



Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: [www.elsevier.com/locate/jare](http://www.elsevier.com/locate/jare)

## Original Article

Two mutations at *KRT74* and *EDAR* synergistically drive the fine-wool production in Chinese sheep

Benmeng Liang<sup>a,b,c,1</sup>, Tianyou Bai<sup>a,c,1</sup>, Yuhetian Zhao<sup>a,c,1</sup>, Jiangang Han<sup>a,c,d</sup>, Xiaohong He<sup>a,c</sup>, Yabin Pu<sup>a,c</sup>, Chunxin Wang<sup>e</sup>, Wujun Liu<sup>f</sup>, Qing Ma<sup>g</sup>, Kechuan Tian<sup>h,i</sup>, Wenxin Zheng<sup>i</sup>, Nan Liu<sup>j</sup>, Jianfeng Liu<sup>b,\*</sup>, Yuehui Ma<sup>a,c,\*</sup>, Lin Jiang<sup>a,c,\*</sup>

<sup>a</sup> National Germplasm Center of Domestic Animal Resources, Ministry of Technology, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China

<sup>b</sup> National Engineering Laboratory for Animal Breeding and MOA Key Laboratory of Animal Genetics and Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

<sup>c</sup> Key Laboratory of Livestock and Poultry Resources (Cattle) Evaluation and Utilization, Ministry of Agriculture and Rural Affairs, China

<sup>d</sup> Animal Genomics Laboratory, UCD School of Agriculture and Food Science, UCD College of Health and Agricultural Sciences, University College Dublin, Belfield, Dublin D04 V1W8, Ireland

<sup>e</sup> Institute of Animal Sciences, Jilin Academy of Agricultural Sciences, Gongzhuling 136100, China

<sup>f</sup> College of Animal Science, Xinjiang Agriculture University, Urumqi, Xinjiang, China

<sup>g</sup> Institute of Animal Science, Ningxia Academy of Agriculture and Forestry Sciences, Yinchuan 75002, Ningxia, China

<sup>h</sup> Institute of Animal Science and Veterinary Medicine, Shandong Academy of Agricultural Sciences, Jinan, China

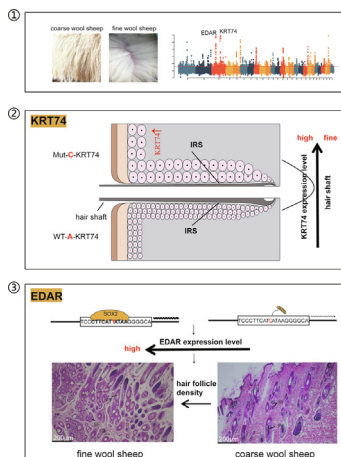
<sup>i</sup> Xinjiang Academy of Animal Science, China

<sup>j</sup> College of Animal Science and Technology, Qingdao Agricultural University, China

## HIGHLIGHTS

- We find two functional mutations driving the finer and denser wool production.
- One C/A missense variant of *KRT74* drive finer wool production in Chinese fine wool sheep.
- The C-KRT74 can activate the KRT74 protein and affect wool fineness.
- One T/C SNP in the regulatory region upstream of *EDAR* drive denser wool production in fine wool sheep.
- The C-to-T mutation can upregulate *EDAR* mRNA expression and potentially affect wool density.

## GRAPHICAL ABSTRACT



## ARTICLE INFO

## Article history:

Received 2 November 2022

Revised 13 April 2023

Peer review under responsibility of Cairo University.

\* Corresponding authors at: National Germplasm Center of Domestic Animal Resources, Ministry of Technology, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (CAAS), Beijing, China (Y. Ma and L. Jiang); National Engineering Laboratory for Animal Breeding and MOA Key Laboratory of Animal Genetics and Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China (J. Liu)

E-mail addresses: [liujf@cau.edu.cn](mailto:liujf@cau.edu.cn) (J. Liu), [yuehui.ma@263.net](mailto:yuehui.ma@263.net) (Y. Ma), [jianglin@caas.cn](mailto:jianglin@caas.cn) (L. Jiang).

<sup>1</sup> These authors contributed equally to this work.

<https://doi.org/10.1016/j.jare.2023.04.012>

2090-1232/© 2023 The Authors. Published by Elsevier B.V. on behalf of Cairo University.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Please cite this article as: B. Liang, T. Bai, Y. Zhao et al., Two mutations at *KRT74* and *EDAR* synergistically drive the fine-wool production in Chinese sheep *KRT74* and *EDAR* synergistically drive the fine-wool production →, Journal of Advanced Research, <https://doi.org/10.1016/j.jare.2023.04.012>

Accepted 14 April 2023  
Available online xxxx

#### Keywords:

Chinese fine wool sheep  
Whole genome sequencing  
Genomic selection signature  
Follicle density  
Missense  
Regulatory

**Objectives:** This study aims to clarify the underlying genetic basis for the denser and finer wool phenotype in fine-wool breeds.

**Method:** Whole-genome sequences of 140 samples, Ovine HD630K SNP array data of 385 samples, including fine, semi-fine, and coarse wool sheep, as well as skin transcriptomes of nine samples were integrated for genomic selection signature analysis.

**Results:** Two loci at *keratin 74* (*KRT74*) and *ectodysplasin receptor* (*EDAR*) were revealed. Fine-scale analysis in 250 fine/semi-fine and 198 coarse wool sheep narrowed this association to one C/A missense variant of *KRT74* (OAR3:133,486,008,  $P = 1.02E-67$ ) and one T/C SNP in the regulatory region upstream of *EDAR* (OAR3:61,927,840,  $P = 2.50E-43$ ). Cellular over-expression and ovine skin section staining assays confirmed that C-*KRT74* activated the *KRT74* protein and specifically enlarged cell size at the Huxley's layer of the inner root sheath ( $P < 0.01$ ). This structure enhancement shapes the growing hair shaft into the finer wool than the wild type. Luciferase assays validated that the C-to-T mutation upregulated *EDAR* mRNA expression via a newly created SOX2 binding site and potentially led to the formation of more hair placodes.

**Conclusions:** Two functional mutations driving finer and denser wool production were characterized and offered new targets for genetic breeding during wool sheep selection. This study not only provides a theoretical basis for future selection of fine wool sheep breeds but also contributes to improving the value of wool commodities.

© 2023 The Authors. Published by Elsevier B.V. on behalf of Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

Sheep (*Ovis aries*) was the first domesticated livestock species in the fertile crescent 10,500 years BP [1]. Initially, sheep were reared for meat; however, during the fifth millennium BP, specialization for secondary products, such as milk and wool, is believed to have occurred [2]. Currently, approximately 1.76 million tons of clean wool are produced by over one billion sheep globally per year and Australia, China and New Zealand are top three wool producers (FAOSTAT 2021) [3]. Although China accounts for more than 20% of the total wool production, it still needs to import large amount of sheep wool per year, for example, ~0.24 million tons imported in 2022 [4]. Therefore, it is important to continuously improve the wool production to meet the industry need. Previous study found that hair follicle density was positively correlated with wool yield ( $0.35 \pm 0.19$ ) and negatively correlated with wool fiber diameter ( $-0.65 \pm 0.12$ ), showing the density of hair follicles affects the fineness and yield of wool [5]. Fine wool breeds have an over three-fold higher ratio of secondary versus primary hair follicles as well as a fiber diameter that is twice as small compared to other breeds [6]. Merino sheep, the most renowned fine wool breed for high rate of fine wool production, thus have been introduced into various countries to improve the performance of native sheep breeds [7]. Furthermore, Romney, Lincoln, and Corriedale sheep have substantially influenced the formation of semi-fine wool breeds [8]. From the perspective of fine-wool breeding, the identification of gene and molecular genetic markers related to the fineness and density of wool can greatly facilitate the production and quality of fine wool sheep.

Some studies have been conducted on this phenotype variation to determine the genetic basis and molecular mechanism underpinning wool formation in sheep; however, very limited reliable loci have been found in the sheep genome. For instance, the genome wide association analysis (GWAS) of French Romane, a composite breed of the short woolly sheep and long hairy Romanov sheep, revealed an insertion of the anti-sense EIF2S2 retrogene into the 3'UTR of the *IRF2BP2* gene for the woolly traits [9]. Genomic analysis revealed that the *KRT* gene family and desmoglein 4 (*DSG4*) gene were the significant selection signatures in 47 sheep (including the Chinese Merino, Altay, Tan, and Shetland sheep breeds) [10]. In addition, integrated analysis of transcriptome, methylation, and GWAS in Merino sheep demonstrated that *SPHK1* is associated with the coefficient of variation of fiber diameter and

hair follicle development [11]. Although such studies have contributed to improving knowledge of the genetic basis of fine wool traits in sheep, the underlying causal mutations and the corresponding mechanisms remain unclear due to the low density of SNP arrays and the limited number of sheep breeds.

In this study, higher density Ovine HD630K SNP array data, whole genome-sequencing assays, and skin transcriptome analysis were integrated into the most comprehensive panel of over 800 sheep samples to elucidate the molecular basis of wool trait variations for Chinese fine and coarse wool sheep breeds. First, the genomic selection signature analysis of whole genome-sequencing assays of 140 sheep and Ovine HD630K SNP array data of 385 sheep representing 16 populations, along with the integration of skin transcriptomes were performed to show extensive wool production variation. Second, the validation of the identified genes was conducted in an extra 450 sheep samples. Then, the statistical association between genetic and phenotypic variation was done to narrow down the results to one missense SNP within *KRT74* and one *cis*-regulatory SNP upstream *EDAR* in 250 fine wool and 198 coarse sheep. Third, these two variants were both functionally validated by using over-expression assays in two independent cell lines and immunohistochemistry analysis of the ovine skin sections. Finally, the mutant *KRT74* allele and *EDAR* upstream mutation were found to synergistically contribute to the smaller fiber diameter and the denser hair follicles of the Merino sheep.

## Materials and methods

### Sample collection

A total of 836 independent individuals from 21 sheep populations, including 343 samples from nine fine wool sheep populations, 180 samples from five semi-fine wool sheep populations, 310 samples from seven coarse wool sheep populations, and three samples from one wild sheep population were collected for analysis in this study (Supplementary Table S1). Sampling of the fine and semi-fine wool sheep is the most widespread and largest survey to date. The unrelatedness of the individual sheep was guaranteed via the knowledge of local herders and family trees. The wool type was ascertained by visual observation and/or a summary of early studies (Supplementary Table S1). Genomic DNA extraction was completed by Niuqin Bio-enterprise (Beijing, China) according

to a paramagnetic particle protocol (LGC) and met the standards for SNP chip data detection.

The scapular skin samples of the adult Xinji Merino and Tan sheep were taken from Yanchi (Ningxia) and Songyuan (Jilin), respectively. Three biological replicates were collected per breed. The skin samples were stored in RNALater (QIAGEN) and 4% paraformaldehyde (Solarbio) for q-PCR and immunohistochemical research. Total RNA extraction from skin tissue was performed in strict accordance with the standard steps of the kit (TIANGEN).

#### Ethics statement

The animal experiments of our study fulfilled the ethical requirement of the Animal Care and Use Committee of our institute, the Chinese Academy of Agricultural Sciences, Institute of Animal Science (IAS2019-57).

#### SNP BeadChip genotyping and quality control

A total of 385 samples from 185 Merino-derived sheep, 88 semi-fine wool sheep, and 112 coarse wool sheep were genotyped using the Thermo Fisher/Affymetrix Genomic Geneseeq Profiler Ovine HD630K BeadChip (**Supplementary Table S2**). Three wild sheep were out-group (**Supplementary Table S2**). The 185 Merino sheep-derived samples were from nine populations and included 27 Subo Merino from Gongnaisi (SuBo), 28 Xinjiang Merino from Ziniquan (ZiNiQ), 20 Xinjiang Merino from Baicheng (BaiCh), 19 Xinjiang Merino from Urumqi (Urumq), 20 Qinghai Wool-mutton Type sheep from Sanjiaocheng (QinHai), 21 Gansu Alpine Merino from Huangcheng (GanSuA), 20 Aohan Merino from Aohan (AoHan), 20 Xinji Merino from Songyuan (Xinji) and 10 Merino from Australia (AusMer). The 88 semi-fine wool sheep individuals were from four sheep populations and included 21 Liangshan semi-fine wool sheep (LiaSh), 27 Guizhou semi-fine wool sheep (GuiZh), 20 Yunnan semi-fine wool sheep (YunNa), and 20 Romney sheep (Romn). The 112 coarse wool sheep samples were from 37 Qinghai Mongolian sheep (QinHaiM), 15 Altay sheep (ALT), and 60 Tan sheep (Tan). The genotyping for extracted DNA was mainly completed using the Niuqin Bio-enterprise (Beijing, China) based on the Ovine HD630K SNP array. A total of 385 individual sheep were genotyped and 499,098 SNPs were generated (based on genome Oar\_v4.0). Three wild sheep individuals were included as out-group SNP data. SNP data quality control was conducted using PLINK v1.90 software and SNPs that met any of the following three criteria were retained: (1) the detection rate of sites and individuals was greater than or equal to 90%; (2) MAF was not smaller than 0.01; (3) Hardy-Weinberg equilibrium ( $p < 1e-5$ ). Finally, 388 individuals and 473,419 common SNPs were used for the diversity, population structure, and phylogenetic analysis, while 385 sheep individuals and 472,773 common SNPs were used for the genomic selection signature analysis.

#### Whole-genome sequencing and variant calling

Based on the BeadChip SNP data, this study newly sequenced the genomes of 120 samples including those from 27 fine wool sheep, 28 semi-fine wool sheep, and 65 coarse wool sheep. These were combined with downloaded genomes from 20 Tan and Australian Merino sheep (SRP066883) [12] for the whole genome sequencing analysis. The 27 Merino derived sheep samples were from 13 Subo Merino from Gongnaisi (SuBo) and 14 Xinjiang Merino from Ziniquan (ZiNiQ). The 28 semi-fine wool sheep samples were from 10 Liangshan semi-fine wool sheep (LiaSh), nine Guizhou semi-fine wool sheep (GuiZh), and nine Yunnan semi-fine wool sheep (YunNa). The 65 coarse wool sheep samples were from 15 Altay sheep and 50 Tan sheep (**Supplementary Table S9**). Gen-

ome library construction and sequencing were completed based on the Hiseq XTen platform (2x150 double-ended mode) and utilized the Illumina normal process (Illumina Inc.) via the BGI Genomics and Compass Biotechnology Company (Beijing, China). The size of the insert fragment was 350 bp. The sheep reference genome was Oarv4.0 [13] and was used for sequence alignment according to previously published methods [14]. In low-quality raw and adapter sequences were removed using the Trimmomatic-0.33 [15]. The reference genome (Oar\_v4.0) was indexed and sequence alignment was done using the command of index and BWA-mem by Burrows-Wheeler Algorithm (BWA) software, respectively [16]. The MarkDuplicates command in Samtools was used to filter paired reads matched to the precise identical location on the Oar\_v4.0 sheep reference genome. Sequence data of an average  $10.27 \times$  depth, 95.40% coverage, and 34,224,113 SNPs were obtained (**Supplementary Table S10**). SNP variations identified in each sample and SNP calling of multiple samples were completed via previously published methodology [17]. SNPs that did not fit the following four criteria were removed: (1) average sequencing depth  $\geq 3 \times$  or  $\leq 30 \times$ ; (2)  $MAF \geq 0.05$ ; (3) missing rate  $\leq 0.1$ ; and (4) the locus was biallelic. After quality control, 139 individual sheep and 23,026,028 common SNPs were retained for downstream analysis.

#### Population structure, phylogenetic, and diversity analysis

All 473,419 BeadChip SNPs were filtered using PLINK 1.90 based on the parameter (--indep-pairwise 1000 5 0.5). A total of 277,375 independent SNPs were obtained for subsequent analyses of population structure, phylogenetics, and diversity. PLINK 1.90 was also used for the Principal Component Analysis (PCA) and Neighbor-Joining (NJ) tree. The visualization of the NJ tree results was performed using the Interactive Tree Of Life (ITOL, <https://itol.embl.de>). The observed  $H_o$ , together with the expected  $H_e$  heterozygosity was calculated using the filtered SNPs to assess the within-population genetic diversity (Supplementary Table S3). Linkage disequilibrium (LD) was calculated between pairs of autosomal SNPs by LDdecay ([github.com/BGI-shenzhen/PopLDdecay](https://github.com/BGI-shenzhen/PopLDdecay)).  $\theta_\pi$  values and population-differentiation values ( $F_{ST}$  fixation) for diversity estimation for each population were calculated based on a sliding window size of 100 kb.

#### Genomic selection signatures

This study detected the genomic selection signatures associated with fine wool traits both in the BeadChip SNP data and the whole-genome sequences. For the BeadChip SNP data, the fine and semi-fine wool sheep included 273 samples from 13 sheep populations (AusMer = 10, Urumq = 19, BaiCh = 20, ZiNiQ = 28, SuBo = 27, Xinji = 20, Aohan = 20, GanSuA = 21, QinHai = 20, LiaSh = 21, GuiZh = 27, YunNa = 20, and Romn = 20) and 112 coarse wool sheep samples from three sheep populations (QinHaiM = 37, ALT = 15, and Tan = 60) (**Supplementary Table S2**). For the whole genome sequences, the fine and semi-fine wool sheep included 64 samples from six sheep populations (AusMer = 10, ZiNiQ = 13, SuBo = 13, LiaSh = 10, GuiZh = 9, and YunNa = 9) and 75 coarse wool sheep samples from two sheep populations (Altay = 15 and Tan = 60) (**Supplementary Table S9**). The genetic differentiation,  $F_{ST}$  fixation [18], the nucleotide diversity ratio,  $\theta_\pi$  (coarse/fine) [19], and the transformed heterozygosity value, ZHP [20] were then calculated to identify the positive selection signals in the fine wool sheep.  $F_{ST}$ ,  $\theta_\pi$  (coarse/fine), and ZHP were calculated via VCFtools with parameters set to 100 kb window and 15 kb step [21,22]. The top 1% overlapping windows among the three statistics were further annotated as potential targets according to the

BioMart database (<https://www.biomart.org/>). The Database for Annotation, Visualization, and Integrated Discovery (DAVID v6.7) was applied to functional enrichment (<https://david.abcc.ncifcrf.gov/summary.jsp>). The enrichment significance was measured using the false discovery rate (FDR) and Benjamin-Ochberg and Bonferroni multiple testing correction.

#### Functional SNPs

To zoom in on the top selected regions, Vcftools was used for the single SNP calculation of  $F_{ST}$ ,  $\theta_{\pi}$ , and Tajima's D. The basic parameters were set to a 20-kb window with a 20-kb sliding step. To predict functional candidates, the evolutionary conservation scores of the most significant variants among 22 mammals were calculated. The PolyPhen online tool ([genetics.bwh.harvard.edu/pph2/](http://genetics.bwh.harvard.edu/pph2/)) was then used to estimate the probable functional effect for each nonsynonymous variant.

#### Validation in an extended population

An additional 448 individuals from 21 sheep populations were used for extended genotyping of the identified variations. Of the 21 sheep populations, nine were fine wool, five were semi-fine wool, and seven were Chinese native coarse populations (**Supplementary Table S24**). The Kompetitive Allele-Specific PCR (KASP, China Golden Marker Beijing Biotech Co., Ltd.) as a high throughput strategy was adopted for genotyping the *KRT74* missense mutation and the *EDAR* upstream mutation (*KRT74*-SNP2 and *EDAR*-SNP1; **Supplementary Table S23**) in of the above 448 sheep. The principle of the KASP primer design was that it must carry standard FAM and HEX tails while carrying the target SNP at the 3' end (**Supplementary Table S37**).

#### Cell maintenance

The A549 (adenocarcinoma human alveolar basal epithelial cells) and MCF10A (human breast epithelial cell) were purchased from Peking Union Medical College Hospital. The A549 cells were cultured in Roswell Park Memorial Institute 1640 (RPMI1640) medium supplemented with 10% FBS and 1% penicillin and streptomycin (Gibco, USA). The MCF10A were cultured in DMEM/F12 medium (Gibco) supplemented with 5% horse serum and 10  $\mu$ g/ml insulin, 20 ng/ml epidermal growth factor, 100 ng/ml cholera toxin, 0.5  $\mu$ g/ml hydrocortisone, and 1% penicillin & streptomycin. The A549 and MCF10A cells were incubated in a Cell Thermostat Carbon Dioxide Culture Box (Thermo Fisher) at 37 °C and 5% CO<sub>2</sub>.

#### Generation/Construction of expression vectors

The pEGFP-N1 vector was provided by Shanghai Genaray Biotech Co Ltd. The amplified coding sequences of the wild-type *KRT74* or the mutant (p.H123N) were cloned into *NheI* and *EcoRI* sites of the mammalian expression vector pEGFP-N1. The generated vectors were designated pEGFP-N1-Wt-A-*KRT74* and pEGFP-N1-Mut-C-*KRT74*, respectively. The Qiagen Endotoxin-Free Plasmid Extraction Kit was used to extract these recombinant plasmids (Qiagen).

#### Transfection

A549 and MCF10A cells were each plated in a 3.5 cm petri dish (Nalge Nunc International, Rochester, NY) 24-h before transfection. The recombinant plasmids were transfected into A549 and MCF10A cells utilizing Lipofectamine 3000 (Invitrogen, America), according to the manufacturer's guidelines. First, 3.75  $\mu$ L Lipofec-

tamine 3000 Reagent (Invitrogen, America) and 2.5  $\mu$ L Endofree plasmid were mixed in 125  $\mu$ L Opti-MEM® I Reduced Serum medium, GlutaMAX™ (Thermo Fisher Scientific, USA). The 1:1 combination solution was incubated for 15 min at RT before it was added to the cells in the 3.5 cm plates. The cells were then incubated at 37 °C and 5% CO<sub>2</sub> for approximately 24–48 h. A confocal microscope (Leica TCS SP8, Germany) was then used to observe the morphology of the cells. In parallel, the cell size was measured by ImageJ software following the directions provided by the manufacturer.

#### qPCR quantification

The cells were washed in PBS and collected, and the total RNA was extracted using an RNA extraction kit (Promega) following the standard procedure provided by the kit. Reverse transcription was then performed using the PrimeScript TR reagent Kit (Takara). Quantitative PCR analysis of *KRT74* (forward primer: ATGAGCCGG-CAACTGAATGT; reverse primer: CTGTAGAGGCTTCGGCTTCC) was run on an ABI7500 sequence detection system (Life Technologies, Germany). The ubiquitous  $\beta$ -actin gene (forward primer: GAAGAT-CAAGATCATTGCTCCT; reverse primer: TACTCTGCTTGCTGATCCA) and *GAPDH* (forward primer: TGATGCTGGTGCTGAGTACG; reverse primer: GGTTACGCCCCATCACAAC) served as reference genes ([icg.big.ac.cn/index.php/Homo\\_sapiens#Internal\\_Control\\_Genes\\_3](http://icg.big.ac.cn/index.php/Homo_sapiens#Internal_Control_Genes_3)). The qPCR was carried out using a Power SYBR Green PCR reagent kit (Applied Biosystems). The primers are shown in **Supplementary Table S36**, and the qPCR conditions are shown in **Supplementary Table S38**. Each sample underwent three biological replicates, and the average value was utilized for subsequent analysis. Finally, a classical  $2^{-\Delta\Delta CT}$  means was applied to determine fold expression changes [23].

#### Cell proliferation detection

Cell proliferation detection assays with Alexa Fluor 555 were done using the BeyoClick™ EdU Cell Proliferation Kit (Beyotime) with six biological replicates per group. After 16 h of transfection, the cell proliferation assay was performed. The 2X EdU working solution was first prepared, and the volume of EdU working solution and culture medium was added to the 96-well plate so that the final concentration of EdU was 10  $\mu$ M (1X). The transfected A549 and MCF10A were pulse-labeled with EdU for 2 and 4 h, respectively. Untreated cells were used as a control group. Detection of EdU incorporation into cellular DNA was accomplished as required using the Click-iT® EdU Alexa Fluor® 555 Cell Proliferation Assay Kit (Molecular Probes, Invitrogen). Briefly,  $1 \times 10^6$  collected cells were washed twice with PBS and then fixed in 4% paraformaldehyde. The cells were then incubated for 15 min at RT in the dark, washed three times in the permeabilization reagent, and incubated with the wash buffer for 15 min. Finally, the cells were incubated in the pre-prepared Click-iT® EdU reaction cocktail at RT for 30 min in the dark, washed three times, and submitted for microscopy detection.

#### RNA-seq data analysis

The embryonic and postpartum transcriptome data of sheep skin tissue were downloaded from the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) database, and mainly included six periods: E105, E135, P7, P30, P90, and P360. The accession numbers of BioProject were PRJNA705554 [11] and PRJNA765722 [24]. The detailed analysis steps were performed following the previous methodology [24,25]. The version of the ovine genome assembly was Oar\_v4.0 (GCA\_000298735.2,



[https://www.ncbi.nlm.nih.gov/assembly/GCA\\_000298735.2](https://www.ncbi.nlm.nih.gov/assembly/GCA_000298735.2)

### Dual-luciferase reporter analysis

The 171-bp sequence upstream of the *EDAR* gene containing the wild-type (C), mutant (T), and 20-bp deletion were inserted into the recombination sites of *MluI* and *BglII* in the luciferase reporter pGL3 vector (Promega). The recombinant plasmids, named pGL3-basic-T, pGL3-basic-C, and pGL3-basic-Del, were extracted using the Endotoxin-Free Plasmid Extraction Kit (Qiagen). The control empty vector was pGL3-basic. Lipofectamine 3000 (Invitrogen, America) was used for the transfection. After 24 h of transfection, the Dual-Luciferase Reporter Assay System (Promega) was applied to the cells and the luciferase activity of the lysates was detected on a multifunction plate analyzer (Tecan Infinite 200 Pro). The relative luciferase fluorescence value was obtained by calculating the ratio of firefly versus renilla. In this experiment, both A549 and MCF10A cells were employed and at least three separate independent runs of each experiment were repeated.

### Immunofluorescence assay

MCF10A cells were plated in 3.5 cm chamber slides (Nalge Nunc International, Rochester, NY) 24-h before transfection. In line with previous methods, Lipofectamine 3000 (Invitrogen, America) was used to transfect the plasmid into MCF10A cells for approximately 24 h at 37 °C and 5% CO<sub>2</sub>. After 20-min permeabilization with 0.5% Triton X-100 and 60-min blocking with 10% normal goat serum, the cells were incubated with 1.5 ml diluted primary rabbit polyclonal antibody KRT74 (1:300, bs-16833R, Bioss) overnight at 4 °C and then incubated in diluted goat anti-rabbit IgG H&L/Cy5 antibody for 60 min and 5 µg/ml DAPI in the dark at RT. Finally, cells were detected using a Leica confocal microscope (TCS SP8, Germany).

### Hematoxylin-eosin staining

A 4% paraformaldehyde solution was used to fix the skin tissues, and paraffin wax was used to encase them. The Leica RM2255 Automated Rotary Microtome (Germany) was employed, and blocks of the paraffin-embedded tissues were sliced into 3-µm thick sections and stained with HE. The dried slices were put in xylene for dewaxing and then were moved into anhydrous alcohol and then into 90, 80, and 70% alcohol in turn. Next, the slides were placed in a hematoxylin solution and then in distilled water. The color was separated by placing the slides in acid water and then ammonia water for several sec each. The slides were then rinsed with water for 60 min and moved into 70 and 90% alcohol for approximately 10 min for dehydration. The slides were subsequently stained with an alcohol eosin solution for 2–3 min and moved into anhydrous alcohol and xylene for dehydration and transparency, respectively. Finally, the neutral gum was dripped onto the sections, and a clean glass coverslip was carefully and slowly placed to seal each section. Morphological observations were performed at 100 × and 200 × magnification using an Olympus BX51 microscope (Olympus, Japan). In parallel, the average hair follicle area was measured by ImageJ software.

### Immunohistochemical analysis

According to the previously published protocol [26]. The 3-µm thick sections were incubated with 100% xylene-xylene, 100% alcohol, 95% alcohol, and 80% alcohol in turn, with the xylene reagent for 10 min and alcohol for 1 min. After dewaxing in clean water, washing three times in distilled water, and adding citric acid buf-

fer, the sections were submitted for heat repair at high pressure for 3 min and then cooled down to RT. After removing endogenous catalase, tissue sections were blocked in 10% goat serum for 15 min and then incubated with the primary antibody (1:400, BS-16833R, Bioss) at RT. After incubating with the secondary antibody (1:400, bs-40295G-HRP, Bioss) for 30 min at RT, sections were counterstained with hematoxylin and detected using diaminobenzidine developer.

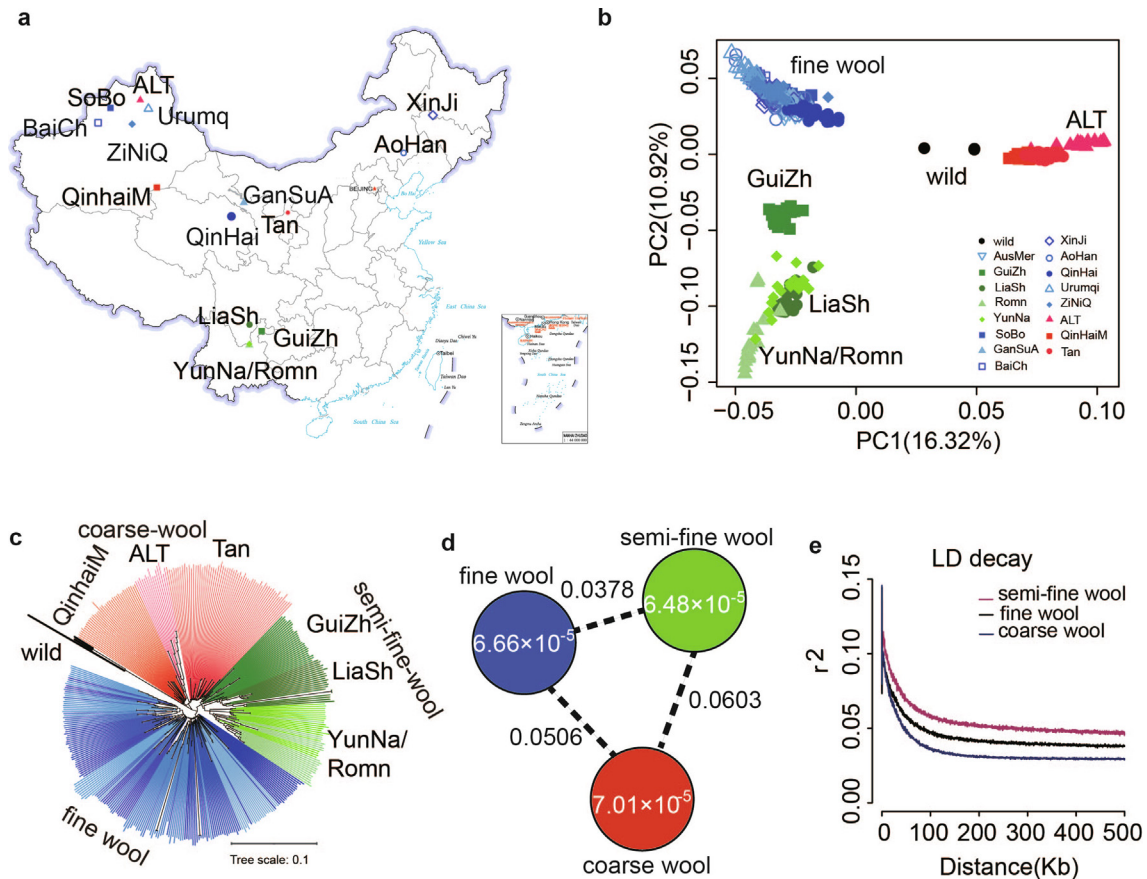
## Results

### Genomic variation, population structure, and genetic diversity analysis

For Ovine HD630K SNP array analysis, a total of 385 individual sheep from sixteen geographic regions of China (Fig. 1a), including seven fine-wool breeds, four semi-fine-wool breeds (SFS), and three coarse-wool breeds (Supplementary Table S2) were recruited. Three wild sheep were used as the out-group. After stringent quality filtering, a final set of 385 individual sheep and 472,773 common SNPs were retained in the downstream genomic selection signature analysis. Principal component analysis (PCA) was used, based on the first two components that explained 27.24% of the total variation, to divide the sheep breeds into four clusters, 'fine-wool' (blue), 'semi-fine-wool' (green), 'coarse-wool' (red), and 'wild sheep' (black) groups (Fig. 1b). PC3 was used to divide the 'semi-fine-wool' group into three large clusters, GuiZh, LiaSh, and YunNa-Romn (Supplementary Fig. S2). The YunNa and Romn semi-fine-wool sheep were grouped, consistent with the fact that the YunNa SFS breed was derived from the Romn SFS breed. Notably, the 'wild sheep' group was next to the 'coarse-wool' group, in line with the fact that wild sheep exhibit the coarse wool phenotype. The phylogenetic analysis using the distance-based neighbor-joining tree supported the four-clade population structure (Fig. 1c). Two summary statistics,  $\theta_{\pi}$ , and  $F_{ST}$  values were calculated genome-wide to estimate genetic diversity. It is noteworthy that when comparing the diversity between fine/semi-fine and coarse sheep, a lower  $\theta_{\pi}$  value was observed in the fine or semi-fine wool sheep breeds compared with the coarse sheep populations. The pairwise genetic differentiation value ( $F_{ST}$ ) was 0.04 between the fine versus semi-fine wool sheep while the value was 0.05–0.06 between the fine and semi-fine wool versus coarse wool sheep, indicating that the fine and semi-fine wool breeds are highly differentiated from the coarse wool breeds (Fig. 1d). This study consistently found that fine and semi-fine wool sheep have a lower level of genome-wide heterozygosity and a higher level of linkage disequilibrium (LD) decay than coarse wool sheep (Fig. 1e; Supplementary Table S3). These results reflect the intensive selection for the fine wool phenotype during the past breeding history.

### Integrated analyses reveal the candidate genes underlying the wool phenotype

This study first scanned the 385 sheep genomes to detect the positive selection signature that underpins the overall wool phenotype. To achieve this, the pairwise genetic differentiation ( $F_{ST}$  index), nucleotide diversity ( $\theta_{\pi}$  ratio), and transformed heterozygosity score (ZHp) were calculated based on a 100-kb sliding window with a 15-kb step size (Fig. 2a). To minimize the false positives, the top-1% threshold for all three statistics was taken into account ( $F_{ST} \geq 0.170449$ ,  $\theta_{\pi}$  ratio  $\geq 1.603$  and  $|ZHp| \geq 3.50743$ ) and 253 protein-coding genes were obtained under putative selective 372 windows (Supplementary Fig. S3a; Supplementary Tables S4–7). Such a selection scan again showed remarkably reduced genetic diversity in fine/semi-fine wool sheep, but not in



**Fig. 1. Geographic distribution and genetic structure of Chinese fine wool, semi-fine, and coarse sheep breeds.** (a) The geographic distribution of 16 Chinese fine wool, semi-fine wool, and coarse wool sheep populations. The blue color represents fine wool sheep breeds (N = 185; Urumq, Xinjiang Merino from Urumqi, 19; BaiCh, Xinjiang Merino from Baicheng, 20; ZiNiQ, Xinjiang Merino from Ziniquan, 28; SoBo, Subo Merino from Gongnaisi, 27; XinJi, Xinji Merino from Songyuan, 20; AoHan, Aohan Merino from Aohan, 20; GanSuA, Gansu Alpine Merino from Huangcheng, 21; QinHai, Qinghai Wool-mutton Type sheep from Sanjiaocheng, 20; AusMer, Australia Merino sheep). The green color represents semi-fine wool sheep breeds (N = 88; LiaSh, Liangshan, 21; GuiZh, Guizhou, 27; YunNa, Yunnan, 20; Romn, Romney, 20). The red color represents the coarse wool sheep breeds (N = 112; QinHaiM, Qinghai Mongolian sheep, 37; ALT, Xinjiang Altay sheep; 15; Tan, Tan sheep from Yanchi, 60). The shape of symbols indicates that the geographic regions are the same as those in the PCA plots and phylogenetic trees. (b) PCA plots of the first two components of all sheep samples containing one wild sheep population. The fraction of the total variance explained is reported on each axis between parentheses. (c) The Neighbor-joining tree of the sheep breeds, with wild sheep as the out-group. (d) Nucleotide diversity ( $\pi$ ) and population differentiation ( $F_{ST}$ ) among the fine wool, semi-fine, and coarse wool sheep populations. The value in each circle indicates the level of nucleotide diversity ( $\pi$ ) for each population and the value on each line represents population divergence ( $F_{ST}$ ) between the two populations. (e) Decay of LD in fine wool, semi-fine, and coarse wool sheep populations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

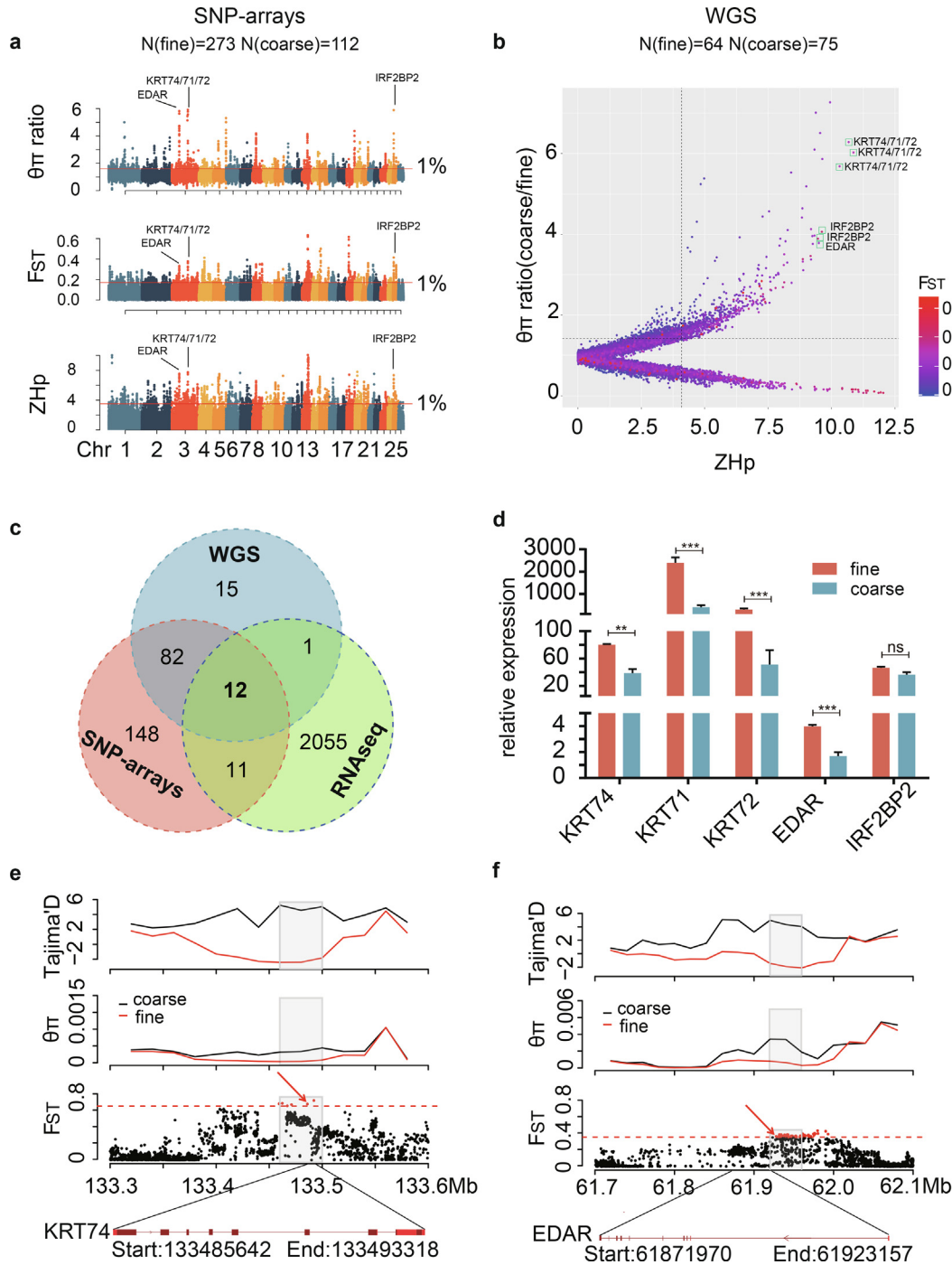
coarse-wool sheep, suggesting stronger positive selection in the fine-wool-related traits (Fig. 2a). This direction of selection is consistent with the heterozygosity and LD decay analysis. Enrichment for Gene Ontology terms revealed that the most over-represented categories were the structural constituent of the epidermis (GO: 0030280, FDR =  $2.35E-04$ ), intermediate filament organization (GO: 0045109, FDR =  $8.23E-03$ ), keratinization (GO: 0031424, FDR =  $1.45E-02$ ), and the keratin filament (GO: 0045095, FDR =  $5.06E-02$ ), which are the critical biological processes during wool growth (Supplementary Table S8). Importantly, the top significant regions contained *keratin-74/71/72/2* (*KRT74/71/72/2*, OAR3:133.3 Mb-133.6 Mb), *ectodysplasin A1 receptor* (*EDAR*, OAR3:61.7 Mb-62.1 Mb), *interferon regulatory factor 2 binding protein 2* (*IRF2BP2*, OAR25:7.3 Mb-7.4 Mb), and *protein lin-28 homolog B* (*LIN28B*, OAR8:32.0 Mb-32.2 Mb) according to the ranks in all three strategies (Supplementary Table S7).

To exclude the existence of other potential causative SNPs not covered by the high-density SNP arrays, this study repeated calculations of the selection signature using the whole genome sequencing data from the 120 sheep that were newly sequenced in this study along with whole genome sequencing data from 20 sheep that was downloaded. This covered 37 fine wool, 28 semi-fine

wool, and 75 coarse wool individuals (Fig. 2b; Supplementary Table S9, S11-13). These calculations identified 202 significant regions containing 110 protein-coding genes, among which, three of the top five significant regions (*IRF2BP2*, *KRT74/71/72*, and *EDAR*) remained the same as in the SNP array analysis (Supplementary Fig. S3b; Supplementary Table S14).

In addition, previous skin transcriptome analysis between fine and coarse wool sheep identified a total of 2,079 differentially expressed genes (DEGs) [24]. By overlapping these DEGs with the selected genes detected by whole genome sequencing and HD630K high-density SNP arrays, 12 common genes (Fig. 2c) were obtained. Out of these genes, *KRT74*, *KRT71*, *KRT72*, and *EDAR* were the top four candidate genes with both the strongest selection and significantly high expression in the fine/semi-fine wool sheep breeds (Fig. 2d; Supplementary Table S15 and S16).

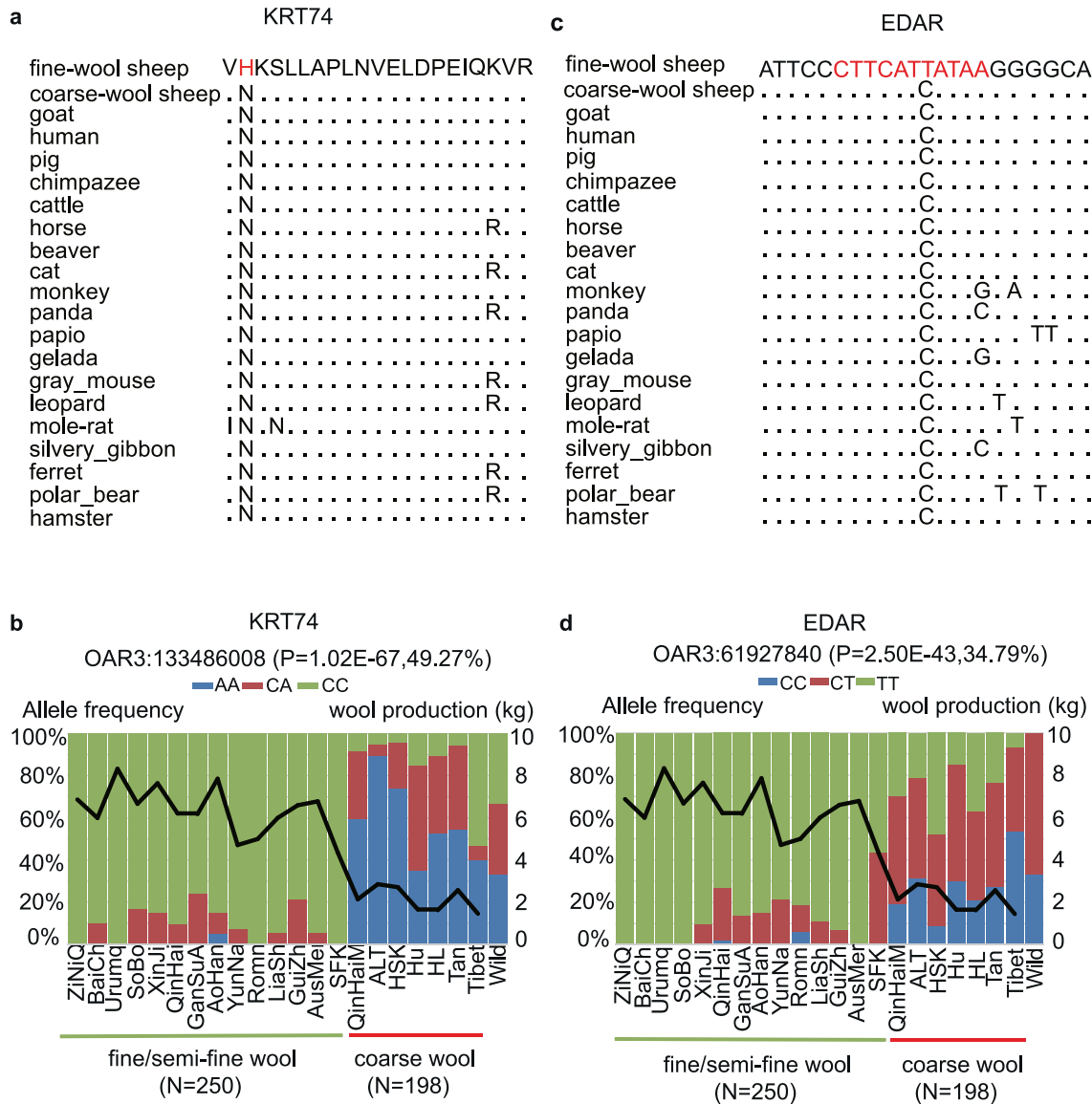
However, *IRF2BP2* and its retrogene *IRF2BP2<sup>asEIF2S2</sup>*—previously found to be associated with the wool phenotype in Romane lambs [27] and in Australian Merino [28]—were no longer detected as the most promising candidate genes in this study. To validate this, PCR amplification and Sanger sequencing were used to screen for the presence of this retrogene in a panel of 66 randomly selected samples including 27 fine wool sheep, 24 semi-fine-wool sheep, and 15



**Fig. 2. Genome-wide statistical analysis.** (a) Positive selection scans for wool fineness with SNP-arrays data. The population genetic differentiation  $F_{ST}$  values, the nucleotide diversity  $\theta_{\pi}$  ratios (coarse/fine), and the transformed heterozygosity score ZHp were calculated within 100 kb sliding windows (step size = 15 kb). The significance threshold of the selection signature was arbitrarily set to the top 1% percentile outliers for each test and is indicated with red horizontal full lines. (b) Selective signature analysis in the 'fine' group and 'coarse' group with whole-genome resequencing data. The 'fine' group included three fine wool and three semi-fine wool populations, a total of 64 samples. The 'coarse' group included ALT and Tan sheep populations, a total of 75 individuals. The top-1% quantiles of all three statistics are shown in the top-right corner defined by black dotted lines. (c) Venn diagrams. Numbers in the intersection regions are the observed overlapping genes among the found candidate genes based on whole genome sequencing, SNP-arrays, and RNAseq analysis. (d) The mRNA expression of five genes in fine sheep (orange) and coarse sheep (green) are based on RNAseq data. \*\* and \*\*\* display the statistical significance of two-tailed Student's  $t$  tests ( $P < 0.01$  and  $0.001$ ), respectively. (e) The strongest positive selection signatures around the KRT74 peak (OAR3:133.3 Mb – 133.6 Mb). (f) The strongest positive selection signatures around the EDAR peak (OAR3:61.7 Mb–62.1 Mb). Both the red dot lines represent the significant threshold of  $F_{ST}$  value per SNP  $> 0.65$  (e) and SNP  $> 0.35$  (f). Both  $\pi$  ratio and Tajima's D values were based on a 20 kb window and a 20 kb step.  $F_{ST}$  were calculated at each SNP position. The gray columns represent the strongest positive selection signatures in the region considered. The small red boxes and short lines represent the gene structure of KRT74 and EDAR. The top two SNPs are noted by red arrows. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

coarse wool sheep. The retrogene in 8 out of 51 Merino and semi-fine-wool sheep was not detected (Supplementary Fig. S4), suggesting that it is not yet fixed in fine wool breeds and there exist

other causative genes. Therefore, the subsequent analysis was focused on the top four candidates, divided into two regions, KRT74/71/72 and EDAR.



**Fig. 3. Annotation and validation of the KRT74-SNP2 and EDAR-SNP1 showing positive selection signatures.** KRT74-SNP2 and EDAR-SNP1 were successfully genotyped using the KASP technology in a total of 448 animals belonging to 14 fine wool sheep/semi-fine wool and seven coarse sheep populations ( $N = 448$  sheep). (a) KRT74 protein sequence analysis. The protein coordinates are based on the ENSOART00020017310.1 Ensemble protein. Protein sequence polymorphisms present in 20 vertebrates are provided. (b) Allelic frequencies at KRT74-SNP2 (OAR3:133486008). The A allele is in blue and the C allele is in green. (c) EDAR DNA sequence analysis. DNA sequence polymorphisms present in 20 vertebrates are provided. (d) Allelic frequencies at EDAR-SNP2 (OAR3:61927840). The C allele is in blue and the T allele is in green. KRT74-SNP2 ( $P = 1.02E-67$ ) and EDAR-SNP1 ( $P = 2.50E-43$ ) showed a strong correlation with sheep wool production, explaining 49.27% and 34.79% of the phenotypic variation, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Further study refined the top selection regions (KRT74/71/72 and EDAR) into the single SNP level using the whole genome sequencing data and three different statistical values,  $\theta_\pi$ , Tajima's D, and  $F_{ST}$ , in a 20-kb sliding window (Supplementary Tables S17-22). Two missense mutations were successfully found in KRT74, H123N (named SNP2, OAR3:133,486,008) and KRT72, I106V (named SNP3, OAR3:133,462,372), as well as one cis-regulatory mutation in EDAR, T/C (named SNP1, OAR3:61927840) (Fig. 2e-f; Supplementary Table S23).

Of these missense variants, SNP2, located in a highly conserved domain of KRT74, showed close to 100% sequence similarity among 20 vertebrate species (Fig. 3a), higher than the missense SNP3 within KRT72 (Supplementary Fig. S5). The PolyPhen-2 score of SNP2 (0.999) was remarkably higher than that of SNP3 (0.229), indicating the former SNP has a stronger functional effect (Supplementary Table S23; Supplementary Fig. S6). Thus, the effect of

SNP2 was further validated in a population of 448 sheep individuals within a total of 21 breeds (min. two individuals per breed, Supplementary Table S24). This finding supports a strong correlation ( $P = 1.02E-67$ ) between the allele frequency of SNP2 and the wool production (Fig. 3b), with the variant C predominant in the Chinese fine/semi-fine wool breeds ( $N = 250, 94.0\%$ ) and the ancestral allele A predominant in the Chinese coarse wool breeds ( $N = 198, 73.5\%$ ). The association analysis showed that the KRT74 mutation explains  $\sim 50\%$  of the wool production variation ( $\sim 2.2$  kg).

The upstream variant SNP1 is present in EDAR and located in a highly conserved element showing  $\sim 100\%$  sequence similarity across 20 vertebrate species (Fig. 3c). The possible correlation between the allele frequency of SNP1 and wool production was extremely significant ( $P = 2.50E-43$ ) (Fig. 3d). The predominant variant of EDAR was "T" in fine wool breeds ( $N = 250, 92.2\%$ ), while



the major allele was “C” in coarse wool sheep ( $N = 198$ , 49.2%). The association analysis showed that the variation at *EDAR* explains ~35% of the wool production variation (~1.6 kg) in this panel. Furthermore, the transcription factor binding site prediction found that the upstream variant T generates a new binding site of a key transcription factor in the early embryonic placode [29], *SOX2* (‘CTTCATTATAA’, **Supplementary Table S25**). These results suggest that both the C/A missense variant of *KRT74* (named SNP2, OAR3:133,486,008) and the T/C SNP in the regulatory region upstream of *EDAR* (named SNP1, OAR3:61,927,840) are the potential causative mutations for the fine wool phenotype, though this needs further functional validation.

#### Functional effects of the *KRT74* missense mutation

The *keratin 74* gene (*KRT74*) forms epithelial (soft) keratin intermediate filaments, specific to the Huxley layer of the inner root sheath (IRS) of the hair follicle [30]. The classical Rex (Re/+) mutant mice with wavy coats exhibit irregular Huxley cell layers and this morphological abnormality shapes the wavy hair shaft elongation [31]. A missense variant in the helix initiation motif of human *KRT74* (p.Asn148Lys) causes woolly hair by influencing the keratin intermediate filament (KIF) network [30]. Therefore, the C/A missense variant of *KRT74* (named SNP2, OAR3:133,486,008), identified in the comprehensive genomic selection detection, is hypothesized to be the causal mutation for the fine wool traits. Thus, the GFP tagged wildtype A-*KRT74* and mutant C-*KRT74* protein in the MCF10A cells (human breast epithelial), as well as A549 cells (human adenocarcinoma alveolar basal epithelial) were over-expressed in this study ( $P = 1.85\text{E-}06$  and  $P = 1.14\text{E-}06$ , **Fig. 4a, 4b**; **Supplementary Figs. S7a, 7b**; **Supplementary Table S26**). ImageJ measurements of the GFP-stained keratin intermediate filaments revealed that mutant C-*KRT74* overexpression led to at least a two-fold enlargement of cell size compared with the wildtype protein (**Fig. 4c**; **Supplementary Fig. S7c**; **Supplementary Table S27**). Cell proliferation detection assays, based on the EdU staining, showed that the mutant C-*KRT74* caused a 1.47-fold increase in fluorescence intensity in MCF10A cells ( $P = 4.08\text{E-}10$ ) and a 2.32-fold increase in A549 cells ( $P = 1.38\text{E-}05$ ) compared with A-*KRT74* (**Fig. 4d**, **Supplementary Fig. S7d**, and **Supplementary Table S28**). Thus, the mutant C-*KRT74* promotes both cell enlargement and proliferation suggesting it is a gain-of-function mutation.

Furthermore, skin transcriptome analysis of the fine wool sheep across two embryonic 105 (E105) and 135 (E135) d and four post-natal days (P7, P30, P90, and P360) demonstrated that postnatal expression of *KRT74* increased and peaked at P360 (**Fig. 4e**; **Supplementary Table S30**). This is consistent with the 20% decrease in the average hair follicle area in the fine wool sheep at P360 compared with P90 ( $P = 5.09\text{E-}10$ , **Fig. 5a-b**; **Supplementary Table S31**), indicating the crucial role of *KRT74* during the postnatal formation of secondary hair follicles.

Subsequent study via q-PCR analysis showed that adult fine-wool sheep carrying the homozygous *KRT74* mutant C allele had more than a 2-fold higher *KRT74* expression in skin samples ( $P = 9.00\text{E-}04$ ) than the adult coarse wool sheep carrying the wild-type A allele (**Fig. 4f** and **Supplementary Table S29**). Immunohistochemistry analysis using a *KRT74* antibody showed *KRT74* expression specifically in the nuclei of the Huxley layer (**Fig. 4g**). The skin section in fine wool sheep showed a pronounced *KRT74* overexpression and enhanced keratin intermediate filament network, resulting in the cell enlargement at the Huxley layer compared with the coarse wool sheep (**Fig. 4g** and **Supplementary Fig. S7g**). This led to the thickened Huxley layer, but finer hair shaft in the adult fine wool sheep.

#### Functional effects of the upstream variant of *EDAR*

*EDAR* is expressed during the formation of hair placodes [32] and its overexpression increases the density of hair follicles [33]. Furthermore, during the period of hair follicle morphogenesis, *EDAR* and *SOX2* have pronounced expression in the maturing hair follicles placodes (Pc) and dermal condensate (DC). DC generates the secondary dermal signal, triggering Pc progenitor proliferation, HF downgrowth, and dermal papilla formation [29]. Thus one T/C variant in the regulatory region upstream of *EDAR* (named SNP1, OAR3:61,927,840), identified in this study, is speculated to be responsible for the hair follicle density likely via *SOX2* transcriptional activation.

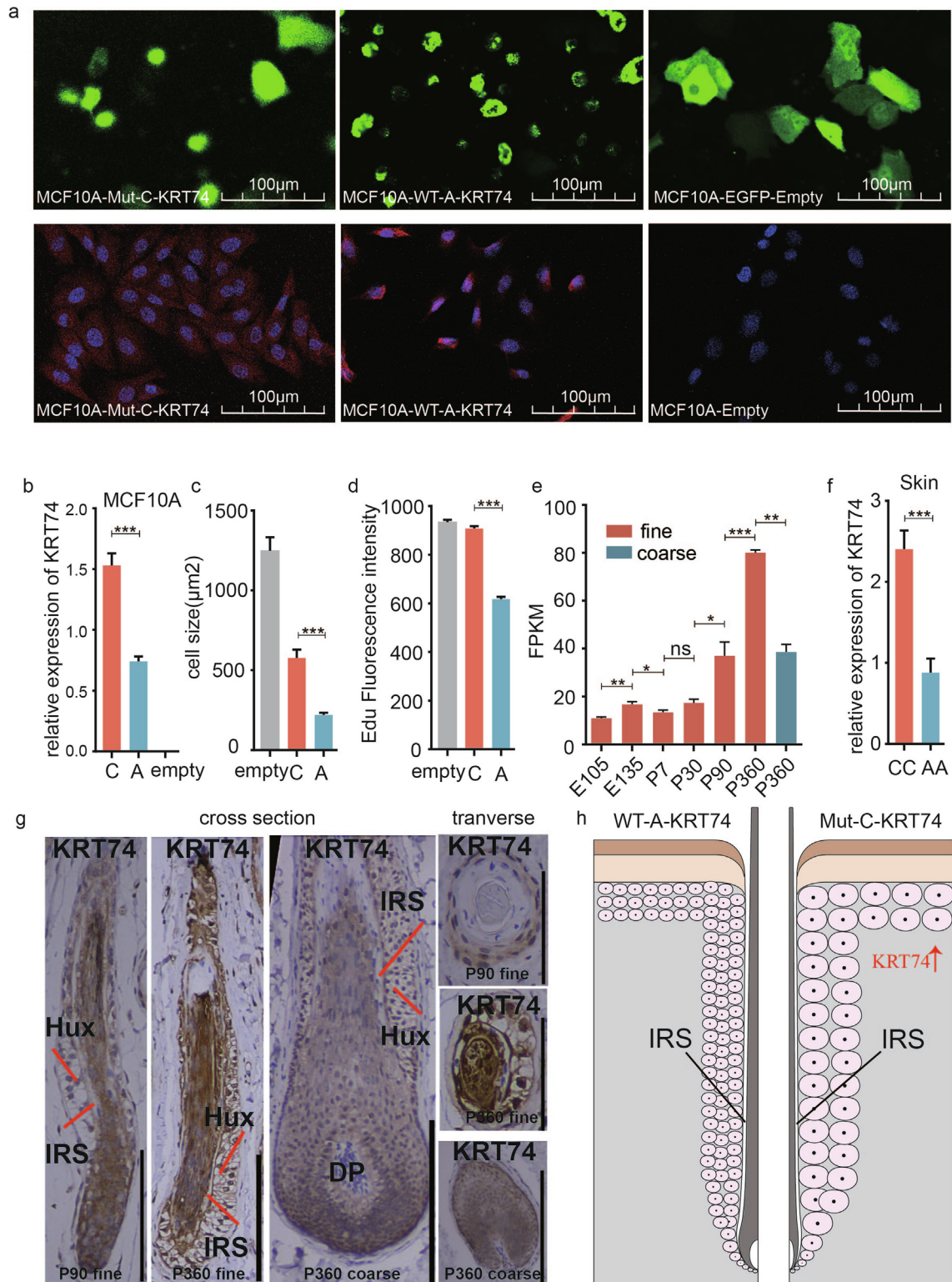
To confirm the above, HE staining was initially performed and the ratio of secondary to primary hair follicles was found to be increased 6-fold in adult fine-wool sheep versus coarse wool sheep (**Fig. 5c**; **Supplementary Table S32**). This result is in association with the finding that the adult fine-wool sheep have a 2.93-fold higher expression of *EDAR* than the adult coarse-wool sheep ( $P = 2.86\text{E-}02$ ) (**Fig. 5d** and **Supplementary Table S34**). Furthermore, the *EDAR* expression first decreased from embryonic E105 to postnatal P90, but then increased from P90 to P360, which closely corresponds to the change in hair follicle density and the decrease of hair follicle area from P90 to P360 (**Fig. 5e**; **Supplementary Table S35**), indicating the crucial role of *EDAR* during the postnatal formation of secondary hair follicles.

Next, a 171-bp sequence containing either wild-type (C), mutant (T), or a 20-bp deletion of the *SOX2* binding site was cloned into the pGL3-basic luciferase reporter vector and assays in both MCF10A and A549 cells were performed. The *SOX2* binding site carrying the fine wool variant (T) led to an increase of luciferase intensity of 1.19-fold ( $P = 1.99\text{E-}02$ ) and 1.42-fold ( $P = 1.39\text{E-}4$ ) compared with the ancestral (C) allele in MCF10A and A549 cells, respectively (**Fig. 5f**; **Supplementary Fig. S7e**; **Supplementary Table S33**). Interestingly, the similar low firefly intensity among the ancestral C allele, the 20-bp deletion, and the empty vector suggest that this single nucleotide mutation (T) generated a transcription factor binding site activating the downstream expression (**Fig. 5f**). The newly created *SOX2* binding site enhances *EDAR* transcriptional expression and the high expression level of *EDAR* probably enhances the formation of follicle placodes and increases the number of hair follicles.

#### Discussion

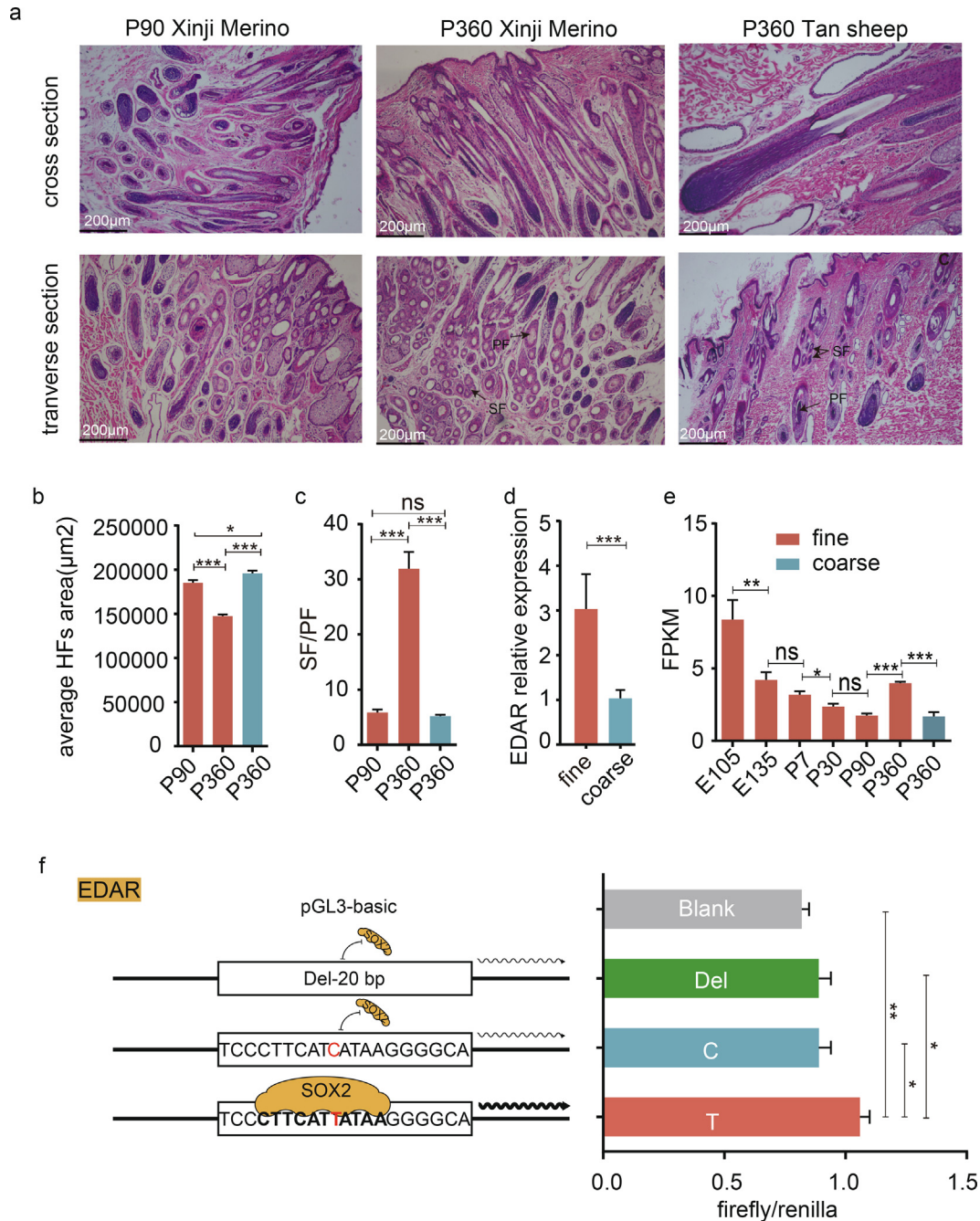
In the present study, whole-genome re-sequencing and Ovine HD630K SNP array data were integrated to identify high-confident candidate regions underlying the fine wool traits among Chinese fine and coarse wool sheep breeds. It is noteworthy that two such regions encompassing two genes (*KRT74* and *EDAR*) were under the strongest positive selection for the fine wool phenotype. These two genes are well known in the formation of keratin filament and hair follicle development as crucial biological processes for wool growth [34,35]. This result demonstrates the precision of the integrative approach.

This study further refined these two regions into two variants: one C/A missense variant of *KRT74* (named SNP2, OAR3:133,486,008) and one T/C SNP in the regulatory region upstream of *EDAR* (named SNP1, OAR3:61,927,840). The large-scale genetic analyses in over 450 sheep individuals showed that the C/A missense variant of *KRT74* (named SNP2, OAR3:133,486,008) could explain as much as ~50% of the variation in wool production (~2.2 kg) in this panel, which showed a strong statistical association between the C/A missense variant of *KRT74* with wool production variation in fine wool sheep. In addition,



**Fig. 4. Functional validation of the KRT74 missense SNP showing positive selection signatures.** (a) MCF10A cells were transfected with the p.His123Asn mutant(Mut)-C-KRT74, wild-type(WT)-A-KRT74 recombinant plasmids and EGFP for 24 h, followed by cell morphology observation and immunofluorescence assay. (b) KRT74 mRNA expression in the transfected MCF10A cells. \*\*\* display the statistical significance of two-tailed Student's *t* tests ( $P < 0.001$ ). (c) Semi-quantification analysis of cell size in the transfected MCF10A cells. (d) The transfected MCF10A cell proliferation detection. (e) KRT74 mRNA expression across two embryonic days (E105 and E135) and four postnatal days (P7, P30, P90, and P360) from the skin tissue of fine wool sheep (orange) and one postnatal day (P360) from the skin tissue of coarse wool sheep skin samples (green). (f) KRT74 mRNA expression in fine wool sheep (orange) and coarse wool sheep skin samples (green). (g) Immunohistochemical analysis of fixed skin tissue from Xinji Merino (P90 and P360) and Tan sheep (P360) shown at 200× magnification. IRS: internal root sheath; Hux: Huxley's layer. (h) Schematic presentation of KRT74 expression in the hair bulb. The red arrow represents increased KRT74 expression. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)





**Fig. 5. Functional validation of the EDAR upstream SNP showing positive selection signatures.** (a) Hematoxylin-eosin (HE) staining analysis of fixed skin tissue from Xinji Merino and Tan sheep shown at 100× magnification. (b) The average hair follicles (HF) area in Xinji Merino and Tan sheep. (c) The ratio of secondary hair follicles relative to primary hair follicles numbers. (d) *EDAR* mRNA expression in adult Tan sheep (green) and adult Xinji fine sheep skin samples (orange). (e) *EDAR* mRNA expression across two embryonic days (E105 and E135) and four postnatal days (P7, P30, P90, and P360) from the skin tissue of fine wool sheep (orange) and one postnatal day (P360) from the skin tissue of coarse wool sheep skin samples (green). (f) Dual-Luciferase reporter assay. Recombinant PGL3-Basic plasmids carrying the 171-bp region with the T (orange) or C allele (light blue) or 20-bp deletion (green) at the EDAR-SNP1 locus was constructed and expression levels were measured following transfecting into MCF10A cell lines. Control expression levels were measured following the transfection of empty vectors (black). \*, \*\*, and \*\*\* display the statistical significance of two-tailed Student's *t* tests ( $P < 0.05$ ,  $0.01$ , and  $0.001$ ), respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the mutant C-KRT74 caused a substitution of amino acid (p. His123Asn), which is located in a highly conserved N-terminal head domain of KRT74 among 20 vertebrate species. Furthermore, the substitution of amino acid of KRT74 (p.His123Asn) was predicted to alter the function of protein, according to PolyPhen-2. As expected, cellular overexpression and immunohistochemical analysis in the current study found that the mutant C-KRT74 remarkably activated the mRNA and protein expression level of this keratin and enhanced the intermediate filament (KIF) network.

This led to significant cell enlargement and proliferation *in vitro*. These results strongly suggest that the C/A missense variant of *KRT74* is a gain-function mutation for the fine wool traits. The expression of *KRT74* is specific to the Huxley layer of the inner root sheath (IRS) [36]. A previous heterozygous mutation in human *KRT74* caused the woolly hair phenotype by destroying KIF formation in a dominant-negative manner [30]. Early exploration in *Re* mice revealed that the thicker Huxley's cell layer shaped hair shaft growth and influenced the diameter of the hair shaft [31]. There-

fore, the overexpressed mutant C-KRT74 identified in the current study likely enhanced the KIF network and promoted the enlargement and proliferation of Huxley cells, thus thickening the Huxley layers at the inner root sheath and shaping the growing hair shaft into finer wool hair.

Meanwhile, the second variant is a C/T SNP within the upstream regulatory region of *EDAR* (named SNP1, OAR3:61,927,840), was completely conserved among 20 vertebrate species. This variation could explain as much as ~35% of the wool production variation (~1.6 kg). HE staining and luciferase assays revealed that the single-nucleotide T mutation found predominantly in the fine wool sheep upregulated *EDAR* mRNA expression and was associated with a six-fold higher density of hair follicles. Altogether, these studies demonstrated that this variation in *EDAR* is responsible for the denser hair follicles in fine wool sheep breeds. Several previous studies have shown that high *EDAR* expression promotes more hair follicles in *EDAR*-hyperactivated mice [33] and gene-edited Cashmere goats [37]. In addition, hair formation of *EDA*-deficient mice was induced by administering anti-*EDAR* Agonist Monoclonal Antibodies, which revealed a strong association between *EDAR* and hair follicle density [38]. *EDAR*370A (1540T/C, 370Val/Ala) knock-in mice showed reduced *EDAR* expression and exhibited a larger proportion of thick hairs compared to wild-type mice [39,40]. The above suggests that higher *EDAR* expression results in greater hair follicle density and more wool hairs. The single-nucleotide T mutation at the upstream regulatory region of *EDAR* (named SNP1, OAR3:61,927,840) consistently created one known important transcription factor binding site for *SOX2*. The *SOX2* gene is associated with the early formation of hair placodes during hair follicle morphogenesis, possibly regulating hair follicle growth together with *EDAR* [29]. Therefore, these studies support a model (Fig. 5f) in which this T mutation of *EDAR* creates a new *SOX2* binding site and facilitates the activation of *EDAR* transcription and thus the formation of more hair follicles.

In Merino sheep, the development of secondary follicles can effectively improve follicle density and wool production [41]. Previous studies have identified several candidate genes during the early developmental stage of the hair follicle, from the embryonic period to postnatal 30 d. A recent study revealed that there is no significant difference in hair follicle density between postnatal 30 d and the embryonic period (SF/PF = ~1–3) [11], while the ratio of secondary to primary follicles (SF/PF) reached 20 in 6-month-old Merino sheep [6], showing the remarkable increase of the secondary hair follicles after postnatal 30 d. This six-fold increase shows that the crucial developmental period of secondary hair follicles is not during the embryonic stage but between postnatal 30 and 180 d. However, the development stage of secondary hair follicles, especially between postnatal 30 d and adult, remained elusive [42,43,44]. Comparative analysis in the current study found that the crucial developmental period for secondary hair follicles is between postnatal 90 and 360 d in fine wool sheep. Both *KRT74* and *EDAR* showed peak expression during this period. However, the accurate time point for the key developmental process of secondary hair follicles needs further investigation.

Overall, the current study unveiled the genetic basis and molecular mechanism underlying fine wool formation in fine wool sheep breeds. This also provided two new genetic markers that help the wool industry produce finer and denser wool, which drives wool production in China.

## Conclusion

In this study, two causative mutations responsible for the finer wool and denser secondary hair follicles were detected through an analysis of over 380 collected sheep accessions, covering Chinese

fine wool, semi-fine-wool, and coarse wool breeds. The present study not only generated new insights into the genetic basis of fine wool generation but also provided new breeding targets for wool-producing animals.

## CRedit authorship contribution statement

**Benmeng Liang** : Conceptualization, Methodology, Software, Data curation, Visualization, Writing – review & editing. **T ianyou Bai**: Methodology, Data curation, Writing – review & editing. **Yuhetian Zhao**: Visualization, Writing – review & editing. **Jian-gang Han**: Investigation. **Xiaohong He** : Investigation, **Yabin Pu**: Investigation, **Chunxin Wang**: Investigation. **Wujun Liu** : Investigation, **Qing Ma** : Investigation. **Kechuan Tian** : Investigation. **Wenxin Zheng** : Investigation. **Nan Liu** : Investigation. **J ianfeng Liu** : Writing – review & editing. **Yuehui Ma**: Funding acquisition, Supervision. **Lin Jiang**: Conceptualization, Methodology, Writing – review & editing.

## Compliance with Ethics Requirements

All Institutional and National Guidelines for the care and use of animals were followed.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We are grateful to Guoqing Shi, Qionghua Hong, Haizhou Sun, Bohui Yang, Mingxin Zhang, and other experts for their help in phenotype identification. Sample collection was supported by the semi-fine wool sheep test stations in Guizhou, Sichuan, and Yunnan and the fine wool sheep test stations in Xinjiang, Qinghai, Gansu, Inner Mongolia, and Jilin.

## Funding

This project was supported by the National Natural Science Foundation of China (Nos. 32222079, 31961143021); the earmarked fund for the Modern Agro-industry Technology Research System (CARS-39-01); the Science and Technology Innovation Project of the Chinese Academy of Agricultural Sciences (ASTIP-IAS01).

## Availability of data and materials

The original whole-genome sequencing data are deposited in NCBI Bioproject under accession No. PRJNA873900. The original SNP chip data that support the findings of this study have been deposited into the CNGB Sequence Archive (CNSA) [45] of China National GeneBank DataBase (CNGBdb) [46] with accession number CNP0004132. All the raw FASTQ transcriptome data can be downloaded from the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) database under BioProject Nos.PRJNA705554 and PRJNA765722.

## Ethics approval and consent to participate

Not applicable.



## Consent for publication

Not applicable.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jare.2023.04.012>.

## References

- Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, de Villemereuil P, et al. Convergent genomic signatures of domestication in sheep and goats. *Nat Commun* 2018;9(1):813. doi: <https://doi.org/10.1038/s41467-018-03206-y>.
- Debono Spiteri C, Gillis RE, Roffet-Salque M, Castells Navarro L, Guilaïne J, Manen C, et al. Regional asynchronicity in dairy production and processing in early farming communities of the northern Mediterranean. *PNAS* 2016;113(48):13594–9. doi: <https://doi.org/10.1073/pnas.1607810113>.
- FAOSTAT: Food and Agriculture Organisation of United Nations. Rome: <https://www.fao.org/faostat/zh/#data/QCL/visualize>; 24 March 2023.
- Wenxin Zheng HX, Min Zhang. The development status, future development trend and suggestions of China's sheep industry for wool. *J China Animal Husbandry* 2023;59(03):300–306+315. doi: <https://doi.org/10.19556/j.0258-7033.20230306-04>.
- Barton S, Purvis I, Brewer H: Are wool follicle characteristics associated with wool quality and production in hogget and adult sheep? In *Proceedings of the Association for the Advancement of Animal Breeding and Genetics*. 2001: 289–292.
- Zhang L, Sun F, Jin H, Dalrymple BP, Cao Y, Wei T, et al. A comparison of transcriptomic patterns measured in the skin of Chinese fine and coarse wool sheep breeds. *Sci Rep* 2017;7(1):14301. doi: <https://doi.org/10.1038/s41598-017-14772-4>.
- Granero A, Anaya G, Demyda-Peyrás S, Alcalde MJ, Arrebola F, Molina A. Genomic Population Structure of the Main Historical Genetic Lines of Spanish Merino Sheep. *Animals (Basel)* 2022;12(10). doi: <https://doi.org/10.3390/ani12101327>.
- Wang W. Breeding preservation and utilization of Yunnan semi-fine wool sheep. *Yunnan Animal Husbandry and Veterinary* 2009;000(021):42–3. doi: <https://doi.org/10.3969/j.issn.1005-1341.2009.z1.018>.
- Demars J, Cano M, Drouilhet L, Plisson-Petit F, Bardou P, Fabre S, et al. Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep. *Mol Biol Evol* 2017;34(7):1722–9. doi: <https://doi.org/10.1093/molbev/msx114>.
- Li X, Yang J, Shen M, Xie XL, Liu GJ, Xu YX, et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat Commun* 2020;11(1):2815. doi: <https://doi.org/10.1038/s41467-020-16485-1>.
- Zhao B, Luo H, He J, Huang X, Chen S, Fu X, et al. Comprehensive transcriptome and methylome analysis delineates the biological basis of hair follicle development and wool-related traits in Merino sheep. *BMC Biol* 2021;19(1):197. doi: <https://doi.org/10.1186/s12915-021-01127-9>.
- Pan Z, Li S, Liu Q, Wang Z, Zhou Z, Di R, et al. Whole-genome sequences of 89 Chinese sheep suggest role of RXFP2 in the development of unique horn phenotype as response to semi-fertilization. *GigaScience* 2018;7(4). doi: <https://doi.org/10.1093/gigascience/giy019>.
- Archibald AL, Cockett NE, Dalrymple BP, Faraut T, Kijas JW, Maddox JF, et al. The sheep genome reference sequence: a work in progress. *Anim Genet* 2010;41(5):449–53. doi: <https://doi.org/10.1111/j.1365-2052.2010.02100.x>.
- Liu X, Zhang Y, Liu W, Li Y, Pan J, Pu Y, et al. A single-nucleotide mutation within the TBX3 enhancer increased body size in Chinese horses. *Curr Biol* 2021;32(2):480–7. doi: <https://doi.org/10.1016/j.cub.2021.11.052>.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30(15):2114–20. doi: <https://doi.org/10.1093/bioinformatics/btu170>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25(16):2078–9. doi: <https://doi.org/10.1093/bioinformatics/btp352>.
- Liu X, Zhang Y, Li Y, Pan J, Wang D, Chen W, et al. EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. *Mol Biol Evol* 2019;36(11):2591–603. doi: <https://doi.org/10.1093/molbev/msz158>.
- Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 1984;38(6):1358–70. doi: <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>.
- Nei M, Li WH. Mathematical model for studying genetic variation in terms of restriction endonucleases. *PNAS* 1979;76(10):5269–73. doi: <https://doi.org/10.1073/pnas.76.10.5269>.
- Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwachow D, et al. Strong signatures of selection in the domestic pig genome. *PNAS* 2012;109(48):19529–36. doi: <https://doi.org/10.1073/pnas.1217149109>.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics* 2011;27(15):2156–8. doi: <https://doi.org/10.1093/bioinformatics/btr330>.
- Fernandes V, Brucato N, Ferreira JC, Pedro N, Cavadas B, Ricaut FX, et al. Genome-Wide Characterization of Arabian Peninsula Populations: Shedding Light on the History of a Fundamental Bridge between Continents. *Mol Biol Evol* 2019;36(3):575–86. doi: <https://doi.org/10.1093/molbev/msz005>.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>(Delta Delta C(T)). *Method* 2013;25(4):402–8.
- Bai T, Liang B, Zhao Y, Han J, Pu Y, Wang C, Ma Y, Jiang L. Transcriptome Analysis Reveals Candidate Genes Regulating the Skin and Hair Diversity of Xinji Fine-Wool Sheep and Tan Sheep. 2022; 12(1):15.
- Xue Z, Ansari AR, Zhao X, Zang K, Liang Y, Cui L, et al. RNA-Seq-Based Gene Expression Pattern and Morphological Alterations in Chick Thymus during Postnatal Development. *Int J Genomics* 2019;2019(6905194). doi: <https://doi.org/10.1155/2019/6905194>.
- Ansari AR, Arshad M, Masood S, Huang HB, Zhao X, Li N, et al. Salmonella infection may alter the expression of toll like receptor 4 and immune related cells in chicken bursa of Fabricius. *Microb Pathog* 2018;121:59–64. doi: <https://doi.org/10.1016/j.micpath.2018.05.019>.
- Allain D, Foulquié D, Autran P, François D, Bouix J. Importance of birthcoat for lamb survival and growth in the Romane sheep breed extensively managed on rangelands. *J Anim Sci* 2014;92(1):54–63. doi: <https://doi.org/10.2527/jas.2013-6660>.
- Ly FH, Cao YH, Liu GJ, Luo LY, Lu R, Liu MJ, et al. Whole-Genome Resequencing of Worldwide Wild and Domestic Sheep Elucidates Genetic Diversity, Introgression, and Agronomically Important Loci. *Mol Biol Evol* 2022;39(2). doi: <https://doi.org/10.1093/molbev/msab353>.
- Saxena N, Mok KW, Rendl M. An updated classification of hair follicle morphogenesis. *Exp Dermatol* 2019;28(4):332–44. doi: <https://doi.org/10.1111/exd.13913>.
- Shimomura Y, Wajid M, Petukhova L, Kurban M, Christiano AM. Autosomal-dominant woolly hair resulting from disruption of keratin 74 (KRT74), a potential determinant of human hair texture. *Am J Hum Genet* 2010;86(4):632–8. doi: <https://doi.org/10.1016/j.ajhg.2010.02.025>.
- Tanaka S, Miura I, Yoshiki A, Kato Y, Yokoyama H, Shinogi A, et al. Mutations in the helix termination motif of mouse type I IRS keratin genes impair the assembly of keratin intermediate filament. *Genomics* 2007;90(6):703–11. doi: <https://doi.org/10.1016/j.ygeno.2007.07.013>.
- Headon DJ, Overbeek PA. Involvement of a novel Tnf receptor homologue in hair follicle induction. *Nat Genet* 1999;22(4):370–4. doi: <https://doi.org/10.1038/11943>.
- Mou C, Jackson B, Schneider P, Overbeek PA, Headon DJ. Generation of the primary hair follicle pattern. *PNAS* 2006;103(24):9075–80. doi: <https://doi.org/10.1073/pnas.0600825103>.
- Wasif N, Naqvi SK, Basit S, Ali N, Ansari M, Ahmad W. Novel mutations in the keratin-74 (KRT74) gene underlie autosomal dominant woolly hair/hypotrichosis in Pakistani families. *Hum Genet* 2011;129(4):419–24. doi: <https://doi.org/10.1007/s00439-010-0938-9>.
- Wu Z, Wang Y, Han W, Yang K, Hai E, Ma R, et al. EDA and EDAR expression at different stages of hair follicle development in cashmere goats and effects on expression of related genes. *Arch Anim Breed* 2020;63(2):461–70. doi: <https://doi.org/10.5194/aab-63-461-2020>.
- Langbein L, Rogers MA, Praetzel S, Winter H, Schweizer J. K6irs1, K6irs2, K6irs3, and K6irs4 represent the inner-root-sheath-specific type II epithelial keratins of the human hair follicle. *J Invest Dermatol* 2003;120(4):512–22. doi: <https://doi.org/10.1046/j.1523-1747.2003.12087.x>.
- Hao F, Yan W, Li X, Wang H, Wang Y, Hu X, et al. Generation of Cashmere Goats Carrying an EDAR Gene Mutant Using CRISPR-Cas9-Mediated Genome Editing. *Int J Sci Biol* 2018;14(4):427–36. doi: <https://doi.org/10.7150/ijbs.23890>.
- Kowalczyk C, Dunkel N, Willen L, Casal ML, Mauldin EA, Gaide O, et al. Molecular and therapeutic characterization of anti-ectodysplasin A receptor (EDAR) agonist monoclonal antibodies. *J Biol Chem* 2011;286(35):30769–79. doi: <https://doi.org/10.1074/jbc.M111.267997>.
- Kamberov YG, Wang S, Tan J, Gerbault P, Wark A, Tan L, et al. Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* 2013;152(4):691–702. doi: <https://doi.org/10.1016/j.cell.2013.01.016>.
- Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L, et al. A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum Mol Genet* 2008;17(6):835–43. doi: <https://doi.org/10.1093/hmg/ddm355>.
- Rogers GE. Biology of the wool follicle: an excursion into a unique tissue interaction system waiting to be re-discovered. *Exp Dermatol* 2006;15(12):931–49. doi: <https://doi.org/10.1111/j.1600-0625.2006.00512.x>.
- Tian Y, Yang X, Du J, Zeng W, Wu W, Di J, et al. Differential Methylation and Transcriptome Integration Analysis Identified Differential Methylation Annotation Genes and Functional Research Related to Hair Follicle Development in Sheep. *Front Genet* 2021;12. doi: <https://doi.org/10.3389/fgene.2021.735827>.
- He J, Zhao B, Huang X, Fu X, Liu G, Tian Y, et al. Gene network analysis reveals candidate genes related with the hair follicle development in sheep. *BMC Genomics* 2022;23(1):428. doi: <https://doi.org/10.1186/s12864-022-08552-2>.
- Brook AH, Short BF, Lyne AG. Formation of new wool follicles in the adult sheep. *Nature* 1960;185:51. doi: <https://doi.org/10.1038/185051a0>.
- Guo X, Chen F, Gao F, Li L, Liu K, You L, Hua C, Yang F, Liu W, Peng C, et al. CNSA: a data repository for archiving omics data. *Database (Oxford)* 2020;2020:baaa055. doi: <https://doi.org/10.1093/database/baaa055>.
- Chen FZ, You LJ, Yang F, Wang LN, Guo XQ, Gao F, et al. CNGBdb: China National GeneBank DataBase. *Yi Chuan* 2020;42(8):799–809. doi: <https://doi.org/10.16288/j.yczz.20-080>.