Deletion of an enhancer in FGF5 is associated with ectopic expression in goat hair follicles and the cashmere growth phenotype

Yefang Li[1,2¶], Shen Song[1,2¶,#a], Xuexue Liu[1,2], Yanli Zhang[1,2], Dandan Wang[1,2], Xiaohong He[1,2], Qianjun Zhao[1,2], Yabin Pu[1,2], Weijun Guan[1,2], YuehuiMa[1,2*], Lin Jiang [1,2*]

[1] Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, P. R. China

[2] CAAS-ILRI Joint Laboratory on Livestock and Forage Genetic Resources, Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), Beijing, 100193, P. R. China

[#a]Current Address: State Key Laboratory of Cardiovascular Disease Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100037, P. R. China

Running title: FGF5 indel association with cashmere growth

* Corresponding author:

E-mail: jianglin@caas.cn (Lin Jiang)

E-mail: yuehui.ma@263.net (Yuehui Ma)

[¶]These authors contributed equally to this work.

Abstract

Research on cashmere growth has a significant effect on the production of cashmere and a profound influence on cashmere goat breeding. Whole-genome sequencing is a powerful platform to rapidly gain novel insights into the identification of genetic mechanisms underlying cashmere fiber growth. Here, we generated whole-genome sequences of 115 domestic goats from China, Nepal and Pakistan, including 51 cashmere goats and 64 non-cashmere goats. We found genetically distinct clusters according to their geographic locations but genetic admixture or introgression may have occurred between the Chinese and Nepalese goats. We identified that the *fibroblast growth factor 5* gene (*FGF5*) shows a strong signature for positive selection in the cashmere goat. The 505-bp indel variant at the *FGF5* gene locus appeared to be strongly associated with cashmere growth. Functional validation showed that the insertion variant may serve as an enhancer for transcription factor binding, resulting in increased transcription of the upstream *FGF5* gene in non-cashmere goats. Our study provides useful information for the sustainable utilization and improved conservation of goat genetic resources and demonstrates that the indel mutation in the *FGF5* gene could potentially serve as a molecular marker of cashmere growth in cashmere goat breeding.

Author summary

Cashmere goats have been selected for thousands of years and have become economically significant livestock in China and other central Asian countries. The

mechanism of cashmere growth is not well understood because most studies have focused on the investigation of candidate genes. Here, we conducted a comprehensive whole-genome analysis for selection signatures in a total of 115 goats from 15 genetically diverse goat breeds. The results revealed a strong selection signature at the *FGF5* gene locus associated with the cashmere growth phenotype. A 505-bp indel was located in the downstream region of *FGF5* and significantly separated in the cashmere goats versus non-cashmere goats. Functional effect analysis of the indel revealed that it may act as an enhancer by specifically binding transcription factors to mediate quantitative changes in *FGF5* mRNA expression. Our study illustrates how a structural mutation of the *FGF5* gene has contributed to the cashmere growth phenotype in domestic goats.

Introduction

Cashmere wool, usually simply known as cashmere, is a fiber obtained from cashmere goats, pashmina goats, and some other breeds of goat. This fiber has been used to make yarn, textiles and clothing for hundreds of years. Cashmere is closely associated with the Kashmir shawl; the word "cashmere" is derived from an anglicization of Kashmir, which occurred in the 19th century when the Kashmir shawl reached Europe from Colonial India [1]. Common usage defines the fiber as wool, but it is finer, stronger, lighter, softer and approximately three times more insulating than sheep wool [2].

Cashmere has been manufactured in China, Mongolia, Nepal and Pakistan for

thousands of years. China has the largest number and richest variety of cashmere goats, such as the Inner Mongolia cashmere, Liaoning and Tibetan varieties, and has become the largest producer of raw cashmere, estimated at 15,438 metric tons (in hair) per year [3]. Nepal has a sizeable indigenous goat population with many nondescript goats. Nepalese goat breeds exhibit enormous variations in fecundity; meat, milk and fibre production; disease resilience; and nutritional requirements. Pakistan is the fourth largest goat-producing country after China, India and Nigeria (FAOSTAT, http://www.fao.org/faostat). The major purposes of Pakistani goats are milk, meat and hair [4]. Moreover, some studies have suggested that a second domestication event for cashmere breeds took place in Pakistan [5].

Cashmere goats grow a double coat composed of the guard hair produced by primary hair follicles (PHFs) and the cashmere produced by the secondary hair follicles (SHFs) [6,7]. The staple length and diameter of hair fibers are the main indicators used to evaluate the value of cashmere. Therefore, identification of related genes and molecular mechanisms that regulate cashmere traits is of great significance. In recent years, based on the goat reference genome, several studies have attempted to characterize genetic variations of cashmere fiber traits in different goat populations using a whole-genome sequencing strategy. For instance, the genes *PRDM6, FGF5* [8], *LHX2, FGF9, WNT2* [9], *SGK3, IGFBP7, OXTR* [10], and so on, showed a strong selection signature for cashmere growth and length in Chinese goat populations. However, to our knowledge, few studies have identified the causative mutations of the

goat *FGF5* gene that underlie cashmere growth in goats. Moreover, the sample size in cashmere goat studies has been limited to Chinese goat breeds, which may not comprehensively analyze cashmere traits.

Here, we sequenced the whole genomes of 115 goats representing 15 breeds from habitats in China, Nepal and Pakistan. To identify the genetic basis for cashmere growth trait in cashmere goats, we performed genomic analysis of selection signatures of goats and identified that the *FGF5* gene showed some of the strongest signatures for positive selection in the cashmere goat genome. Further exploration of the *FGF5* genotypes and functional validation assays indicated that a 505-bp indel mutation located downstream from *FGF5* gene may act as an enhancer, resulting in increased mRNA expression of the *FGF5* gene; moreover, deletion of this enhancer is strongly associated with cashmere growth in cashmere goats.

Results

Genomic variants

We used the Illumina HiSeq platform to generate whole-genome sequencing data of 115 goats sampled from China, Nepal and Pakistan (**Fig 1A**). The median genome coverage achieved across the full data set was ~5X (min=2.11X, max=7.78X), representing ~87.44% (min=78.54%, max=89.88%) base coverage per individual genome (**S2 Table**). The alignment ratio of reads to the genome was 93.57%-99.93% (**S2 Table**). Strict read alignment and genotype calling procedures allowed us to obtain a total of 17,534,538 single nucleotide polymorphisms (SNPs). The majority of

the autosomal SNPs were located within intergenic (11,821,678, 45.847%) and intronic (11,347,329, 44.007%) regions, with only 0.858% (221,209) located in exonic regions. Approximately 9% were present in downstream or upstream gene regulatory regions. A total of 61,164 missense SNPs and 126,811 synonymous SNPs resulted in a nonsynonymous/synonymous ratio of 0.482 (**S3 Table**).

Genetic diversity

Compared to other goats, Tibetan cashmere goats were found to show the highest genome-wide heterozygosity levels, fewer runs of homozygosity (ROHs) and the lowest linkage disequilibrium (LD) decay. However, among northern Chinese cashmere goat breeds, Liaoning cashmere goat (LiNi), Alashan cashmere goat (ALS) and Arbus cashmere goat (ABS), exhibited lower genome-wide heterozygosity, more ROHs and more LD levels, while Erlangshan cashmere goat (ELS) exhibited the opposite. This result may be related to more intensive selection breeding. Imported dairy goats and Xiangdong black (XiDo) goat showed similarly high levels of genetic diversity. The genome-wide heterozygosity levels, ROHs and LD decay were found to be lower in Nepalese highland goats but higher in Nepalese lowland goats. Compared to Chinese and Nepalese goat breeds, the Pakistani goat breeds showed less level of genetic diversity, especially the Bugi Toori goat, which is likely a consequence of its inbreeding history [11] (**S4 Table, S1 and S2 Figs**).

Phylogenetic analyses

Principal component analysis (PCA) of ~3.4 million unlinked SNPs revealed that the

Chinese, Nepalese and Pakistani goats can be separately clustered by the first PCA axis. One Pakistani goat breed (BTR) was genetically more distant from other goat breeds (**Fig 1B, upper panel**). Restricting the analysis to Chinese goat breeds revealed four major clusters (**Fig 1B, main panel**), with the first PCA axis separating the Toggenburg (TGB) and Laoshan (LaSh) dairy goat populations, the second PCA axis separating the Tibetan cashmere goat breeds, and the third PCA axis separating the XiDo black goat from southern China (**S3 Fig**). PC1, PC2 and PC3 were able to explain genetic differences of 3.98%, 3.08% and 2.90%, respectively. The admixture analysis results were largely consistent with the PCA results as well as the similar genetic makeup among the Chinese, Nepalese and Pakistani goats (**S4 Fig**). When K=4, the dairy goats, Chinese goats, BTR goat and the remaining goats were genetically distinct; when K=6 and K=7, the Chinese native goats were divided into the XiDo goat from southern Chinese, the northern Chinese cashmere goats and the Tibetan cashmere goats.

We next investigated the ML-TreeMix tree [12] and the distance-based neighbor-joining tree [12] using *Capra ibex* as the outgroup among all the goats. The distance-based neighbor-joining tree displayed six clades according to location or specific goat trait, which were consistent with the PCA and admixture analysis results (**Fig 1C**). The reliability of the neighbor-joining tree was estimated by 100 bootstrap pseudoreplicates. The ML-TreeMix tree without migration events (ML=0) inferred from the TreeMix analysis divided the 115 goats into six clusters, which were

consistent with the neighbor-joining tree results (**S5A Fig**). When M=1 and M=2, the results suggested that gene exchange occurred between wild and Nepalese and Pakistani goats (**S5B and S5C Fig**); when M=3, genetic materials of LiNi flowed to XiDo (S5D Fig); and when M=4-6, genes flowed from XiDo to the TGB, LaSh and Nepalese lowland (NPL) goat breeds (**Fig 1D, S5D-S5F Fig**).

Genome-wide selection scans for cashmere growth

To detect the positive selection signatures within 100 kb sliding windows, we next scanned the cashmere goat (including TBG, TRT, LiNi, ABS, ELS, ALS and NPH) genomes relative to non-cashmere goat (including TGB, LaSh, XiDo, NPL, BTR, KMR, PTR and TPR) genomes by using three statistical methods, namely, F_{ST}, θ_{π} ratio ($\theta_{\pi\text{-noncash}}/\theta_{\pi\text{-cash}}$) and ZHp (**Fig 2**). The top-5% selection candidates that were common to all three statistical methods identified 982 windows, which annotated a total of 378 protein-coding genes (**S5-S8 Tables, S6 Fig**). Enrichment analyses for Gene Ontology terms revealed that the melanocortin receptor activity (GO:0004977, *P*-value = 2.33E-06), response to stimulus (GO:0050896, *P*-value = 4.70E-07), developmental process (GO:0032502, *P*-value = 2.07E-06) and cellular metabolic process (GO:0044237, *P*-value = 1.07E-06) categories were significantly overrepresented (**S9 Table**). Notably, the genome window containing the *FGF5* locus was under higher selection (chromosome 6, **Fig 2**). This gene is known to be involved in hair growth. Moreover, several genes under strong selection signals are plausibly related to metabolism, inflammation, melanin precipitation and high-altitude

adaptation. For example, *STIM1* is an endoplasmic reticulum calcium sensor involved in regulating Ca2+ and metabotropic glutamate receptor signaling in the nervous system [14,15]. The CERT protein mediates the START pathway of ceramide transport in a nonvesicular manner and the amphiphilic cavity of the START domain is optimized for specific binding of natural ceramides [16,17]. *NOP14* plays significant roles in the proliferation and migration of pancreatic cancer cells [18]. The mutations in the *SGCB* gene can lead to a loss of functional protein and result in limb-girdle muscular dystrophy disease . MYCBP2 is a member of the PHR protein family and an E3 ubiquitin ligase, and it was shown to have important functions in developmental processes, such as axon termination and synapse formation [20]. *WARS2* has low enzyme activity, and inhibition of *WARS2* in endothelial cells reduces angiogenesis [21,22]. *MC1R* and *KIT* genes have been implicated in human and animal hair pigmentation, reflecting a role in the development and function of melanocytes [23,24]. The *DSG3* gene is responsible for the high-altitude adaptation of the Tibetan goat [25,26].

Annotation of variants under positive selection in *FGF5*

We next sought to further refine the selection targets within the *FGF5* locus by using three different methods, namely, the θ_π ratio ($\theta_{\pi\text{-noncash}}/\theta_{\pi\text{-cash}}$), Tajima's D and F_{ST}, and detecting the read depth. We noticed a 505-bp deletion within the most significant selection region in the *FGF5* gene (**Fig 3, S10-S13 Tables**), located at position 95,454,689-95,455,189 of chromosome 6 (**Fig 4A**). The deletion variant is present in

192 cashmere goats, suggesting that it arose in an independent genetic background.

193 We designed primers spanning the breakpoint of the deletion to genotype the indel
194 variant by gel electrophoresis, generating a 267-bp fragment in all cashmere goats but
195 a 772-bp fragment in non-cashmere goats (**Fig 4A**). To further identify the possible
196 functional consequences of the deletion variant, we investigated whether it showed
197 any association with cashmere growth,extending our analysis to a more
198 comprehensive panel of 288 goats originating from 20 populations (**S14 Table**). The
199 results confirmed a remarkable correlation between the frequencies of the indel
200 variant and cashmere growth. Cashmere goat breeds showed the highest allelic
201 frequencies of deletion (>0.9), whereas non-cashmere goat breeds showed higher
202 allelic frequencies of insertion (nearly 0.8, **Fig 4B**).

203 Furthermore, the insertion fragment of the *FGF5* locus in humans showed a high
204 conservation score in the 100-vertebrate animals alignment (e.g., goat, mouse, cat,
205 dog, sheep, yak and donkey); by using UCSC database, we also found that the
206 insertion fragment contains EP300, FOS and CEBPB transcription factors among
207 different species [27,28] (**Fig 4C**). This finding indicated that the indel fragment has
208 cis-regulatory effects for *FGF5* gene transcription.

209 **Biological significance of the indel variant**

210 Since the indel variant contains an extremely conserved FOS transcription factor
211 binding site, we sought to functionally verify the effect of the binding site in this
212 variant. A pair of biotin-labeled probes were designed, namely, a wildtype probe

containing TGAGTCA and a mutant probe excluding the site. After the binding reaction with nucleoprotein fractions from NIH/3T3 cells derived from mouse embryonic fibroblasts, the wildtype probe resulted in an obvious positive protein complex band but the mutant probe did not. The addition of either 80-fold or 160-fold cold probes to the reaction system produced a significantly thinner binding band and significantly weakened grayscale compared to those with wildtype probe. The weaker binding reaction may be due to the lower cold probe concentration or higher nucleoprotein concentration making the competitive reaction incomplete. After adding c-FOS antibody to the reaction system, a complex was obviously retained at the top of the gel (**Fig 5A**). We thus hypothesized that the indel variant can specifically bind to the FOS transcription factor, which may have important regulatory effects on upstream target genes.

To confirm our hypothesis, we constructed dual-luciferase recombinant plasmids either including (pGL4.23-ins) or excluding (pGL4.23-del) the indel fragment. These two recombinant vector plasmids and an empty vector were each transfected into NIH/3T3 cells together with an internal luciferase control (pGL4.74) to measure the luciferase activities. We observed significantly higher luciferase activity in cells expressing the indel fragment than both empty cells and cells expressing pGL4.23-del (**Fig 5B**). This assay suggested that the indel fragment of the *FGF5* gene functions as an enhancer to which certain transcription factors specifically bind to upregulate *FGF5* gene expression.

We next examined whether the indel mutation could alter the transcriptional response to cashmere length using RT-qPCR assays. The mRNA expression level of cashmere goats carrying the deletion was significantly decreased compared to that of non-cashmere goats carrying the insertion ($P<0.01$, **Fig 5C**). Thus, the results of the gel shift experiment, dual-luciferase assay and RT-qPCR assays confirmed that the deletion variant disrupts the binding of transcription factors (e.g. FOS) and leads to lower expression of the *FGF5* gene in the skin of cashmere goats, while the insertion variant serves as an enhancer element that amplifies the transcriptional activity mediated by *FGF5* in non-cashmere goats.

Discussion

In this study, we sequenced the genomes of 115 goats from 15 breeds originating from China, Nepal and Pakistan. The genome data set allowed us to identify a total of ~17.5 million SNP variants, which helped us reveal the genetic diversity and population structure of these goats. In this study, most of goat breeds showed a higher diversity than others, such as the Tibetan cashmere goats, dairy goats, Erlangshan cashmere goat and Xiangdong black goat. However, the Liaoning, Alashan and Arbus cashmere goats showed lower diversity than other Chinese goats, in line with previous work [10]. The three cashmere goat breeds are famous worldwide for their fine, long fibres. This fact indicated that these breeds may have been subject to stronger intensive selection for cashmere growth. An interspecies comparison showed that Pakistani goats had a lower diversity than others. Notably, similar to previous work

255 based on the goat 50K SNP chip [10], the Bugi Toori (BTR) goat breed showed the
256 lowest genetic diversity and a great differentiation from other Pakistani goat breeds.
257 There may have been a historical bottleneck in history or an in-flight phenomenon in
258 the BTR goat breed (**S1 and S2 Figs, S3 Table**).

259 Population structure analysis revealed that all the goats evaluated in this study were
260 divided into six clusters, namely, dairy goat, southern Chinese goat, northern Chinese
261 goat, Tibetan goat, Nepalese goat and Pakistani goat. When potential migration edges
262 were added to the ML-TreeMix tree, gene exchange between the wild goats and
263 Nepalese goats as well as Pakistani goats was detected among the clusters. This result
264 may indicate that the local goats had a hybridization event with wild goats in the past.
265 However, we observed migration edges between the Xiangdong black goat and
266 Liaoning cashmere goat, as well as the dairy goat and Nepalese goat. There is a lower
267 possibility of genetic admixture or introgresssion between the breeds because of the
268 geographically distance between their inhabited regions. Therefore, to determine
269 whether the southern Chinese goats underwent gene exchange with northern Chinese
270 goat or Nepalese goats, the inclusion of more southern Chinese goat breeds is
271 required. (**Fig 1B-1D, S3-S5 Figs**).

272 Compared with other domestic animals, goats are more adaptable to extreme
273 environments. In China, cashmere goats are mainly distributed in the northern China
274 and the Tibetan Plateau, where they have adapted well to the cold environment. More
275 importantly, the fluff produced by cashmere goats provides good, warm materials for

the native population. Therefore, different cashmere traits are continuously formed under natural and artificial selection. This study compared the genomes of cashmere goats including those from habitats in Liaoning, Inner Mongolia, Tibetan areas and Nepalese highland areas bordering the Tibetan region with various non-cashmere goats from different areas. Scanning the genome of cashmere goat breeds for signatures of positive selection revealed the *FGF5* gene among the top candidates (**Fig 2**). The *FGF5* gene participates in the FGF pathway, which plays a central role in hair growth. Studies on the *FGF5* gene demonstrated the relation to coat hair length in mice [29], dogs [29], cats [31], humans [32], donkeys [33] and alpacas [34]. The same selection target has been described in a number of cashmere goat investigations [10,35]. In addition, disruption of the *FGF5* gene via the CRISPR/Cas9 system in cashmere goats increased the number of secondary hair follicles and enhanced the fiber length [36]. Previous studies have found a few SNP variants of the *FGF5* gene that may be associated with hair length, including a missense SNP (c.284G> T) in dogs, four SNPs (c.194C>A, c.182T>A, c.474delT and c.475A>C) in cats, two SNPs (c.433_434delAT and c.245G> A) in donkeys and a missense SNP (c.499C>T) in alpacas [30,31,33,34]. Recently, one SNP (c.253G>A) in the 5'-UTR of *FGF5* resulted in a start codon that could lead to a premature/dysfunctional protein in Tibetan cashmere goats [35].

At the molecular level, our work did not reveal any missense SNPs in the exons of *FGF5* but instead revealed a significant indel variant in the region downstream of the

297 *FGF5* gene locus (**Fig 3**). Interestingly, the result of the expanded population
298 verification showed that the 505-bp indel variant was significantly separated in
299 cashmere goats versus non-cashmere goats. The cashmere goats mainly exhibited a
300 deletion mutation (> 0.9), whereas non-cashmere goats mainly exhibited an insertion
301 mutation (~ 0.8). Some of the goat breeds have a small number of hybrids, such as the
302 Nepalese lowland goat, Xiangdong black goat and Laoshan dairy goat, which may be
303 caused by crossbreeding or altitude factors. Therefore, this result indicated that the
304 indel variant can serve as a genetic marker for the cashmere growth trait.
305 Furthermore, the indel mutation was found to contain a conserved binding site for the
306 FOS transcription factor located in the mutation array and to be highly conserved in
307 various mammals (**Fig 4**). In humans, the mutation is located downstream from the
308 *FGF5* locus and has been identified as an enhancers according to the FANTOM5
309 Human Enhancers database (http://slidebase.binf.ku.dk/human_enhancers/).
310 Therefore, it is speculated that the indel variant plays a potential enhancer role in
311 *FGF5* gene transcription.

312 Thus, an electrophoretic mobility shift assay (EMSA), and a novel dual-luciferase
313 reporter assay based on the expression of firefly and Renilla luciferase and mRNA
314 expression levels of *FGF5* in goats were performed to explore the relevance of the
315 indel variant to the *FGF5* gene. EMSA is a powerful tool for evaluating DNA-protein
316 or RNA-protein interaction and is often used to detected the activated transcription
317 factors (TF) that bind with DNA or RNA in the nucleus [37]. EMSAs based on

NIH/3T3 nuclear extracts revealed that the protein complex bound to the biotin-probe containing the wildtype FOS binding site, but did not bind to the probe that contained the mutant FOS binding site. Efficient competition for protein complex formation was observed with the inclusion of a wildtype cold probe, and a clear supershift occurred when anti-c-FOS antibody was added, which further confirmed that the FOS transcription factor can specifically bind to the indel variant. The c-FOS protein is a member of the FOS protein family [38]. The dual-luciferase reporter gene assay is widely used to study promoter activity, transcription factors, intracellular signaling, protein interactions [39], miRNA regulation [40], and target site recognition [41]. The dual reporter gene assay based on firefly (*Photinus pyralis*) and sea kidney (*Renilla reniformis*, also known as marine pansy) luciferases can improve experimental accuracy by normalizing results and reducing technical differences [42]. The results of this assay showed that the insertion mutation can significantly enhance promoter transcription and increase gene expression thereby verifying its enhancer function. Finally, we detected significant differences in the expression levels of the *FGF5* gene from skin tissues of cashmere goats compared with non-cashmere goat, further confirming that the insertion variant may serve as an enhancer by binding to a transcription factor to result in increased transcription of its upstream *FGF5* gene target (**Fig 5**).

In conclusion, our study provides a whole-genome sequence analysis of Chinese, Nepalese and Pakistani goat breeds. It includes a total of 115 individual genomes

spread across 15 goat breeds. The phylogenetic relationship of the 115 individuals revealed genetically distinct clusters according to their geographic locations, but genetic admixture or introgression may have occurred between Chinese and Nepalese goats. Genomic regions showing signatures of positive selection in cashmere goats revealed that the *FGF5* gene was the top candidate for the cashmere growth trait. Genotyping data from a large panel of 288 cashmere and non-cashmere goats revealed that a 505-bp indel variant, located downstream from the *FGF5* gene, is strongly associated with the cashmere length phenotype; furthermore, the deletion fragment reached close-to-fixation (~90%) frequencies in cashmere goats. Functional assays demonstrated that the insertion variant may act as an enhancer by binding to transcription factor, ultimately causing increased transcription of the upstream *FGF5* gene target. Our study provides useful information for the sustainable utilization and improved conservation of goat genetic resources. The valuable genetic marker that we identified will contribute to cashmere goat breeding to improve cashmere growth in the future.

Materials and Methods

All sample collection were approved by the Animal Welfare and ethics Committee of Institute of Animal Science, Chinese Academy of Agricultural Sciences (Permit number: IAS2019-61).

Sample information

In this study, We collected a total of 91 goats representing 44 Chinese native goats, 16

Nepalese goats and 31 Pakistani goats for whole-genome sequencing. In addition, we downloaded genomic dataset of 24 Chinese goats from the Sequencing Read Archieve (https://www.ncbi.nlm.nih.gov/) under accession code PRJNA338022. 51 of the total 115 goats are cashmere goats, including 8 Liaoning (LiNi), 6 Arbus (ABS), 7 Erlangshan (ELS), 3 Alashan (ALS), 10 Tibetan Bange (TBG), 10 Tibetan Ritu (TRT) in China and 7 Nepalese highland (NPH) in Nepal. 64 native goats produce little cashmere, including 10 Toggenburg dairy (TGB), 7 Laoshan dairy (LaSh), 7 Xiangdong black (XiDo) in China, 9 Nepalese lowland (NPL) in Nepal, 6 Bugi Toori (BTR), 9 Kamori (KMR), 11 Pateri (PTR) and 5 Tapri (TPR) in Pakistan (**S1 Table, Fig 1A**). A minimum of two separate flocks were sampled for each breed or location, and parent/offspring pairs were excluded.

Whole genome sequencing analysis

DNA extraction was conducted by Wizard® Genomic DNA Purification Kit (Promega). About 3μg of genomic DNA from each collected sample was sequenced on Illumina HiSeq 2000 instruments at BerryGenomics Company (Beijing, China). The 350bp sequencing library with paired-end sequencing was constructed using Illumina's standard protocol. At least 5x genome coverage and no less than 15Gb sequencing data were gained per individual. After trimming low-quality bases and adapter sequences, the clean reads were aligned against the latest goat reference genome assembly ARS1(GCF_001704415.1) [43] using mem algorithm in Burrows-Wheeler Aligner (BWA) software [44,45]. Then the mapping results were

381 converted to BAM format by SAMtools (Version: 1.1) [46]and sorted by SortSam
382 tools in Picard packages (picard.sourceforge.net, Version: 1.86). Only properly paired
383 reads both aligned to the reference were retained for subsequent analysis (**S2 Table**).
384 The BamCoverage (https://github.com/BGI-shenzhen/BamCoverage/) was used to
385 compute the coverage and depth of sequence alignments, with the "statistics
386 Coverage" parameter.

387 **Variant calling**

388 The program Genome Analysis Toolkit (GATK) [47]and SAMtools v0.1.19 [48] was
389 used to identify SNPs, short insertions and deletions (indels). Reads were realigned
390 around indels using the Realigner Target Creator and Indel Realigner tools from
391 GATK, before calling SNPs with the GATK Unified Genotype and SAMtools
392 mpileup modules, separately. SNPs were retained if matching the following five
393 criteria: (1) the SNP confidence score (QD) was greater than or equal to 20; (2) the
394 Phred-scaled P-value of the Fisher's exact test to detect strand bias (FS) was inferior
395 to equal to 10; (3) the Z-score of the Wilcoxon rank sum test of Alt vs. Ref read
396 position bias (ReadPosRankSum) was greater or equal to -8; (4) the Qualscore of each
397 individual SNP was larger-than-average; (5) SNPs showed only two possible alleles
398 and a minimal allele frequency of 5%. The variants with sequence coverage and
399 base-level values lower than the average of all sites were filtered. The 115 individual
400 SNP VCF files were combined into the merged dataset of 17,534,538 autosomal SNPs
401 and this merged SNP dataset was further phased to impute its own missing positions

using BEAGLE software [49,50].

Annotation

SNP variants were classified into protein coding regions (overlapping a coding exon), 5'UTRs and 3'UTRs (overlapping untranslated region), intronic regions (overlapping with an intron), or intergenic regions using the goat genome GTF file downloaded from Ensembl 94 (ftp://ftp.ensembl.org/pub/release-100/gtf/capra_hircus/) and the SNPEff software (Version: 4.0) [51]. SNPs located within protein coding regions were further binned into synonymous and non-synonymous SNPs **(S3 Table**).

Diversity analysis

The within-population genetic diversity for goat populations was assessed using the filtered SNPs and various metrics, including observed (Ho) and expected heterozygosity (He) **(S4 Table)**. The runs of homozygosity (ROH) for each goat population, including the number of ROHs and the total size within ROHs for each individual, were calculated by the command of "--homozyg-window-snp 50 --homozyg-window-het 1 --homozyg-kb 500 –homozyg-density 1000" using the program PLINK v1.90b [52]. Linkage disequilibrium (LD) was computed for each population via the squared correlation coefficient (r^2) between pairwise SNPs by the command of "-MaxDist 500 -MAF 0.005 -Het 0.9 –Miss 0.25" using the software PopLDdecay (https://github.com/BGI-shenzhen/PopLDdecay)

Phylogenetic analysis

All SNPs were pruned using PLINK (Version:1.90b) and considering window sizes of 1000 variants, a step size of 5, and a pairwise r^2 threshold of 0.5 (--indep-pairwise 1000 5 0.5). The principal component analysis (PCA) was carried out using the GCTA 1.91 software [53]. The neighbor-joining tree [13] was constructed using PHYLIP 3.68 (evolution.genetics.washington.edu/phylip.html). MEGA7 software [54] was used to visualize the phylogenetic trees. The population structure was examined via calculating Cross Value with an expectation maximization algorithm implemented in the software ADMIXTURE [55]. The number of assumed genetic clusters K ranged from 2 to 7. The population-level admixture analysis was conducted by TreeMix v.1.12 [12]. The program inferred the ML tree for 15 goat breeds (117 individuals) and an outgroup (wild goat). The command was '-I input –bootstrap -k 10000 –root outgroup -o output'. From one to 6 migration events were gradually added to the ML tree, and the command was '-i input -bootstrap -k 10000 -m migration events -o output'.

Selective sweeps

We scanned the cashmere goat genome for signatures of positive selection by combing three selection signature tests of the population-differentiation statistic (F_{ST}) [12], the relative nucleotide diversity (θ_π ratio, $\theta_{\pi\text{-Noncash}}/\theta_{\pi\text{-Cash}}$) [57] and the transformed heterozygosity score (ZHp). Genomic evidence for positive selection in response to cashmere growth was evaluated by contrasting differentiation indices between the cashmere goats versus the other goats. F_{ST} and nucleotide diversity (θ_π)

were calculated by VCFTools [58]. The window-based ZHp approach was calculated as previously described [59]. Each test was based on a 100-kb window with 10-kb increment. We considered top 1% level for empirical percentile (F_{ST}>0.153, θ_π ratio>1.547, |ZHp| >3.210) windows as candidate outliers in strong selective sweeps. To annotation candidate genes harbored in these selective regions, we used Rscript to map genes in selective windows. The overlapping windows shared by top 5% highest all three tests were considered as conservative candidate selection targets and were further annotated by the genomic database BioMart (http://www.biomart.org/). To detect the genomic loci that are associated with cashmere length around the *FGF5* gene, we also calculated the θ_π ratios, Tajima'*D* [60] and pairwise F_{ST} in 2 Mb windows between the cashmere breeds and the non-cashmere breeds. The Gene Ontology (GO) enrichment analysis of the annotated candidates were performed by using both the online G:profiler.

Validation in the extended population

In order to predict functional candidates, this 505 bp indel variant returning the most significant signature was classified according to their evolutionary conservation scores among other mammals. Primers were designed according to the indel region of FGF5 gene: FGF5-indel-F: 5'-GGTGATAAGCCACACGTTCAAA-3', FGF5-indel-R: 5'-TGGCTGTGATCAAACTTACAACC -3'. The indel region was genotyped by PCR amplification using the reaction condition of the 5-min pre-denaturation, 30s-denaturation, 30s-annealing and 45s-extension for 40 cycles.

The genotype results were visualized by agarose gel electrophoresis. The indel of *FGF5* gene were successfully genotyped in the extended population of 288 goats, including 153 cashmere goats and 135 non-cashmere goats.

Electrophoretic mobility shift assay

The crude nuclear protein was extracted from NIH/3T3 cells using the Nuclear/Cytoplasmic Protein Extraction Kit (SINP001, Viagene) and protein concentration was determined by the Enhanced BCA Protein Assay Kit (CHEM001, Viagene). The oligonucleotide probe of the wild allele was 5'-ATGACTC**TGAGTCA**GTCTCCTCC-3', while the oligonucleotide probe of the mutant allele was 5'-ATGACTCGTCTCCTCC-3'. The probes were synthesized by Viagene Biotech company. EMSA was performed using a non-radioactive EMSA kit (SIDET101, Viagene) with biotin-probes, according to the user's manual instruction. Briefly, 4 μg nuclear protein was incubated with poly dI:dC for 20 min at room temperature in binding reaction buffer. Then biotin-probe was added to and incubated with the mixture at room temperature for at least 20 min. The reaction mixtures were separated by electrophoresis by 8% non-denaturing polyacrylamide gel in 0.5× Tris-borate-EDTA buffer at 120V for 1h. The gel was transferred onto a pre-soaked nylon-membrane at 390 mA for 40min and afterward, the energy of 800 mJ is applied for crosslinking DNA to nylon membrane using CL-1000 Ultraviolet Crosslinker (UVP, UK). Finally, the complexes bands were visualized by chemiluminescent detection. Competition reaction with a 80-fold and 160-fold molar excess of unlabeled

485 oligonucleotide were performed to confirm the specificity of the DNA-protein
486 complex. For the supershift experiment, 2μg of anti-c-fos antibodies (sc-8047X, Santa
487 Cruz) were added to the mixture.

488 **Dual-luciferase reporter assay**

489 NIH/3T3 cells were propagated in the medium of Roswell Park Memorial Institute
490 1640 (RPMI 1640), supplemented with 10% heat-inactivated fetal bovine serum and
491 penicillin (0.2 U/ml)/streptomycin (0.2 μg/ml)/L-glutamine (0.2 μg/ml) (Gibco, USA).
492 The 772-bp and 267-bp fragment of *FGF5* indel region were cloned into the pGL4.23
493 vector (Promege, USA) expressing Firefly luciferase gene, respectively. We thus
494 generated two recombinant vectors pGL4.23-ins (containing 505-bp insertion) and
495 pGL4.23-del (containing the deletion). Each plasmid was co-transfected into NIH/3T3
496 cells with the internal control vector pGL4.74 expressing Renilla luciferase gene by
497 Lipofectamine^{TM} 3000 (Invitrogen, America) according to the manufacturer's
498 instruction. The firefly luminescence signal (FiLuc) and Renilla luciferase signal
499 (hRLuc) of NIH/3T3 cells were measured for each transfection on a multi-function
500 microplate reader (Tecan Infinite 200 Pro) using the Dual-Luciferase Reporter Assay
501 System (E1910, Promega) after 24h transfection.

502 **RT-qPCR quantification**

503 The total RNA was extracted from Inner Mongolia Alashan cashmere goats and Dazu
504 black goats using RNA extraction Kit (Promega), and RNA quality and concentration
505 were measured on an Agilent 2100 Bioanalyzer (Germany). The RIN value of

506 samples greater than 8.0 were used for RT-qPCR analysis.The cDNA was synthesized
507 using PrimeScript RT reagent kit with gDNA Eraser (Takara, Dalian, China) in a 20
508 μl reaction mixture following the manufacturer's instruction. The expression levels of
509 FGF5 gene was normalized against UBC reference genes [61,62]. The primers used in
510 the RT-qPCR experiment were showed in **S11 Table**. The RT-qPCR was performed
511 using TB Green Premix Ex Taq (Takara, Dalian, China). The qPCR reaction program
512 was set as follow: 95 ℃ 30 s, 40 cycles of 5 s at 95 ℃ and 34 s at 60 ℃. The qPCRs
513 were run of both technical and biological replicates (n=3) using an ABI7500 sequence
514 detection system (Applied Biosystems by Life Technologies, Darmstadt, Germany).
515 Fold expression changes were determined using a standard $2^{-\Delta\Delta CT}$ method that
516 compares C_T (cycle threshold) values of a reference gene to the gene of interest for
517 the ΔC_T calculation and compares the ΔC_T value of a reference sample with the
518 sample of interest for the $\Delta\Delta C_T$ calculation [63].

519 ## Acknowledgments

520 The authors are grateful to all goat owners and breeding organizations who donated
521 samples. We thank members of the Nextgen project for sharing their data. We thank
522 the National Germplasm Center of Domestic Animal Resources of IAS.

523

References:

1. Britannica E. cashmere: Encyclopædia Britannica; 2008.

2. Von Bergen W. What the manufacturer requires in raw wool. 1963.

3. National Bureau of Statistics of China. China statistical yearbook 2019. China Statistics Press. 2019.

4. Khan MS, Rehman ZU, Khan MA, Ahmad S. Genetic resources and diversity in Pakistani goats. Pak Vet J. 2008; 28.

5. Porter V. Goats of the world.: Farming Press; 1996.

6. Stenn KS, Paus R. Controls of hair follicle cycling. Physiol Rev. 2001; 81: 449-494.

7. Ansari-Renani HR, Ebadi Z, Moradi S, Baghershah HR, Ansari-Renani MY, Ameli SH. Determination of hair follicle characteristics, density and activity of Iranian cashmere goat breeds. Small Ruminant Res. 2011; 95: 128-132.

8. Zhang B, Chang L, Lan X, Asif N, Guan F, Fu D, et al. Genome-wide definition of selective sweeps reveals molecular evidence of trait-driven domestication among elite goat (Capra species) breeds for the production of dairy, cashmere, and meat. Gigascience. 2018; 7: y105.

9. Wang X, Liu J, Zhou G, Guo J, Yan H, Niu Y, et al. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. Sci Rep-Uk. 2016; 6: 38932.

10. Li X, Su R, Wan W, Zhang W, Jiang H, Qiao X, et al. Identification of selection signals by

543 large-scale whole-genome resequencing of cashmere goats. Sci Rep-Uk. 2017; 7: 1-10.

544 11. Kumar C, Song S, Dewani P, Kumar M, Parkash O, Ma Y, et al. Population structure, genetic
545 diversity and selection signatures within seven indigenous Pakistani goat populations. Anim
546 Genet. 2018; 49: 592-604.

547 12. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele
548 frequency data. Plos Genet. 2012; 8: e1002967.

549 13. SAITOU N. The neighbor-joining methods: A new method for reconstructing phylogenetic
550 trees. Mol.biol.evol. 1987; 4.

551 14. Hartmann J, Karl RM, Alexander RP, Adelsberger H, Brill MS, Rühlmann C, et al. STIM1
552 controls neuronal Ca^{2+} signaling, mGluR1-dependent synaptic transmission, and cerebellar
553 motor behavior. Neuron. 2014; 82: 635-644.

554 15. Majewski Ł, Maciąg F, Boguszewski PM, Wasilewska I, Wiera G, Wójtowicz T, et al.
555 Overexpression of STIM1 in neurons in mouse brain improves contextual learning and
556 impairs long-term depression. Biochim Biophys Acta Mol Cell Res. 2017; 1864: 1071-1087.

557 16. Kudo N, Kumagai K, Tomishige N, Yamaji T, Wakatsuki S, Nishijima M, et al. Structural basis
558 for specific lipid recognition by CERT responsible for nonvesicular trafficking of ceramide.
559 Proc Natl Acad Sci U S A. 2008; 105: 488-493.

560 17. Bandet CL, Mahfouz R, Véret J, Sotiropoulos A, Poirier M, Giussani P, et al. Ceramide
561 Transporter CERT Is Involved in Muscle Insulin Signaling Defects Under Lipotoxic
562 Conditions. Diabetes. 2018; 67: 1258-1271.

563 18. Zhou B, Wu Q, Chen G, Zhang TP, Zhao YP. NOP14 promotes proliferation and metastasis of
564 pancreatic cancer cells. Cancer Lett. 2012; 322: 195-203.

565 19. Giugliano T, Fanin M, Savarese M, Piluso G, Angelini C, Nigro V. Identification of an
566 intragenic deletion in the SGCB gene through a re-evaluation of negative next generation
567 sequencing results. Neuromuscul Disord. 2016; 26: 367-369.

568 20. Pierre S, Zhang DD, Suo J, Kern K, Tarighi N, Scholich K. Myc binding protein 2 suppresses
569 M2-like phenotypes in macrophages during zymosan-induced inflammation in mice. Eur J
570 Immunol. 2018; 48: 239-249.

571 21. Wang M, Sips P, Khin E, Rotival M, Sun X, Ahmed R, et al. Wars2 is a determinant of
572 angiogenesis. Nat Commun. 2016; 7: 12061.

573 22. Agnew T, Goldsworthy M, Aguilar C, Morgan A, Simon M, Hilton H, et al. A Wars2 Mutant
574 Mouse Model Displays OXPHOS Deficiencies and Activation of Tissue-Specific Stress
575 Response Pathways. Cell Rep. 2018; 25: 3315-3328.

576 23. Chen S, Zhu B, Yin C, Liu W, Han C, Chen B, et al. Palmitoylation-dependent activation of
577 MC1R prevents melanomagenesis. Nature. 2017; 549: 399-403.

578 24. Moss KG, Toner GC, Cherrington JM, Mendel DB, Laird AD. Hair depigmentation is a
579 biological readout for pharmacological inhibition of KIT in mice and humans. J Pharmacol
580 Exp Ther. 2003; 307: 476-480.

581 25. Kumar C, Song S, Jiang L, He X, Zhao Q, Pu Y, et al. Sequence Characterization of DSG3 Gene
582 to Know Its Role in High-Altitude Hypoxia Adaptation in the Chinese Cashmere Goat. Front

583 Genet. 2018; 9: 553.

584 26. Song S, Yao N, Yang M, Liu X, Dong K, Zhao Q, et al. Exome sequencing reveals genetic
585 differentiation due to high-altitude adaptation in the Tibetan cashmere goat (Capra hircus).
586 Bmc Genomics. 2016; 17: 122.

587 27. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K, Cheng C, et al. Architecture of the
588 human regulatory network derived from ENCODE data. Nature. 2012; 489: 91-100.

589 28. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, et al. Sequence features and
590 chromatin structure around the genomic regions bound by 119 human transcription factors.
591 Genome Res. 2012; 22: 1798-1812.

592 29. Hébert JM, Rosenquist T, Götz J, Martin GR. FGF5 as a regulator of the hair growth cycle:
593 evidence from targeted and spontaneous mutations. Cell. 1994; 78: 1017-1025.

594 30. Housley D, Venta PJ. The long and the short of it: evidence that FGF5 is a major determinant of
595 canine 'hair'-itability. Anim Genet. 2006; 37: 309-315.

596 31. Drögemüller C, Rüfenacht S, Wichert B, Leeb T. Mutations within the FGF5 gene are
597 associated with hair length in cats. Anim Genet. 2007; 38: 218-221.

598 32. Higgins CA, Petukhova L, Harel S, Ho YY, Drill E, Shapiro L, et al. FGF5 is a crucial regulator
599 of hair length in humans. Proceedings of the National Academy of Sciences. 2014; 111:
600 10648-10653.

601 33. Legrand R, Tiret L, Abitbol M. Two recessive mutations in FGF5 are associated with the
602 long-hair phenotype in donkeys. Genet Sel Evol. 2014; 46: 1-7.

603 34. Pallotti S, Pediconi D, Subramanian D, Molina MG, Antonini M, Morelli MB, et al. Evidence of
604 post-transcriptional readthrough regulation in FGF5 gene of alpaca. Gene. 2018; 647: 121-128.

605 35. Guo J, Zhong J, Li L, Zhong T, Wang L, Song T, et al. Comparative genome analyses reveal the
606 unique genetic composition and selection signals underlying the phenotypic characteristics of
607 three Chinese domestic goat breeds. Genet Sel Evol. 2019; 51: 70.

608 36. Wang X, Cai B, Zhou J, Zhu H, Niu Y, Ma B, et al. Disruption of FGF5 in cashmere goats using
609 CRISPR/Cas9 results in more secondary hair follicles and longer fibers. Plos One. 2016; 11:
610 e164640.

611 37. Garner MM, Revzin A. The use of gel electrophoresis to detect and study nucleic acid—protein
612 interactions. Trends Biochem Sci. 1986; 11: 395-396.

613 38. Milde-Langosch K. The Fos family of transcription factors and their role in tumourigenesis. Eur
614 J Cancer. 2005; 41: 2449-2461.

615 39. Jia S, Peng J, Gao B, Chen Z, Zhou Y, Fu Q, et al. Relative quantification of protein-protein
616 interactions using a dual luciferase reporter pull-down assay system. Plos One. 2011; 6:
617 e26414.

618 40. Jin Y, Chen Z, Liu X, Zhou X. Evaluating the microRNA targeting sites by luciferase reporter
619 gene assay. Methods Mol Biol. 2013; 936: 117-127.

620 41. Jin Y, Chen Z, Liu X, Zhou X. Evaluating the microRNA targeting sites by luciferase reporter
621 gene assay. MicroRNA protocols: Springer. 2013. pp. 117-127.

622 42. Stables J, Scott S, Brown S, Roelant C, Burns D, Lee MG, et al. Development of a dual

623 glow-signal firefly and Renilla luciferase assay reagent for the analysis of G-protein coupled
624 receptor signalling. J Recept Sig Transd. 1999; 19: 395-410.

625 43. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule
626 sequencing and chromatin conformation capture enable de novo reference assembly of the
627 domestic goat genome. Nat Genet. 2017; 49: 643-650.

628 44. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform.
629 Bioinformatics. 2009; 25: 1754-1760.

630 45. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv
631 preprint arXiv:1303.3997. 2013.

632 46. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence
633 alignment/map format and SAMtools. Bioinformatics. 2009; 25: 2078-2079.

634 47. McCormick RF, Truong SK, Mullet JE. RIG: recalibration and interrelation of genomic
635 sequence data with the GATK. G3: Genes, Genomes, Genetics. 2015; 5: 655-665.

636 48. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, et al. Resequencing 302 wild and cultivated
637 accessions identifies genes related to domestication and improvement in soybean. Nat
638 Biotechnol. 2015; 33: 408-414.

639 49. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference
640 for whole-genome association studies by use of localized haplotype clustering. The American
641 Journal of Human Genetics. 2007; 81: 1084-1097.

642 50. Browning BL, Browning SR. Genotype imputation with millions of reference samples. The

643 American Journal of Human Genetics. 2016; 98: 116-126.

644 51. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating
645 and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. 2012; 6: 80-92.

646 52. Chang CC, Chow CC, CAM TL, Shashaank V, Purcell SM, Lee JJ. Second-generation PLINK:
647 rising to the challenge of larger and richer datasets. Gigascience. 2015: 1.

648 53. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait
649 analysis. Am J Hum Genet. 2011; 88: 76-82.

650 54. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0
651 for bigger datasets. Mol Biol Evol. 2016; 33: 1870-1874.

652 55. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated
653 individuals. Genome Res. 2009.

654 56. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated
655 individuals. Genome Res. 2009; 19: 1655-1664.

656 57. Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction
657 endonucleases. Proc Natl Acad Sci U S A. 1979; 76: 5269-5273.

658 58. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, Depristo MA, et al. The variant call
659 format and VCFtools. Bioinformatics. 2011; 27: 2156-2158.

660 59. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, et al. Whole-genome
661 resequencing reveals loci under selection during chicken domestication. Nature. 2010; 464:

587-591.

60. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989; 123: 585-595.

61. Bai WL, Yin RH, Yin RL, Jiang WQ, Wang JJ, Wang ZY, et al. Selection and validation of suitable reference genes in skin tissue of Liaoning cashmere goat during hair follicle cycle. Livest Sci. 2014; 161: 28-35.

62. He X, Chao Y, Zhou G, Chen Y. Fibroblast growth factor 5-short (FGF5s) inhibits the activity of FGF5 in primary and secondary hair follicle dermal papilla cells of cashmere goats. Gene. 2016; 575: 393-398.

63. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001; 25: 402.

Supporting information

S1 Fig. Distribution of mean size and number of ROH.

TGB, Toggenburg dairy goat from Heilongjiang; LaSh, Laoshan dairy goat from Shandong; XiDo, Xiangdong black goat from Hunan; LiNi, Liaoning cashmere goat

681 from Liaoning; ABS, Arbus cashmere goat; ELS, Erlangshan cashmere goat; ALS,
682 Alashan cashmere goat from Inner Mongolia; TBG, Tibetan Bangor cashmere goat;
683 TRT, Tibetan Ritu cashmere goat from Tibet; NPH, Nepalese Highland goat; NPL,
684 Nepalese Lowand goat; TPR, Tapri goat; KMR, Kamori goat; PTR, Pateri goat; BTR,
685 Bugi Toori goat.

686 **S2 Fig. Decay of LD in the goat genome for each breed.**

687 **S3 Fig. PCA result of the first and third components of Chinese goats.**

688 **S4 Fig. Genetic population structure of the 115 goats conducted by Admixture.**

689 The length of each colored segment represents the proportion of the individual
690 genome inferred from ancestral populations (K=2-7).

691 **S5 Fig. Migration analysis of 115 goats by Treemix software.**

692 (A-F) panels represent models of population affinities assuming 0-3 and 5-6 migration
693 edges in TreeMix, respectively. The inferred migration weight is provided by the
694 color of the arrow displayed.

695 **S6 Fig. The overlapped regions for selection signatures.**

696 The overlapped regions by top 5% highest F_{ST} and θ_{π} ratio (noncash/cash) and ZHp
697 for cashmere goats.

698 **S1 Table. Population distribution.**

699 **S2 Table. Mapping coverage and depth.**

S3 Table. SNP summary statistics in the goat breeds.

S4 Table. Genome-wide heterzygosity of goat breeds.

S5 Table. F_{ST} selection signatures with the windows by top 1% highest.

S6 Table. θπ selection signatures with the windows by top 1% highest.

S7 Table. ZHp selection signatures with the windows by top 1% highest.

S8 Table. Overlapped regions by top 5% highest F_{ST}, θπ and ZHp selection signatures.

S9 Table. Enrichment analyses of GO terms.

S10 Table. Populations for validation.

S11 Table. Primers for qPCR.

S12 Table. θπ values of genomic regions around gene *FGF5*.

S13 Table. Tajima'*D* values of genomic regions around gene *FGF5*.

S14 Table. F_{ST} values of genomic regions around gene *FGF5*.

S15 Table. Depth values of genomic regions around gene *FGF5*.

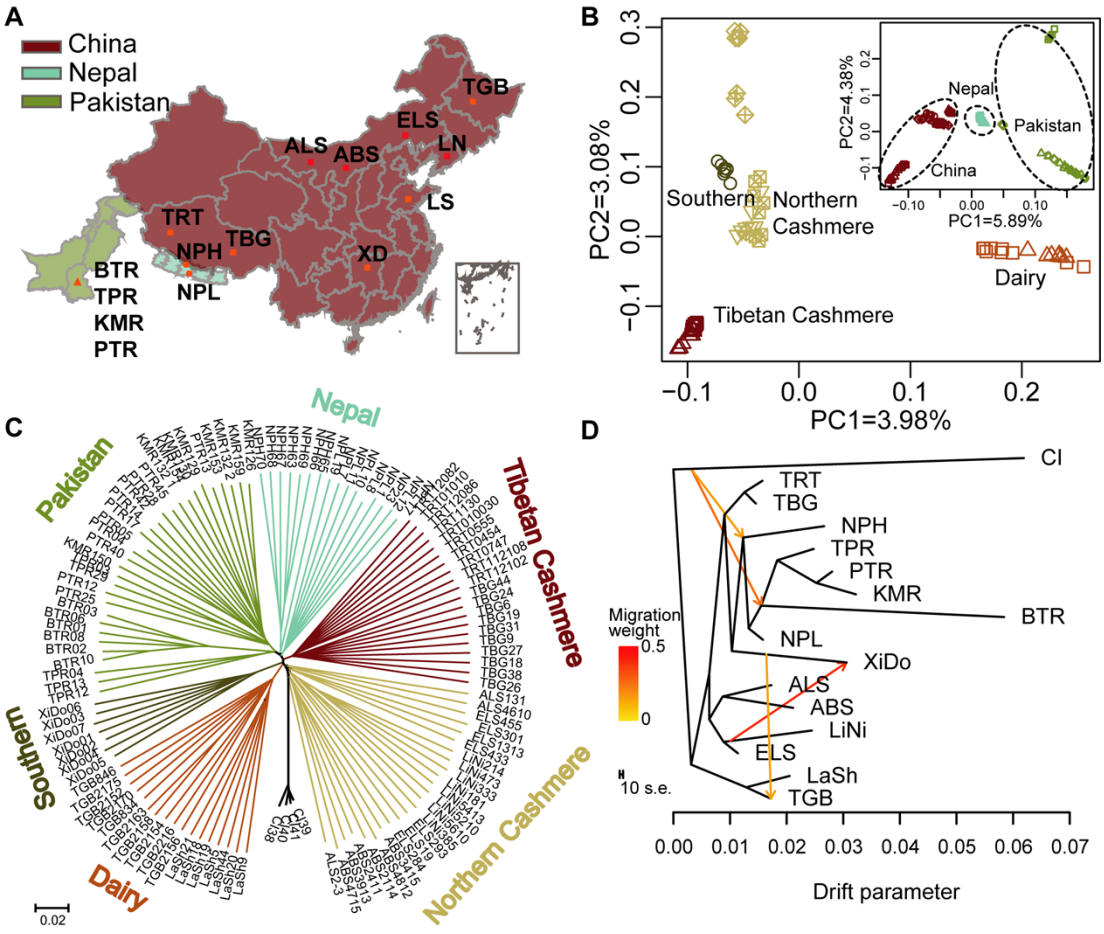
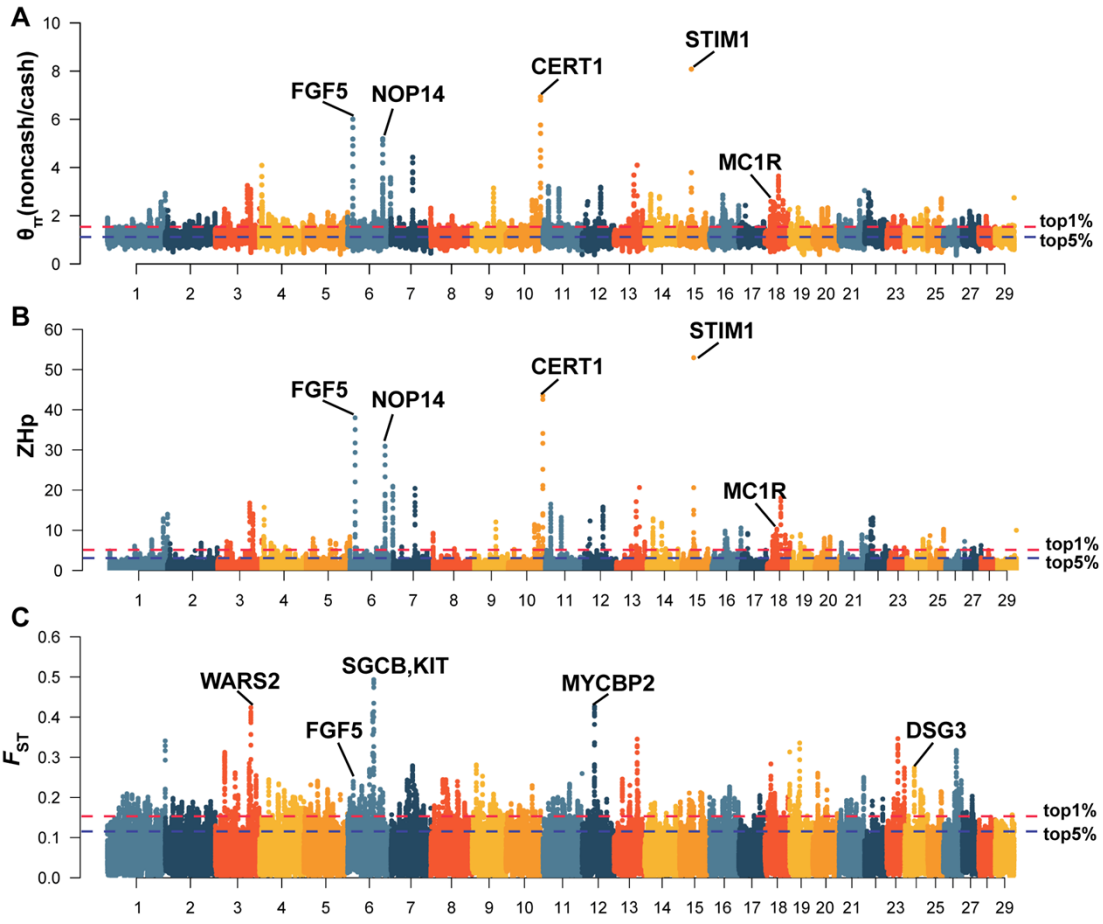


Fig 1. Geographic distribution, genetic structure of Chinese, Nepalese and Pakistani goat breeds.

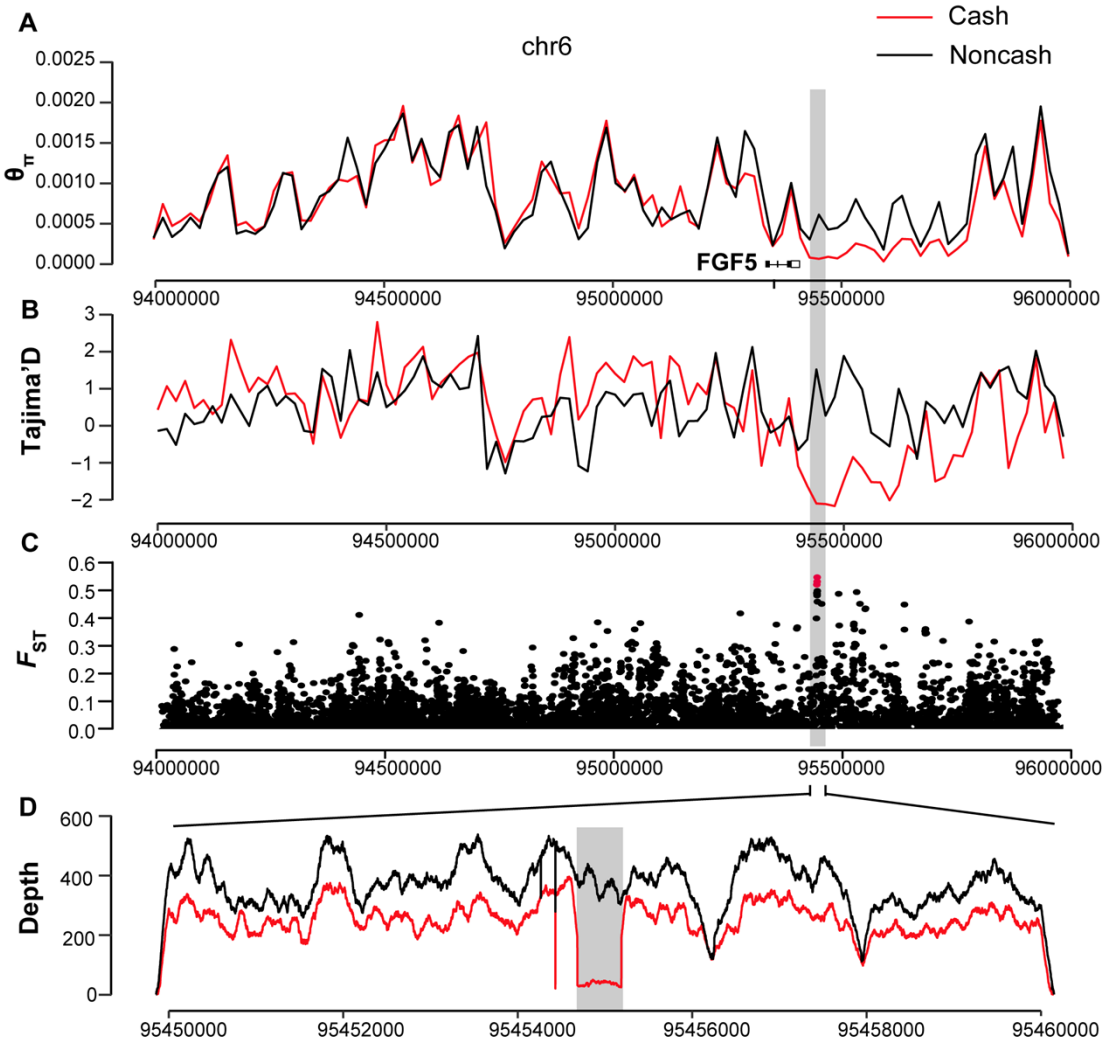
(A) The geographic distribution of 15 goat populations. The red color represent Chinese goats (TGB, Toggenburg dairy goat from Heilongjiang; LaSh, Laoshan dairy goat from Shandong; LiNi, Liaoning cashmere goat from Liaoning; ABS, Arbus cashmere goat; ELS, Erlangshan cashmere goat; ALS, Alashan cashmere goat from Inner Mongolia; TBG, Tibetan Bangor cashmere goat; TRT, Tibetan Ritu cashmere goat from Tibet; XiDo, Xiangdong black goat from Hunan). The blue color represents Nepalese goats (NPH, Nepalese Highland goat; NPL, Nepalese Lowand goat). The green color represents Pakistani goats (BTR, Bugi Toori goat; KMR, Kamori goat; PTR, Pateri goat; TPR, Tapri goat). (B) PCA plots of the first two components of all goats

(inner plot) and Chinese goats (outer plot). The fraction of the total variance explained is reported on each individual axis between parentheses. (C) The neighbor-joining tree of the goat breeds, with *Capra ibex* as the outgroup. Bootstrap reported was close to 100%. (D) The ML-TreeMix tree of all goats, with *Capra ibex* as the outgroup, assuming four migration events. Migration arrows are colored according to their weights. Horizontal branch lengths are proportional to the amount of genetic drift parameter that has occurred on the branch. The drift parameter measures the variance in allele frequency estimated along each branch of the tree. The yellow and orange lines indicate the instantaneous admixtures, whereas arrows denote continuous (unidirectional) genes flow.



735 **Fig 2. Positive selection scans for cashmere growth.**

736 Cashmere goats are compared with non-cashmere goats. The nucleotide diversity θπ ratio
737 (θπ-noncash/θπ-cash) (A),the transformed heterozygosity score ZHp (B) and the population
738 genetic differentiation F_{ST} values (C) are calculated within 100 kb sliding windows (step size 10
739 kb). The significance threshold of a selection signature was arbitrarily set to the top 5% percentile
740 outliers for each individual test and is indicated with blue horizontal dashed lines. The red
741 horizontal dashed lines delineate the top 1% quantile.

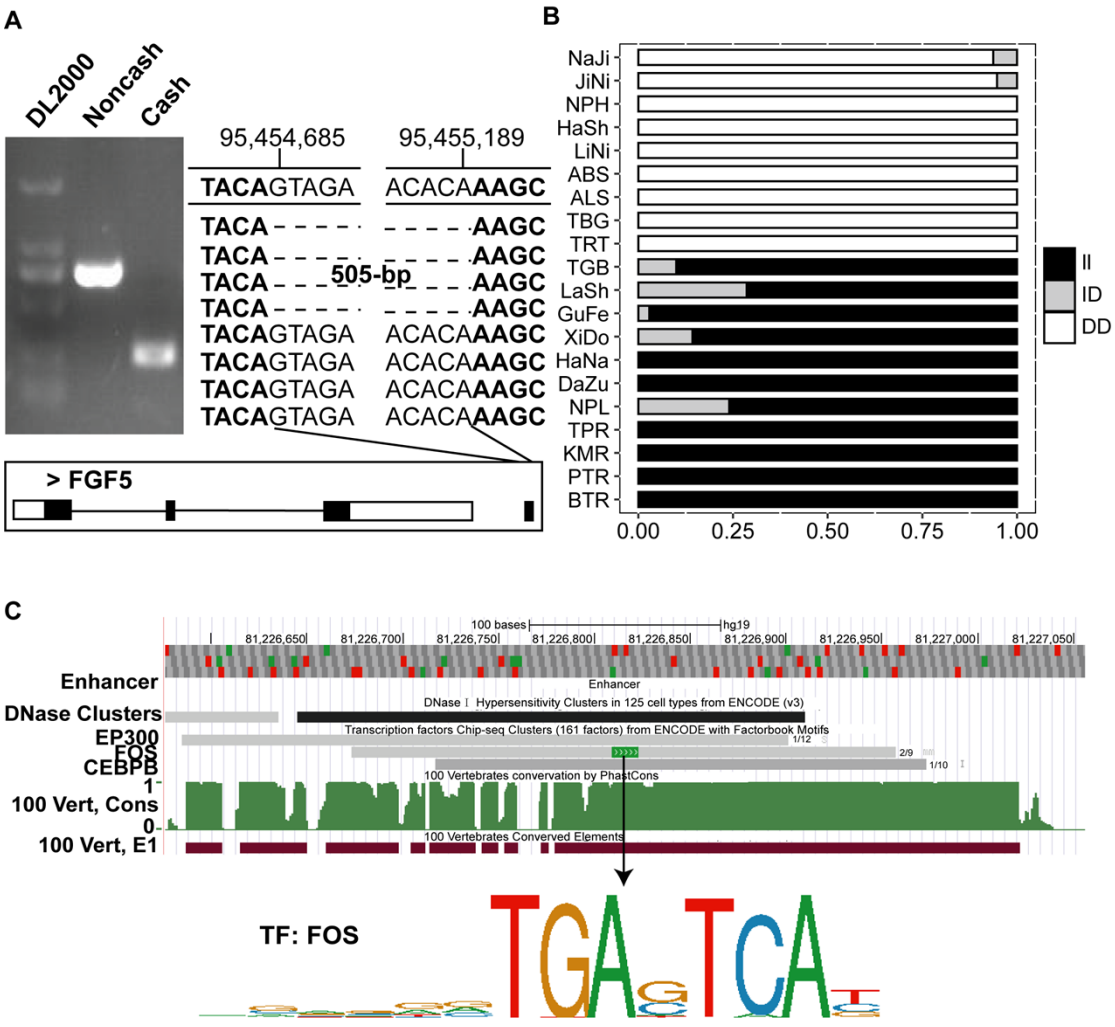


742

743

744 **Fig 3. The strongest positive selection signature around the *FGF5* peak.**

745 The θ_π ratio (A), Tajima's *D* (B) and F_{ST} value (C) are plotted against the peak position from
746 94.0 Mb to 96.0 Mb, and the read depth value (D) is plotted against the peak position from 95.45
747 Mb to 95.46 Mb on chromosome 6. Both θ_π ratio and Tajima's *D* values were based on a 20 kb
748 window and a 20 kb step. The red and the black lines represent cashmere and non-cashmere goats,
749 respectively. The gray columns represent the strongest positive selection signature in the region
750 considered. The small black boxes and short lines represent the gene structure of *FGF5*.

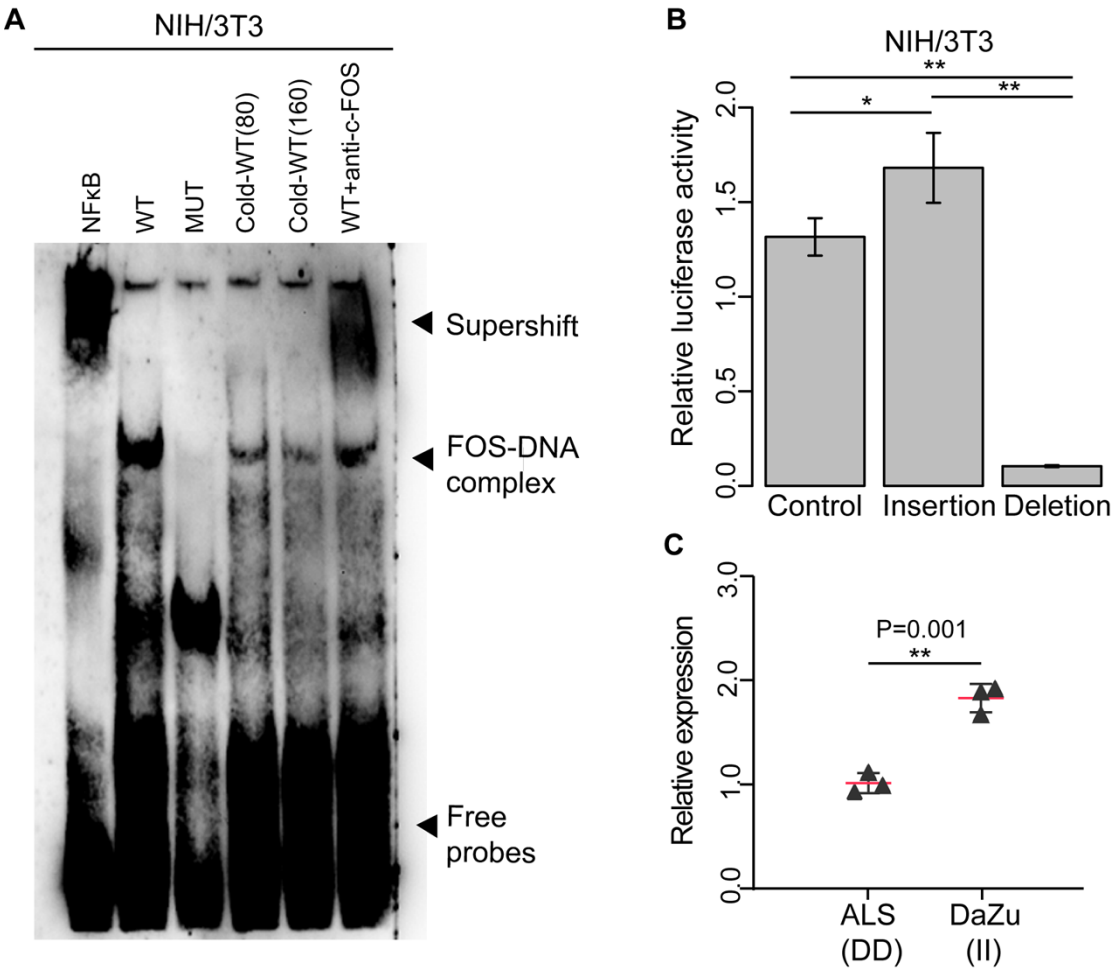


751

752 **Fig 4. Annotation of the indel variant in the *FGF5* gene showing positive selection**

signatures.

(A) The PCR amplification of 505-bp indel variant, generating a 267-bp fragment in all cashmere goats while a 772-bp fragment in non-cashmere goats. The indel is located at position 95,454,689-95,455,189 of chromosome 6 in the downstream of *FGF5* gene. (B) Genotypes of indel were determined in a larger population (N = 288 goats). II represents homozygous insertion genotype; ID represents heterozygous indel genotype; DD represents homozygous deletion genotype. Cashmere goat breeds include NaJi (Nanjiang cashmere goat), JiNi (Jining grey goat), NPH (Nepalese highland goat), HaSh (Hanshan white cashmere goat), LiNi (Liaoning cashmere goat), ABS (Arbus cashmere goat), ALS (Alashan cashmere goat), TBG (Tibetan Bangor cashmere goat) and TRT (Tibetan Ritu cashmere goat). Non-cashmere goat breeds include TGB (Toggenburg dairy goat), LaSh (Laoshan dairy goat), GuFe (Guangfeng goat), XiDo (Xiangdong black goat), HaNa (Hainan black goat), DaZu (Dazu black goat), NPL (Nepalese lowland goat), BTR (Bugi Toori goat), KMR (Kamori goat), PTR (Pateri goat) and TPR (Tapri goat). (C) The insertion fragment of *FGF5* gene in humans contains a highly conserved FOS transcription factor binding site (TGAGTCA) in the UCSC database.



768

769 **Fig 5. Validation of the indel variant in the *FGF5* gene.**

770 (A) Electrophoretic mobility shift assays (EMSAs) using the nuclear protein from NIH/3T3 cells.

771 NFκB acts as a positive control (lane 1). WT and MUT represent the probe containing the

772 wildtype FOS binding site and the probe excluding the FOS binding site (lane 2 and 3),

773 respectively. The cold competitions of the protein complex formation by 80 and 160 fold over that

774 of wildtype probe (lane 4 and 5). The clear supershift with anti-c-FOS antibody mixtured to

775 wildtype react (lane 6). (B) Dual-luciferase activity assay of the NIH/3T3 cell lysates

776 cotransfected with the pGL4.74 internal reference plasmid and the pGL4.23 empty vector as

777 control, the pGL4.23 recombinant plasmids of the insertion or the deletion variant. (C) The qPCR

778 gene expression of the *FGF5* gene in the skin of ALS (Alashan cashmere goat) and DaZu black

779 goats (non-cashmere goat). * and ** displayed the statistical significance of *P-values* <0.05 and

780 0.01, respectively.

781