scientific reports

Check for updates

OPEN

Exploring the polygenic landscape of wool traits in Turkish Merinos through multi-locus GWAS approaches: middle Anatolian Merino

Yalçın Yaman[1]✉, A. Taner Önaldi[2], Şükrü Doğan[2], Mesut Kirbaş[2], Sedat Behrem[3] & Yavuz Kal[2]

This study investigates the genetic underpinnings of wool traits, specifically fibre diameter (FD) and staple length (SL), in Middle Anatolian Merino sheep using multi-locus genome-wide association study (GWAS) approaches. Representing the first attempt to examine these polygenic traits with multi-locus methods, the analysis employed four techniques: mrMLM, FASTmrMLM, FASTmrEMMA, and ISIS EM-BLASSO. A total of 18 Quantitative Trait Nucleotides (QTNs) were identified for FD, with 7 co-detected by multiple methods, and 14 QTNs were identified for SL, with 5 co-detected by multiple methods. Post-hoc power analysis revealed high statistical power for both traits (FD: 0.95, SL: 0.91). Notably, three candidate genes—*PTPN3*, *TCF4*, and *ZBTB8A*—were found to be consistent with prior studies. Gene enrichment and pathway analyses reaffirmed the complex and multifactorial molecular mechanisms governing wool traits. These findings enhance our understanding of the polygenic nature of wool traits, shedding light on the intricate genetic regulation and pinpointing genomic regions potentially influencing wool physiology. By identifying specific QTNs associated with FD and SL, this research provides a foundation for elucidating the genetic mechanisms underlying these economically significant traits. Upon validation in diverse populations, these findings hold substantial promise for the application of marker-assisted selection (MAS) to improve wool traits.

Keywords Middle Anatolian Merino, Wool traits, Dual-purpose breeding, Multi-locus GWAS, Growth performance

Throughout history, textile production has relied heavily on fibers from domesticated sheep fleece, with wool maintaining its significance as a premier animal-based material in textile manufacturing[1]. This natural fiber, derived from ovine coats, has been utilized for millennia, with felt likely representing the earliest form of woolen fabric. Wool's versatility and unique properties have ensured its continued economic importance in the textile industry, despite the rise of synthetic alternatives[2]. Sheep wool has long been a staple in textile production, offering a natural and renewable resource for a wide array of products. Its applications span from carpets and garments to curtains, covers, and bedding, showcasing its adaptability and enduring appeal[3]. While synthetic fibers have diminished wool's overall market prominence, it retains significant value in specific textile niches, testament to its irreplaceable qualities and the industry's recognition of its distinct advantages. This enduring relevance of wool in textiles underscores not only its historical importance but also its continued role in modern manufacturing, balancing tradition with contemporary demands for sustainable and high-performance materials[2].

Beyond its significance in the textile industry, sheep wool has gained recognition as an exceptional insulation material, prized for its thermal, hygroscopic, and acoustic properties. With a thermal conductivity (λ) of 0.04 W/m·K, it performs nearly as well as popular synthetic insulators like expanded polystyrene (EPS),

[1]Department of Genetics, Faculty of Veterinary Medicine, Siirt University, Siirt 56000, Turkey. [2]Bahri Dagtas International Agricultural Research Institute, Konya 42000, Turkey. [3]Department of Genetics, Faculty of Veterinary Medicine, Aksaray University, Aksaray 68000, Turkey. ✉email: yalcinyaman@gmail.com; yalcin.yaman@siirt.edu.tr

polyurethane (PUR), and cellulose insulation, which have λ values of 0.038 and 0.033 W/m·K respectively. In the construction sector, sheep wool aligns perfectly with green building principles. Its eco-friendly nature, abundance, renewability, and full recyclability make it an ideal choice for sustainable building practices. Moreover, when compared to conventional insulation materials, sheep wool offers significant advantages in terms of sustainability. It not only reduces production costs for insulation materials but also minimizes environmental impact, making it a win-win solution for both economic and ecological considerations[4,5]. Additionally, as a potential insulation material in construction, sheep wool offers significant advantages in case of fire. Wool fibers burn slowly and, in the event of a fire, produce minimal sputtering compared to polystyrene, which is highly flammable and emits toxic fumes[6]. Studies have also demonstrated that sheep wool has the potential to be used as an effective acoustic insulation material[7].

Furthermore, recent research has revealed promising agricultural applications for sheep wool beyond its traditional use in textiles. Experiments have demonstrated the potential of wool as a substrate in dendro-horticultural crops, showing enhanced plant growth in trays containing washed wool or a wool-soil mixture[8]. This natural, durable, and biodegradable material not only improved plant development but also proved to be an environmentally safe fertilizer. The study suggests that using sheep wool as a substrate in dendro-horticultural practices can simultaneously boost plant growth and align with economic and environmental protection goals, offering a sustainable solution for agricultural practices.

In recent decades, there has been a growing emphasis on dual-purpose sheep production systems, aiming to optimize both meat and wool outputs to enhance overall economic viability. Research institutions are focusing on improving sheep's production potential for wool, mutton, and milk. This adaptive approach aims to develop breeds excelling in both wool and meat production, creating more resilient and profitable farming systems while addressing waste management and market challenges.

Turkey initiated such a dual-purpose sheep production program in the 1950s at the Bahri Dagdas International Agricultural Research Institute (BDIARI), operating under the country's Ministry of Agriculture and Forestry. This strategic effort involved crossbreeding the indigenous Akkaraman sheep, known for its hardiness and adaptation to local conditions, with German Mutton Merino rams. When achieving the G3 crossbred line, the nucleus herds were closed to backcrossing at least three decades ago and continued to be bred within the G2 and G3 lines. The composite breed obtained as a result of backcrossing was named Middle Anatolian Merino, and due to its approximately 70-year history, it has been registered as a native breed by the Turkish Ministry of Agriculture and Forestry. The primary objectives of this breeding program were to enhance growth characteristics and improve fleece quality, thereby creating a more versatile and economically valuable sheep breed. This initiative exemplified Turkey's proactive approach to modernizing its agricultural practices and optimizing livestock production to meet evolving market demands has been implemented in the nucleus herds since closing the backcrossing, particularly focusing on growth performance and classical breeding methods.

The Middle Anatolian Merino has gained considerable popularity among sheep farmers and has become widely established throughout Turkey's Central Anatolia region. Its dual-purpose characteristics and adaptation to local conditions have made it a preferred choice for many breeders in the area. In addition to classical breeding efforts, our team conducted a multi-locus Genome-Wide Association Study (GWAS) to investigate the polygenic effects influencing growth traits (unpublished data). Despite this research, there is still a notable absence of genetic studies specifically targeting wool traits for synchronized genomic selection in dual-purpose breeds. This gap highlights the need for further exploration into the genetic basis of wool traits to enhance breeding strategies aimed at improving both meat and wool production simultaneously in these versatile livestock breeds.

In the study described here, we investigated the genetic underpinnings of wool quality traits, such as fibre diameter (μ) and staple length (mm), in Middle Anatolian Merino sheep. To delve into these complex quantitative traits—fibre diameter (FD) and staple length (SL)—we employed advanced GWAS approaches. Our research utilized five multi-locus GWAS methods: Multi-Locus Random-Single Nucleotide Polymorphism (SNP)-Effect Mixed Linear Model (mrMLM)[9], Fast Multi-Locus Random-SNP-Effect Mixed Linear Model (FastmrMLM)[10], Iterative sure independence screening EM-Bayesian LASSO (ISIS EM-BLASSO)[11], fast multi-locus random-SNP-effect EMMA (FASTmrEMMA)[12], and polygenic-background-control-based least angle regression plus empirical Bayes (pLARmEB)[13]. These multi-locus GWAS techniques enable the identification of polygenic effects from genetic markers and potential candidate genes associated with traits of interest, including FD and SL. Such traits are typically under polygenic control, reflecting the complex genetic basis of wool characteristics in these dual-purpose sheep breeds.

Results

While moderately high SNP-based heritability was observed for FD ($h^2 = 0.461$), the coefficient of variation for the same trait was low (CV = 7.57%). Conversely, heritability for SL was calculated to be low ($h^2 = 0.191$), but a higher coefficient of variation (CV = 24.437%) was detected for this trait (Table 1).

Traits	*n*	Mean	Min	Max	SD	SE	CV %	h^2	h^2 SE	*p*-value
Fibre diameter (μ)	654	21.5180	16.6000	25.5900	1.6290	0.0637	7.5704	0.4607	0.0978	< 0.0001
Staple length (mm)	654	24.3280	9.3750	45.3780	5.9450	0.2320	24.4369	0.1911	0.0734	< 0.0001

Table 1. Descriptive statistics, calculated SNP-based heritability, and coefficient of variation (%) for FD and SL traits. *SD* standard deviation, *SE* standard error, *CV* Coefficient of variation, h^2 heritability.

The genetic and phenotypic correlations between wool traits and growth traits were calculated to assess the relationships between these traits in the studied dual-purpose population. The analysis revealed a statistically significant positive correlation for both genetic and phenotypic measures between FD-WW and FD-ADG (Table 2).

The assessment of normality for FD and SL data was conducted using Anderson-Darling tests[14]. prior to correlation analyses. The results demonstrated varying distributions across the datasets: For the FD data, the histogram and probability plot indicated a normal distribution, which was statistically confirmed by the Anderson-Darling test ($p = 0.11$; Fig. 1a). However, the SL data exhibited significant deviation from normality ($p < 0.01$; Fig. 1b). The histogram displayed a notably non-normal distribution pattern with irregular peaks, and the probability plot showed clear departures from the theoretical normal line. To address this non-normality in the SL data, a Rank-based Inverse Normal Transformation was applied. This transformation successfully normalized the data, as evidenced by both the symmetrical bell-shaped histogram and the linear probability plot ($p > 0.15$; Fig. 1c). This systematic approach to testing and transforming the data ensured the validity of subsequent correlation analyses by meeting the fundamental assumption of normality.

Using four out of the five multi-locus GWAS methods (mrMLM, FastmrMLM, ISIS EM-BLASSO, and FASTmrEMMA), a total of 18 QTNs were found to be significantly associated with FD at the LOD ≥ 5 threshold. No QTN associated with the FD trait that met the LOD ≥ 5 threshold were detected using the pLARmEB method (Table 3; Fig. 2a). Some QTNs were co-detected by multiple multi-locus GWAS methods.

A total of 14 QTNs associated with the SL trait, meeting the significance criterion of LOD ≥ 5, were identified using the multi-locus GWAS methods mrMLM, FastmrMLM, and ISIS EM-BLASSO (Table 4; Fig. 2b). However, FASTmrEMMA and pLARmEB failed to detect any SL-associated QTNs. Similar to the FD trait, some QTNs were co-detected for SL.

No obvious inflation or deflation was detected in the QQ plots, indicating that possible systematic bias resulting from cryptic relatedness or population stratification was effectively controlled (Fig. 3a and b).

A post hoc statistical power analysis was conducted using the median values of r^2 and MAF associated with the identified QTNs related to FD and SL. The median values of r^2 and MAF for FD were calculated to be 2.003 and 0.390, respectively, while for SL, the median values of r^2 and MAF were calculated to be 1.669 and 0.316, respectively. The statistical power was calculated to be 0.95 for FD and 0.91 for SL (Fig. 4a and b).

Genes in Linkage Disequilibrium (LD) with four intergenic QTNs that met the adopted D' and r^2 threshold values (D' > 0.7, $r^2 > 0.5$) were identified. Attempts to detect additional linked protein-coding genes using lower D' and r^2 values did not reveal any genes in LD with other intergenic QTNs (Table 5). All SNPs in LD with trait-associated QTNs are provided in Supplementary Table S1.

Gene annotation analysis revealed a significant number of protein-coding genes within the ± 200 KB range of the associated QTNs. Additionally, eight QTNs related to FD (Table 6) and three QTNs associated with SL (Table 7) were located within the intronic regions of these genes, suggesting that these particular QTNs and their harbouring genes are especially worth further investigation for their potential role in influencing wool traits.

Discussion

The data on FD and SL reveal different patterns of heritability and variability. FD shows moderately high heritability ($h^2 = 0.461$) but low variability (CV = 7.570%), suggesting that while genetic factors significantly influence this trait, there is limited variation within the population. Conversely, SL exhibits low heritability ($h^2 = 0.191$) but high variability (CV = 24.437%), indicating that environmental factors may play a larger role, yet there is more diversity in this trait among the sheep. For FD, genetic improvement could be effective but may be limited by the low variability. On the other hand, genetic improvement might be more challenging due to low heritability for SL, but the high variability provides a potential for change if effective selection methods are used.

The most notable findings from genetic and phenotypic correlation analyses were the moderate positive genetic correlations between FD and both weaning weight (WW) (rG = 0.385, $p < 0.001$) and average daily gain (ADG) (rG = 0.455, $p < 0.001$). These relationships were also reflected in the phenotypic correlations, albeit to a lesser extent (rP = 0.168 and 0.187, respectively; $p < 0.001$ for both). These results suggest that lambs with higher FD wool have faster-growing speed and higher weaning weights. This finding aligns with some previous studies[15–17] but contradicts others[18], underscoring the challenge faced by breeders in simultaneously improving both growth rates and wool fineness.

Traits	*n*	rG	SE (rG)	*p*-value	rP	*p*-value
FD-SL	654	−0.0433	0.1435	0.3829	−0.0103	0.7920
FD-BW	654	−0.1574	0.1091	0.0785	0.0071	0.8550
FD-WW	654	0.3852	0.1108	0.0003	0.1683	0.0001
FD-ADG	654	0.4552	0.1123	0.0000	0.1865	0.0001
SL-BW	654	0.2123	0.1591	0.0919	0.0502	0.2000
SL-WW	654	0.1373	0.1847	0.2305	0.0192	0.6240
SL-AGD	654	0.1226	0.1988	0.2718	0.0107	0.7840

Table 2. The genetic and phenotypic correlations between wool traits and growth traits. *FD* Fibre Diameter, *SL* Staple Length, *BW* Birth weight, *WW* Weaning weight, *ADG* Average Daily Gain, *rG* Genetic correlation, *rP* Phenotypic correlations.

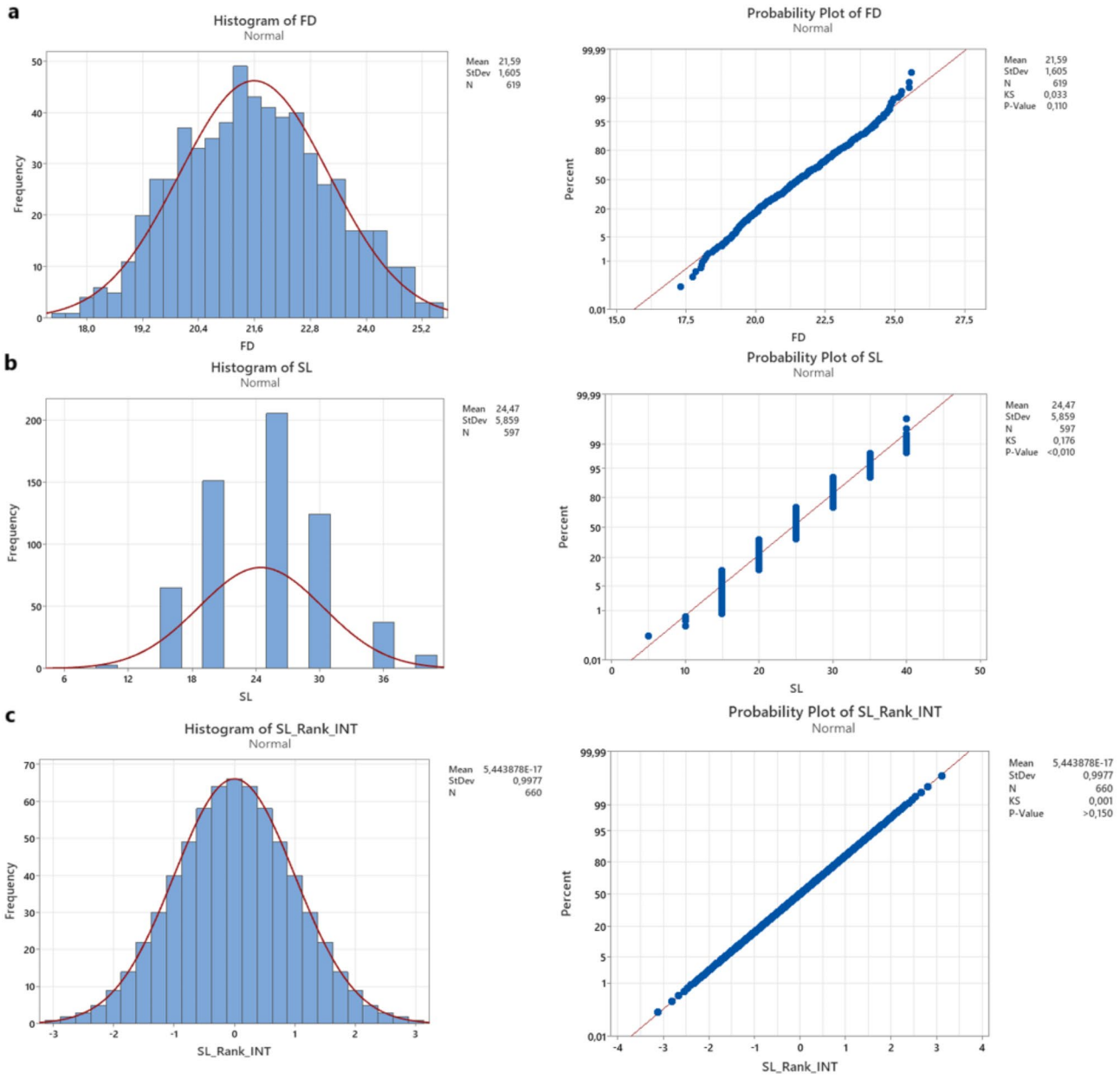

Fig. 1. Distribution analysis of FD and SL measurements before and after normalization. In the figure, (**a**) Histogram and QQ plot of raw FD data showing normal distribution. The red curve in the histogram represents the theoretical normal distribution, while the QQ plot demonstrates good alignment with the theoretical normal line. (**b**) Histogram and QQ plot of raw SL data showing non-normal distribution. The deviation from normality is evident in both the histogram's shape and the QQ plot's departure from the theoretical normal line. (**c**) Histogram and QQ plot of SL data after Rank-Based Inverse Normal Transformation. The transformation successfully normalized the data, as demonstrated by both the symmetrical bell-shaped histogram and the linear alignment in the QQ plot.

SL exhibited weak and non-significant correlations with all growth traits examined (BW, WW, and ADG). The genetic correlations ranged from 0.123 to 0.212, while phenotypic correlations were even weaker (0.011 to 0.050), with none reaching statistical significance. This relative independence of SL trait from growth traits suggests that selection for improved staple length could potentially be achieved without significantly impacting growth characteristics, or vice versa.

The weak and non-significant negative genetic correlation between FD and SL (rG = -0.043, $p = 0.383$) and the negligible phenotypic correlation (rP = -0.010, $p = 0.792$) indicate that these two wool traits are largely independent. Previous studies have reported both positive[19] and negative correlations[16] between these traits. The present findings contradict earlier studies, emphasising the need for breed-specific analyses in wool trait selection programmes.

Marker	Method	CHR	QTN Effect	LOD Score	-log10 (*p*)	r^2 (%)	MAF	Allele
s05784.1	mrMLM	1	−0.386	5.326	6.135	2.627	0.381	A
	FASTmrMLM		−0.348	5.34	6.149	2.211		
OAR2_13110080.1	mrMLM	2	−0.411	5.182	5.985	3.578	0.424	A
	FASTmrMLM		−0.319	5.468	6.282	2.221		
	ISIS EM-BLASSO		−0.311	5.803	6.63	2.118		
OAR2_234923968.1	mrMLM	2	−0.362	5.093	5.893	0.908	0.232	T
s08473.1	mrMLM	3	−0.355	7.511	8.39	1.482	0.276	A
	FASTmrMLM		−0.386	6.696	7.551	1.816		
s36241.1	FASTmrEMMA	3	0.737	5.674	6.496	2.449	0.399	A
ilmnseq_rs414098117	ISIS EM-BLASSO	3	0.291	6.732	7.588	0.994	0.286	G
OAR5_73412004.1	FASTmrMLM	5	−0.503	5.698	6.52	0.19	0.083	C
OAR5_74249563.1	ISIS EM-BLASSO	5	−0.3	6.106	6.942	1.345	0.326	G
OAR6_70593653.1	ISIS EM-BLASSO	6	−0.255	5.304	6.112	1.538	0.441	C
OAR7_50086827.1	mrMLM	7	−0.336	7.009	7.874	3.072	0.49	A
	ISIS EM-BLASSO		−0.274	6.155	6.993	2.112		
s72477.1	FASTmrEMMA	7	0.756	6.071	6.907	2.076	0.272	G
OAR8_70323257.1	mrMLM	8	0.431	5.896	6.726	2.839	0.348	A
OAR8_34350484.1	FASTmrEMMA	8	0.721	6.071	6.906	2.552	0.436	C
	ISIS EM-BLASSO		0.314	6.859	7.719	2.43		
s63912.1	ISIS EM-BLASSO	14	0.321	7.75	8.635	2.946	0.498	G
OAR15_4227328.1	mrMLM	15	0.337	5.314	6.122	2.449	0.421	A
	ISIS EM-BLASSO		0.339	9.005	9.921	2.559		
OAR16_38312498.1	FASTmrEMMA	16	0.612	5.521	6.338	1.599	0.315	A
	ISIS EM-BLASSO		0.255	5.475	6.289	0.937		
OAR23_23465186.1	FASTmrEMMA	23	−0.638	5.331	6.14	1.929	0.418	A
OAR25_42272912.1	ISIS EM-BLASSO	25	0.284	5.24	6.046	1.619	0.41	C

Table 3. Fiber diameter-associated QTNs identified using multi-locus GWAS methods. *QTN* Quantitative trait nucleotide, r^2 The proportion of phenotypic variance explained by each significant QTN (%), *MAF* Minor Allele Frequency, *Allele* alternative (minor) allele.

The multi-locus GWAS approaches employed in this study have revealed significant genetic associations for both FD and SL traits in Middle Anatolian Merino sheep. For FD, 18 QTNs met the significance threshold (LOD ≥ 5) across 12 chromosomes. The identification of these QTNs by four out of five GWAS methods, along with the co-detection of some QTNs, underscores the robustness of our findings. The most significant association was observed for the OAR15_4227328.1 marker on chromosome 15 (LOD = 9.005, -log10P = 9.921), detected by the ISIS EM-BLASSO method. The same QTN was also co-detected by the mrMLM model with a LOD score of 6.122. A total of seven QTNs (s05784.1, OAR2_13110080.1, s08473.1, OAR7_50086827.1, OAR8_34350484.1, OAR15_4227328.1, and OAR16_38312498.1) were simultaneously detected by two or more multi-locus methods (Fig. 5a). The range of QTN effects (-0.638 to 0.756) indicates both positive and negative influences on FD, reflecting the complex nature of this trait.

Similarly, for SL trait, 14 significant QTNs were identified across 11 chromosomes, albeit by fewer methods (3 out of 5). The strongest association for SL was found for the OAR3_41348258.1 marker on chromosome 3 (LOD = 7.757, -log10(P) = 8.643), detected by the mrMLM method. In the analyses for SL, a total of five QTNs (OAR1_117144883.1, s34469.1, OAR7_95970158.1, OAR20_13168385_X.1, and s23479.1) were simultaneously detected by two or more multi-locus methods (Fig. 5b).

The magnitude of QTN effects for SL (-1.786 to 1.427) indicate substantial genetic variation underlying this trait. The variation in minor allele frequencies (MAF) of the identified QTNs, ranging from 0.083 to 0.498 for FD and 0.172 to 0.489 for SL, indicates a mix of common and relatively rare variants associated with wool traits. It is noteworthy that the proportion of phenotypic variance explained by individual QTNs (r^2) was relatively small, ranging from 0.190 to 3.578% for FD and 0.429–3.947% for SL. This observation is consistent with the polygenic model of inheritance for complex traits, where many loci with small effects collectively contribute to phenotypic variation. Such genetic architecture presents both challenges and opportunities for selective breeding strategies.

The post-hoc statistical power analysis, using the median r^2 and MAF values of the associated QTN groups, indicated that the statistical power for both traits clearly exceeded the 0.80 threshold. Specifically, the power for the FD trait was 0.95, while the SL trait showed a power of 0.91. These findings suggest a reliable level of accuracy in the genotype-phenotype associations (Fig. 4).

Gene annotation analysis identified several QTNs associated with FD and SL traits, along with nearby protein-coding genes that may influence these characteristics. The FD-associated QTNs—OAR2_13110080.1, ilmnseq_rs414098117, s36241.1, s63912.1, OAR15_4227328.1, OAR16_38312498.1, OAR23_23465186.1, and OAR25_42272912.1—are located within the introns of several genes, including *PTPN3*, *ALMS1*, *ABTB3*, *ZNF536*,

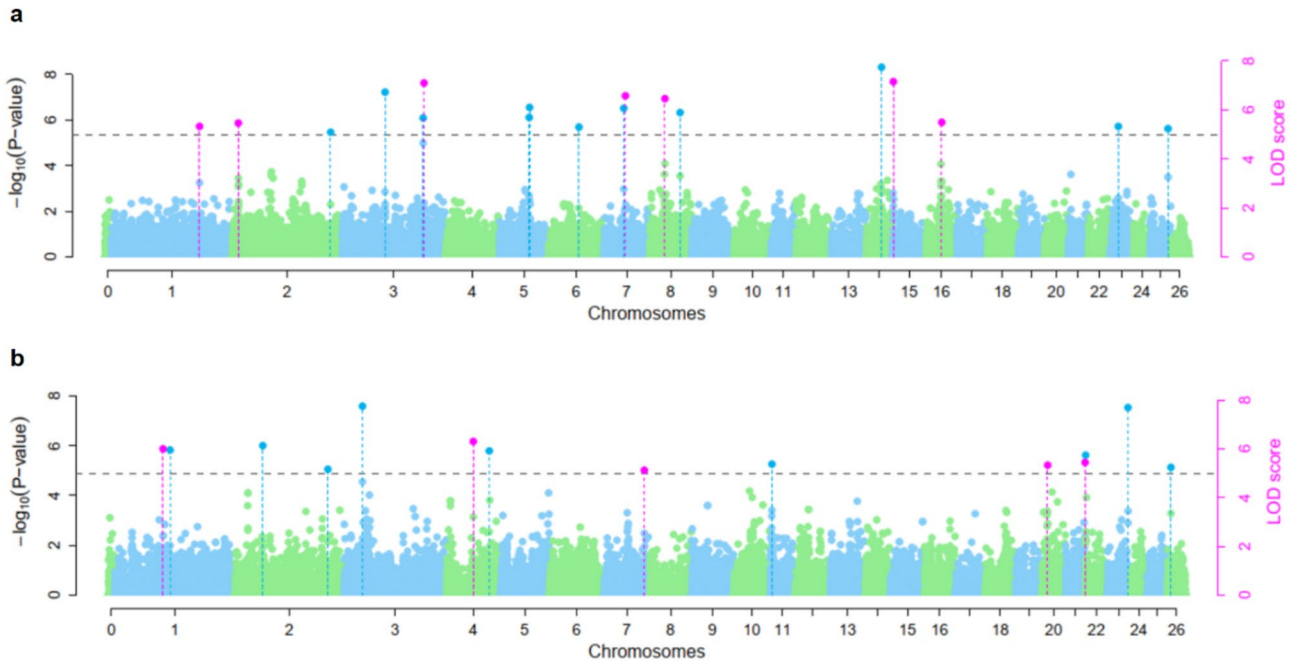

Fig. 2. Manhattan plots illustrating QTNs associated with FD and SL traits. Manhattan plots displaying the distribution significant of QTNs across chromosomes. The x-axis represents chromosomal positions (Chr 0 indicates SNPs with undetermined genomic coordinates), while the dual y-axes show −log 10 (p-values) (left) and LOD scores (right) for each QTN. The horizontal dashed line represents the significance threshold of LOD ≥ 5. The blue plots indicate QTNs detected by a single method, while the pink plots indicate QTNs identified by two or more methods.

Marker	Method	CHR	QTN Effect	LOD Score	−log10 (*p*)	r^2 (%)	MAF	Allele
OAR1_117144883.1	mrMLM	1	1.427	5.766	6.591	1.487	0.257	G
	FASTmrMLM		1.26	6.245	7.086	1.166		
OAR1_135202799.1	ISIS EM-BLASSO	1	−1.056	5.958	6.79	1.806	0.402	A
OAR2_221790517.1	mrMLM	2	−1.512	5.17	5.973	2.654	0.346	A
OAR2_67867687.1	ISIS EM-BLASSO	2	−1.321	6.138	6.975	0.831	0.223	A
OAR3_41348258.1	mrMLM	3	−1.471	7.757	8.643	2.623	0.368	G
s34469.1	mrMLM	4	−1.545	6.443	7.29	1.203	0.201	G
	FASTmrMLM		−1.468	6.182	7.021	1.094		
s59388.1	ISIS EM-BLASSO	4	−1.11	5.923	6.754	2.062	0.424	A
OAR7_95970158.1	mrMLM	7	−1.194	5.121	5.922	2.923	0.489	G
	FASTmrMLM		−1.031	5.135	5.936	2.194		
s47477.1	ISIS EM-BLASSO	11	−1.163	5.378	6.189	1.112	0.286	G
OAR20_13168385_X.1	mrMLM	20	−1.786	5.314	6.122	3.947	0.373	G
	FASTmrMLM		−1.245	5.361	6.171	1.931		
s23479.1	mrMLM	21	−1.624	5.476	6.291	1.532	0.255	A
	FASTmrMLM		−1.296	5.418	6.231	0.982		
OAR22_1408563.1	ISIS EM-BLASSO	22	1.306	5.749	6.574	0.429	0.189	A
OAR23_58625279.1	ISIS EM-BLASSO	23	−1.357	7.701	8.585	2.27	0.349	G
OAR26_10480109.1	FASTmrMLM	26	−1.448	5.248	6.054	0.597	0.172	G

Table 4. Staple length-associated QTNs identified using multi-locus GWAS methods. *QTN* Quantitative trait nucleotide, r^2 The proportion of phenotypic variance explained by each significant QTN (%), *MAF* Minor Allele Frequency, *Allele* alternative (minor) allele.

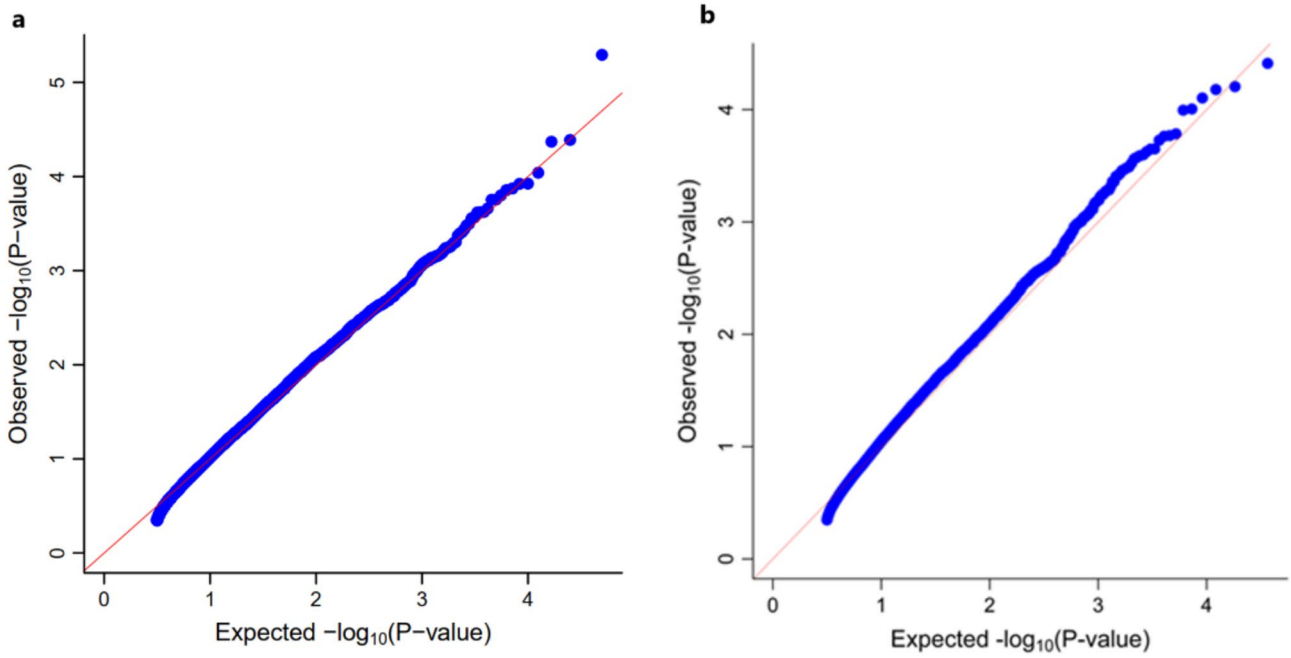

Fig. 3. QQ plots to evaluate possible inflation or deflation in the test statistics. QQ plots showing the distribution of observed versus expected – log 10 (p) values from genome-wide association analyses. The red line represents the null hypothesis of no true associations, where observed values would perfectly match expected values. The close alignment of points with the diagonal line in the lower portions of both plots indicates good control of population stratification and other potential systematic biases. The QQ plot was generated using the multi-locus mrMLM method GWAS results from the mrMLM R package.

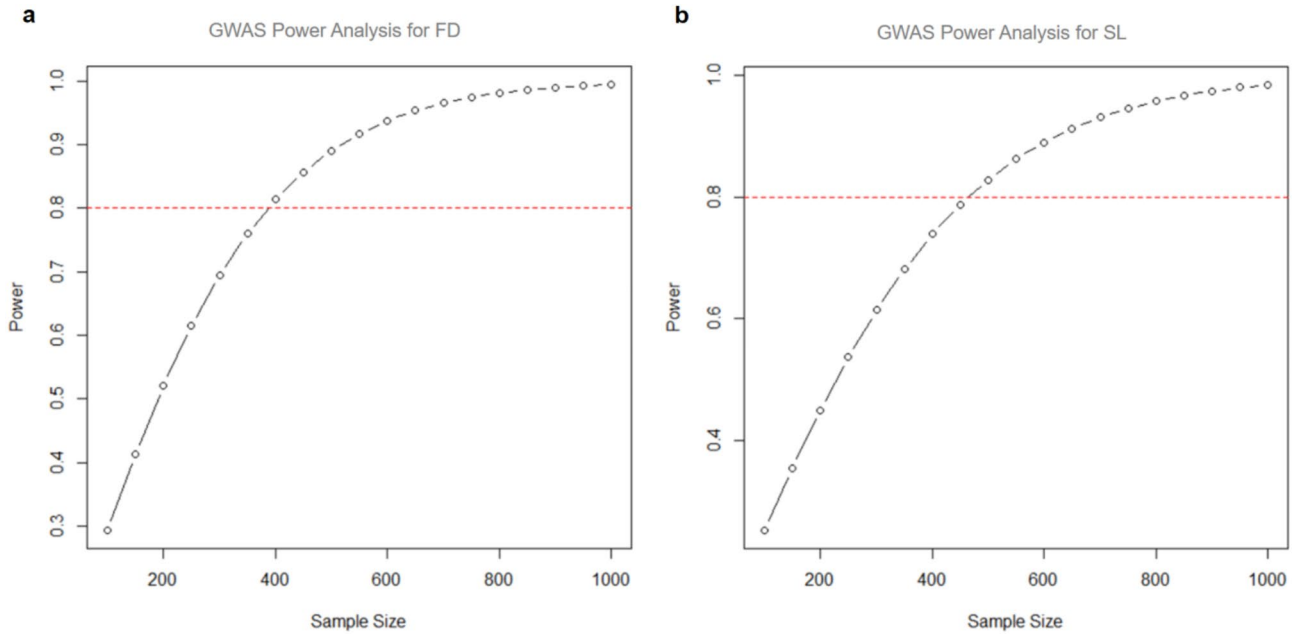

Fig. 4. Statistical power analysis and sample size for FD and SL traits. Power analysis plots showing the relationship between sample size (x-axis) and statistical power (y-axis) for FD and SL analyses, respectively. The red dashed horizontal line represents the conventional statistical power threshold of 0.8 (80%). Sample sizes where the curve intersects with or exceeds the 0.8 threshold line indicate the minimum number of samples needed to achieve adequate statistical power for detecting genetic associations.

Marker ID	rs	Chr	Linked marker	r^2 value	D' value	Distance (bp)	Location of the linked marker	Linked gene
s05784.1	rs407503444	1	rs601924329	0.72	0.87	− 20,414	Downstream	*RTP4*
s34469.1	rs429219260	4	rs421758222	0.89	1.00	− 115,850	Exon	*SEPTIN7*
s72477.1	rs408019749	7	rs409968221	0.90	0.98	− 17,131	Exon	*ENSOARG00020023429*
OAR8_34350484.1	rs418459259	8	rs424629240	0.88	1.00	− 59,064	Intron	*PREP*

Table 5. LD analysis for intergenic QTNs to explore genes in linkage disequilibrium with them.

Marker	CHR	bp	rs	Location	Genes within ± 200 kb range	Distance kb (respectively)
s05784.1	1	213,823,256	rs407503444	Intergenic	*RTP4; MASP1; SST1*	46.5; 145.6; 200
OAR2_13110080.1	2	13,110,080	rs409333407	Intron	*PTPN3*	–
OAR2_234923968.1	2	234,923,968	na	Intergenic	*SYNC; RBBP4; ZBTB8OS; ZBTB8A; NHSL3; YARS1; S100PBP; FNDC5; TMEM54; HPCA; RNF19B*	25.5; 49.5; 64.6; 91.6; 22.5; 49.6; 82.6; 127.6; 145.6; 151.7; 193.7
ilmnseq_rs414098117	3	95,450,579	na	Intron	*ALMS1*	–
s08473.1	3	190,562,060	rs424841795	Intergenic	*LARGE1*	73.6
s36241.1	3	188,517,813	rs399360626	Intron	*ABTB3*	–
OAR5_73412004.1	5	73,412,004	rs427830404	Intergenic	*CLINT1;LSM11; THG1L; SOX30*	6.6; 102.1; 114.1; 147.1
OAR5_74249563.1	5	74,249,563	rs404567493	Intergenic	*EBF1*	51.1
OAR6_70593653.1	6	70,593,653	rs422444939	Intergenic	*na*	na
OAR7_50086827.1	7	50,086,827	rs414621088	Intergenic	*VPS13C*	93.1
s72477.1	7	45,373,665	rs408019749	Intergenic	*ENSOARG00020023429; TRIM9*	9; 102.1
OAR8_34350484.1	8	34,350,484	rs418459259	Intergenic	*PREP; ENSOARG00020031450*	22.5; 138.1
OAR8_70323257.1	8	70,323,257	rs401908606	Intergenic	na	na
s63912.1	14	40,602,275	rs427297087	Intron	*ZNF536*	–
OAR15_4227328.1	15	4,227,328	rs410362423	Intron	*DYNC2H1*	–
OAR16_38312498.1	16	38,312,498	rs424510982	Intron	*FYB1*	–
OAR23_23465186.1	23	23,465,186	rs413627200	Intron	*MAPRE2*	–
OAR25_42272912.1	25	42,272,912	rs405821739	Intron	*GRID1*	–

Table 6. Gene annotation for FD associated QTNs.

Marker	CHR	bp	rs	Location	Genes within ± 200 kb range	Distance kb (respectively)
OAR1_117144883.1	1	117,144,883	rs408825301	Intergenic	*ENSOARG00020009040; CADM3; ACKR1*	3; 133.6; 169.7
OAR1_135202799.1	1	135,202,799	rs404367775	Intergenic	*GRIG1; BACH1*	39; 197.2
OAR2_221790517.1	2	221,790,517	rs402996646	Intron	*PTH2R*	–
OAR2_67867687.1	2	67,867,687	rs403919282	Intron	*TMC1*	–
OAR3_41348258.1	3	41,348,258	rs402255442	Intron	*AAK1*	–
s34469.1	4	64,733,484	rs429219260	Intergenic	*SEPTIN7*	22.0
s59388.1	4	100,969,544	rs425217180	Intron	*ENSOARG00020033807*	–
OAR7_95970158.1	7	95,970,158	rs425058811	Intergenic	*ENSOARG00020040529; ENSOARG00020035024*	51.1; 135.2
s47477.1	11	4,572,362	rs429260631	Intron	*HLF*	–
OAR20_13168385_X.1	20	13,168,386	na	Intergenic	*KCNK17; KCNK16; KCNK5; SAYSD1*	9; 21; 54; 176.7
s23479.1	21	53,766,578	na	na	na	na
OAR22_1408563.1	22	1,408,563	rs415311733	Intergenic	na	na
OAR23_58625279.1	23	58,625,279	rs398292302	Intron	*TCF4*	–
OAR26_10480109.1	26	10,480,109	rs402089881	Intergenic	na	na

Table 7. Gene annotation for SL associated QTNs.

DYNC2H1, *FYB1*, *MAPRE2*, and *GRID1*, respectively (Table 6). Similarly, for SL, QTNs OAR2_221790517.1, OAR2_67867687.1, OAR3_41348258.1, s59388.1, s47477.1, and OAR23_58625279.1 are found within the intronic regions of *PTH2R*, *TMC1*, *AAK1*, *ENSOARG00020033807*, *HLF*, and *TCF4*, respectively (Table 7). Additionally, the QTNs s05784.1, s34469.1, s72477.1, and OAR8_34350484.1, which are located in intergenic

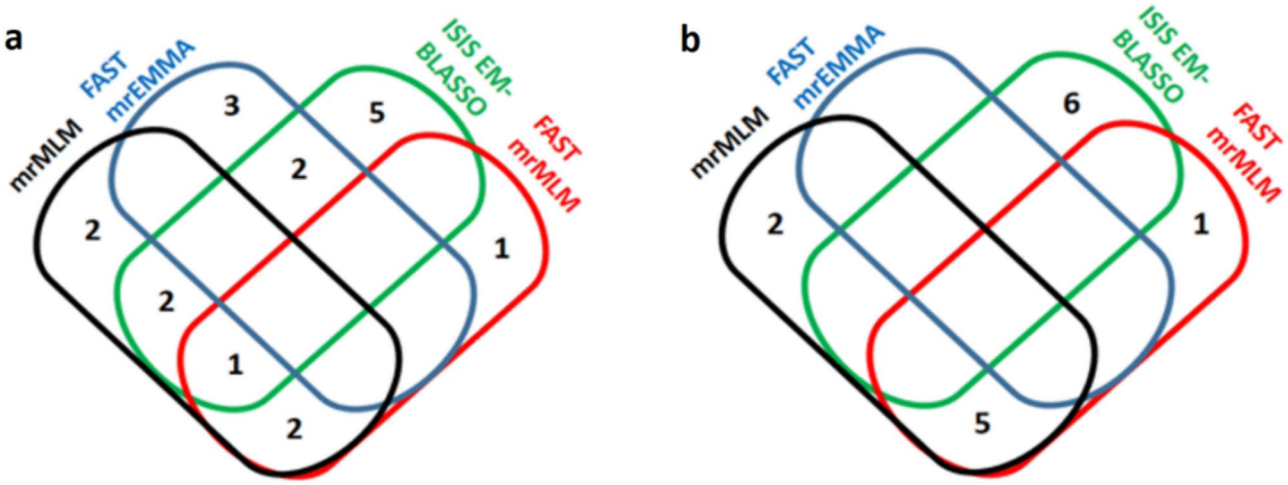

Fig. 5. Venn diagrams showing single and co-detection of QTNs for the FD (**a**) and SL (**b**) traits. The Venn diagrams represent the unique and overlapping QTNs detected by each model for FD and SL traits. The numbers within the overlapping regions indicate the count of QTNs detected by multiple models, highlighting the consistency across approaches.

regions, have been found to be in strong LD (D' > 0.87, r^2 > 0.7) with SNPs located in the downstream regions, exons, or introns of the *RTP4*, *SEPTIN7*, *ENSOARG00020023429*, and *PREP* genes, respectively.

The intronic positioning of these QTNs, or the QTNs (rs601924329, rs421758222, rs409968221, and rs424629240) that are in strong LD with protein-coding genes, suggests that they may influence the genetic mechanisms underlying fleece quality by modulating the expression levels of these genes. This regulatory potential could ultimately impact wool quality in sheep.

Over 140 candidate genes have been identified in various studies as potentially associated with wool traits[20–26] in sheep breeds (Supplementary table S2). Across the eight studies analysed, including the current investigation, concordance was observed for only two genes—*ZHX2* and *HAS2*—between the studies conducted by Wang et al.[20] and Bolormaa et al.[22]. Similarly, alignment was identified for two genes—*PTPN3* and *TCF4*—between Wang et al.[20] and the present study, while a single gene—*ZBTB8A*—was shared between Zhao et al.[24] and the current research.

Given that over 1,000 documented sheep breeds[27] have adapted to a wide range of pastoral environments worldwide since domestication, it is plausible to propose that the genetic foundations of wool traits in various breeds have been distinctly moulded by millennia of adaptation to specific climatic and geographical conditions. This highlights a complex interaction between genetic makeup and environmental factors, which has driven the diverse phenotypic expressions in wool traits observed among sheep populations.

Howewer; it is particularly noteworthy that the three genes reported in the present study, *PTPN3*, *TCF4*, and *ZBTB8A* were also reported by Wang et al.[20] and Zhao et al.[24]. Among them, the *PTPN3* gene, encoding protein tyrosine phosphatase non-receptor type 3, is pivotal in regulating cellular processes essential for maintaining homeostasis, signal transduction, and intercellular interactions. It plays a critical role in preserving tight junction integrity and modulating cytoskeletal dynamics, thereby supporting cell structure and function[28]. Belonging to the protein tyrosine phosphatase (PTP) family, *PTPN3* is characterised by its capacity to dephosphorylate tyrosine residues on proteins. This activity is fundamental for regulating signaling pathways involved in cell growth, differentiation, and apoptosis. *PTPN3* is particularly significant in focal adhesions, where it interacts with cytoskeletal components to influence adhesion and migration[28,29]. In addition to its structural roles, *PTPN3* modulates immune responses by regulating T cell receptor (TCR) signaling. It dephosphorylates immunoreceptor tyrosine-based activation motifs (ITAMs), which are crucial for T cell activation and proliferation. This function underscores its importance in maintaining immune balance and preventing excessive immune responses that could result in autoimmunity[30]. *PTPN3* also participates in various signaling pathways, including the mitogen-activated protein kinase (MAPK) cascade. Its interaction with MAPK components suggests a role in integrating signals from diverse pathways, thereby influencing cellular responses to multiple stimuli. Additionally, *PTPN3* contributes to cellular metabolism and energy regulation. Emerging evidence indicates its involvement in pathways related to oxidative stress and mitochondrial function[29].

The other overlapping gene, *TCF4*, which encodes transcription factor 4, is a member of the basic helix-loop-helix (bHLH) transcription factor family and plays a crucial role in neurodevelopment. It is integral to neuronal differentiation, synaptic function, and the regulation of gene expression during nervous system development[31–33]. *TCF4* primarily functions as a transcriptional regulator by binding to E-box sequences located in the promoter regions of target genes. Depending on the cellular context and interacting proteins, this binding can either activate or repress gene expression[34,35]. The protein's ability to form homodimers or heterodimers with other bHLH proteins expands its regulatory potential, enabling it to influence diverse signaling pathways critical for cell differentiation and survival[36]. Beyond its neurodevelopmental roles, *TCF4* is implicated in epithelial-mesenchymal transition (EMT), a fundamental process in tissue development and repair. It is a key regulator

within the Wnt signaling pathway, where it interacts with β-catenin to enhance the transcription of EMT-related target genes. This function underscores its significance in embryonic development and tissue homeostasis[37–39].

The third overlapping gene, *ZBTB8A*, also known as BOZF1, encodes a member of the BTB/POZ domain-containing transcription factors, which are pivotal in regulating various biological processes, including the cell cycle, differentiation, and oncogenesis. Its functionality is attributed to two C2H2-type zinc fingers that enable DNA binding and transcriptional regulation[40]. The activity of *ZBTB8A* is intricately modulated through its interactions with chromatin remodelers and histone chaperones, which are essential for chromatin maintenance and gene expression regulation during cellular processes such as differentiation[41]. Moreover, epigenetic modifications, particularly DNA methylation, influence its expression. Computational analyses suggest that *ZBTB8A* is part of a broader family of ZBTB proteins that engage with diverse transcription factors and chromatin-modifying complexes[42]. The gene exhibits tissue- and condition-specific expression patterns, with notable associations observed in autoimmune diseases like type 1 diabetes, where altered *ZBTB8A* expression has been reported[43]. During embryonic development, *ZBTB8A* likely plays a critical role in maintaining stem cell identity, as it is expressed at early developmental stages. Its regulation by key pluripotency factors such as Nanog and Pou5f1 highlights its involvement in preserving stem cell characteristics and facilitating differentiation, emphasizing its importance in developmental biology[44].

The analysis of pathway enrichment data highlights potential molecular mechanisms that may contribute to the regulation of wool traits (Fig. 6a and b). For instance, the "Peptidase 59, serine active site" pathway likely plays a critical role in protein modification processes during wool development. Serine proteases, known for their involvement in protein processing and regulation[45,46], may influence the structural and functional aspects of wool fibre formation and differentiation. Additionally, the identification of the "Histone-binding protein (RBBP4, N-terminal)" pathway suggests the involvement of epigenetic regulatory mechanisms. As a histone-binding protein, RBBP4 is implicated in chromatin remodeling[47,48], potentially affecting the expression of genes associated with wool growth and characteristics.

These findings offer preliminary insights into the molecular framework underlying wool development, indicating that protein modifications and epigenetic regulation may act in concert to influence wool traits. However, further experimental validation is essential to establish causal relationships between these pathways and specific wool characteristics. Functional studies would be necessary to confirm these associations and to elucidate the precise roles of these molecular mechanisms.

In conclusion, this study contributes to our understanding of the genetic factors influencing FD and SL traits in sheep. The identified QTNs offer a starting point for more precise breeding strategies and provide insights into the biological mechanisms behind wool traits. However, given the complex nature of FD and SL, focusing on multiple genetic markers through genomic selection may be more effective than targeting individual QTNs. Future studies should validate these findings in other populations and explore the biological significance of the identified regions. Additionally, the distribution of significant QTNs across various chromosomes suggests that multiple biological pathways could be involved in controlling FD and SL. Investigating the genes and regulatory elements near these QTNs could enhance our understanding of wool fiber development. Combining these findings with gene expression data, proteomics, and functional genomics will be key to building a more complete picture of wool biology and will offer potential for the use of the reported QTNs in marker-assisted selection efforts aimed at improving wool traits.

Materials and methods

Animals

The study focused on Middle Anatolian Merino ewes, encompassing 654 individuals approximately 18 months of age. These ewes were maintained at the Bahri Dagdas International Agricultural Research Institute (BDIARI) under uniform climatic and management conditions. The research adhered to guidelines set by the Local Ethics Committee for Animal Experiments of the Sheep Breeding and Research Institute, Turkey (Approval no: 04.10.2021/049). The authors also ensured compliance with the ARRIVE guidelines.

Wool quality and quantity assessment

A standardized protocol was employed for quantifying fiber diameter and staple length. Fleece samples of about 50 g were collected from each ewe's caudal scapular region using an electric shearing device. To preserve sample integrity, each specimen was stored in a labeled and sealed plastic bag. The Ankara Sheep and Goat Breeders Association laboratory conducted fiber analysis using an OFDA 2000 optical fiber diameter analyzer (BSC Electronics, Ardross, Australia). The analytical process involved carefully aligning and straightening a representative fiber subset from each sample before inserting it into the device. The OFDA 2000 then performed automated optical measurements, providing quantitative data on fiber diameter (in microns, μ) and staple length (in millimeters, mm). All sheep in the study were of the same age and born in the same birthing season.

Descriptive statistics

The GCTA software[49] was utilized to calculate SNP-based heritability (h^2) for each trait and genetic correlations between traits. Growth performance data, including birth weight (BW), weaning weight (WW), and average daily gain (ADG), were also available for the studied population. Given the dual-purpose nature of the Middle Anatolian Merino breed and ongoing classical breeding efforts for both meat production and wool quality, genetic and phenotypic correlations between wool quality traits and growth performance characteristics were calculated. This approach aimed to determine the reciprocal effects of selection for these traits. Coefficients of variation (CV%) were determined by dividing the standard deviation of each dataset by its mean and multiplying by 100. Minitab v21.4 software was employed to calculate descriptive statistics and phenotypic correlations (rP) between phenotypes.

Fig. 6. The identified pathways involve the genes within the ± 200 kb range of associated QTNs for FD (**a**) and SL (**b**) traits. In the figure, point size: Indicates Number of Genes. Larger points represent higher gene counts. Color scale: Represents "− log10(FDR)" values. Panel (**a**) shows a single pink shade (1.40), while panel (**b**) displays a purple-to-red gradient scale (2–5). Blue bars indicate statistically less significant results. Fold Enrichment: The x-axis (0–600) represents the degree of enrichment of a particular feature compared to what would be expected by chance.

Genotyping and quality controls

DNA samples extracted from whole blood were genotyped using the OVINE 64 K BeadChip on the ILLUMINA platform, featuring 64,734 markers. Prior to quality control (QC) procedures, SNPs on sex chromosomes and mitochondrial DNA were excluded. QC criteria were applied as follows: SNP Call rate ≥ 0.95, Minor Allele Frequency (MAF) ≥ 0.05, animal missing genotype rates (mind) ≤ 0.1, and Hardy-Weinberg equilibrium threshold ≥ 10E-5. Plink v1.07 software[50] was used for QC procedures. Post-QC, 654 animals (319 rams and 335 ewes) met the criteria, with 49,945 SNPs retained for further analysis.

Association study

Phenotype data for FD and SL were tested for normality prior to conducting genotype-phenotype multi-locus association analyses. The SL dataset deviated from a normal distribution; therefore, a Rank-based inverse normal

transformation was applied using the "RNOmni" R package. Five multi-locus genome-wide association (GWAS) methods—mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, and ISIS EM-BLASSO—were implemented using the "mrMLM" R package[51], adhering to default parameters. To address potential type 1 errors arising from unaccounted cryptic relatedness, kinship matrices were constructed within the mrMLM package and incorporated into the statistical model for each multi-locus method. In these analyses, marker significance was determined by Logarithm of Odds Ratio (LOD) scores. While markers with LOD ≥ 3 are traditionally considered genome-wide significant[9,11,12,52,53], a more stringent criterion of LOD ≥ 5 was implemented to enhance result reliability. This threshold indicates that the likelihood of a true association is 100,000 times greater than the likelihood of no association[54].

A post-hoc statistical power analysis was performed using the "pwr" package in the R platform to ensure that the study had adequate power (i.e., above the commonly accepted threshold of 0.80) to detect meaningful associations[55,56]. Given that a total of 32 QTNs were identified in association with FD and SL traits, conducting a separate power analysis for each QTN was impractical. Instead, the analysis was based on the median r^2 and MAF values from the QTN groups associated with each trait.

Linkage disequilibrium analysis

Linkage disequilibrium (LD) analysis is a fundamental component of genetic studies, particularly in elucidating the genetic architecture of complex traits. LD refers to the non-random association of alleles at different loci, and its quantification is primarily based on two metrics: D' and r^2 values. The D' value measures the degree of association between alleles, whereas the r^2 value estimates the correlation between alleles at two loci. In this study, LD analysis was performed on intergenic QTNs to identify genes in LD with these QTNs. A D' value greater than 0.7 is widely regarded as indicative of strong LD, suggesting that alleles are co-inherited more frequently than expected by chance[57,58]. Accordingly, a stringent threshold of D' ≥ 0.7 was applied. To account for the dependency of r^2 on allele frequency and to ensure comprehensive identification of linked SNPs within a 500 kb window, moderate of $r^2 \geq 0.5$ was also adopted[59,60]. All LD analyses were performed using the LD calculator, available in the Ensembl online tool (https://www.ensembl.org).

Genomic annotation, gene ontology, and pathway analysis

Gene annotations were sourced from Ensembl (https://www.ensembl.org/index.html), NCBI's Genome Data Viewer (https://www.ncbi.nlm.nih.gov/gdv), UCSC Genome Browser (https://genome.ucsc.edu/) and, Research Collaboration Network (RCN; https://www.animalgenome.org) based on the ARS-UI_Ramb_v2.0 and Oar_v3.1 assemblies for sheep. ShinyGO, a graphical tool utilizing all available gene sets for the domestic sheep (Ovis aries) genome, was employed for gene enrichment and network analysis[61]. A false discovery rate threshold of 0.05 was set for determining significant gene enrichment. ShinyGO was also used to generate visual representations of the results.

Data availability

Microarray genotype data that support the findings of this study have been deposited on https://figshare.com with the accession link https://figshare.com/s/9448de9f2e41df87f528.

Received: 3 October 2024; Accepted: 19 March 2025
Published online: 27 March 2025

References

1. Brandt, L. Ø. et al. Characterising the potential of sheep wool for ancient DNA analyses. *Archaeol. Anthropol. Sci.* **3** (2), 209–221. https://doi.org/10.1007/s12520-011-0055-2 (2011).
2. Rippon, J. A. et al. Wool: structure, properties, and processing. *Encyclopedia Polym. Sci. Technol.* 1–46. https://doi.org/10.1002/0471440264.pst402.pub2 (2016).
3. Korjenic, A., Klarić, S., Hadžić, A. & Korjenic, S. Sheep wool as a construction material for energy efficiency improvement. *Energies* **8** (6), 5765–5781. https://doi.org/10.3390/en8065765 (2015).
4. Parlato, M. C. & Porto, S. M. Organized framework of main possible applications of sheep wool fibers in Building components. *Sustainability* **12** (3), 761. https://doi.org/10.3390/su12030761 (2020).
5. Porubská, M., Koóšová, K. & Braniša, J. The application of sheep wool in the Building industry and in the removal of pollutants from the environment. *Processes* **12** (5), 963. https://doi.org/10.3390/pr12050963 (2024).
6. Zach, J., Hroudova, J. & Brozovsky, J. Study of hydrothermal behavior of thermal insulating materials based on natural fibers. *Int. Sch. Sci. Res. Innov.* **8**, 9 (2014).
7. Del Rey, R., Uris, A., Alba, J. & Candelas, P. Characterization of sheep wool as a sustainable material for acoustic applications. *Materials* **10** (11), 1277. https://doi.org/10.3390/ma10111277 (2017).
8. Adi, M. & Păcurar, I. Study regarding the use of sheep wool in Dendro-Horticultural. *Bull. UASVM.* **73** (1). https://doi.org/10.15835/buasvmcn-agr:12002 (2016).
9. Wang, S. et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* **6** (1). https://doi.org/10.1038/srep19444 (2016).
10. Tamba, C. L. & Zhang, Y. A fast mrMLM algorithm for multi-locus genome-wide association studies (Preprint). *bioRxiv (Cold Spring Harbor Laboratory)*. (2018). https://doi.org/10.1101/341784
11. Tamba, C. L., Ni, Y. & Zhang, Y. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput. Biol.* **13** (1), e1005357. https://doi.org/10.1371/journal.pcbi.1005357 (2017).
12. Wen, Y. et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* **19** (4), 700–712. https://doi.org/10.1093/bib/bbw145 (2016).
13. Zhang, J. et al. pLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity* **118** (6), 517–524. https://doi.org/10.1038/hdy.2017.8 (2017).
14. Anderson, T. W. & Darling, D. A. Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *Ann. Math. Stat.* **23** (2), 193–212. https://doi.org/10.1214/aoms/1177729437 (1952).

15. Notter, D. R. & Hough, J. D. Genetic parameter estimates for growth and fleece characteristics in Targhee sheep. *J. Anim. Sci.* **75** (7), 1729. https://doi.org/10.2527/1997.7571729x (1997).
16. Jafari, S. & Hashemi, A. Genetic analysis of fleece and post-weaning body weight traits in Makuie sheep. *Genet. Mol. Res.* **13** (1), 1079–1087. https://doi.org/10.4238/2014.february.20.9 (2014).
17. Mortimer, S. I. et al. Genetic parameters for wool traits, live weight, and ultrasound carcass traits in Merino sheep1. *J. Anim. Sci.* **95** (5), 1879–1891. https://doi.org/10.2527/jas.2016.1234 (2017).
18. Wuliji, T., Dodds, K., Land, J., Andrews, R. & Turner, P. Response to selection for ultrafine Merino sheep in new Zealand. *Livest. Prod. Sci.* **58** (1), 33–44. https://doi.org/10.1016/s0301-6226(98)00195-x (1999).
19. Ramos, Z. et al. Genetic and phenotypic relationships between ewe reproductive performance and wool and growth traits in Uruguayan Ultrafine Merino sheep. *Journal of Animal Science*, *101*. (2023). https://doi.org/10.1093/jas/skad071
20. Wang, Z. et al. Genome-Wide association study for wool production traits in a Chinese Merino sheep population. *PLoS ONE.* **9** (9), e107101. https://doi.org/10.1371/journal.pone.0107101 (2014).
21. Rong, E. G. et al. Association of methionine synthase gene polymorphisms with wool production and quality traits in Chinese Merino population12. *J. Anim. Sci.* **93** (10), 4601–4609. https://doi.org/10.2527/jas.2015-8963 (2015).
22. Bolormaa, S. et al. A conditional multi-trait sequence GWAS discovers pleiotropic candidate genes and variants for sheep wool, skin wrinkle and breech cover traits. *Genet. Selection Evol.* **53** (1). https://doi.org/10.1186/s12711-021-00651-0 (2021).
23. Zhao, B. et al. Integration of a single-step genome-wide association study with a multi-tissue transcriptome analysis provides novel insights into the genetic basis of wool and weight traits in sheep. *Genet. Selection Evol.* **53** (1). https://doi.org/10.1186/s12711-021-00649-8 (2021).
24. Zhao, H. et al. Whole-genome re-sequencing association study on yearling wool traits in Chinese fine-wool sheep. *J. Anim. Sci.* **99** (9). https://doi.org/10.1093/jas/skab210 (2021).
25. Arzik, Y. et al. Genome-Wide scan of wool production traits in Akkaraman sheep. *Genes* **14** (3), 713. https://doi.org/10.3390/genes14030713 (2023).
26. Becker, G. M., Woods, J. L., Schauer, C. S., Stewart, W. C. & Murdoch, B. M. Genetic association of wool quality characteristics in united States Rambouillet sheep. *Front. Genet.* 13. https://doi.org/10.3389/fgene.2022.1081175 (2023).
27. Kijas, J. W. et al. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE.* **4** (3), e4668. https://doi.org/10.1371/journal.pone.0004668 (2009).
28. Saadatabadi, L. M. et al. Signature selection analysis reveals candidate genes associated with production traits in Iranian sheep breeds. *BMC Vet. Res.* **17** (1). https://doi.org/10.1186/s12917-021-03077-4 (2021).
29. Genera, M. et al. Interactions of the protein tyrosine phosphatase PTPN3 with viral and cellular partners through its PDZ domain: insights into structural determinants and phosphatase activity. *Front. Mol. Biosci.* 10. https://doi.org/10.3389/fmolb.2023.1192621 (2023).
30. Hsu, E. et al. Suppression of hepatitis B viral gene expression by protein-tyrosine phosphatase PTPN3. *J. Biomed. Sci.* **14** (6), 731–744. https://doi.org/10.1007/s11373-007-9187-x (2007).
31. Forrest, M., Waite, A., Martin-Rendon, E. & Blake, D. Knockdown of human tcf4 affects multiple signaling pathways involved in cell survival, epithelial to mesenchymal transition and neuronal differentiation. *Plos ONE.* **8** (8), e73169. https://doi.org/10.1371/journal.pone.0073169 (2013).
32. Hennig, K. M. et al. WNT/B-Catenin pathway and epigenetic mechanisms regulate the Pitt-Hopkins syndrome and schizophrenia risk gene TCF4. *Complex. Psychiatry*. **3** (1), 53–71. https://doi.org/10.1159/000475666 (2017).
33. Torshizi, A. D. et al. Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. *Sci. Adv.* **5** (9). https://doi.org/10.1126/sciadv.aau4139 (2019).
34. Kim, S. et al. CpG methylation in exon 1 of transcription factor 4 increases with age in normal gastric mucosa and is associated with gene Silencing in intestinal-type gastric cancers. *Carcinogenesis* **29** (8), 1623–1631. https://doi.org/10.1093/carcin/bgn110 (2008).
35. Liu, Z. et al. Expression and functional analysis of tcf4 isoforms in human glioma cells. *Mol. Med. Rep.* https://doi.org/10.3892/mmr.2018.8553 (2018).
36. Jeong, J., Lee, J. & Lee, S. TCF4 is a molecular target of Resveratrol in the prevention of colorectal cancer. *Int. J. Mol. Sci.* **16** (5), 10411–10425. https://doi.org/10.3390/ijms160510411 (2015).
37. Xiong, Y., Liu, Y., Song, Z., Fei, H. & Yang, X. Identification of WNT/β-catenin signaling pathway in dermal papilla cells of human scalp hair follicles: TCF4 regulates the proliferation and secretory activity of dermal papilla cell. *J. Dermatol.* **41** (1), 84–91. https://doi.org/10.1111/1346-8138.12313 (2013).
38. Rannals, M. D. et al. Psychiatric risk gene transcription factor 4 regulates intrinsic excitability of prefrontal neurons via repression of SCN10A and KCNQ1. *Neuron* **90** (1), 43–55. https://doi.org/10.1016/j.neuron.2016.02.021 (2016).
39. Gao, J. et al. Association between a TCF4 polymorphism and susceptibility to schizophrenia. *Biomed. Res. Int.* **2020**, 1–6. https://doi.org/10.1155/2020/1216303 (2020).
40. Kim, M. & Hur, M. Functional characterization of novel BTB/POZ-domain protein ZBTB8A. *FASEB J.* **24** (S1). https://doi.org/10.1096/fasebj.24.1_supplement.678.10 (2010).
41. Olivieri, D. et al. The BTB-domain transcription factor ZBTB2 recruits chromatin remodelers and a histone chaperone during the exit from pluripotency. *J. Biol. Chem.* **297** (2), 100947. https://doi.org/10.1016/j.jbc.2021.100947 (2021).
42. Overs, A. et al. The detection of specific hypermethylated WIF1 and NPY genes in Circulating DNA by crystal digital PCRTM is a powerful new tool for colorectal cancer diagnosis and screening. *BMC Cancer*. **21** (1). https://doi.org/10.1186/s12885-021-08816-2 (2021).
43. Hamilton-Williams, E. E. et al. Fine mapping of type 1 diabetes regions Idd9.1 and Idd9.2 reveals genetic complexity. *Mamm. Genome*. **24** (9–10), 358–375. https://doi.org/10.1007/s00335-013-9466-y (2013).
44. Gaspar, J., Doss, M., Winkler, J., Wagh, V., Hescheler, J., Kolde, R., … Sachinidis,A. (2012). Gene expression signatures defining fundamental biological processes in pluripotent, early, and late differentiated embryonic stem cells. Stem Cells and Development, 21(13), 2471–2484. https://doi.org/10.1089/scd.2011.0637.
45. Lockridge, O. & Schopfer, L. M. Review of tyrosine and lysine as new motifs for organophosphate binding to proteins that have no active site Serine. *Chemico-Biol. Interact.* **187** (1–3), 344–348. https://doi.org/10.1016/j.cbi.2010.03.002 (2010).
46. Mathew, C., Gunathilaka, R. & D., &, M. S Production, purification and characterization of a thermostable alkaline Serine protease from Bacillus lichniformis NMS-1. *Int. J. Biotechnol. Mol. Biology Res.* **6** (3), 19–27. https://doi.org/10.5897/ijbmbr2014.0199 (2015).
47. Zhi, S. et al. Quantitative proteomics of HFD-induced fatty liver uncovers novel transcription factors of lipid metabolism. *Int. J. Biol. Sci.* **18** (8), 3298–3312. https://doi.org/10.7150/ijbs.71431 (2022).
48. Mu, W. et al. RBBP4 dysfunction reshapes the genomic landscape of H3K27 methylation and acetylation and disrupts gene expression. *G3 Genes Genomes Genet.* **12** (6). https://doi.org/10.1093/g3journal/jkac082 (2022).
49. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88** (1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011 (2010).
50. Purcell, S. et al. PLINK: a tool set for Whole-Genome association and Population-Based linkage analyses. *Am. J. Hum. Genet.* **81** (3), 559–575. https://doi.org/10.1086/519795 (2007).
51. Zhang, Y. et al. MrMLM V4.0.2: an R platform for multi-locus genome-wide association studies. *Genomics Proteom. Bioinf.* **18** (4), 481–487. https://doi.org/10.1016/j.gpb.2020.06.006 (2020).

52. Ren, W., Wen, Y., Dunwell, J. M. & Zhang, Y. pKWmEB: integration of Kruskal–Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity* **120** (3), 208–218. https://doi.org/10.1038/s41437-017-0007-4 (2017).
53. Cui, Y., Zhang, F. & Zhou, Y. The application of Multi-Locus GWAS for the detection of Salt-Tolerance LOCI in rice. *Front. Plant Sci.* **9** https://doi.org/10.3389/fpls.2018.01464 (2018).
54. Nyholt, D. R. All LODs are not created equal. *Am. J. Hum. Genet.* **67** (2), 282–288. https://doi.org/10.1086/303029 (2000).
55. Ahn, C. Sample size and power Estimation in Case-Control genetic association studies. *Genomics Inf.* **4** (2), 51–56 (2006).
56. Hong, E. P. & Park, J. W. Sample size and statistical power calculation in genetic association studies. *Genomics Inf.* **10** (2), 117. https://doi.org/10.5808/gi.2012.10.2.117 (2012).
57. Isnard, A., Kouriba, B., Doumbo, O. & Chevillard, C. Association of rs7719175, located in the IL13 gene promoter, with schistosoma haematobium infection levels and identification of a susceptibility haplotype. *Genes Immun.* **12** (1), 31–39. https://doi.org/10.1038/gene.2010.43 (2010).
58. Liu, H. et al. Association of single-nucleotide polymorphisms in TLR4 gene and gene–environment interaction with primary open angle glaucoma in a Chinese Northern population. *J. Gene. Med.* **22** (1). https://doi.org/10.1002/jgm.3139 (2019).
59. Wray, N. & R Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Res. Hum. Genet.* **8** (2), 87–94. https://doi.org/10.1375/1832427053738827 (2005).
60. Berihulay, H., Islam, R., Jiang, L. & Ma, Y. Genome-Wide linkage disequilibrium and the extent of effective population sizes in six Chinese goat populations using a 50K single nucleotide polymorphism panel. *Animals* **9** (6), 350. https://doi.org/10.3390/ani9060350 (2019).
61. Ge, S. X., Jung, D. & Yao, R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36** (8), 2628–2629. https://doi.org/10.1093/bioinformatics/btz931 (2019).

Acknowledgements

The research adhered to guidelines set by the Local Ethics Committee for Animal Experiments of the Sheep Breeding and Research Institute, Turkey (Approval no: 04.10.2021/049). The authors also ensured compliance with the ARRIVE guidelines.The authors express their gratitude to the General Directorate of Agricultural Research and Policies for invaluable support. The corresponding author dedicates this article to our greatest leader, Mustafa Kemal Atatürk, who said, "Peace at home, peace in the world."

Author contributions

Y.Y. conceptualized the experiment. S.B. performed wool quality analyses. Ş.D. contributed to the data analysis. A.T.Ö., M.K., and Y.K. contributed to interpreting the data and were involved in project management. All authors reviewed the manuscript.

Funding

Funding for this research was provided by the Republic of Turkey Ministry of Agriculture and Forestry, General Directorate of Agricultural Research and Policies (TAGEM) (Project no: TAGEM/HAYSÜD/E/20/A4/P2/2141).

Declarations

Competing interests

The authors declare no competing interests.

Ethical statement

This study was conducted in compliance with the guidelines of the Local Ethics Committee for Animal Experiments of the Sheep Breeding and Research Institute, Turkey (Approval no: 04.10.2021/049) and the authors complied with the ARRIVE guidelines.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-95099-3.

Correspondence and requests for materials should be addressed to Y.Y.