Journal of Zhejiang University-SCIENCE B in press
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.B2600483

Functional SNP in lncPYGL promoter regulates its expression by altering ZNF322 binding and is associated with litter size in goats

Haiyan YANG[1], Yifan ZHANG[1], Chunxiao LI[1], Yaohui WANG[1], Ebadu AREB[1], Yangyang BAI[1], Haijing ZHU[2], Xianyong LAN[1]✉, Chuanying PAN[1]✉

[1]*College of Animal Science and Technology, Northwest A&F University, Shaanxi Key Laboratory of Molecular Biology for Agriculture, Yangling, Shaanxi 712100, China*
[2]*Shaanxi Provincial Engineering and Technology Research Center of Cashmere Goats, Life Science Research Center, Yulin University, Yulin, Shaanxi 719000, China*

Abstract: Litter size is regulated by complex genetic networks, involving long non-coding RNAs (lncRNAs) that play key roles in mammalian reproductive processes. This study aimed to characterize the promoter region of lncPYGL, identify genetic variants associated with first-parity litter size, and investigate the involvement of potential transcription factors in its regulation in Shaanbei white cashmere goats. Through dual-luciferase reporter assays, we identified the core promoter region of lncPYGL, located between −1032 to −669 bp relative to the transcription start site. Three SNPs were pinpointed within this region, of which SNP2 and SNP3 were significantly associated with litter size. Mechanistic studies demonstrated that ZNF322 binds to the lncPYGL promoter and enhances its transcriptional activity. Notably, the wild-type SNP3-C allele exhibited stronger ZNF322 binding and higher promoter activity than the mutant A allele. Consistently, ovarian lncPYGL expression was significantly higher in goats carrying the CC genotype than those carrying the CA or AA genotype. These findings provide evidence that SNP3 may modulate lncPYGL transcription by affecting ZNF322 binding, although the causal relationship between lncPYGL expression and litter size remains to be established. SNP3 represents a candidate functional regulatory variant, whereas the independent predictive value of SNP2 requires further validation. Overall, this study provides valuable insights into the transcriptional regulation of lncRNAs and identifies a potential regulatory mechanism associated with reproductive traits in Shaanbei white cashmere goats.

Key words: Goat; Long non-coding RNA (lncRNA); Promoter; Single nucleotide polymorphism (SNP); ZNF322

unedited

1. Introduction

Litter size directly determines the scale of herd expansion, thus affecting the productivity and profitability of farms (Meza-Herrera et al., 2024). It is a complex trait with low heritability and is controlled by multiple minor-effect genes (Feng et al., 2026). Traditional breeding methods based on pedigree information and phenotypic performance are characterized by long breeding cycles, slow generation turnover, and low selection efficiency (Hassanine et al., 2025). However, the development of marker-assisted selection (MAS) has provided a key technical approach to overcome these challenges (Georges et al., 2019). In sheep, significant progress has been made by identifying molecular genetic markers strongly associated with important economic traits and

✉ Xianyong LAN, lanxianyong79@126.com;
✉ Chuanying PAN, chuanyingpan@126.com;
Xianyong LAN, https://orcid.org/0000-0003-2254-5805;
Chuanying PAN, https://orcid.org/0009-0002-3871-4730;
Haiyan YANG, https://orcid.org/0009-0004-2579-4019

Received July 4, 2026; Revision accepted Nov. 2, 2026;
Crosschecked xxx. xx, 20xx

© Zhejiang University Press 2026

applying them to early selection. For example, major genes, such as *BMP15, GDF9*, and *BMPR1B,* have been successfully used to enlarge litter size (Han et al., 2025). However, the effects of these major genes in goats are relatively limited or unstable (Gholizadeh, Mohammadnezhad, 2024). This suggests that the genetic regulatory mechanisms governing reproductive traits in goats may differ. Therefore, there is a need to identify new species-specific functional markers for these animals.

Long non-coding RNAs (lncRNAs) are key regulators of gene expression. They are widely involved in core reproductive processes in mammals, including follicular development, hormone synthesis, and ovulation regulation (Wang et al., 2024). In the ovaries of high-fecundity goats, lncRNAs, such as ENSCHIT00000005909, may positively regulate *IL1R2* gene expression, thereby influencing ovarian function (Lian et al., 2020). In Leizhou black goats, MSTRG.3782 participates in follicular development by regulating API5 expression (Ling et al., 2017). On the other hand, 24 lncRNAs overlapping with cis-regulatory elements were identified in Anhui white goats. These cis-regulatory elements are involved in progesterone-mediated oocyte maturation, steroid biosynthesis, oocyte meiosis, and the gonadotropin-releasing hormone (GnRH) signaling pathway (Ling et al., 2017). Although functional studies of lncRNAs in reproductive biology have advanced considerably in recent years, most of them remain largely limited to downstream regulatory mechanisms, such as those acting as competing endogenous RNAs (ceRNAs) or investigating protein complexes. However, the biological functions of lncRNAs are highly dependent on their spatiotemporal expression levels in specific tissues or developmental stages (Usman et al., 2024). Currently, little is known about the transcriptional regulation of lncRNAs. In particular, the genetic basis underlying the marked expression divergence between high- and low-fecundity breeds remains poorly elucidated. Thus, investigating this upstream regulatory network is essential for fully understanding the mechanisms that drive high-fecundity traits.

In the gene transcriptional regulatory network, the promoter acts as a master switch controlling the spatiotemporal expression and transcriptional activity of a gene (Zhang et al., 2022). Genetic variants located in crucial gene regulatory regions—such as promoters, especially single nucleotide polymorphisms (SNPs)—can directly influence the expression level of target genes by altering the binding affinity of specific transcription factors (Degtyareva et al., 2021). For example, pig and sheep breeding studies have demonstrated that variants at key promoter sites can alter the binding of transcription factors, such as SP1 and CREB, which in turn drives the differential expression of downstream fecundity-promoting genes (McDonald et al., 2016; Zhang et al., 2022). Functional SNPs located in regulatory regions can be ideal genetic markers for molecular breeding. Therefore, it is necessary to focus on the promoter regions of key reproduction-related lncRNAs, identify functional mutations within them, and elucidate their molecular effects.

Our research team previously identified lncPYGL as a key candidate and highly associated it with goat litter size. This study aimed to identify the core promoter of lncPYGL, characterize functional promoter SNPs within the promoter region, determine their effects on transcription factor binding and promoter activity, as well as evaluate their association with litter size in goats.

2. Materials and methods

2.1 Animals and tissue collection

Ear and ovarian tissue samples from Shaanbei white cashmere goats were collected in Yulin City, Shaanxi Province, China. Ear tissues were obtained from goats raised at Shaanxi Haoli Cashmere Goat Technology Development Co., Ltd. in Zizhou County, whereas ovarian tissues were obtained from a local slaughterhouse (the collection of all ovarian samples took place from the same slaughterhouse on the same day). After collection, all samples were immediately placed in liquid nitrogen for temporary storage and then transferred to a −80°C freezer for long-term preservation.

Under Test

2.2 Vector construction

The sequence of lncPYGL from -2000 bp upstream to +300 bp downstream was selected from the NCBI database. Potential promoter active regions were predicted using https://services.healthtech.dtu.dk/services/Promoter-2.0/. Based on the prediction results, four pairs of primers for truncated promoter fragments of different lengths were designed, with these fragments sharing a common reverse primer. The primer sequences are shown in Table 1. The amplified products were cloned into the pGL3-Basic vector.

Three SNP-mutant luciferase reporter constructs, pGL3-SNP1-G, pGL3-SNP2-T and pGL3-SNP3-A, were synthesized by Sangon Biotech (Shanghai, China). The coding sequences of ZNF322 and PBX3 were retrieved from the NCBI database. To generate Flag-tagged overexpression constructs, the Flag-tag coding sequence (5′-GCTACAAAGACGATGACGACAAGGATTACAAGGATGACGACGATAAAGACTACAAGG ACGATGATGATAAA-3′) was introduced at the 5′ end of each coding sequence. The resulting overexpression constructs were synthesized by Sangon Biotech (Shanghai, China).

2.3 Cell culture and transfection

HEK293T cells, CHO-K1 cells and goat ovarian granulosa cells (GCs) stored frozen were thawed and cultured in complete medium containing 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin at 37℃ in humidified incubator with 5% CO_2. When the cells reached approximately 90% confluence, they were trypsinized and seeded into 48-well plates, then cultured until they reached approximately 50% confluence before transfection.

For transfection, the empty pGL3-Basic plasmid, the recombinant plasmids pGL3-pro1, pGL3-pro2, pGL3-pro3, pGL3-pro4, and the pGL3-Control were each co-transfected with pRL-TK into the cells. The pGL3-Control, pRL-TK and empty pGL3-Basic served as the positive control, internal reference plasmid and negative control, respectively. To normalize the variation in transfection efficiency, the recombinant plasmids and pRL-TK were co-transfected at a ratio of 50:1.

2.4 Dual-luciferase reporter assay

Based on the instructions of the Dual-Luciferase Reporter Assay Kit (Yeasen, Shanghai, China), cell lysis buffer was added to the 48-well plate and the cell lysates were collected. The fluorescence intensity was measured using a BHP9504 microplate luminometer under light-protected condition. Using pRL-TK as an internal reference, the relative luciferase activity was calculated as the ratio of firefly luciferase activity to Renilla luciferase activity. Each sample was measured in triplicate and the results were expressed as mean ± standard deviation.

2.5 Genomic DNA extraction and SNP genotyping

A total of 365 healthy, unrelated, adult female Shaanbei white cashmere goats of similar age were randomly selected for tissue sample collection and their first parity litter size data were gathered. Genomic DNA was extracted from the ear tissue samples using a high-salt method. After measuring the DNA concentration with a UV spectrophotometer (Thermo Fisher Scientific, USA), the samples were diluted to 20 ng/μL and stored at −40℃ for future use. Equal amounts of 30 DNA samples were pooled to form a single DNA pool, which was used to detect genetic polymorphisms in the promoter-active region. The primer sequences used are shown in Table 1.

Under Review

Table 1 Primer sequence and length information

Gene name		Primers	Size (bp)
PGL3-pro1	F	CGGGGTACCGCGAGCGTGTATTCTATGGC	2049
	R	ACCACTGATACCCTGAAAC	
PGL3-pro2	F	CGGGGTACCTTCCGGTAGAGGCCTAGTCA	1543
	R	ACCACTGATACCCTGAAAC	
PGL3-pro3	F	CGGGGTACCTGGTGGTGCGAGGATACAGA	1119
	R	ACCACTGATACCCTGAAAC	
PGL3-pro4	F	CGGGGTACCCCGCCAGCTCTCTCTAGATTA	756
	R	ACCACTGATACCCTGAAAC	
lncPYGL-SNP	F	TGGTGGTGCGAGGATACAGA	495
	R	ACTCTAACCACTGCGCTTCC	
lncPYGL-qRT-PCR	F	GGGTGAATCTTCTTGGCGGT	83
	R	CCCGCTTCCTCATTTCCACA	
GAPDH	F	AAAGTGGACATCGTCGCCAT	116
	R	CCGTTCTCTGCCTTGACTGT	
Negative Control	F	CTCACTCTGAGGAATATTTTGC	130
	R	ATGGGTAAATCCTGGAGACG	
lncPYGL-binding	F	AACCTTCAGCTCCCCACTCCCC	129
	R	TAGAGAGAGCTGGCGGCTCCAG	
ZNF322-qRT-PCR	F	CACCAGGCAACTCACACTCA	141
	R	GAGGAAGCCACTCTCACACG	
PBX3-qRT-PCR	F	ATCATGACCATCACCGACCAG	172
	R	TCATTAGCTGGGGATCGGGA	

unedited

2.6 RNA extraction and qRT-PCR

Total RNA was extracted from the ovarian tissues of Shaanbei white cashmere goats using the TRIzol method (Wang et al., 2025). RNA purity and integrity were assessed by UV spectrophotometry and 1.0% agarose gel electrophoresis. Reverse transcription was performed using the Evo M-MLV RT Kit II (AG, Hunan, China) following the manufacturer's instructions to synthesize complementary DNA (cDNA). This was diluted to 50 ng/μL and stored at −20°C for later use.

Primers for both lncPYGL and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) were designed to detect their expression using NCBI Primer-BLAST. GAPDH was employed as the internal reference gene. For the analysis of genotype-dependent expression of lncPYGL in the ovarian tissues, such tissues were collected from 121 adult female goats at the same slaughterhouse on the same day. Genomic DNA was extracted from the ovarian tissues for SNP3 genotyping. Subsequently, six individuals were randomly selected from each SNP3 genotype group (CC, CA and AA), resulting in 18 ovarian tissue samples for the RT-qPCR analysis of lncPYGL expression, with six independent biological replicates per genotype. The expression of lncPYGL in individual goats with different genotypes was detected using the ChamQ SYBR qPCR Master Mix (Vazyme, Nanjing, China).

2.7 Western blot

Total cellular proteins were extracted using RIPA lysis buffer, and protein concentrations were determined using a bicinchoninic acid (BCA) protein assay kit. Equal amounts of protein were mixed with 5× protein loading buffer and denatured at 100℃ for 10 min. The protein samples were separated by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto polyvinylidene difluoride (PVDF) membranes. After treatment with a rapid blocking buffer at room temperature for 20 min, the membranes were washed three times with Tris-buffered saline containing Tween-20 (TBST) for 5 min each and incubated with the corresponding primary antibodies overnight at 4℃. The following day, the membranes were washed three times with TBST for 5 min each, followed by incubation with the appropriate horseradish peroxidase (HRP)-conjugated secondary antibodies at room temperature for 2 h. Further washing followed, then protein signals were detected using enhanced chemiluminescence (ECL) reagents and captured by a chemiluminescence imaging system (Yang et al., 2026). Flag-tagged ZNF322 and PBX3 proteins were detected using an anti-DYKDDDDK tag mouse monoclonal antibody (1:10,000; Cat. No. LF303; Yazyme Biotech, Shanghai, China). The primary antibody against GAPDH was purchased from Wanleibio (1:1,000; Cat. No. WL01114; Shenyang, China). HRP-conjugated goat anti-mouse IgG (H+L) (1:20,000; Cat. No. 62-6520) and goat anti-rabbit IgG (H+L) (1:20,000; Cat. No. 31460) secondary antibodies were purchased from Thermo Fisher Scientific.

2.8 Chromatin immunoprecipitation coupled with quantitative real-time PCR (ChIP-qPCR)

ChIP was performed using the PureBinding® Chromatin Immunoprecipitation Kit (Cat. No. P0301; Geneseed Biotech, Guangzhou, China) according to the manufacturer's instructions. Chromatin was cross-linked, fragmented, and immunoprecipitated using an anti-ZNF322 antibody. The enriched DNA fragments were purified and analyzed by qPCR to evaluate ZNF322 binding to the lncPYGL promoter region containing SNP3.

2.9 Statistical analysis

Based on the SNP genotyping results, allele frequency, genotype frequency, expected heterozygosity (He), observed homozygosity (Ho), effective number of alleles (Ne), and polymorphism information content (PIC) were calculated to evaluate the genetic diversity of SNP loci in the population. Associations between SNP genotypes and first-parity litter size were analyzed using the general linear model (GLM) procedure in SPSS Statistics 23.0 (IBM Corp., Armonk, NY, USA), with genotype included as a fixed effect and first-parity litter size as the dependent variable. Least-squares means (LSMs) and their standard errors (SEs) were estimated for each genotype and reported as LSM ± SE. Pairwise comparisons among genotypes were performed using Tukey's adjustment for multiple comparisons. Statistical significance was set at $P < 0.05$.

Each recombinant plasmid was co-transfected with the internal reference plasmid pRL-TK into cells. After 48 hours, the firefly luciferase activity and Renilla luciferase activity were measured following the kit instructions. The ratio of firefly luciferase activity to Renilla luciferase activity was calculated as the relative luciferase activity. Three replicate wells were prepared for each group, and the experiment was independently repeated three times. The results were expressed as the mean ± standard deviation of three independent experiments. One-way ANOVA was performed using SPSS 23.0 software, and multiple comparisons were conducted using the Dunnett method (or LSD method) to compare the differences between each truncated promoter plasmid and the empty vector (or control).

Unedited

3. Results

3.1 Identification of the core promoter region of lncPYGL

On the basis of our previous whole-transcriptome sequencing analysis of ovarian tissues from consistently single-kidding and consistently multiple-kidding Shaanbei white cashmere goats, a long non-coding RNA (lncRNA)—LOC108636992—was identified. Genomic mapping revealed its location on goat chromosome 10 (NC_030817.1) overlapping with the glycogen phosphorylase, liver form (PYGL) gene on the opposite strand, indicating that it is transcribed antisense to PYGL. Based on its genomic location and orientation relative to PYGL, LOC108636992 was designated as lncPYGL in this study (RefSeq accession number: XR_001918688.1). Subsequently, RT-qPCR analysis was performed to validate the expression pattern of lncPYGL in ovarian tissues from consistently single-kidding and consistently multiple-kidding goats. The results showed that lncPYGL expression was significantly higher in the ovaries of consistently multiple-kidding goats than those of consistently single-kidding goats (Fig. 1a). To further investigate the transcriptional regulation of lncPYGL, four progressively truncated fragments of its upstream promoter region were cloned into the pGL3-Basic vector. The resulting recombinant plasmids were designated as pGL3-pro1 (−1962 bp to +87 bp), pGL3-pro2 (−1456 bp to +87 bp), pGL3-pro3 (−1032 bp to +87 bp), and pGL3-pro4 (−669 bp to +87 bp) (Figs. 1b and 1c). HEK293T cells, CHO-K1 cells and goat GCs were used to detect the promoter region of lncPYGL. Relative luciferase activity for each fragment was calculated based on the ratio of firefly luciferase activity to Renilla luciferase activity. The results showed that the luciferase activity of pGL3-Control was consistently higher than that of pGL3-Basic, in line with the expected roles of the positive and negative controls. As the promoter sequence was progressively shortened from pGL3-pro1 to pGL3-pro4, a significant decrease in luciferase activity occurred when the fragment was shortened from pGL3-pro3 to pGL3-pro4. This indicated that the active region of the promoter was located between pGL3-pro3 and pGL3-pro4, corresponding to the region from −1032 bp to −669 bp upstream of the transcription start site (Figs. 1d-1f).

unedited

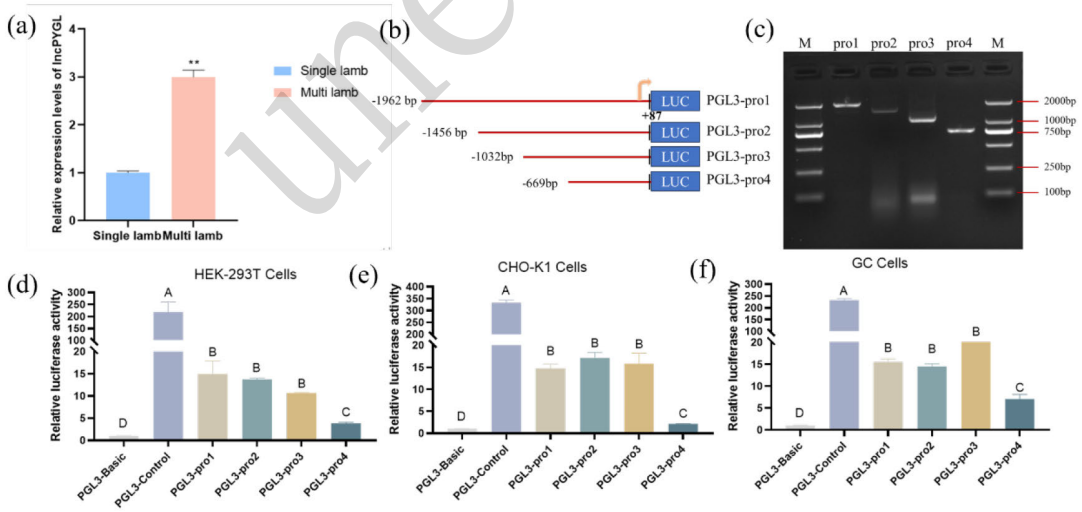


Fig. 1 Expression analysis of lncPYGL and identification of its core promoter region. a. Relative expression levels of lncPYGL in ovarian tissues from consistently single-lamb and consistently multiple-lamb goats. b. Schematic diagram of different truncated fragments. c. Agarose gel electrophoresis of PCR products from different truncated fragments. d-f. Identification of the core promoter region of lncPYGL using dual-luciferase reporter assays.

3.2 SNPs identification and prediction of potential transcription factor binding sites

Genetic variants located in the active region of a promoter may inhibit or enhance the binding affinity of transcription factors to this region, thereby regulating the transcription level of the gene. Therefore, we designed PCR primers covering the detected promoter region of lncPYGL (−1032 bp to −669 bp). PCR amplification was performed using a DNA pool from 30 Shaanbei white cashmere goat DNA samples. The amplified PCR products were subjected to Sanger sequencing. The comparisons of the sequencing results with the -2000 bp upstream reference genomic sequence (NC_030817.1) from NCBI identified three SNPs within the promoter region (SNP1: g.59264584 C>G; SNP2: g.59264385 G>T; SNP3: g.59264327 C>A) (Fig. 2).

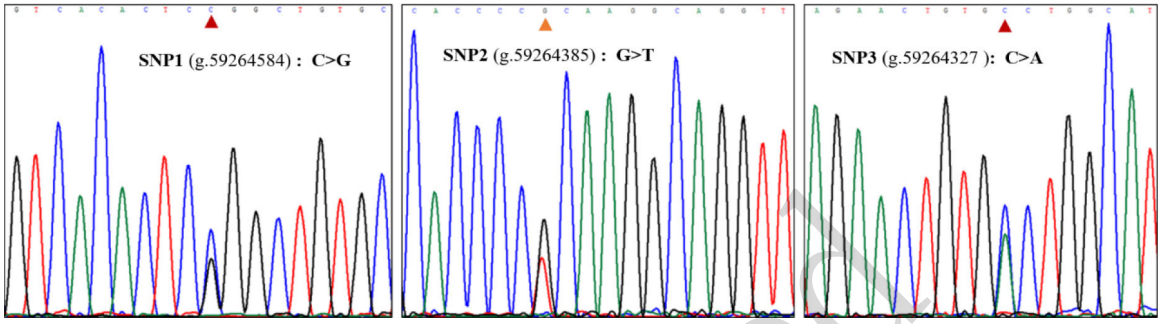


Fig. 2 Identification of SNPs in the promoter region of lncPYGL

To investigate whether the identified SNPs affected lncPYGL promoter activity, we constructed mutant pGL3-Basic vectors for each of the three SNP loci and performed dual-luciferase reporter assays. Mutations at SNP1 and SNP2 caused no significant change in promoter activity (Figs.3a-3f). In contrast, mutation at SNP3 significantly decreased promoter activity, as indicated by the decreased luciferase activity (Figs. 3g-3i). These results suggest that the SNP3 mutation may alter the binding affinity of transcription factors to the lncPYGL promoter, thereby affecting its transcriptional activity.

Because SNP3 genotypes significantly affect promoter activity, qRT-PCR was performed to evaluate lncPYGL expression levels in goats with different SNP3 genotypes. The results showed that lncPYGL expression levels in the wild-type SNP3 genotype were significantly higher than those in the mutant and heterozygous genotypes (Fig. 4). Since lncPYGL overlaps with the protein-coding gene PYGL on the opposite strand, we further investigated whether SNP3 genotypes were associated with PYGL expression to consider the possibility that this genetic variant also influences the neighboring protein-coding gene. However, no significant differences in PYGL mRNA expression were observed among the three SNP3 genotypes (Fig. S1). These findings suggest that SNP3 is associated with changes in lncPYGL expression without detectable differences in PYGL mRNA abundance under the conditions examined.

Uncorrected

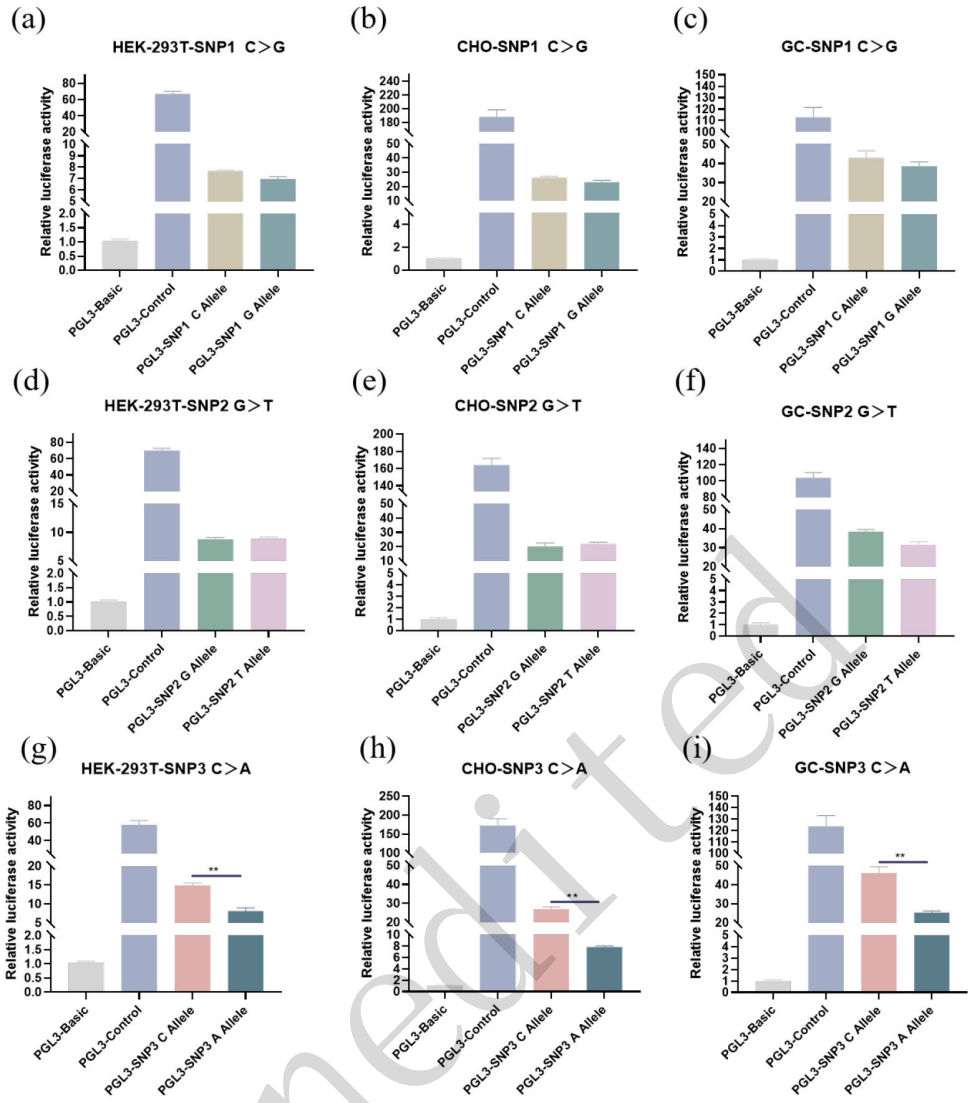


Fig. 3 Effects of SNP genetic variation in the active region of promoter activity. a-c. Effects of different genotypes of SNP1. d-f. Effects of different genotypes of SNP2. g-i. Effects of different genotypes of SNP3.

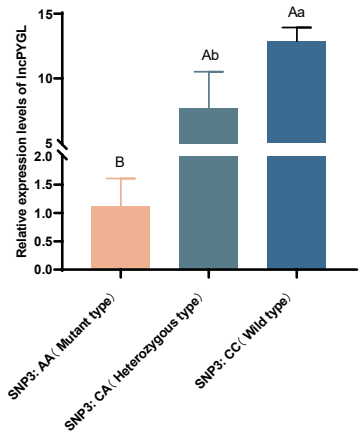


Fig. 4 lncPYGL mRNA expression levels in goats with different genotypes of SNP3

Unedited

3.3 Functional SNP alters the binding affinity of transcription factor ZNF322

To investigate the molecular mechanisms regulating lncPYGL promoter activity, online tools were used to predict transcription factors that could potentially bind to the SNP3 site. On the basis of the reported biological functions of candidate transcription factors and the predicted binding of their motifs to the DNA sequence containing SNP3, ZNF322 and PBX3 were selected for further analysis (Table 2). Overexpression vectors for these were then constructed, and dual-luciferase reporter assays were performed to evaluate their effects on lncPYGL promoter activity. The results showed that ZNF322 primarily bound to the CC genotype of SNP3 in the lncPYGL promoter-active region and significantly enhanced its transcriptional activity. In contrast, PBX3 bound to both the CC and AA genotypes of SNP3 in the lncPYGL promoter active region and suppressed their transcriptional activity (Fig. 5). Subsequently, ChIP-qPCR was performed to investigate the binding of ZNF322 to the lncPYGL promoter region (Fig. 6). The results revealed that ZNF322 preferentially bound to the lncPYGL promoter region harboring the SNP3-C allele, indicating allele-dependent binding specificity.

To further investigate the regulatory effects of ZNF322 and PBX3 on lncPYGL, each transcription factor was overexpressed. The overexpression efficiency and lncPYGL expression levels were determined by qRT-PCR (Fig. 7). The results showed that ZNF322 overexpression significantly promoted lncPYGL expression, whereas PBX3 overexpression significantly inhibited lncPYGL expression.

Table 2 Transcription factors that may be bound to DNA sequences containing SNP3 in the active region of lncPYGL promoter

Location Name	Transcription factor name	Binding motif
lncPYGL-Pro-SNP3-WT	Tcf3	CACCTG
	Ahr	GCGTG
	ZNF322	GAGGCGGGAC
	AR	GGTAC
	TEAD1	GGAATG
	NFYA	CCAAT
lncPYGL-Pro-SNP3-Mut	JUN	TGACTCA
	NFE2L1	TGCTGAGTCAT
	PBX3	TGAT

unedited

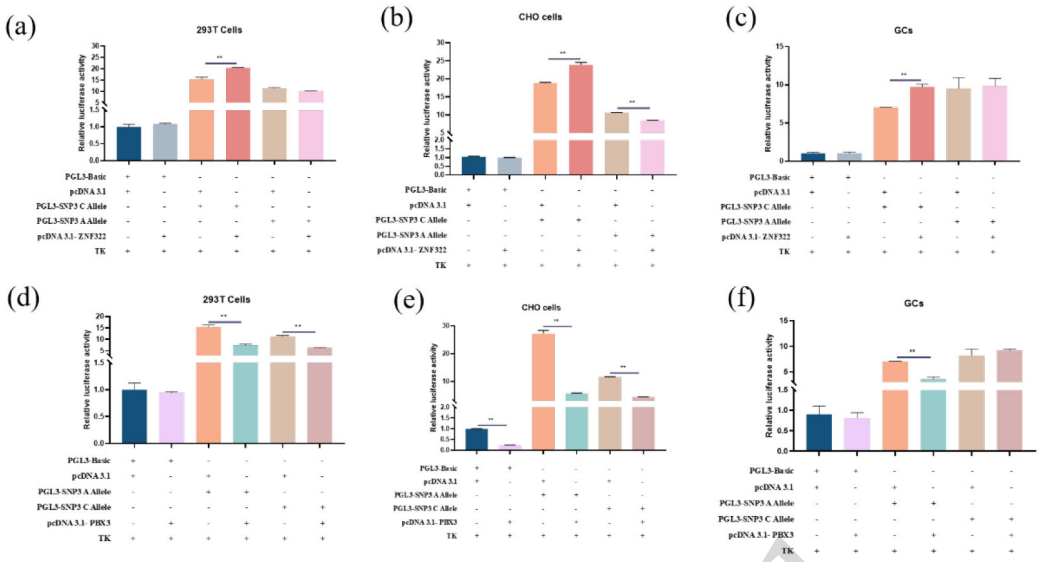


Fig. 5 Detection of luciferase activity following the co-transfection of ZNF322 and PBX3 with different SNP3 genotypes. a-c. Fluorescence activity of ZNF322. d-f. Fluorescence activity of PBX3.

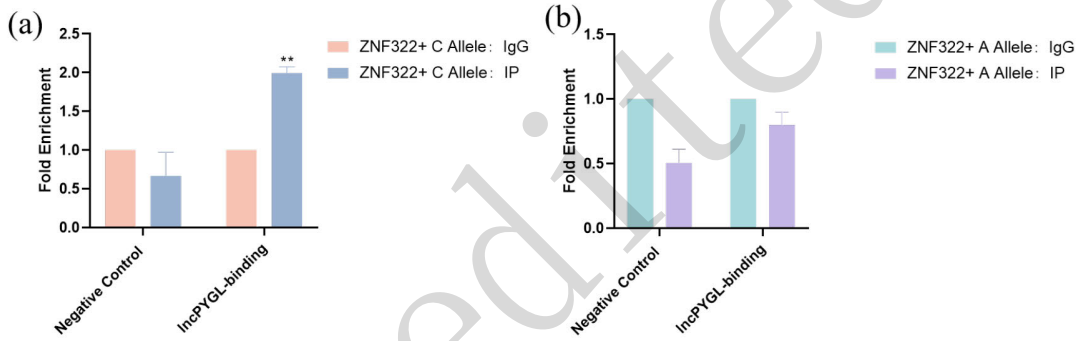


Fig. 6 ChIP-qPCR analysis of ZNF322 enrichment at the lncPYGL promoter carrying different SNP3 alleles. a. SNP3-C; b. SNP3-A.

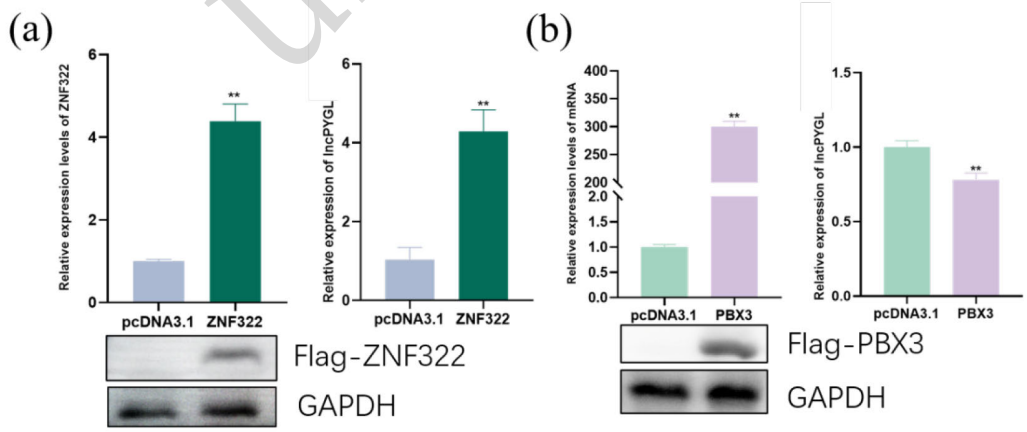


Fig. 7 Effects of ZNF322 and PBX3 overexpression efficiency on lncPYGL expression levels. a. Analysis of ZNF322 mRNA and protein expression levels, as well as lncPYGL mRNA expression levels. b. Analysis of PBX3 mRNA and protein expression levels, as well as lncPYGL mRNA expression levels.

unedited

3.4 Association analysis between SNP and goat litter size

To investigate the relationship between genetic variants in the identified promoter region and goat litter size, we first analyzed the resequencing data of 120 Shaanbei white cashmere goats generated by our research group. All three SNP loci were polymorphic in the population; therefore, these SNPs were further genotyped in 365 Shaanbei white cashmere goats using Sanger sequencing. The analysis of genetic parameters revealed that for SNP1, the frequency of the CC genotype was higher than that of the GC and GG genotypes; for SNP2, the frequency of the GG genotype was lower than that of the GT and TT genotypes; for SNP3, the frequency of the CC genotype was higher than that of the CA and AA genotypes. Chi-square test analysis indicated that all SNPs deviated from the Hardy-Weinberg equilibrium ($P < 0.05$). Based on the PIC values, SNP1 and SNP2 exhibited moderate polymorphism ($0.25 < \text{PIC} \leq 0.50$), while SNP3 exhibited low polymorphism ($\text{PIC} < 0.25$) (Table 3).

To determine whether the genetic variants in the lncPYGL promoter active region were associated with litter size in Shaanbei white cashmere goats, an association analysis was performed between the three SNPs and litter size. The results showed no significant association between SNP1 genotypes and litter size. For SNP2, the GG genotype had a significantly higher litter size than the GT and TT genotypes; for SNP3, the AA genotype had a significantly higher litter size than the CC and CA genotypes. These finding suggest that SNP2 and SNP3 may serve as potential molecular markers associated with litter size in Shaanbei white cashmere goats (Table 4).

Table 3 Genetic parameters of lncPYGL promoter active region SNP loci in the Shaanbei white cashmere goat population

SNP mutation	Genomic location	Genotype	Frequencies		Population genetic parameters				HWE
			Genotype	Allele frequencies	Ho	He	Ne	PIC	*P*-value
SNP1 C > G	g. 59264584	CC(n=264)	0.7233	0.8118(C)	0.6945	0.3055	1.440	0.2589	2.2188E-15
		GC(n=65)	0.1781						
		GG(n=36)	0.0986	0.1882(G)					
SNP2 G > T	g. 59264385	GG(n=106)	0.2904	0.4658(G)	0.5023	0.4977	1.9907	0.3738	1.6793E-08
		GT(n=128)	0.3507						
		TT(n=131)	0.3589	0.5342(T)					
SNP3 C > A	g. 59264327	CC(n=270)	0.7397	0.8411 (C)	0.7327	0.2673	1.3648	0.2316	3.9354E-06
		CA(n=74)	0.2027						
		AA(n=21)	0.0575	0.1589(A)					

Table 4 Association analysis between SNP sites in the active region of lncPYGL promoter and litter size

SNP mutation name	Genotype	Litter sizes	*P*-Value
SNP1 C > G	CC(n=264)	1.633 ± 0.035	0.972306
	GC(n=65)	1.615 ± 0.071	
	GG(n=36)	1.639 ± 0.096	
SNP2 G > T	GG(n=106)	1.755 ± 0.055^{a}	0.018906
	GT(n=128)	1.547 ± 0.050^{b}	
	TT(n=131)	1.611 ± 0.050^{b}	
SNP3 C > A	CC(n=270)	1.600 ± 0.034^{B}	0.008182
	CA(n=74)	1.635 ± 0.066^{B}	
	AA(n=21)	2.000 ± 0.124^{A}	

Unedited

To further investigate the genetic relationships among the three SNPs identified in the lncPYGL promoter region, linkage disequilibrium (LD) and haplotype analyses were performed using genotype data from 365 Shaanbei white cashmere goats. Pairwise LD analysis revealed high D' values among the three SNPs, namely 0.966, 0.917 and 1.000 for SNP1 – SNP2, SNP1–SNP3, and SNP2–SNP3, respectively. However, the corresponding r^2 values were relatively low (0.247, 0.037 and 0.217, respectively), indicating limited allelic correlations despite the high D' values (Fig. S2). Three-locus haplotype analysis identified four major haplotypes, namely CTC (53.08%), GGC (18.18%), CGA (15.64%) and CGC (12.51%), with alleles ordered as SNP1–SNP2–SNP3. Among these, CTC was the most frequent haplotype in the study population (Fig. S3). Exploratory haplotype-based association analysis was performed using the estimated haplotype dosages of the three SNPs. Taking CTC as the reference haplotype, CGA showed a positive association with first-parity litter size, whereas GGC and CGC exhibited no statistically significant associations (Table S1). However, these findings should be interpreted as preliminary because the analysis did not fully account for uncertainty in haplotype phase estimation or potential population structure.

4. Discussion

The role of non-coding RNAs in regulating reproductive performance in female livestock has long attracted considerable attention. However, most studies have focused on their conservation and intracellular functions, such as the regulation of competing endogenous RNA (ceRNA) and protein interactions (Liu et al., 2024). Research has found that lincRNA HEOE regulates follicular dynamics in pigs through the miR-16-ZMAT3 axis and PGF2α, thereby influencing the estrous cycle and fertility (Mattick et al., 2023). The lncRNA NORSF acts as a ceRNA for miR-339: it relieves the regulatory effect of miR-339 on CYP19A1, which encodes aromatase (P450arom), thereby affecting estradiol secretion (Wang et al., 2023). However, relatively little attention has been paid to the regulatory mechanisms at the level of transcriptional initiation. This study provides new insights into the upstream genetic basis of differential lncPYGL expression between high- and low-fecundity goats, contributing to a more comprehensive understanding of lncRNA functional studies from the perspective of transcriptional regulation.

Our experimental results showed that the transcriptional activity of lncPYGL is primarily regulated by its core promoter region (−1032 bp to −669 bp). Similarly, promoter variants critically influence reproduction-related genes. For example, studies on the FSHR promoter have found that SNP variants can significantly affect transcriptional activity (Simoni et al., 2020). Variants in the CYP19A1 promoter region affect aromatase expression levels (Hashemain et al., 2022). These subtle transcriptional differences may be amplified during follicular development through a cascade amplification effect within the concerned signaling pathways. Simultaneously, significant feedback regulatory mechanisms exist during follicular development (Li et al., 2021). Therefore, the differential expression of lncPYGL may influence reproductive function in animals by altering the dosage-dependent effects of non-coding regulatory factors.

Transcription factors are the core determinants of gene expression. The transcription factor ZNF322 identified in this study has been shown to participate in the regulation of cell proliferation and differentiation. For example, in cancer research, ZNF322 was found to form a protein complex with c-Jun, which regulates cell cycle-related genes and promotes cell proliferation. As a typical C2H2-type zinc finger protein, ZNF322 modulates target gene expression by binding to GC-rich DNA regions (Jen et al., 2016). Luciferase reporter and transcription factor binding site determination analysis confirmed that ZNF322 specifically recognizes the lncPYGL promoter region.

Notably, in this study, SNP3 was associated with differences in ZNF322 occupancy at the lncPYGL promoter. Our ChIP-qPCR analysis revealed significant ZNF322 enrichment at the

Unedited

promoter in samples carrying the CC genotype, whereas no significant enrichment was detected in AA samples. Together with the dual-luciferase reporter assays, these results support a role of SNP3 in modulating the ZNF322-mediated transcriptional regulation of lncPYGL. However, ChIP-qPCR does not directly quantify DNA–protein binding affinity, and the observed differences in promoter occupancy may also be influenced by the local chromatin environment. SNP-transcription factor (SNP–TF) interaction patterns have been widely reported in other species. For example, Degtyareva et al. pointed out that a large number of disease-associated SNPs are located within transcription factor binding sites and can affect gene expression (Degtyareva et al., 2021). In the present study, the SNP3-C allele exhibited stronger ZNF322 binding and higher promoter activity than the SNP3-A allele, suggesting that this nucleotide substitution may modulate lncPYGL transcription by altering ZNF322 binding. Therefore, the SNP–TF–lncRNA regulatory framework identified in this study provides functional evidence linking a non-coding genetic variant to transcriptional regulation, although its contribution to reproductive phenotypes remains to be fully established. Importantly, the apparent inverse relationship was inferred from genotype-group comparisons rather than a direct correlation analysis between ovarian lncPYGL expression and litter size in the same individuals. Therefore, these findings do not establish that higher lncPYGL expression directly leads to reduced prolificacy.

An important finding of this study is the apparent inverse relationship between SNP3-associated lncPYGL expression and litter size. Consistent with the allele-specific differences in ZNF322 binding and promoter activity, ovarian lncPYGL expression was significantly higher in goats carrying the wild-type CC genotype than those carrying the heterozygous CA or mutant AA genotype. These opposing trends suggest that increased lncPYGL expression associated with the SNP3-C allele may be linked to reduced reproductive performance, raising the possibility that lncPYGL negatively regulates litter size.

Nevertheless, the above interpretation should be treated with caution. Our previous transcriptomic analysis revealed higher overall lncPYGL expression in ovarian tissues from consistently multiple-kidding goats than those from consistently single-kidding goats, which appears inconsistent with the inverse relationship observed across SNP3 genotypes. This discrepancy suggests that the relationship between lncPYGL expression and reproductive performance may be more complex than a simple negative association. Differences in genetic background, ovarian physiological status, or transcriptional regulatory mechanisms may all contribute to these observations. Moreover, lncPYGL expression may be influenced by regulatory factors other than SNP3, and its biological effects may depend on specific transcript isoforms and cellular contexts. Importantly, the present study does not directly establish a causal relationship between lncPYGL expression and litter size. Therefore, although the SNP3–ZNF322–lncPYGL regulatory axis provides a plausible mechanism for the observed genotype-associated expression differences, its contribution to reproductive performance remains to be elucidated. Future studies integrating genotype-stratified expression analysis, isoform-specific functional validation, and independent population-based investigations are needed to clarify the biological role of lncPYGL in goat prolificacy.

Given the close physical proximity of the three SNPs, linkage disequilibrium and haplotype analyses were performed to further evaluate their genetic relationships. Although SNP2 and SNP3 are separated by only 58 bp, they exhibited a high D' but relatively low r^2 value, indicating limited allelic correlation despite strong normalized linkage disequilibrium. All three SNPs showed significant deviations from the Hardy–Weinberg equilibrium (HWE). Although the SNPs were identified through resequencing and further validated by Sanger sequencing, potential genotyping errors cannot be completely excluded without systematic genotype concordance assessment. In addition, non-random mating, artificial selection, genetic relatedness, and population substructure may have contributed to the observed HWE deviations. All 365 Shaanbei white cashmere goats originated from the same breeding farm and birth cohort and were maintained under generally

Unedited

comparable feeding and management conditions, which helped reduce environmental variation associated with farm origin, birth year and birth season. However, this relatively homogeneous sampling design cannot eliminate potential confounding effects arising from family structure, genetic relatedness, or differences in reproductive management. Due to the lack of detailed pedigree information and individual-level management records, these factors could not be adequately accounted for in the statistical analysis. Consequently, the observed associations between SNP genotypes and litter size should be interpreted with caution. Further studies incorporating systematic genotyping quality control, comprehensive pedigree and management information, appropriate mixed-effects models, and validation in independent populations are necessary to confirm the robustness of these associations and evaluate the potential breeding utility of the identified SNPs.

In molecular breeding, genetic markers identified through association studies may reflect linkage disequilibrium with underlying causal variants rather than direct functional effects (Tam et al., 2019). Herein, functional assays provided evidence that SNP3 participates in the ZNF322-mediated transcriptional regulation of lncPYGL. Nevertheless, its contribution to litter size and its predictive performance across different genetic backgrounds remain uncertain. Given the significant deviations from HWE and the potential influence of population structure and genetic relatedness, SNP3 should currently be considered a candidate regulatory variant rather than a validated marker for breeding applications. Further evaluation in independent populations is prompted before its use in marker-assisted selection can be recommended.

This study focused on elucidating the upstream genetic basis of differential lncPYGL expression rather than its downstream biological functions. Although it has been shown that lncRNAs can participate in the regulation of cellular processes through ceRNA networks or protein interactions, functional interpretations often lack genetic support in the absence of a clear transcriptional regulatory mechanism. The present findings provide evidence that the SNP3 C>A substitution is associated with altered ZNF322 binding, lncPYGL promoter activity, and ovarian lncPYGL expression. Nevertheless, the downstream molecular pathways connecting these transcriptional changes to reproductive phenotypes are yet to be uncovered.

In summary, the SNP–TF–lncRNA regulatory model proposed in this study provides a new framework for investigating complex reproductive traits in goats. Specifically, the SNP3-C allele favors the ZNF322-mediated transcriptional activation of lncPYGL, whereas the A allele is associated with reduced promoter activity and lower lncPYGL expression. The opposite trends observed between SNP3-associated lncPYGL expression and litter size suggest a potential negative association, although the underlying causal relationship remains to be established. Future studies should further explore whether lncPYGL participates in follicular development by regulating classic pathways, such as TGF-β or BMP signaling.

5. Conclusions

This study systematically investigated the transcriptional regulatory mechanisms of lncPYGL and its association with reproductive performance in Shaanbei white cashmere goats. The core promoter region of lncPYGL was located between −1032 bp to −669 bp upstream of the transcription start site. Three SNPs (SNP1: g.59264584 C>G; SNP2: g.59264385 G>T; SNP3: g.59264327 C>A) were identified in the promoter region, of which SNP2 and SNP3 were significantly associated with litter size. Notably, the SNP3-C allele exhibited stronger ZNF322 binding and higher promoter activity than the A allele, consistent with the higher ovarian lncPYGL expression observed in goats with the CC genotype than those with the CA or AA genotype. However, the AA genotype was associated with a higher mean litter size than the CC genotype, suggesting a potential inverse relationship between SNP3-associated lncPYGL expression and litter size. These findings suggest that SNP3 may regulate lncPYGL transcription by altering ZNF322

Unedited

binding, although a causal relationship between lncPYGL expression and litter size has not yet been established (Fig. 8).

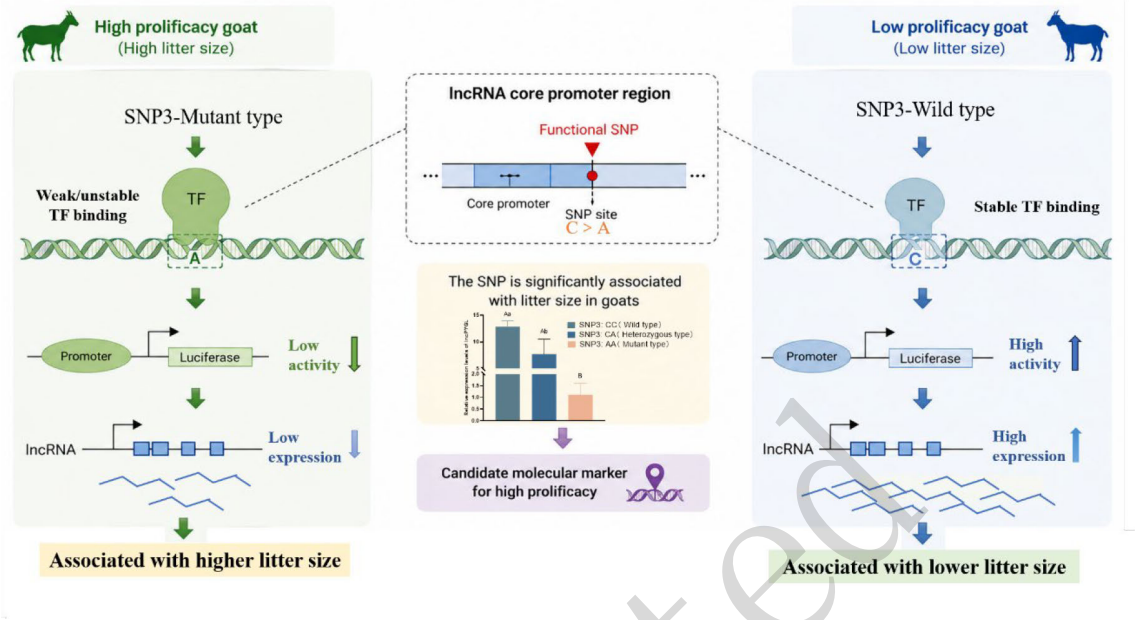


Fig. 8 Schematic model of the transcriptional regulatory mechanism of lncPYGL and its association with litter size

Unedited

Data availability statement: Data are available from the corresponding author upon reasonable request and institutional approval.

Acknowledgments: This work was supported by the National Natural and Science Foundation of China (No. 32272850). The National Key R&D Program of China (2022YFD1300200; 2022YFF1000100).

Author contributions: Validation, Haiyan Yang, Chunxiao Li, and Yifan Zhang; Investigation, Haiyan Yang, Yaohui Wang; Resources, Haijing Zhu, Xianyong Lan; Data Curation, Haiyan Yang, Yifan Zhang; Writing – Original Draft Preparation, Haiyan Yang; Writing – Review & Editing, Xianyong Lan, Chuanying Pan and Ebadu Areb; Visualization, Chunxiao Li, Yifan Zhang, and Yangyang Bai; Supervision, Xianyong Lan, Chuanying Pan, and Haijing Zhu; Project Administration, Xianyong Lan, Chuanying Pan; Funding Acquisition, Chuanying Pan.

Compliance with ethics guidelines:

(1) Conflicts of Interest: Haiyan YANG, Yifan ZHANG, Chunxiao LI, Yaohui WANG, Ebadu AREB, Yangyang BAI, Haijing ZHU, Xianyong LAN, and Chuanying PAN declare that they have no conflicts of interest.

(2) Ethics Statement: All institutional and national guidelines for the care and use of laboratory animals were followed. Sample collection followed China's national standard for Laboratory Animal Welfare and Ethical Review (GB/T 35892-2018). The experimental procedures were authorized by the Review Committee for the Use of Animal Subjects at the Northwest A&F University (Approval No. IACUC2024-1108; Date: 8 March 2024), conducted in compliance with the ethics commission guidelines.

Declaration on the use of generative AI tools: During the preparation of this work, the authors used DeepSeek and ChatGPT to improving language fluency and checking grammatical errors in the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Degtyareva, A.O., Antontseva, E.V., Merkulova, T.I., 2021. Regulatory SNPs: altered transcription factor binding sites implicated in complex traits and diseases. International journal of molecular sciences, 22(12):6454.http://dx.doi.org/10.3390/ijms22126454

Feng, D., Wei, C., Hu, S.-Y., *et al.*, 2026. Decoding Complex Traits in Goats Through Genome-Wide Association Studies: Progress, Challenges, and Perspectives. International Journal of

Molecular Sciences, 27(7):2945.http://dx.doi.org/10.3390/ijms27072945

Georges, M., Charlier, C., Hayes, B., 2019. Harnessing genomic information for livestock improvement. Nature Reviews. Genetics, 20(3):135–156.http://dx.doi.org/10.1038/s41576-018-0082-2

Gholizadeh, M., Mohammadnezhad, M., 2024. Genetic insights into litter size in goats: a meta-analysis of KISS1 and BMP15 SNP variants. Animal Reproduction Science, 271:107632.http://dx.doi.org/10.1016/j.anireprosci.2024.107632

Han, Y., Cao, G., Chen, W., *et al.*, 2025. The role of TGF-β signaling pathway in determining small ruminant litter size. Biology, 14(7):786.http://dx.doi.org/10.3390/biology14070786

Hashemain, Z., Amiri-Yekta, A., Khosravifar, M., *et al.*, 2022. CYP19A1 promoters activity in human granulosa cells: a comparison between PCOS and normal subjects. Cell Journal, 24(4):170–175.http://dx.doi.org/10.22074/cellj.2022.7787

Hassanine, N.N.A.M., Saleh, A.A., Essa, M.O.A., *et al.*, 2025. Candidate genes, markers, signatures of selection, and quantitative trait loci (QTLs) and their association with economic traits in livestock: genomic insights and selection. International journal of molecular sciences, 26(16):7688.http://dx.doi.org/10.3390/ijms26167688

Jen, J., Lin, L.-L., Chen, H.-T., *et al.*, 2016. Oncoprotein ZNF322A transcriptionally deregulates alpha-adducin, cyclin D1 and p53 to promote tumor growth and metastasis in lung cancer. Oncogene, 35(18):2357–2369.http://dx.doi.org/10.1038/onc.2015.296

Li, L., Shi, X., Shi, Y., *et al.*, 2021. The signaling pathways involved in ovarian follicle development. Frontiers in Physiology, 12:730196.http://dx.doi.org/10.3389/fphys.2021.730196

Lian, Z., Zou, X., Han, Y., *et al.*, 2020. Role of mRNAs and long non-coding RNAs in regulating the litter size trait in chuanzhong black goats. Reproduction in Domestic Animals = Zuchthygiene, 55(4):486–495.http://dx.doi.org/10.1111/rda.13642

Ling, Y., Xu, L., Zhu, L., *et al.*, 2017. Identification and analysis of differentially expressed long non-coding RNAs between multiparous and uniparous goat (capra hircus) ovaries. PLOS One, 12(9):e0183163.http://dx.doi.org/10.1371/journal.pone.0183163

Liu, M., Chen, J., Liu, S., *et al.*, 2024. LH-stimulated periodic lincRNA HEOE regulates follicular dynamics and influences estrous cycle and fertility via miR-16-ZMAT3 and PGF2α in pigs. International Journal of Biological Macromolecules, 281(Pt 3):136426.http://dx.doi.org/10.1016/j.ijbiomac.2024.136426

Mattick, J.S., Amaral, P.P., Carninci, P., *et al.*, 2023. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nature Reviews. Molecular Cell Biology, 24(6):430–447.http://dx.doi.org/10.1038/s41580-022-00566-8

McDonald, E.A., Smith, J.E., Cederberg, R.A., *et al.*, 2016. Divergent activity of the gonadotropin-releasing hormone receptor gene promoter among genetic lines of pigs is partially conferred by nuclear factor (NF)-B, specificity protein (SP)1-like and GATA-4 binding sites. Reproductive Biology and Endocrinology: RB&E, 14(1):36.http://dx.doi.org/10.1186/s12958-016-0170-0

Meza-Herrera, C.A., Navarrete-Molina, C., Macias-Cruz, U., *et al.*, 2024. Dairy Goat Production Systems: A Comprehensive Analysis to Reframe Their Global Diversity. Animals: an open access journal from MDPI, 14(24):3717.http://dx.doi.org/10.3390/ani14243717

Simoni, M., Brigante, G., Rochira, V., *et al.*, 2020. Prospects for FSH treatment of Male infertility. Journal of Clinical Endocrinology and Metabolism, 105(7):dgaa243.http://dx.doi.org/10.1210/clinem/dgaa243

Tam, V., Patel, N., Turcotte, M., *et al.*, 2019. Benefits and limitations of genome-wide association studies. Nature Reviews. Genetics, 20(8):467–484.http://dx.doi.org/10.1038/s41576-019-0127-1

Usman, M., Li, A., Wu, D., *et al.*, 2024. The functional role of lncRNAs as ceRNAs in both ovarian processes and associated diseases. Non-coding RNA Research, 9(1):165–177.http://dx.doi.org/10.1016/j.ncrna.2023.11.008

Under Peer Review

Wang, J., Chen, H., Zhang, Y., *et al.*, 2024. Long non-coding RNA Loc105611671 promotes the proliferation of ovarian granulosa cells and steroid hormone production upregulation of CDC42. Frontiers in Veterinary Science, 11:1366759.http://dx.doi.org/10.3389/fvets.2024.1366759

Wang, J., Li, Q., Wang, Z., *et al.*, 2025. LncRSFD alternative splicing modulates the proliferation and differentiation of adipocytes through sponging miRNAs. Animal Research and One HEALTH.http://dx.doi.org/10.1002/aro2.70040

Wang, M., Wang, Y., Yang, L., *et al.*, 2023. Nuclear lncRNA NORSF reduces E2 release in granulosa cells by sponging the endogenous small activating RNA miR-339. BMC Biology, 21(1):221.http://dx.doi.org/10.1186/s12915-023-01731-x

Yang, Y., Zhang, S., Chen, L., *et al.*, 2026. CircMED17 regulates bovine adipogenesis and fat deposition via mediating the proliferation and differentiation of bovine adipocytes. Journal of Zhejiang University-SCIENCE B.http://dx.doi.org/10.1631/jzus.B2500442

Zhang, D., Zhang, Q., Wang, L., *et al.*, 2022. Alternative splicing isoforms of porcine CREB are differentially involved in transcriptional transactivation. Genes, 13(8):1304.http://dx.doi.org/10.3390/genes13081304

unedited