wild goat genome revealed a total of 2,823 high impact single nucleotide variations and small insertions and deletions, which may be associated with the evolution of Pashmina goats. The Pashmina goat genome sequence provided in this study may improve our understanding of complex traits found in Pashmina goats, such as annual fiber cycling, defense mechanism against hypoxic, survival secret in extremely cold conditions, and adaptation to a sparse diet. In addition, the genes identified from *de novo* assembly could be utilized in differentiating Pashmina fiber from other fibers to avoid falsification at marketing practices.

Keywords: Pashmina goat, whole genome sequence, goat SNP, Pashmina fiber, cold stress

BACKGROUND

The domestic goat (*Capra hircus*) is an Asian animal distributed across all ecologies ranging from cold arid to hot humid. It serves as an important source of meat, milk, skin, fiber, and manure. Goats exhibit several traits and diseases that are similar to those of humans, which is the reason why goats are being extensively used as an animal model for biomedical research (Fulton et al., 1994). Modern domestic goats have been domesticated from *Capra aegagrus*

(Dong et al., 2015; Sheikh et al., 2016). Over centuries, different goat breeds have been established through evolution and genetic selection, exhibiting traits such as (1) coat color variations, (2) adaptability to different climatic conditions, (3) change in size, (4) development of fine fiber (Dong et al., 2015).

An important breed of extremely cold Temperate Himalayan region of India is Changthangi Pashmina goat which produces world's most sought-after natural animal fiber, **Pashmina** (Bhat B. et al., 2019). These goats are also referred as Pashmina goats or Cashmere goats. Pashmina goats are double hair coated with an outer coat of long coarse guard hair and an inner coat of shorter fine Pashmina fiber. The guard hair develops from primary hair follicles (PHFs) and Pashmina from secondary hair follicles (SHF) (Ansari-Renani et al., 2011). These goats survive in extreme climatic conditions [$+35^{\circ}\text{C}$ (short summer) and -40°C (long winter)] at an altitude of 4,000–5,500 m above mean sea level (AMSL) under cold and hypoxic conditions. To adapt to the harsh climatic conditions like cold, hot, arid, dry, and poor grazing conditions, these goats have acquired many special abilities and attributes. The growth of a double hair coat in Pashmina goats may be one of the mechanisms of protection from cold, and the result of the response triggered by the thermoregulatory center of the brain.

It is believed that the Pashmina goats were originated in the Himalayan ranges and migrated to different regions of central Asia ranging from China, Mongolia, Iran, Russia, Afghanistan, and India (Ryder, 1966). China is the leading Pashmina producers followed by Mongolia, Iran, New Zealand, and Britain. Even though India—Kashmir contributes $<1\%$ of the world's Pashmina production, because of its fineness and quality it holds a unique position in the world's Pashmina trade (Shakyawar et al., 2013). Globally Pashmina is sold under the geographical indication (GI) tract of *Cashmere*, which belongs to India—**Kashmir**. Changthangi Pashmina goats produce the world's finest and the most costly Pashmina fiber with an average fineness of 11–13 microns (Bumla et al., 2012).

Pashmina fiber is the primary source of income for the nomadic population of Ladakh (Sheikh et al., 2016). Apart from the economic value of the Pashmina goat, this goat serves as an important source of food (milk and meat), skin, and manure in the region. This goat is a safe and secure form of investment and stable means of income in the cold and arid deserts of Ladakh. These animals often provide the only practical means of utilizing vast areas of natural grasslands in the areas where crop production is uneconomical. Adapted to the harsh environmental conditions, the Changthangi goat is a unique genetic resource of the country. However, due to lack of breeding policy and population structure over a long period of time, it is suspected that inbreeding in these indigenous breeds may pose serious threat (Ganai et al., 2011).

In this study, we report the first Pashmina goat genome sequence. We generated 294.8 GB ($>100\text{X}$ coverage) raw reads from a Changthangi Pashmina goat using Illumina HiSeq 2500 sequencer. We deciphered 2.94 GB of the Pashmina goat genome, revealing 26,687 protein-coding genes, 842 miRNAs, 188 lncRNAs, and 1879 snRNAs. We compared the Pashmina

goat genome with the wild goat genome revealed important genetic variants related to the evolution of Pashmina goats.

METHODS

Ethics Statement

This study was approved by the Institutional Animal Welfare and Ethics Committee of the Sher-e-Kashmir University of Agricultural Science and Technology of Kashmir (SKUAST-K). All experiments and methods were performed in accordance with relevant guidelines and regulations. Experimental goats were housed in SKUAST-Kashmir goat farm located in the northern Himalayas (Satakna, Ladakh) at an altitude of 5,000 m AMSL.

Sampling, Genome Sequencing, and Assembly

Genomic DNA was extracted from the blood of a 26 months old male Changthangi Pashmina goat. A whole-genome sequencing (WGS) library was prepared with the Illumina-compatible NEXTFlex Rapid DNA sequencing kit (BIOO Scientific, Austin, Texas, U.S.A.) as per the manufacturer's guidelines. Genomic DNA was sheared using Covaris S2 sonicator (Covaris, Woburn, Massachusetts, USA) to generate approximate fragment size distribution from 200 to 400 bp. Here, the fragment size distribution was checked on Agilent Bioanalyzer and subsequently purified using Hiprep magnetic beads (Magbio). Purified fragments were end-repaired, adenylated, and ligated to Illumina multiplex barcode adaptors as per NEXTFlex Rapid DNA sequencing kit protocol. Adapter-ligated DNA was purified and size selected using Hiprep beads. Resultant fragments were amplified for four cycles of PCR using Illumina-compatible primers provided in the NEXTFlex Rapid DNA sequencing kit. The final PCR product (i.e., sequencing library) was purified with Hiprep beads, followed by a library-quality control check. Illumina compatible sequencing library was initially quantified by Qubit fluorometer (Thermo Fisher Scientific, MA, USA) and its fragment size distribution was analyzed on Agilent TapeStation. Sequencing and base calling were performed according to the Illumina recommendations.

Using the Illumina HiSeq 2500 platform, a total of 294.8 GB (150 bp reads) high-quality data (100X coverage of the estimated genome size) were generated. The raw reads were pre-processed to remove the adapter sequences, low-quality reads, and low-quality bases filtration toward 3'- end using cutadapt program v3.1 (Martin, 2011). Filtered reads were mapped to *C. hircus* reference genome assembly ARS1 downloaded from National Center for Biotechnology Information (NCBI) using bowtie2 v2.4.2 (Langmead and Salzberg, 2012). The Unmapped paired-end reads to reference were used for contig assembly using Abyss *de novo* assembler v2.0 (Jackman et al., 2017). The complete assembly was obtained by merging the reference-assisted and *de novo* assembled consensus sequences. The completeness and correctness of genome-assembly were evaluated with the Benchmarking Universal Single-Copy Orthologs (BUSCO) program v5.1.2 (Simão et al., 2015).

Genome Annotation

The repetitive elements were identified in the final assembled draft genome using RepeatMasker v4.0.9 (Tarailo-Graovac and Chen, 2009). The transfer RNAs (tRNAs) were predicted using the tRNAScan-SE program v2 (Lowe and Chan, 2016), with default parameters for eukaryotic genomes. Ribosomal RNAs (rRNAs) were collected based on homology information from *Homo sapiens*, *Bos taurus*, *Bos mutus* and *Ovis aris* rRNAs using the BlastN program v2.11.0+ (Ye et al., 2006). Other non-coding RNA were predicted from the assembled genome with Infernal v1.1.2 (Nawrocki and Eddy, 2013) using the Rfam database v13 (Griffiths-Jones et al., 2003), with an *E*-value cutoff of 0.001. Only non-truncated CM hits detected in the first pass of the pipeline with an inclusion threshold of ≤ 0.001 and score ≥ 45 were reported. Overlapping hits with lower score hits were filtered out.

Protein-coding gene identification was performed using Augustus (Stanke and Waack, 2003) and Maker v2.31.10 (Cantarel et al., 2008) programs. A total of 26,687, protein-coding genes were identified using a 2-pass schema. The first round of gene prediction was carried out using AUGUSTUS with *H. sapiens*, *B. taurus*, *B. mutus*, and *O. aris* as reference models for the hard-masked assembled genome. The predicted gene model from AUGUSTUS was taken as input along with transcripts and the available gene model of *C. hircus* for the second round of gene prediction in the MAKER tool. The MAKER tool provides evidence-based gene modeling. After performing the gene prediction through MAKER, we selected the following filtering criteria for the selection of the final gene model.

- Predicted gene should have a start and stop codon.
- Length of the gene should be >300 bps.
- Gene does not contain more than 1% of N's.

Predicted proteins were annotated using homology-based prediction by searching against mammalian gene sequences utilizing the BLAST program.

- At first the protein sequences were similarity searched against the UniProt Bovidae (91,402) and Caprine family (91,402) protein database using BLASTP program with an *e*-value of 0.00001 for gene ontology (GO) and annotation (Consortium, 2014; Shaik et al., 2019).
- The unmapped genes were homology searched against NCBI *C. hircus* (GCF-001704415.1) proteins (42,687) using the BLASTP program with an *E*-value cut-off of 0.00001.
- The unannotated sequences were further annotated against NCBI non-redundant (NR) database and Pfam database with default parameters. A total of 90% of predicted genes were annotated.

The predicted proteins were uploaded to the KEGG (Kyoto Encyclopedia of Genes and Genomes)—KAAS (KEGG Automatic Annotation Server) (Moriya et al., 2007) server for pathway identification using *B. taurus*, *B. mutus*, *C. hircus*, *Bos indicus* and *Ovis aries* as reference organisms.

Variant Detection and Annotation

Variant identification was done using the Genome Analysis Toolkit (GATK) v4.0.7.0 (McKenna et al., 2010) applied on

bowtie output. As recommended by the GATK practice, Picard tools v2.25.1 (Pic, 2019) were used to add read group information, mark duplicates, and index a sorted BAM file. As per the GATK pipeline following steps were performed for variant calling from genomic data; split and trim to reassign mapping quality, local realignment, InDel realignment, and BaseQualityScore recalibration. For variant discovery, HaplotypeCaller and GenotypeGVCFs were used followed by filtering variants. Identified variants were divided into different functional classes based on their genomic distribution. *C. hircus* gene annotation file was used to determine if a SNP is located within mRNA start and end positions (genic), CDS, 5'UTR, or 3'UTR. The variants identified were filtered for the coverage of more than 20 (more than 20 reads covering the position) and quality of 30 (Phred scaled quality of more than 30) to include high confidence variants. Variants were annotated using SnpEff program v4.3T (Cingolani et al., 2012). To further evaluate the biological significance of the genes with high impact variations, the pathway analysis were performed using KEGG-KAAS server (Moriya et al., 2007; Shaik et al., 2019).

RESULTS AND DISCUSSION

Genome Sequence and Annotation

We sequenced genome DNA from a 26-month-old male Changthangi Pashmina goat. High-quality DNA extracted from blood was used to construct paired-end sequencing libraries. Using the Illumina HiSeq 2500 platform, a total of 294.8 GB raw data were generated. Out of the total cleaned reads (965,981,627 reads), 80% reads were mapped to the reference genome, covering 97.84% genome. To identify genes specific to Pashmina goats, *de novo* assembly of the unaligned reads were performed. The final genome assembly was obtained by merging the reference-assisted and *de novo* assembled consensus sequences. The final gap closer was executed using GapCloser program (Simpson and Durbin, 2012) with PE-LI libraries which generated a final draft genome of 2.94 GB with an N50 value of 102582650 (**Supplementary Table 1**). Calculation of the completeness of genome assembly using BUSCO program suggested 99.3% of the assembled genome was complete (85.6% complete and single-copy, 13.7% complete and duplicated) remaining 0.7% of genome were fragmented.

A total of 18,656 contigs (**Supplementary Table 1**) were assembled from *de novo* assembly, which may be specific to the Pashmina goat breed. To interpret the biological implication of sequences extracted from *de novo* assembly, all sequences were mapped to the KEGG database using KASS servers (using cut-off FDR corrected *q*-value <0.01). KEGG pathways analysis deduced that the sequences are predicted to be involved in metabolism pathways like amino-acids (cysteine, methionine, threonine, serine, and glycine), nucleotides metabolism, fatty acid metabolism; and signaling pathways like Calcium signaling, Notch signaling, cAMP signaling, Rap1, and PI3K-Akt signaling. Amino-acids and energy play a vital role in continuous Pashmina fiber growth and follicle initiation. Enrichment analysis suggests the potential role of cAMP, PI3K-Akt, and calcium signaling in regulating cold stress responses in Pashmina goats. Further

TABLE 1 | Repeat element statistics from Pashmina goat genome assembly.

	No. of elements	Length occupied (bp)	%age in genome
SINEs	2,045,717	296,111,061	10.07
MIRs	391,052	55,830,025	1.90
LINEs	1,316,415	746,218,268	25.37
LINE1	567,547	331,707,256	11.28
LINE2	246,431	61,914,088	2.10
L3/CR1	33,442	6,797,367	0.23
RTE	467,971	345,644,042	11.75
LTR elements	349,139	102,247,773	3.48
ERVL	72,022	28,212,673	0.96
ERVL-MaLRs	120,773	39,051,305	1.33
ERV_classI	39,090	13,144,887	0.45
ERV_classII	101,109	18,070,827	0.61
DNA elements:	278,549	55,919,821	1.90
hAT-Charlie	160,945	29,848,276	1.01
ToMar-Tigger	43,361	11,562,004	0.39
Unclassified	4,534	803,290	0.03
Small RNA	250,244	39,202,683	1.33
Satellites	137,840	149,989,331	5.10
Simple repeats	823,810	89,185,768	3.03
Low complexity:	105,341	8,887,201	0.30
Total repeats:		1,450,391,539 bp	49.30

transcriptomic or proteomic studies are required to identify specific genes and pathways mediating cold stress response in Pashmina goats.

A total of 1,450,391,539 bp (49.30% of the genome) of the Pashmina goat genome contain repetitive elements (**Table 1**) which is in accordance with the earlier studies of cattle (*B. taurus*) genome (50%) (Sequencing et al., 2012). In our study, we identified that long interspersed elements cover 25.37% of the whole genome, which is equitable to that of cattle (23%), human (21%), and horse (20%) (Sequencing et al., 2012). The percentage of short interspersed elements is 10.07% which is lower than that of humans (13%) and cattle's (18%) and greater than that of mice (8%) and horses (7%). 40.83% of the total Pashmina goat genome contains different types of interspersed repeats. Also, 8.44% of the total Pashmina genome contains small RNAs, microsatellites, and simple repeats; which should be useful in quantitative trait locus mapping or marker-assisted breeding in modulating economically important traits in specialty animal fiber like Pashmina.

A multi-way approach was carried out for protein-coding gene prediction from the final assembled genome. We have used AUGUSTUS and MAKER tools for gene prediction. The first round of gene prediction was carried out using AUGUSTUS with humans as a reference model for the hard-masked assembled genome. The predicted gene model from AUGUSTUS was taken as input along with transcripts and the available gene model of *C. hircus* for the second round of gene prediction in the MAKER tool. A total of 26,687 protein-coding genes were

TABLE 2 | Pathways affected by high impact SNPs and InDels in Pashmina goat genome.

Pathways	Genes
Cytokine-cytokine receptor interaction	BMP3, FAS, LEPR, CCR3, IL3RA, TNFRSF11A, IFNGR2, IL1A, CXCL17
ECM-receptor interaction	COL6A5, TNN, ITGB5, COL4A1
Notch signaling pathway	PTCRA, DLL3
MAPK signaling pathway	RASGRP2, CACNA1I, FAS, MAPK11, IL1A, CACNA1S
AMPK signaling pathway	ACACB, CPT1B, LEPR
Calcium signaling pathway	CACNA1I, ATP2B3, CACNA1S, MYLK
PPAR signaling pathway	CPT1B, PLIN2
mTOR signaling pathway	GRB10, ATP6V1G1, WDR59
Neuroactive ligand-receptor interaction	GRID1, CHRNB3, TAAR8, LEPR, RLN3, C3
Circadian entrainment	PER3, CACNA1I
TNF signaling pathway	FAS, MAPK11
Neurotrophin signaling pathway	IRAK2, MAPK11
Jak-STAT signaling pathway	IL3RA, IFNGR2, LEPR
VEGF signaling pathway	MAPK11
Wnt signaling pathway	APC2, ROR2

identified by combining reference and *de novo* assembled genome (**Supplementary Table 2**).

Variant Detection

GATK workflow identified 14,270,872 SNVs. The transitions to transversions (Ts/Tv) ratio was 2.35 and homozygosity (4,965,247) and heterozygosity (9,306,235) percentage were 34.79 and 65.21%, respectively, both in accordance with the earlier studies of mammalian genome analysis. Of the SNPs, 35% are intronic, 63% intergenic, 0.6% were in 3' and 5' UTR regions. 206,385 SNVs were identified in coding regions with 96,092 as synonymous and 50,355 as non-synonymous variants. Distributions of SNVs and InDels in different chromosomes of the Pashmina goat genome are presented in **Supplementary Figure 1**. A total of 1,423 high-impact SNVs (start lost-103, stop gained-790, and stop lost-530) were also identified which may be associated with the evolution of the Pashmina goat (**Supplementary Table 3**). We identified 1,126,239 InDels, which consisted of 484,468 insertions and 641,871 deletions. The length distribution of identified InDels ranged from -28 to +28 bp and homozygosity (621,032) and heterozygosity (505,207) percentages were 55 and 45%, respectively. Of all InDels, 61.7 and 37% were in intergenic and intronic regions, respectively. 0.01% InDels were divided between UTR regions of genes. A total of 1,400 high-impact InDels were identified which contained 1,176 frameshift, 20 and 5 stop lost, and stop gained InDels, respectively (**Supplementary Table 4**).

Functional analysis suggest genes with high impact SNPs were involved mainly in signaling pathway like Notch, MAPK, AMPK,

Calcium, PPAR, mTOR, TNF, Neurotrophin, Jak-STAT, VEGF, and Wnt (Table 2). These signaling pathways are involved in diverse biological processes and their specific role in Pashmina goats needs to be further elucidated.

Genome-Wide Identification of lncRNAs

Long non-coding RNAs (lncRNAs) are emerging as key regulators for a myriad of biological processes. In the current lncRNA databases, most of the identified lncRNAs are derived from mice and humans (Volders et al., 2014). Several recent studies on *B. taurus* (Huang et al., 2012), *Gallus gallus* (Li et al., 2012), and *Sus scrofa* (Tang et al., 2017) have increased the information pool of lncRNAs. However, meager information is available on lncRNAs for *Caprines* species. To our knowledge, this is the first study to report genome-wide identification of lncRNAs from any goat species. In this study, we systematically identified a comprehensive list of lncRNA (Supplementary Table 5) identified from the Pashmina goat genome and their role in regulating different biological processes.

Recent studies suggest the role of lncRNAs in modulating immunity; in our study, we identified three major classes of immune regulatory lncRNAs. (1) HOX antisense intergenic RNA myeloid 1 (*HOTAIRM1*) is known to be associated with granulocytes, and is a key regulator in myeloid transcriptional regulation, by modulating *HOXA* expression in cis-configuration (Mumtaz et al., 2017). (2) *HOX* transcript antisense RNA (*HOTAIR*) is an oncogenic long non-coding RNA overexpressed in various carcinomas. It recruits various chromatin-modifying enzymes and regulates gene silencing (Bhan and Mandal, 2015). (3) Nuclear Enriched Abundant Transcript 1 (*NEAT1*) known to the innate immune response to viral infection (Mumtaz et al., 2017).

lncRNA also plays a critical role in mediating gene expression during different developmental and differentiation processes. We identified six classes of lncRNA, which are known for mediating developmental traits in different animal models. (1) *KCNQ1* overlapping transcript 1 (*KCNQ1OT1*) plays crucial role in the transcriptional silencing of the *KCNQ1* locus by regulating histone methylation. *KCNQ1OT1* gene inactivation results multiple growth defects in mice (Fatica and Bozzoni, 2014). (2) *HOXA* transcript at the distal tip (*HOTTIP*) knockdown of these genes results in retinal cell development in mice and altered limb morphology in chickens (Fatica and Bozzoni, 2014). (3) *H19* has a role in cell proliferation; *H19* limits body growth by regulating *IGF2* expression. Mice with the loss of *H19* function show an overgrowth phenotype. (4) Metastasis associated lung adenocarcinoma transcript 1 (*MALAT1*) is majorly involved in neural development, in cultured mice hippocampal neurons *MALAT1* knockdown shows decreased dendritic growth and decreased synaptic density. (5) X-inactive specific transcript (*XIST*) acts as a regulator of X-chromosome inactivating in mammals. *XIST* deletion in mice causes a loss of X-chromosome inactivation and female-specific lethality. (6) Myocardial infarction-associated transcript (*MIAT*) is associated with retinal cell fate and myocardial infarction in humans. Other identified ncRNAs (rRNA, snRNA, tRNA,

lncRNAs, and miRNAs) including FASTA sequences are listed in Supplementary Table 5.

Genome-Wide Identification of miRNAs

MicroRNAs (miRNAs) are small (22 nt long), non-coding regulatory RNAs, which can evoke post-translational repression of mRNA levels of target genes. miRNAs are not only involved in transcriptional/post-transcriptional regulation but also regulate response to environmental stresses. In this study, using high-throughput genome sequencing we identified a total of 338 mature miRNAs in the Changthangi Pashmina goat genome. The length of mature miRNAs varies from 17 to 25 nucleotides with an average of 21 nucleotides. The major class of miRNAs 91% falls within the range of 20–23 nucleotides (Supplementary Figure 2). The highest number of miRNAs are observed in the mir-1255 family followed by mir-1302, mir-544, mir-692, mir-154, mir-562, mir-663, mir-684, mir-650, and let-7.

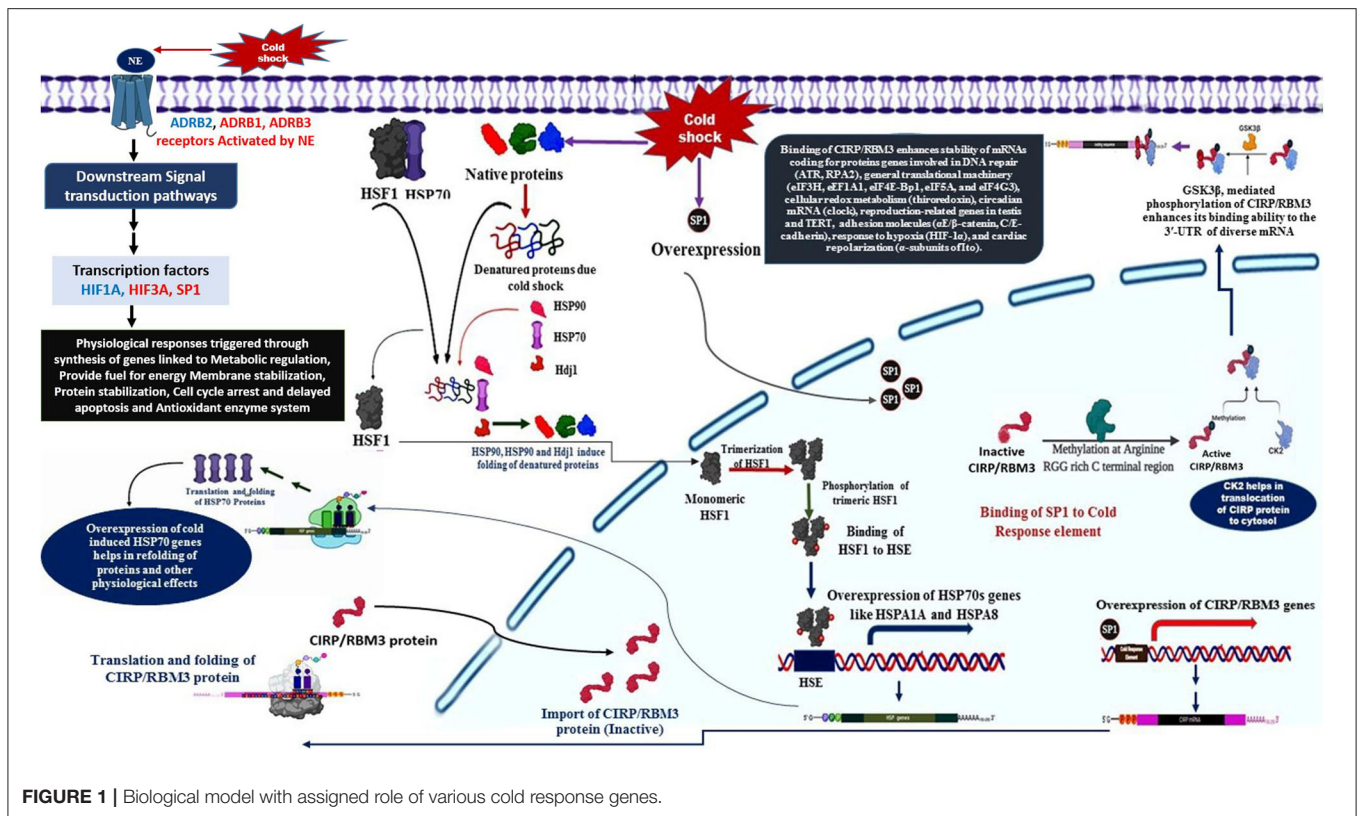
The miR-1255 commonly express in exosome and regulates TGF- β signaling pathway by interacting with *SMAD4* gene (Xin et al., 2020). MicroRNA-214, miR-31 and miR-218 controls skin and hair follicle development by modulating Wnt signaling and β -catenin signaling (Mardaryev et al., 2010; Ahmed et al., 2014; Hu et al., 2020; Bhat et al., 2021). miR-128, miR-148, and miR-301 regulate key genes involved in cholesterol-lipoprotein trafficking (Wagschal et al., 2015). miR-187 family regulates key genes (*TGFBI*, *THBS1*, *ACVR18* and *BMP88*) in TGF-beta signaling pathway (Miao et al., 2016; Bhat S. A. et al., 2019). Complete annotation of miRNAs from Pashmina goat genome were listed in Supplementary Table 5.

Cold Stress Response in Pashmina Goat: A Possible Regime

Cold tolerance in Pashmina goats is an extremely complex phenomenon that is influenced by a large number of physiological, biochemical, and endocrine factors. However, the molecular mechanisms underlying the adaptation to cold stress remain largely unknown. The changes that produce a cold-hardy phenotype under cold tolerance situation would involve acclimation and acclimatization in response to low temperatures, rapid cold-hardening, and cold-induced gene expression (Hansen, 2004).

The primary response to cold stress in animals is suggested to be a neuroendocrine response, which triggers the release of catecholamine hormone, usually nor-epinephrine (NE). The cold signal in the form of NE is perceived by beta-2 Adrenergic receptor, a GPCR (Bhat et al., 2017), predominantly localized to the vascular system in comparison to $\beta 1$ and $\beta 3$. This signal perception by $\beta 2AR$ is known to activate multiple intracellular signal transduction pathways that influence various molecular, biochemical, and physiological processes (Figure 1). NE signaling through $\beta 2AR$ regulates blood pressure, heart, and respiratory rate, and body temperature (Chruscinski et al., 2001).

- The cold signal in the form of NE is perceived by β -adrenergic receptors, such as *ADRB1*, *ADRB2*, and *ADRB3*.



- Activation of $\beta 2AR$ triggers several downstream signaling cascades (DSC).
- The DSC lead to the down-regulation of *HIF3A* and also activation of *SP1* and *HIF1A*.
- The activation of *SP1* leads to the over-expression of *CIRP* and *RBM3* genes with the help of *CK2* and *GSK3 β* (Aoki et al., 2002; Yang et al., 2006; De Leeuw et al., 2007; Sumitomo et al., 2012).
- *CIRP* and *RBM3* proteins are predominantly localized in the nucleus, but can migrate to cytoplasm upon stress condition, and acts as an RNA chaperone regulating mRNA stability through its binding signature site in the 3'-UTR of its targets, which includes genes involved in DNA repair (*ATR*, *RPA2*), cellular redox metabolism (thioredoxin), adhesion molecules ($\alpha E/\beta$ -catenin, *C/E-cadherin*), circadian mRNA (clock), reproduction-related genes in testis and *TERT*, response to hypoxia (*HIF-1 α*), general translational machinery (*eIF3H*, *eEF1A1*, *eIF4E-Bp1*, *eIF5A*, and *eIF4G3*), and cardiac repolarization (α -subunits of Ito). In addition, *CIRP* can also be secreted into extracellular space through lysosome pathway upon stimulation by *LPS* or hypoxia/reoxygenation (Xia et al., 2012).

Another class of protective agents, the heat shock proteins (HSPs), also contribute significantly to the overwintering cold tolerance (Rinehart et al., 2007). Stress factors, like cold, induces over-expression of heat shock genes responsible for the synthesis of molecular chaperones to refold the misfolded of cold-induced

proteins. Later mechanism is triggered by the stress-induced synthesis of HSFs, which bind to heat shock elements (HSE) consisting of the pentanucleotide motif 5'-nGAAAn-3' (Lis and Wu, 1993; Tissieres and Georgopoulos, 1994; Voellmy, 1994).

HSF1 and HSP70 interaction in cold-stress response

- *HSF1* is activated by unfolded proteins resulted due to cold shock, in its inactivated monomeric state *HSF1* is bound to *HSP70* (Santoro, 2000).
- The cold induced unfolded proteins are bound by molecular chaperones, such as *HSP90*, *HSP70*, and *HDJ1* (Santoro, 2000).
- The released *HSF1* is translocated to nucleus, where they undergo trimerization and phosphorylated (Santoro, 2000).
- The phosphorylated trimeric *HSF1* binds to HSE located upstream of HSP genes, resulting in transcriptional activation and synthesis of HSPs, such as *HSP70*, *HSPA1A*, and *HSPA8* genes (Banerjee et al., 2014).
- The transcripts of HSPs are shuttled to the cytosol for translation, the higher expression of HSP, in turn, regulate folding and Activation of a specific class of transcription factors called as heat shock factors (HSFs) (Åkerfelt et al., 2010).

Further, in context to the status of the translation process during cold stress condition, it is suggested that the cAMP/PKA pathway is involved in the dysregulation of translation factors like *EIF6*, *TUFM*, *EIF1AD*, *EIF2B2*, *EIF2B3*, *RPL7*, *EEF2*, *EIF3H*, and *GARS* to modulate the cold stress-responsive protein synthesis. Generally, cold shock results in protein degradation, but an

adaptive response is generated to maintain integrity as well as stability of proteins by selective expression of certain heat shock proteins (e.g., *DNAJA4*, *DNAJB12*, *DNAJC7*, *DNAJC11*, *HSPA8*, and *HSP90B1*).

Cold shock results in changes to the lipid bilayer and composition of the membrane in mammalian cells. The fatty acid and lipid metabolism especially beta-oxidation enzymes (*ACSL1*, *SLC27A1*, *ACADVL*, *HADHA*) are increased to provide fuel in the form *acetyl CoA* for energy generation as well as thermogenesis. Furthermore, HIF1A induced glycolytic enzymes like *G6PD*, *PFKFB3* are up-regulated, resulting in an increase in energy generation during the hypoxic condition. Thus, glucose metabolism and lipid metabolism are regulated in such a way to meet the requirement of energy while maintaining the stability of the lipid membrane. Cold exposure leads to increased production of reactive oxygen species (ROS) which influence HIFs activity and involves in lipid peroxidation (Quirós et al., 2016). Glutathione peroxidase (GPX), catalase (CAT), and peroxiredoxin (PRDX) are activated to neutralize the ROS effect.

Hence, the cross-talk among various signaling pathways mainly hormonal signaling (NE signaling), cAMP/PKA signaling, Src kinase-PI3K/Akt-dependent pathway, Ca²⁺ signaling, ERK1/2 signaling, p38 signaling, and ROS signaling could be responsible for the generation of a stress tolerance response in Pashmina goats by modulating the expression of specific cold stress-responsive genes. **Figure 1** illustrates the role of various cold-responsive genes.

CONCLUSIONS

Here, we report the characterization of the high altitude Pashmina goat genome for the first time. The present study on the genome and annotations may provide Pashmina breeders and other researchers with useful information regarding trait biology and their subsequent improvement. In particular, we highlight pathways that could be involved in cold-stress response and fiber cycling in Pashmina goats. The annotation of coding and non-coding genes provides, for the first time, an understanding of

the gene content in Pashmina goat, which is valuable for future studies on genes, gene structure, and functional genomics.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://www.ncbi.nlm.nih.gov/>, SRR7739516.

ETHICS STATEMENT

The animal study was reviewed and approved by Sher-e-Kashmir University of Agricultural Sciences and Technology-Kashmir (Certificate Number: AU/FVSc/PS-57/2058).

AUTHOR CONTRIBUTIONS

BB and NG designed the study. AS, BB, SA, and NG planned the work-flow. BB performed data analysis. BB, RM, and SM wrote the manuscript. All authors reviewed the manuscript.

FUNDING

This work was supported by grants from the Indian Council of Agricultural Research, under the National Agricultural Science Fund scheme (Grant ID: NASF/GTR5006/2015-16).

ACKNOWLEDGMENTS

The bioinformatics data analysis was supported by National Agricultural Higher Education Project (NAHEP), which is duly acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.695178/full#supplementary-material>

REFERENCES

- (2019). *Picard Toolkit*. Available online at: <http://broadinstitute.github.io/picard/>
- Åkerfelt, M., Morimoto, R. I., and Sistonen, L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell Biol.* 11, 545–555. doi: 10.1038/nrm2938
- Ahmed, M. I., Alam, M., Emelianov, V. U., Poterlowicz, K., Patel, A., Sharov, A. A., et al. (2014). MicroRNA-214 controls skin and hair follicle development by modulating the activity of the WNT pathway. *J. Cell Biol.* 207, 549–567. doi: 10.1083/jcb.201404001
- Ansari-Renani, H., Ebadi, Z., Moradi, S., Baghershah, H., Ansari-Renani, M., and Ameli, S. (2011). Determination of hair follicle characteristics, density and activity of Iranian cashmere goat breeds. *Small Rumin. Res.* 95, 128–132. doi: 10.1016/j.smallrumres.2010.09.013
- Aoki, K., Ishii, Y., Matsumoto, K., and Tsujimoto, M. (2002). Methylation of xenopus CIRP2 regulates its arginine-and glycine-rich region-mediated nucleocytoplasmic distribution. *Nucl. Acids Res.* 30, 5182–5192. doi: 10.1093/nar/gkf638
- Banerjee, D., Upadhyay, R. C., Chaudhary, U. B., Kumar, R., Singh, S., Polley, S., et al. (2014). Seasonal variation in expression pattern of genes under hsp70. *Cell Stress Chaperones* 19, 401–408. doi: 10.1007/s12192-013-0469-0
- Bhan, A., and Mandal, S. S. (2015). LncRNA hotair: a master regulator of chromatin dynamics and cancer. *Biochim. Biophys. Acta* 1856, 151–164. doi: 10.1016/j.bbcan.2015.07.001
- Bhat, B., Ganai, N. A., Andrabi, S. M., Shah, R. A., and Singh, A. (2017). TM-aligner: multiple sequence alignment tool for transmembrane proteins with reduced time and improved accuracy. *Sci. Rep.* 7, 1–8. doi: 10.1038/s41598-017-13083-y
- Bhat, B., Singh, A., Iqbal, Z., Kaushik, J. K., Rao, A., Ahmad, S. M., et al. (2019). Comparative transcriptome analysis reveals the genetic basis of coat color variation in pashmina goat. *Sci. Rep.* 9:6361. doi: 10.1038/s41598-019-42676-y
- Bhat, B., Yaseen, M., Singh, A., Ahmad, S. M., and Ganai, N. A. (2021). Identification of potential key genes and pathways associated with the pashmina

- fiber initiation using RNA-seq and integrated bioinformatics analysis. *Sci. Rep.* 11, 1–9. doi: 10.1038/s41598-021-81471-6
- Bhat, S. A., Ahmad, S. M., Ibeagha-Awemu, E. M., Bhat, B. A., Dar, M. A., Mumtaz, P. T., et al. (2019). Comparative transcriptome analysis of mammary epithelial cells at different stages of lactation reveals wide differences in gene expression and pathways regulating milk synthesis between jersey and kashmiri cattle. *PLoS ONE* 14:e0211773. doi: 10.1371/journal.pone.0211773
- Bumla, N. A., Maria, A., Sasan, J. S., and Khateeb, A. M. (2012). Quality of Indian pashmina fibre in terms of its physico-mechanical properties. *Wayamba J. Anim.Sci.* 459–462. Available online at: <http://www.wayambajournal.com/documents/1348636207.pdf>
- Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., et al. (2008). Maker: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18, 188–196. doi: 10.1101/gr.6743907
- Chruscinski, A., Brede, M. E., Meinel, L., Lohse, M. J., Kobilka, B. K., and Hein, L. (2001). Differential distribution of β -adrenergic receptor subtypes in blood vessels of knockout mice lacking β 1- or β 2-adrenergic receptors. *Mol. Pharmacol.* 60, 955–962. doi: 10.1124/mol.60.5.955
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Consortium, U. (2014). Uniprot: a hub for protein information. *Nucl. Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989
- De Leeuw, F., Zhang, T., Wauquier, C., Huez, G., Kruys, V., and Gueydan, C. (2007). The cold-inducible RNA-binding protein migrates from the nucleus to cytoplasmic stress granules by a methylation-dependent mechanism and acts as a translational repressor. *Exp. Cell Res.* 313, 4130–4144. doi: 10.1016/j.yexcr.2007.09.017
- Dong, Y., Zhang, X., Xie, M., Arefnezhad, B., Wang, Z., Wang, W., et al. (2015). Reference genome of wild goat (*Capra aegagrus*) and sequencing of goat breeds provide insight into genetic basis of goat domestication. *BMC Genomics* 16:1. doi: 10.1186/s12864-015-1606-1
- Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* 15, 7–21. doi: 10.1038/nrg3606
- Fulton, L. K., Clarke, M. S., and Farris Jr, H. E. (1994). The goat as a model for biomedical research and teaching. *Illar J.* 36, 21–29. doi: 10.1093/ilar.36.2.21
- Ganai, T., Misra, S., and Sheikh, F. D., (2011). Characterization and evaluation of pashmina producing changthangi goat of Ladakh. *Indian J. Anim. Sci.* 81, 592–599. Available online at: <https://tinyurl.com/rynfknh6>
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003). Rfam: an RNA family database. *Nucl. Acids Res.* 31, 439–441. doi: 10.1093/nar/gkg006
- Hansen, P. (2004). Physiological and cellular adaptations of zebu cattle to thermal stress. *Anim. Reprod. Sci.* 82, 349–360. doi: 10.1016/j.anireprosci.2004.04.011
- Hu, S., Li, Z., Lutz, H., Huang, K., Su, T., Cores, J., et al. (2020). Dermal exosomes containing miR-218-5p promote hair regeneration by regulating I-catenin signaling. *Sci. Adv.* 6:eaba1685. doi: 10.1126/sciadv.aba1685
- Huang, W., Long, N., and Khatib, H. (2012). Genome-wide identification and initial characterization of bovine long non-coding RNA s from EST data. *Anim. Genet.* 43, 674–682. doi: 10.1111/j.1365-2052.2012.02325.x
- Jackman, S. D., Vandervalk, B. P., Mohamadi, H., Chu, J., Yeo, S., Hammond, S. A., et al. (2017). Abyss 2.0: resource-efficient assembly of large genomes using a bloom filter. *Genome Res.* 27, 768–777. doi: 10.1101/gr.214346.116
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9:357. doi: 10.1038/nmeth.1923
- Li, T., Wang, S., Wu, R., Zhou, X., Zhu, D., and Zhang, Y. (2012). Identification of long non-protein coding RNAs in chicken skeletal muscle using next generation sequencing. *Genomics* 99, 292–298. doi: 10.1016/j.ygeno.2012.02.003
- Lis, J., and Wu, C. (1993). Protein traffic on the heat shock promoter: parking, stalling, and trucking along. *Cell* 74, 1–4. doi: 10.1016/0092-8674(93)90286-Y
- Lowe, T. M., and Chan, P. P. (2016). tRNAscan-se on-line: integrating search and context for analysis of transfer RNA genes. *Nucl. Acids Res.* 44, W54–W57. doi: 10.1093/nar/gkw413
- Mardaryev, A. N., Ahmed, M. I., Vlahov, N. V., Fessing, M. Y., Gill, J. H., Sharov, A. A., et al. (2010). Micro-RNA-31 controls hair cycle-associated changes in gene expression programs of the skin and hair follicle. *FASEB J.* 24, 3869–3881. doi: 10.1096/fj.10-160663
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Miao, X., Luo, Q., Zhao, H., and Qin, X. (2016). Genome-wide analysis of miRNAs in the ovaries of Jining grey and Laiwu black goats to explore the regulation of fecundity. *Sci. Rep.* 6, 1–9. doi: 10.1038/srep37983
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAAS: an automatic genome annotation and pathway reconstruction server. *Nucl. Acids Res.* 35(Suppl. 2), W182–W185. doi: 10.1093/nar/gkm321
- Mumtaz, P. T., Bhat, S. A., Ahmad, S. M., Dar, M. A., Ahmed, R., Urwat, U., et al. (2017). LncRNAs and immunity: watchdogs for host pathogen interactions. *Biol. Proced. Online* 19:3. doi: 10.1186/s12575-017-0052-7
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509
- Quirós, P. M., Mottis, A., and Auwerx, J. (2016). Mitonuclear communication in homeostasis and stress. *Nat. Rev. Mol. Cell Biol.* 17:213. doi: 10.1038/nrm.2016.23
- Rinehart, J. P., Li, A., Yocum, G. D., Robich, R. M., Hayward, S. A., and Denlinger, D. L. (2007). Up-regulation of heat shock proteins is essential for cold survival during insect diapause. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11130–11137. doi: 10.1073/pnas.0703538104
- Ryder, M. (1966). Coat structure and seasonal shedding in goats. *Anim. Sci.* 8, 289–302. doi: 10.1017/S000335610003467X
- Santorio, M. G. (2000). Heat shock factors and the control of the stress response. *Biochem. Pharmacol.* 59, 55–63. doi: 10.1016/S0006-2952(99)00299-3
- Sequencing, T. B. C. G., Wang, Z., Ding, G., Chen, G., Sun, Y., Sun, Z., et al. (2012). Genome sequences of wild and domestic Bactrian camels. *Nat. Commun.* 3:1202. doi: 10.1038/ncomms2192
- Shaikh, N. A., Banaganapalli, B., Elango, R., and Hakkeem, K. R. (Eds.). (2019). *Essentials of Bioinformatics Volume I: Understanding Bioinformatics: Genes to Proteins*. Cham: Springer Publishers.
- Shakyawar, D., Raja, A., Kumar, A., Pareek, P., and Wani, S. (2013). Pashmina fibre-production, characteristics and utilization. *Indian J. Fibre Text. Res.* 38, 207–214. Available online at: <http://nopr.niscair.res.in/bitstream/123456789/19255/1/IJFTR%2038%282%29%20207-214.pdf>
- Sheikh, F., Malik, T. H., Sofi, A., Wani, S., and Ganai, T. (2016). Introduction and performance study of pashmina goats in Kargil district (non traditional area) of Jammu & Kashmir, India. *Indian J. Anim. Res.* 50, 129–132. doi: 10.18805/ijar.7487
- Sim ao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). Busco: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Simpson, J. T., and Durbin, R. (2012). Efficient *de novo* assembly of large genomes using compressed data structures. *Genome Res.* 22, 549–556. doi: 10.1101/gr.126953.111
- Stanke, M., and Waack, S. (2003). Gene prediction with a hidden markov model and a new intron submodel. *Bioinformatics* 19(Suppl. 2), ii215–ii225. doi: 10.1093/bioinformatics/btg1080
- Sumitomo, Y., Higashitsuji, H., Higashitsuji, H., Liu, Y., Fujita, T., Sakurai, T., et al. (2012). Identification of a novel enhancer that binds sp1 and contributes to induction of cold-inducible RNA-binding protein (CIRP) expression in mammalian cells. *BMC Biotechnol.* 12:72. doi: 10.1186/1472-6750-12-72
- Tang, Z., Wu, Y., Yang, Y., Yang, Y.-C. T., Wang, Z., Yuan, J., et al. (2017). Comprehensive analysis of long non-coding RNAs highlights their spatio-temporal expression patterns and evolutionary conservation in *Sus scrofa*. *Sci. Rep.* 7:43166. doi: 10.1038/srep43166
- Tarailo-Graovac, M., and Chen, N. (2009). Using repeat masker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* 25, 4–10. doi: 10.1002/0471250953.bi0410s25
- Tissieres, A., and Georgopoulos, C. (Eds.). (1994). “Structure and regulation of heat shock gene promoters,” in *The Biology of Heat Shock Proteins and Molecular*

- Chaperones* (Morimoto: Cold Spring Harbor Laboratory Press, Plainview), 375–393.
- Voellmy, R. (1994). Transduction of the stress signal and mechanisms of transcriptional regulation of heat shock/stress protein gene expression in higher eukaryotes. *Crit. Rev. Eukar. Gene Express.* 4, 357–401.
- Volders, P.-J., Verheggen, K., Menschaert, G., Vandepoele, K., Martens, L., Vandesompele, J., et al. (2014). An update on lncpedia: a database for annotated human lncRNA sequences. *Nucl. Acids Res.* 43, D174–D180. doi: 10.1093/nar/gku1060
- Wagschal, A., Najafi-Shoushtari, S. H., Wang, L., Goedeke, L., Sinha, S., Andrew, S. D., et al. (2015). Genome-wide identification of microRNAs regulating cholesterol and triglyceride homeostasis. *Nat. Med.* 21, 1290–1297. doi: 10.1038/nm.3980
- Xia, Z., Zheng, X., Zheng, H., Liu, X., Yang, Z., and Wang, X. (2012). Cold-inducible RNA-binding protein (CIRP) regulates target mRNA stabilization in the mouse testis. *FEBS Lett.* 586, 3299–3308. doi: 10.1016/j.febslet.2012.07.004
- Xin, Y., Wang, X., Meng, K., Ni, C., Lv, Z., and Guan, D. (2020). Identification of exosomal miR-455-5p and miR-1255a as therapeutic targets for breast cancer. *Biosci. Rep.* 40:BSR20190303. doi: 10.1042/BSR20190303
- Yang, R., Weber, D. J., and Carrier, F. (2006). Post-transcriptional regulation of thioredoxin by the stress inducible heterogenous ribonucleoprotein A18. *Nucl. Acids Res.* 34, 1224–1236. doi: 10.1093/nar/gkj519
- Ye, J., McGinnis, S., and Madden, T. L. (2006). Blast: improvements for better sequence analysis. *Nucl. Acids Res.* 34(Suppl. 2), W6–W9. doi: 10.1093/nar/gkl164

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Bhat, Ganai, Singh, Mir, Ahmad, Majeed Zargar and Malik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.