Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Investigation of prolificacy across sheep breeds reveals new potential genetic contributors

William POISSON
Université Laval
Cathy MICHAUD
Société des éleveurs de moutons de race pure du Québec
Johanne CAMERON
Société des éleveurs de moutons de race pure du Québec
Amélie ST-PIERRE
Centre d'expertise en production ovine du Québec
Ricardo PERECIN-NOCITI
Université Laval
Isabelle GILBERT
Université Laval
Éric PAQUET
Université Laval
Stéphane FABRE
GenPhySE, Université de Toulouse, INRAE, ENVT
Claude ROBERT
Claude.Robert@fsaa.ulaval.ca

Université Laval

Research Article

Keywords: Ovis aries, Sheep, Prolificacy, Litter Size, Genomics, GWAS, Genotyping, Sequencing, Growth differentiation factor 9

Posted Date: October 21st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11165966/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Prolificacy is a major determinant of reproductive efficiency in sheep, yet its genetic architecture remains incompletely understood, particularly in Canadian populations. Several genetic variants affecting ovine fertility have been reported, notably in *Bone morphogenetic protein 15 (BMP15)* and *Growth differentiation factor 9 (GDF9)*, two genes involved in folliculogenesis. Since 2023, more than 5,000 sheep have been genotyped as part of the development of a genomic selection program in Canada, enabling the investigation of these variants. This study evaluated previously reported prolificacy-associated variants and investigated the genetic basis of exceptional prolificacy in the Canadian Rideau Arcott breed.

Results

Among 30 candidate loci, three were significantly associated with litter size in specific breeds using nearly 100,000 litter size records analyzed with linear mixed-effects models. The *G6* variant in *GDF9* gene on chromosome 5 was associated with lower litter size in homozygous compared with heterozygous Romanov ewes and compared with both heterozygous and non-carrier Dorper ewes, with significant deviations from an additive effect. In Rideau Arcott, the *G7* variant in *GDF9*, also known as *FecGF*, was associated with increases of 0.15 and 0.25 lambs per litter in heterozygous and homozygous animals, respectively ($p < 0.01$). A subsequent genome-wide association analysis on 2,317 litters from Rideau Arcott identified two other loci on chromosome 5, and DNA sequencing revealed another *GDF9* variant, designated *FecGR* herein, which was associated with an increase of 0.21 lambs per litter in heterozygous ewes ($p < 0.01$). After accounting for both *GDF9* variants, the association at the second locus persisted, and haplotype analyses highlighted several candidate genes within a 1.5-Mb region. These included several *SPINK* family members and *PPP2R2B*, which encodes a regulatory subunit of Protein phosphatase 2A (PP2A) complex and has been implicated in reproductive processes.

Conclusions

These findings provide new tools for marker-assisted selection that could complement traditional EPD-based genetic selection, while contributing to a better understanding of the genetic architecture of ovine prolificacy and identifying new candidate genes for future investigation.

Background

In the Canadian province of Quebec, ovine production is structured as a three-tier pyramid, in which purebred breeders produce elite animals to meet the needs of hybridizers and commercial producers. Within this system, breeders actively select animals according to their breed purpose, including prolificacy, maternal ability, and terminal growth and meat traits. Among purebred animals born within

the last 10 years and registered in the national genetic selection program, GenOvis, prolific breeds are currently predominant, with 34% and 28% of registrations corresponding to the Rideau Arcott and Romanov breeds, respectively (GenOvis, personal communication). The Rideau Arcott is a Canadian fixed synthetic breed developed at the Animal Research Institute of Ottawa, Ontario, over a 16-year period beginning in 1972 (1, 2). It was developed from 10 breeds, with Finnsheep, Suffolk, and East Friesian contributing the largest proportions (40%, 20%, and 14%, respectively) (1, 2). Animals resulting from these controlled crosses were primarily selected for prolificacy, with an average ovulation rate of 3.25 ova per ovulation and 2.5 to 2.9 lambs per litter in adult ewes (1). In contrast, the Romanov is an older breed developed in Russia at the end of the 17th century (3). Since then, it has been raised as a purebred or used in crossbreeding programs in many countries because of its high prolificacy, averaging up to 3.5 lambs per litter (4).

In addition to being an important breed purpose, prolificacy is of particular interest to breeders who derive part or most of their income from the sale of breeding stock. Until 2023, selection for prolificacy in Canadian purebred flocks was based exclusively on expected progeny difference (EPDs) and/or phenotypic records, as none of the purebred breeders routinely tested animals for known major prolificacy-associated markers. However, genomic variation underlying prolificacy may already contribute indirectly to selection through EBVs. Indeed, a study of prolificacy-associated genetic markers in the Lacaune breed in France showed that influential genomic variants can affect the value and accuracy of EBVs, thereby resulting in indirect selection on these variants (5). Numerous genetic variants have been reported to significantly affect ovulation rate and litter size in sheep (Additional file 1). These variants exhibit diverse modes and magnitudes of action, with some causing sterility in the homozygous state (6–10), whereas others span a continuum of genetic effects ranging from additive effects to varying degrees of dominance (11–13). Most of these variants have been identified in two transforming growth factor-beta (TGF-β) family genes: 11 in *Bone morphogenetic protein 15* (*BMP15*) and seven in *Growth differentiation factor 9* (*GDF9*). Both genes are expressed in oocytes and are involved in the early stages of folliculogenesis (14, 15). Other variants have been identified in the TGF-β receptor gene *Bone morphogenetic protein receptor 1B* (*BMPR1B*) (16–18), as well as in genes outside the TGF-β pathway. These include *beta-1,4-N-acetyl-galactosaminyl transferase 2* (*B4GALNT2*), which is involved in ovarian protein glycosylation during folliculogenesis (19); the *leptin receptor* gene (*LEPR*), which is associated with the regulation of nutritional pathways (20–22); and *MTNR1A*, which encodes a melatonin receptor and has been associated with effects on the oestrous cycle and fertility, as reviewed elsewhere (23). Furthermore, interactions among variants within the same gene or across different genes have been reported, including interactions involving *GDF9* and *BMP15*, *B4GALNT2* and *BMP15*, and *BMP15* and *BMPR1B* (6, 12, 24–28).

Several of these variants were initially identified through phenotypic observations, either from unusually high litter sizes or from increased variation in litter size associated with the incomplete segregation of prolificacy-associated markers within a population (29–32). Others were identified through genome-wide association studies using genotyping data (25, 26, 33, 34). Sequencing and other omics approaches were subsequently used in some studies to identify the causal or candidate variants among linked

polymorphisms (19, 25, 31, 33). Most of these known markers are now included in SNP array panels, enabling their screening across populations. Since 2023, Canadian purebred breeders have been provided with genomic reports to support marker-assisted selection. These reports provide the status of more than 30 known markers; however, the phenotypic effects of many polymorphic markers have not yet been characterized in several of the breeds raised in Canada. Given the observed variation in litter size within some of these breeds, the objectives of this study were therefore to investigate the potential effects of known prolificacy-associated markers segregating in sheep breeds reared in Canada and to explore the genetic background underlying the hyperprolificacy of Rideau Arcott ewes.

Methods

Animals and phenotypic datasets

Animal and phenotypic data, including flock information, birth information, and litter size records, were extracted from the Canadian sheep genetic evaluation program database (GenOvis). Four main datasets were defined for this study. The numbers of ewes and flocks represented in each breed and dataset are presented in Additional file 2.

The *Pheno dataset* consisted of phenotypic records from all purebred ewes born between 2013 and 2026 in a flock participating in the genomic program, with or without genotyping data. This dataset was used for the phenotypic analyses. Breed codes were attributed as follow : Border Cheviot (BC), Border Leicester (BL), Canadian Arcott (CD), Charolais (CO), Dorper (DO), Dorset Polled (DP), East Friesian (EF), Finnsheep (FN), Hampshire (HA), Ile de France (IF), Katahdin (KA), North Country Cheviot (NC), Oxford (OX), Polypay (PO), Rideau Arcott (RI), Romanov (RV), Southdown (SO), Suffolk (SU), Texel (TX), White Dorper (WD).

The *Marker dataset* included phenotypic and genomic data from breeds represented by at least 10 genotyped purebred ewes per participating flock. This dataset was used to investigate the effects of known genetic variants associated with prolificacy.

The *GWAS dataset* was a subset of the *Marker dataset* and included only genotyped multiparous Rideau Arcott ewes with a minimum of three registered litters. This criterion was applied to obtain more reliable estimates of individual mean litter size. Phenotypic and genotypic data from these animals were used for genome-wide association analyses.

The *WGS dataset* consisted of a selection of Rideau Arcott ewes from the *GWAS dataset*, representing the different carrier-haplotype statuses identified in the population. These animals were whole-genome sequenced to investigate the presence of additional prolificacy-associated variants in the Rideau Arcott breed.

Genotyping data

Genotyping data have been collected since 2023 as part of the development of the Canadian sheep genomic selection program. 3-mm ear punches were collected from animals privately owned by sheep breeders on a voluntary basis, using Tissue Sampling Units (TSU) and following Allflex sampling guidelines (Allflex Livestock Intelligence, Madison, WI, USA). The sampling protocol was approved by the Université Laval animal welfare committee (VRR-23-1362). Genotyping was performed using one of two custom versions of the Illumina OvineSNP50v3 HTS BeadChip. The first and second versions included 65,088 and 65,477 SNPs, respectively. The second version, implemented in 2024, was upgraded to include additional variants previously reported in the ovine literature. Genotyping data were retrieved from the sheep genomic database GenOvine, hosted at the non-human genomic data center of Université Laval (Genovalia, Québec, QC, Canada).

Genomic data filtering was performed using *VCFtools v0.1.16* (35). For all analyses, animals with a call rate below 97.5% were removed. For the known candidate variant analyses using the *Marker dataset*, genotypes corresponding to candidate variants were extracted, and no additional variant-level filtering was applied to avoid excluding variants with low alternative allele frequencies. For the *GWAS dataset*, duplicated SNPs and SNPs with a call rate below 98% or a minor allele frequency (MAF) below 5% were removed. Mitochondrial and Y-chromosome SNPs were also excluded, whereas X-chromosome SNPs were retained to avoid removing variants located in reproduction-related genes, such as *BMP15*. These filtering steps resulted in a dataset of 50,531 SNPs for subsequent analyses. No filtering based on linkage disequilibrium or Hardy–Weinberg equilibrium was performed. Genotype imputation was subsequently performed using *Beagle v5.1* (36) with default parameters. A post-imputation MAF filter removed two additional SNPs, resulting in a final genomic dataset of 50,529 SNPs.

Phenotypic analyses

The relationship between the mean and variance of litter size within breeds was assessed as previously described by Chantepie et al. (26). Briefly, phenotypic records from primiparous and multiparous ewes were considered for breeds with at least 100 registered litters and represented in at least two flocks. Breed-specific mean and variance of litter size were calculated from these records. A second-order polynomial regression was fitted to assess the relationship between the mean and variance of litter size, with the quadratic coefficient constrained to be negative to reproduce the concave relationship previously described (29). Models were fitted using the *lm* function from the *stats* package in R (R Core Team (2026)) and the coefficient of determination (R^2) was obtained from the same package.

Known markers effects analyses

A linear mixed-effects modelwas used to assess the effects of potential reproduction-related variants on litter size, using the *lmer* function from the *lme4* and *lmerTest* packages (38,39). The tested variants and their allele frequencies within each breed are presented in Additional file 1. Correlations among potential fixed and random-effect variables were assessed using the *cor* function from the *stats* package in R (R

Core Team (2026)). Dam age at lambing was removed from the models because it was highly correlated with parity number (r = 0.96). Ewe birth year was also excluded to avoid accounting for the genetic component associated with past selection for litter size, which could potentially influence allele frequencies. For each breed, flock at lambing, parity number, and lambing month were included as fixed effects when sufficient variation was observed. Litter size records were not averaged across parities to avoid potential bias resulting from the unequal number of parities represented among ewes. Therefore, ewe ID was included as a random effect to account for the repeated records of individual ewes.

The variables included in each marker-by-breed analysis are reported in Additional file 3. Genotypes were fitted as categorical factors to avoid imposing a linear mode of inheritance. Global variant effects were assessed using type II ANOVA from the *car* R package (40). P-values were corrected for multiple testing within each breed using the Benjamini–Hochberg false discovery rate (FDR) procedure implemented in the *p.adjust* function. Notably, only one variant from each of the GDF9_R87H/GDF9_E241K and MTNR1A_612/MTNR1A_706 pairs was tested to avoid duplicating the same analysis and introducing bias into the FDR correction.

The *emmeans* and *cld.emmGrid* functions from the *emmeans* v2.0.3 package (41) were used to estimate estimated marginal litter size means for each genotype when a significant global variant effect was detected by ANOVA and to perform Tukey-adjusted post hoc pairwise comparisons. For variants showing a significant global effect and represented by the three genotypes (wild-type, heterozygous, and homozygous carriers), the additive component was further assessed by testing the deviation of the genotype marginal means from a linear inheritance pattern corresponding to 0, 0.5, and 1, using the *contrast* function from the *emmeans* package.

Genome-wide association for prolificacy in Rideau Arcott breed

The association between polymorphisms retained after genotyping-array filtering and litter size in Rideau Arcott ewes from the *GWAS dataset* was conducted in two steps. First, the known prolificacy-associated *FecGF* marker was not included in the model to assess its potential influence on the distribution of GWAS *p*-values. Second, genotypes for the *FecGF* marker were included as an additional fixed effect to account for its effect.

Litter size was modeled using a linear mixed-effects model with the same fixed and random effects as those described previously for the known marker analyses. The residuals from the linear mixed-effects model were extracted and averaged across records for each ewe. These mean residuals, which accounted for the fixed and random effects included in the preliminary model, were then used as the response variable in the GEMMA association analysis (42). To account for genetic relatedness among animals, a VanRaden additive genetic relationship matrix (GRM) was calculated using the *G.matrix* function from the *snpReady* package (43), based on the genotype dosage file generated with *PLINK v1.9*

(44). The resulting GRM was incorporated into the GEMMA model along with the mean residuals. The *Manhattan()* and *qq()* functions from the *qqman* R package (45) were used to visualize the GWAS results. Genome-wide significant ($p = 0.05$) and suggestive ($p = 0.10$) thresholds were adjusted for multiple testing using the Bonferroni correction.

Additive and dominance GRMs were also used to estimate litter size heritability. These GRMs were fitted as random effects together with the previously described fixed and random effects in a linear mixed-effects model implemented using the *mmer* function from the *sommer* R package (46).

Fine-mapping and haplotype analysis

Based on the GWAS signal observed on chromosome 5, fine-mapping was performed to identify haplotypes most likely carrying a variant associated with litter size. A 2-Mb region surrounding the SNP with the lowest *P*-value (s21241.1) was extracted for all animals in the *GWAS dataset* and divided into overlapping 10-SNP windows using a custom R script. For each 10-SNP window, a haplotype library was generated based on the haplotype phase estimated during the previous *Beagle v5.1* imputation step. The association between litter size and each haplotype was then assessed using the same statistical model as that used for the known marker analyses and GWAS, with the known *FecGF* genotype included as a fixed effect.

For each window, haplotypes with *P*-values below 0.10 were retained and merged when the same haplotype was conserved across two consecutive windows based on the nine overlapping SNPs. The ends of merged haplotypes were defined when the *P*-value increased above 0.10, indicating that the suggestive association was no longer maintained. Following concatenation, the two long haplotypes with the lowest *P*-values were selected and cropped at positions where the *P*-value increased twofold, thereby retaining the most strongly associated intervals. These two final long haplotypes (LH1 and LH2) were then included in the statistical model to confirm that their associations remained significant after consolidation.

To select animals for whole-genome sequencing, identity by state (IBS) was calculated using the *snpgdsIBS* function from the *SNPRelate* package (47). Animals were selected to ensure representation of each haplotype while minimizing IBS among selected animals, thereby maximizing the number of genetic lineages represented. A total of 24 Rideau Arcott ewes were selected from the *GWAS dataset* to form the *WGS dataset* and short-read paired-end whole-genome resequenced using an Illumina NovaSeq platform (2 × 150 bp) with PCR-free shotgun DNA library preparation. Raw reads were trimmed using *fastp v0.24.0* (48), and quality control was assessed using *FastQC v0.12.1* (49). Reads were aligned to the ARS-UI_Ramb_v3.0 reference genome (GCF_016772045.2) using *BWA v0.7.18* with default parameters (50). The *GATK v4.6.1.0* pipeline was used to mark duplicate reads, perform base quality score recalibration, and call variants using the *HaplotypeCaller* workflow, specifying the *GVCF* mode (51). Low-quality SNPs were excluded using the *VariantFiltration* function with the following thresholds: QD <

2.0, FS > 60.0, MQ < 40.0, SOR > 3.0, MQRankSum < −12.5, and ReadPosRankSum < −8.0. The mean autosomal sequencing depth was 16.65× (SD = 2.35).

Sanger sequencing and model correction for the new GDF9 variant

To characterize the effect of the newly identified g.42114679C>T variant of GDF9 gene, which is outside the carrier haplotype region, and to correct the model used for the GWAS analyses accordingly, all ewes from the *GWAS dataset* were genotyped by Sanger sequencing. The region was amplified using the following primers: forward 5'-CACTCAAAGGGCTATACTTGG-3' and reverse 5'-GGATGGAGATTGATGTGACG-3', with the forward primer also used for sequencing. The amplified region included the *FecGF* and *GDF9* G6 variants, allowing validation of *FecGF* genotypes and determination of *GDF9* G6 genotypes for animals genotyped using the first version of the genotyping array, which did not include this variant.

The effect of the g.42114679C>T variant was evaluated using the same modeling and statistical procedures described for the known marker analyses. The individual effects of *FecGF*, *G6*, and g.42114679C>T, as well as their interactions and the effects of their multilocus genotype combinations, were also characterized. The best-fitting model, based on the Akaike information criterion (AIC), calculated using the *AIC* function in R (R Core Team, 2026), was used to obtain residuals following the same procedure as that described for the *FecGF* correction step in the GWAS analysis. These residuals were subsequently used to investigate the GWAS peak signals and determine whether the newly identified *GDF9* variant accounted for the association signal within the long haplotypes defined above or represented a separate associated variant.

Exploration of variants within carrier haplotypes

Following the persistence of the association signal at s21241.1 after model correction for the newly identified GDF9 variant, WGS variants located within the previously identified potentially carrier long-haplotype (LH) regions (chromosome 5, g.55708610 to g.57224603), or between the boundaries of these regions and the adjacent polymorphism represented on the genotyping array (g.55653839 and g.57268067), were extracted. Variants with fewer than three copies of either the reference or alternative allele, including monomorphic variants, as well as variants with a call rate below 90%, were filtered out.

Potentially causal variants were prioritized using three complementary metrics, applied separately to LH1 and LH2. The first metric was a correlation-based approach measuring the correlation between each WGS variant and LH status, together with its statistical significance, using the *cor.test* function with the Pearson method (R Core Team, 2026). Variants showing a correlation coefficient significantly different from zero ($P < 0.05$) were retained according to this metric.

The second metric identified variants matching specific genotype patterns according to LH carrier status. Variants were retained if they met the following conditions: (1) all animals carrying two copies of LH2 were homozygous for the alternative allele (no animals were homozygous for LH1 in the dataset); and (2) all animals carrying one copy of either LH were either heterozygous or homozygous for the alternative allele. Only variants fulfilling these conditions were retained according to this metric.

The third metric assessed the allele-frequency gradient between animals carrying at least one copy of the LH and non-carrier animals. An exact Fisher's test was performed for each WGS variant using the *fisher.test* function (R Core Team, 2026). Variants showing a significant positive allele-frequency difference, indicating enrichment of the alternative allele among LH carriers, were considered potential causal variants according to this metric.

The final list of candidate WGS variants included variants fulfilling at least two of the three metrics, regardless of whether the metrics were met for the same LH or for different LHs. An UpSet plot was used to visualize the overlap among variants identified by the different metrics and associated with each long haplotype, using the *UpSetR* package (52). Variant-effect predictions were obtained using the command-line version of *VEP* (53). Finally, the presence of the alternative allele was investigated in six Hampshire and eight Suffolk animals for which WGS data were available.

Results

Phenotypic analysis of litter sizes among breeds reared in Canada

Variation in litter size has led to the discovery of numerous major-effect mutations in sheep. The segregation of these variants within populations increases phenotypic variance, which has previously been used as a tool to identify new mutations (29). The regression of litter size variance on mean litter size across 20 sheep breeds reared in Canada showed a similar relationship to that previously described when breeds with a mean litter size greater than two lambs per litter were excluded (Figure 1, purple line). Including the Romanov breed, whose prolificacy has been hypothesized by some authors to have a polygenic basis (54–56), highlighted the higher-than-expected litter size variance of Rideau Arcott and Finnsheep breeds relative to the regression line (Figure 1, orange line). The litter size variances of these two breeds remained above the regression line when they were included in the regression model (Figure 1, red line). Excluding the prolific breeds resulted in a lower coefficient of determination ($R^2 = 0.89$) compared with the model including all breeds ($R^2 = 0.95$). Conversely, excluding the Rideau Arcott and Finnsheep breeds resulted in a slight numerical improvement in model fit, with an R^2 of 0.98.

Investigation of known markers’ effect on litter size

Thirty variants previously reported to affect ewe prolificacy or potentially contribute to its variation are currently included on the Canadian custom genotyping array (Figure 2). Among these, only ten were polymorphic in at least one of the breeds investigated in the present study. In addition, MTNR1A_612 and MTNR1A_706 and GDF9_R87H and GDF9_E241K were in complete linkage disequilibrium in the studied dataset (Figure 3; Additional file 2). Investigation of the effects of these variants on litter size identified three variants with a statistically significant effect (Table 1; Additional file 3).

Within the *GDF9* gene, the *FecGF* mutation (GDF9_V371M; G7) was significantly associated with increased litter size in Rideau Arcott ewes. Heterozygous and homozygous animals showed 6% and 10% higher estimated marginal means (EMMs), respectively, compared with non-carrier animals. No significant difference was observed between heterozygous and homozygous animals; however, litter size showed a numerical increase in homozygous animals, with no significant deviation from the additive model. Similar numerical differences were observed in other breeds carrying *FecGF*, including Canadian Arcott, East Friesian, and Polypay, although none showed a statistically significant effect.

The *GDF9* G6 variant was significantly associated with litter size in Romanov and Dorper breeds. Interestingly, heterozygous animals had higher EMM litter sizes than homozygous animals in both breeds, resulting in a significant deviation from the additive inheritance pattern. Homozygous animals had lower estimated litter sizes than non-carriers in Dorper ewes, whereas this difference was only numerical in Romanov ewes.

Within the *BMP15* gene, despite the small sample size available for White Dorper, ewes homozygous for the L252P variant had significantly higher EMM litter sizes than heterozygous animals. However, the presence of only one non-carrier animal prevented a clear assessment of the effect of the variant relative to the non-carrier genotype and made the assessment of deviation from additivity unreliable.

Table 1 Genetic candidates with significant effects on litter size.

Candidate	Breed	p-value[1]	Genotype	Number	EMM ± SE[2]	Contrast (D)[3]	p-value[3]
FecGF	RI	0.0008	WT	238	2.48±0.06[a]	0.03	0.4
			HET	371	2.63±0.05[b]		
			HOM	109	2.73±0.07[b]		
GDF9_V332I (G6)	DO	0.008	WT	18	1.54±0.14[b]	0.31	0.01
			HET	14	1.66±0.16[b]		
			HOM	7	1.17±0.17[a]		
	RV	0.008	WT	162	2.97±0.09[ab]	0.18	0.002
			HET	221	3.06±0.09[b]		
			HOM	97	2.79±0.1[a]		

1. False discovery rate correction applied using number of markers tested, per breed.
2. Estimated marginal mean (EMM) of litter size ± standard error (SE). Superscripts represented post-hoc Tukey statistical grouping on EMM.
3. Contrast of EMMs with expected linear regression values and associated deviation's p-value.

Discovery of new potential markers associated with litter size in the hyperprolific Rideau Arcott breed

Following confirmation of the effect of *FecGF* in the Rideau Arcott breed, a genome-wide association analysis (GWAS) was performed, as non-carrier ewes also exhibited high prolificacy. The additive heritability of litter size within the GWAS dataset was estimated at 0.10, which is considered low. The GWAS, performed on litter size records from 462 ewes with at least three registered litters, identified the three most strongly associated SNPs within or near the *GDF9* gene, none of which corresponded to the *FecGF* variant. These results further emphasized the contribution of the *GDF9* region to prolificacy in the Rideau Arcott breed (Figure 4A and B; Additional file 4). No deviation from the expected distribution of p-values was observed in the tail of the quantile–quantile plot, indicating sufficient correction for population structure and kinship. The strong associations of the three variants near *GDF9* were readily detectable (Figure 4C). Including the *FecGF* variant as a fixed effect in the linear mixed-effects model resulted in a sharp reduction of the association signal in this region (Figure 4D and E). However, an increase in the association signal was observed approximately 14 Mb downstream of *GDF9*, with the p-value of the g.56108129C>T variant (designated s21241.1 on the Illumina panel) approaching the

suggestive threshold (Figure 4D and E; Additional file 5). This signal was therefore considered sufficiently strong to warrant further investigation of the variants in this region.

The reconstruction of haplotypes potentially carrying the causal variant led to the definition of two long haplotypes (LH1 and LH2), each comprising 14 variants, including s21241.1, and spanning 1.5 Mb on chromosome 5 (g.55708610–g.57224603). Including both long haplotypes (LH1 and LH2) along with *FecGF* in the linear mixed-effects model improved model fit by 10 AIC units compared with the model including s21241.1 with *FecGF*, and by 22.4 AIC units compared with the model including *FecGF* alone. These results demonstrated the added value of including LH1 and LH2 in the model and confirmed the persistence of a statistically significant association (Additional file 6). The animals selected for whole-genome sequencing consisted of ten non-carriers, six LH1 heterozygotes, five LH2 heterozygotes, the only LH2 homozygote and the only two animals heterozygous for both LH1 and LH2.

In light of the signal neighbouring *FecGF* observed in the initial GWAS, the *GDF9* gene was first investigated using whole-genome sequencing data. This analysis led to the detection of a newly reported mutation (Figure 5A–C). This missense cytosine-to-thymine substitution at position 42114679 (ARS-UI_Ramb_v3.0), located in exon 2 of *GDF9*, results in an amino acid substitution from aspartic acid to asparagine at position 270 (D270N). Subsequent Sanger sequencing of the GWAS animals, combined with previously available sequencing data, identified 94 heterozygous and one homozygous animal, while the remaining 367 animals were non-carriers, corresponding to an alternative allele frequency of 10.4% (Figure 5D). Including this variant as a fixed effect in a linear mixed-effects model revealed a significant association with litter size ($p = 0.002$), with heterozygous ewes having an estimated marginal mean (EMM) litter size 0.21 lambs higher per litter than non-carriers (Additional file 7). The EMM litter size of the single homozygous ewe was 3.21 lambs per litter based on her four registered litters, corresponding to a non-significant difference of an additional 0.65 lambs per litter compared with non-carriers. Following the detection of this significant association for g.42114679C>T, the variant was designated *FecGR* herein, with “R” referring to the Rideau Arcott breed, following the nomenclature commonly used for GDF9 mutations in sheep (6).

To select the best model for extracting residuals for a subsequent conditional GWAS accounting for litter size associated *GDF9* variants, linear mixed-effects models including *FecGF* and *FecGR*, either individually or as combined genotypes (diplotypes), were tested (Additional file 7). The *G6 GDF9* variant was also included in several models to evaluate potential interactions with the two other *GDF9* variants. No significant interactions between the *GDF9* variants were observed. However, the combined *FecGF-FecGR* genotype was associated with a significantly higher estimated marginal mean (EMM) litter size compared with animals carrying neither variant and those carrying only one of the two variants in the heterozygous state (Figure 6). Interestingly, the *FecGF-G6* combination resulted in an EMM litter size approaching that of *FecGF* homozygotes, while animals carrying the *FecGR-G6* combination had EMM values similar to those of *FecGF-FecGR* carriers, although these differences were not statistically significant. Including *FecGF* and *FecGR* in the same model as combined genotypes also increased their estimated effects, as comparisons could then be made with animals carrying neither variant. For

example, the estimated marginal litter size mean increased from 6% to 8.5% and from 8% to 10% in heterozygous ewes carrying *FecGF* and *FecGR*, respectively (Additional file 7). The model including *FecGF* and *FecGR* as individual genotype effects provided the best fit (Additional file 7) and was therefore selected to generate the residuals for the subsequent GWAS, aiming to assess the persistence of the s21241.1 association signal (Figure 4F and G). This conditional GWAS showed an increased association signal for s21241.1, which exceeded the suggestive significance threshold compared with the previous GWAS accounting for *FecGF* alone (Additional file 8).

Fine-mapping of the region surrounding the GWAS peak was performed using an initial set of 17,352 variants meeting the defined region and filtering steps. Among these variants, 8,906 fulfilled the threshold for at least two of the three detection metrics (correlation, genotype pattern restriction, and allele-frequency gradient), while 1,998 and 2,852 variants fulfilled all three metrics for LH1 and LH2, respectively (Figure 7A; Additional file 9). Of these, 431 variants commonly passed the filtering criteria for both haplotypes. Assessment of alternative allele presence in Hampshire and Suffolk whole-genome-sequenced animals further narrowed this set to 16 variants. Of these, 12 were located within introns of *PPP2R2B*, two were located downstream of *SPINK1*, and two were located in intergenic regions (Figure 7B). Additional variants located within other genes and not represented in the Hampshire and Suffolk dataset were also identified, but none were common to both LH1 and LH2.

Discussion

Along with *Bone morphogenetic protein 15* (*BMP15*), *Growth differentiation factor 9* (*GDF9*) is a member of the transforming growth factor beta (TGF-β) superfamily that is expressed in the oocyte and acts during the early stages of folliculogenesis (14, 15, 57). *GDF9* is one of the major genes implicated in sheep prolificacy, with at least seven variants reported to increase ovulation rate or litter size (9, 10, 12, 13, 31, 34, 58). Among the 13 breeds investigated in the present study, three *GDF9* variants, *G6*, *FecGF*, and *FecGR*, were significantly associated with litter size.

The G6 variant was first identified in two Cambridge ewes and results in a nonpolar valine-to-isoleucine substitution at position 332 of the GDF9 protein (6). Since its initial description, this mutation has been reported in several breeds (59, 60), as was also observed in the present study, where it was polymorphic in all but one breed. Our analyses revealed an overdominant pattern in both Romanov and Dorper ewes, with heterozygous animals exhibiting greater adjusted litter size than homozygotes in both Romanov and Dorper breeds. A greater prolificacy of heterozygotes compared with homozygotes has also been reported in the Luzhong breed and in Suffolk × Ujimqin crossbred animals (61, 62). This pattern contrasts with the numerical increase in litter size from non-carriers to homozygotes reported in the first study associating G6 with prolificacy in the Araucana breed (58). The overdominant pattern observed for G6 is consistent with the peculiar reproductive effects reported for several *BMP15* and *GDF9* mutations. For most variants, homozygous ewes are sterile, whereas heterozygous ewes exhibit increased prolificacy compared with non-carriers (6, 7, 9, 11, 25, 28). In contrast, homozygous ewes carrying other mutations, including FecXGR, FecXO, FecXN, FecGE, and FecGF, are hyperprolific (13, 26, 33, 34). These

contrasting phenotypic effects have been proposed to result from the sensitivity of the underlying pathways to alterations in ligand–receptor affinity, as well as from imbalances among the different active homodimeric and heterodimeric forms of GDF9 and BMP15 (33, 57). The previously described structural changes induced by the G6 mutation could therefore alter these interactions, potentially contributing to the higher prolificacy of heterozygotes and the lower, but still elevated, prolificacy of homozygotes. In this context, G6 may represent one of the variants contributing to the prolificacy of the Romanov breed (59, 63, 64), which has previously been hypothesized to have a polygenic architecture (54–56). This could also explain the results presented in Fig. 1, with the Romanov breed showing greater variance in litter size than predicted by extrapolating the regression based only on breeds with a mean litter size below two lambs per litter.

Likewise, the high prolificacy of Finnsheep was also hypothesized to result from a polygenic background as early as 1996 (54). Subsequently, the *FecGF* (*G7*, V371M) mutation in *GDF9* was identified in Belclare, Cambridge, and Norwegian White Sheep (6, 34), all of which were developed with a contribution from Finnsheep. Based on these breed relationships, Våge et al. (34) hypothesized that the variant originated from Finnsheep, a hypothesis that was supported one year later by the detection of the mutation at a mean frequency of 0.30 in a Finnsheep control population (12). In the Finnsheep population reared in Canada, *FecGF* was detected at a frequency of 13% among the 39 genotyped animals, including eight heterozygotes and one homozygote, which is considerably lower than the previously reported frequency. However, given the small number of genotyped animals, no litter size analysis was conducted. The high variance in litter size relative to the average observed in this study could, however, support an effect of this variant in this breed. In contrast, the mutation was detected at a frequency of 40% in the Rideau Arcott breed, for which approximately 40% of the genetic background has been estimated to originate from Finnsheep (1, 2). The high number of genotyped animals and the availability of extensive litter size records allowed us to confirm linear increases of 0.15 and 0.25 lambs per litter compared with non-carrier animals for heterozygous and homozygous ewes, respectively. The increase observed in heterozygous ewes is consistent with the previously reported, although non-significant, increase of 0.13 lambs per litter in heterozygous Belclare ewes, whereas the effect observed in homozygous animals was considerably lower than the increase of 0.69 lambs per litter previously reported based on only two homozygous animals (12). No significant effect of *FecGF* was detected in Canadian Arcott, Polypay, or East Friesian ewes. These results were expected given the small number of heterozygous and the absence of homozygote ewes identified in these breeds. When considering the entire genotyped population, including both males and females, *FecGF* frequencies were 6% in Canadian Arcott and 7% in both Polypay and East Friesian. The presence of the mutation in East Friesian remains unexplained, as the populations investigated originated from different regions and neither breed was developed from the other. Past crosses between these populations could provide one plausible explanation.

As for *FecGF*, the recently reported mutation designated *FecGR* herein (g.42114679C > T) was also identified in the East Friesian genome GCA_018804185.1 (NWAFU_Friesian_1.0; BioProject PRJNA721520) at position g.42458995 (65). The presence of this mutation in an East Friesian genome is consistent with its detection in the Rideau Arcott population, as approximately 14% of the initial breed

composition was derived from East Friesian (1, 2). However, the variant was not detected by Sanger sequencing in the six East Friesian ewes with the highest average litter size based on at least three litters in the *Pheno dataset*. This may be related to the absence of strong selection for prolificacy in this dairy breed, where breeders' income is primarily derived from milk rather than lamb production. Following the discovery of *FecGR* in Rideau Arcott sheep and in the East Friesian genome after investigating its origin as part of the present study, we found that the mutation initially considered novel had already been reported in a previous conference proceeding as being part of GDF9 gene intron (66), whereas its position corresponds to exon 2 according to the current gene annotation. *FecGR* was identified in Australian maternal multibreed/admixed animals involving Border Leicester, Corriedale, and Coopworth, with a suspected origin in Finnsheep (66). The mutation was subsequently also detected in Coopworth and in composite animals from the International Sheep Genome Consortium population. Coopworth is a synthetic breed developed from Border Leicester and Romney (67), with both breeds also contributing to the development of Rideau Arcott (1, 2). Nevertheless, the variant was not detected in either Finnsheep or Border Leicester among the six highest-prolificacy individuals genotyped in Canada, leaving the origin of the *FecGR* mutation in the Rideau Arcott breed unresolved. Phenotypically, one copy of *FecGR* was associated with an increase of 0.21 lambs per litter in Rideau Arcott ewes, consistent with the previously reported increase of 0.22 lambs per litter (66). Further investigation will be needed to characterize the effect of *FecGR* in homozygous ewes, as only a small number of highly prolific homozygous animals were identified in both the present and previous study.

Several studies have documented combined effects of carrying variants in *GDF9*, *BMP15*, *BMPR1B*, or *B4GALNT2* simultaneously (6, 12, 24–28). The significantly higher estimated marginal litter size observed for ewes carrying both *FecGF* and *FecGR* in the present study is consistent with these previous reports. In contrast, the estimated marginal means of litter size for animals carrying *FecGF* and *G6* or *FecGR* and *G6* in the heterozygous state did not differ significantly from those of the other genotype combinations, despite a numerical pattern that could be consistent with an additive effect (Fig. 6). These combinations warrant further investigation, as only a small number of Rideau Arcott ewes carrying G6 in combination with *FecGF* or *FecGR* have been genotyped to date. Furthermore, further characterization of the haplotypes carrying these three mutations would be valuable, as no homozygous animals for one mutation were found to carry another mutation. This pattern suggests that these closely spaced variants may be inherited within specific haplotypes, although the available sample size does not allow the extent of recombination between these sites to be determined.

Previous studies in sheep have identified genes outside the TGF-β superfamily that have major effects on prolificacy (19, 21). For example, heterozygous ewes carrying one copy of the *FecL* variant in *B4GALNT2*, which encodes a glycosyltransferase, have been reported to produce an additional 0.5 lamb per litter on average (5). In Rideau Arcott, the persistence of an association signal around 56 Mb on chromosome 5 after accounting for *FecGF* and *FecGR* in the GWAS suggested that another variant could contribute to the high prolificacy of this breed. Fine-mapping of this region using whole-genome sequencing data and haplotype screening highlighted eight main genes, two of which contained variants fulfilling all three detection criteria in both long haplotypes (LH1 and LH2). Two downstream variants

were identified within *SPINK1*, which encodes serine peptidase inhibitor Kazal type 1, a protein that inhibits serine proteases (68). In reproduction, *SPINK1* has been reported to modulate sperm physiology, including sperm capacitation, by reducing endogenous nitric oxide concentrations (69). However, to our knowledge, no role for *SPINK1* in female reproduction has been reported. Three additional *SPINK* family genes, *SPINK5*, *SPINK6*, and *SPINK14*, are also located within the fine-mapped region. Interestingly, *SPINK5* is expressed in the human female reproductive system, including the ovaries, according to the Human Protein Atlas, and has been reported to interact with *CSGALNACT1* (70), a glycosyltransferase also expressed in human ovaries (71). A potential interaction with inhibin subunit beta A (*INHBA*) is also suggested by the STRING interaction database, which is consistent with the proposed mechanism by which the *FecL* variant in the *B4GALNT2* gene may affect prolificacy, namely through the glycosylation of inhibin subunits (19). These interactions further support *SPINK5* as a potential candidate gene. However, no variants common to both LH1 and LH2 were identified within this gene. The criteria used to prioritize variants, including correlation, genotype pattern restriction, and allele-frequency gradient, may be biased if the causal variant is present across multiple haplotypes. In such a situation, comparing the two carrier haplotypes with the remaining animals could reduce the ability of these metrics to discriminate the causal variant. Conversely, restricting the analysis to variants shared between LH1 and LH2 could also exclude causal variants if different variants independently contribute to the association observed in the two haplotypes. Therefore, other *SPINK* genes, as well as additional genes and long non-coding RNAs located within the haplotypic region, cannot be fully excluded as potential contributors to the observed association.

The second gene commonly retained after haplotype filtering was *Protein Phosphatase 2 Regulatory Subunit B beta* (*PPP2R2B*). PPP2R2B encodes a regulatory subunit of protein phosphatase 2A (PP2A), a serine/threonine phosphatase that dephosphorylates target residues in a site-specific manner (72). The primary role of PPP2R2B within the PP2A complex is to recruit specific substrates, thereby contributing to substrate recognition and protein–protein interactions (72). PPP2R2B RNA has been detected in human oocytes according to the Human Protein Atlas and has also been reported in human (73) and mouse oocyte (74). Recently, PPP2R2B mRNA was identified as interacting with miRNAs involved in a key reproduction-related subnetwork constructed by comparing miRNA expression between sheep breeds with high and low prolificacy (75). The biological plausibility of *PPP2R2B* as a candidate gene is further supported by the established role of PP2A in oocyte meiotic processes and female fertility in mice. More specifically, altering the Aα subunit of PP2A, encoded by *PPP2R1A*, which has a structural role within the PP2A complex (72), was associated with increased aneuploidy and embryonic mortality through effects on meiotic spindle length and chromosome alignment and segregation (76). Consequently, approximately three times fewer embryos successfully implanted compared with controls (76). Inactivation of PP2A catalytic subunits has also been shown to affect kinetochore–microtubule attachment during meiosis I in mice (77). Interestingly, *PPP2R1A* was also identified in a differentially expressed protein network when comparing high- and low-prolificacy sheep breeds in another recent study (78). The identification of PP2A subunit genes in two independent reproduction-related studies in sheep therefore provides additional support for a potential role of PP2A in sheep prolificacy. Although no

coding variants in *PPP2R2B* remained after filtering based on the fine-mapping criteria, 12 intronic variants, including two deletions and one 13-bp insertion, were retained. Intronic variants can influence several aspects of the splicing process, affecting the final transcript (79). As previously reviewed, alternative splicing can occur within genes encoding PP2A regulatory subunits and may affect their subcellular localization and function (72). Altered splicing therefore represents one possible mechanism through which these intronic variants could influence PPP2R2B function and, consequently, reproductive processes associated with prolificacy in sheep.

Altogether, *FecGF*, *FecGR*, and the new locus surrounding *SPINK* genes and *PPP2R2B* may explain part of the high prolificacy of Rideau Arcott sheep, while their incomplete segregation within the breed could explain at least part of the distinctively high variance in litter size observed in this population.

Conclusions

This study underscored the value of breeder-recorded data and data aggregation for investigating markers of interest in sheep. The integration of data from 13 sheep breeds enabled the confirmation of three variants as significantly associated with litter size in terminal and highly prolific breeds. In addition, the dataset enabled the detection and characterization, without prior knowledge, of the recently reported D270N variant in *GDF9*. These results provide additional tools for genetic selection for prolificacy in certain sheep breeds reared in Canada, potentially facilitating increases in litter size and, consequently, farm income for producers pursuing this breeding objective.

The exploration of the genetic background underlying prolificacy in the Rideau Arcott breed also identified new candidate genes potentially involved in sheep prolificacy and provided additional evidence supporting a potential role of PP2A signaling in reproductive processes. Further research will be required to pinpoint the precise variant underlying the association in this Canadian breed.

As allelic status and marker effects are now confirmed, characterizing the consequences of incorporating these markers into EPD calculations may be particularly important, as their use could influence EPD accuracy.

Declarations

Ethics approval and consent to participate

Ear punches used to obtain genotyping and sequencing data were collected from animals owned by sheep breeders, who voluntarily covered the associated sampling costs. The sampling protocol using Allflex Tissue Sampling Units (TSU; Allflex Livestock Intelligence, Madison, WI, USA) was approved by the Université Laval Animal Welfare Committee (VRR-23-1362). The Purebred Sheep Association of Quebec (SEMRPQ) provided consent for the use of the data generated from its represented breeders for this project and its publication.

Consent for publication

Not applicable.

Availability of data and materials

Genomic data (genotyping and sequencing) underlying the analyses were anonymized due to their sensitivity in relation to breeders' livestock selection decisions. The anonymized genomic data from *Marker*, *GWAS* and *WGS* datasets were deposited into the *European Variation Archive* (EVA) database under the project accession number PRJEB128063.

Results generated during the current study are available within the paper and its Supplementary Information.

Competing interests

The authors declare that they have no competing interests**.**

Funding

This project was supported by two project grants attributed to C. Robert from Genome Québec and Genome Canada awarded in the Genomics Integration Program and Genomic Applications Partnership Program (GAPP). The Société des Éleveurs de Moutons de Race Pure du Québec (SEMRPQ) and the Centre d'Expertise en Production Ovine du Québec (CEPOQ) provided in-kind support namely samples, data and expertise. The Natural Sciences and Engineering Research

Council of Canada supported W. Poisson with student scholarships awards.

Authors' contributions

All authors have read and approved the manuscript. WP contributed to grant writing and study design, managed the sampling, genotyping, and sequencing processes, analyzed the genomic data, and wrote, reviewed, and edited the original draft. CM and JC contributed to grant writing and sampling management and acted as liaisons between the research team and the breeders. AS-P compiled the phenotypic data required for the analyses through the national genetic selection program. RP-N contributed to sample management, DNA extraction, and preparation for genotyping and sequencing. IG provided expertise in DNA extraction and preparation and contributed to writing, reviewing, and editing the original draft. EF provided supervision for statistical analyses. SF was a major contributor to study design and provided expertise and supervision throughout the project. He also contributed to writing, reviewing, and editing the original draft. CR was the main contributor to funding acquisition, study design, supervision, project administration, and writing, reviewing, and editing.

Acknowledgements

We particularly acknowledge breeders from Quebec and other provinces of Canada as well as the Société des éleveurs de moutons de races pures du Québec (SEMRPQ) and the Centre d'expertise en production ovine du Québec (CEPOQ) for their main contribution to this project. We acknowledge the contribution of Dr. Laurence Drouilhet for her valuable advice regarding statistical analyses, Dr. Maxim Ben Braiek for his haplotype definition script and Florence Plisson-Petit and Audrey Zita Siewe Tchakam for PCR and Sanger sequencing preparation.

References

1. Fahmy MH, Shrestha JNB. DLS and Arcott sheep: new Canadian breeds. Canada:Minister of Supply and Services Canada; 1993.
2. Shrestha JNB, Heaney DP. Review of Canadian, Outaouais and Rideau Arcott breeds of sheep: 1. Development and characterization. Small Rumin Res. 2003. 10.1016/S0921-4488(03)00070-1.
3. Dmitriev NG, Ėrnst LK. Animal genetic resources of the USSR. In: FAO and UNEP, editors. Fao animal production and health paper. Rome; 1989.
4. Dýrmundsson ÓR, Niżnikowski R. North European short-tailed breeds of sheep: a review. animal. 2010. 10.1017/S175173110999156X.
5. Martin P, Raoul J, Bodin L. Effects of the FecL major gene in the Lacaune meat sheep population. Genet Sel Evol. 2014. 10.1186/1297-9686-46-48.
6. Hanrahan JP, Gregan SM, Mulsant P, Mullen M, Davis GH, Powell R, et al. Mutations in the genes for oocyte-derived growth factors GDF9 and BMP15 are associated with both increased ovulation rate and sterility in Cambridge and Belclare sheep (Ovis aries). Biol Reprod. 2004. 10.1095/biolreprod.103.023093.
7. Lassoued N, Benkhlil Z, Woloszyn F, Rejeb A, Aouina M, Rekik M, et al. FecXBar a Novel BMP15 mutation responsible for prolificacy and female sterility in Tunisian Barbarine Sheep. BMC Genet. 2017. 10.1186/s12863-017-0510-x.
8. Martinez-Royo A, Jurado JJ, Smulders JP, Martí JI, Alabart JL, Roche A, et al. A deletion in the bone morphogenetic protein 15 gene causes sterility and increased prolificacy in Rasa Aragonesa sheep. Anim Genet. 2008. 10.1111/j.1365-2052.2008.01707.x.
9. Nicol L, Bishop SC, Pong-Wong R, Bendixen C, Holm LE, Rhind SM, et al. Homozygosity for a single base-pair mutation in the oocyte-specific GDF9 gene results in sterility in Thoka sheep. Reprod Camb Engl. 2009. 10.1530/REP-09-0193.
10. Souza CJH, McNeilly AS, Benavides MV, Melo EO, Moraes JCF. Mutation in the protease cleavage site of GDF9 increases ovulation rate and litter size in heterozygous ewes and causes infertility in homozygous ewes. Anim Genet. 2014. 10.1111/age.12190.
11. Bodin L, Di Pasquale E, Fabre S, Bontoux M, Monget P, Persani L, et al. A novel mutation in the bone morphogenetic protein 15 gene causing defective protein secretion is associated with both increased ovulation rate and sterility in lacaune sheep. Endocrinology. 2007. 10.1210/en.2006-0764.

12. Mullen MP, Hanrahan JP. Direct evidence on the contribution of a missense mutation in gdf9 to variation in ovulation rate of finnsheep. PLoS ONE. 2014. 10.1371/journal.pone.0095251.
13. Silva BDM, Castro EA, Souza CJH, Paiva SR, Sartori R, Franco MM, et al. A new polymorphism in the Growth and Differentiation Factor 9 (GDF9) gene is associated with increased ovulation rate and prolificacy in homozygous sheep. Anim Genet. 2011. 10.1111/j.1365-2052.2010.02078.x.
14. Juengel JL, Hudson NL, Heath DA, Smith P, Reader KL, Lawrence SB, et al. Growth differentiation factor 9 and bone morphogenetic protein 15 are essential for ovarian follicular development in sheep. Biol Reprod. 2002. 10.1095/biolreprod.102.007146.
15. Kona SSR, Praveen Chakravarthi V, Siva Kumar AVN, Srividya D, Padmaja K, Rao VH. Quantitative expression patterns of GDF9 and BMP15 genes in sheep ovarian follicles grown *in vivo* or cultured *in vitro*. Theriogenology. 2016. 10.1016/j.theriogenology.2015.09.022.
16. Mulsant P, Lecerf F, Fabre S, Schibler L, Monget P, Lanneluc I, et al. Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Mérino ewes. Proc Natl Acad Sci U S A. 2001. 10.1073/pnas.091577598.
17. Souza CJ, MacDougall C, MacDougall C, Campbell BK, McNeilly AS, Baird DT. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. J Endocrinol. 2001. 10.1677/joe.0.169r001.
18. Wilson T, Wu XY, Juengel JL, Ross IK, Lumsden JM, Lord EA, et al. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. Biol Reprod. 2001. 10.1095/biolreprod64.4.122.
19. Drouilhet L, Mansanet C, Sarry J, Tabet K, Bardou P, Woloszyn F, et al. The highly prolific phenotype of lacaune sheep is associated with an ectopic expression of the B4GALNT2 gene within the ovary. PLOS Genet. 2013. 10.1371/journal.pgen.1003809.
20. Haldar A, French MC, Brauning R, Edwards SJ, O'Connell AR, Farquhar PA, et al. Single-nucleotide polymorphisms in the LEPR gene are associated with divergent phenotypes for age at onset of puberty in Davisdale ewes. Biol Reprod. 2014. 10.1095/biolreprod.113.115923.
21. Juengel JL, French MC, O'Connell AR, Edwards SJ, Haldar A, Brauning R, et al. Mutations in the leptin receptor gene associated with delayed onset of puberty are also associated with decreased ovulation and lambing rates in prolific Davisdale sheep. Reprod Fertil Dev. 2016;28(9):1318. 10.1071/RD14382.
22. Macé T, González-García E, Foulquié D, Carrière F, Pradel J, Durand C, et al. Genome-wide analyses reveal a strong association between LEPR gene variants and body fat reserves in ewes. BMC Genomics. 2022. 10.1186/s12864-022-08636-z.
23. Abuzahra M, Wijayanti D, Effendi MH, Mustofa I, Lamid M. Association of melatonin receptor 1 A with litter size in sheep: A review. F1000Research. 2023; 10.12688/f1000research.134890.2
24. Calvo JH, Serrano M, Martinez-Royo A, Lahoz B, Sarto P, Ibañez-Deler A, et al. SNP rs403212791 in exon 2 of the *MTNR1A* gene is associated with reproductive seasonality in the Rasa aragonesa

sheep breed. Theriogenology. 2018. 10.1016/j.theriogenology.2018.02.013.

25. Calvo JH, Chantepie L, Serrano M, Sarto MP, Iguacel LP, Jiménez MÁ, et al. A new allele in the BMP15 gene (FecXRA) that affects prolificacy co-segregates with FecXR and FecXGR in Rasa aragonesa sheep. Theriogenology. 2020. 10.1016/j.theriogenology.2020.01.010.
26. Chantepie L, Bodin L, Sarry J, Woloszyn F, Plisson-Petit F, Ruesche J, et al. Genome-wide identification of a regulatory mutation in BMP15 controlling prolificacy in sheep. Front Genet. 2020. 10.3389/fgene.2020.00585.
27. Davis GH, Dodds KG, Bruce GD. Combined effect of the Inverdale and Booroola prolificacy genes on ovulation rate in sheep. Proc. Assoc. Advmt. Anim. Breed. Genet. 1999;13.
28. Galloway SM, McNatty KP, Cambridge LM, Laitinen MPE, Juengel JL, Jokiranta TS, et al. Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. Nat Genet. 2000. 10.1038/77033.
29. Chantepie L, Bodin L, Sarry J, Woloszyn F, Ruesche J, Drouilhet L, et al. Presence of causative mutations affecting prolificacy in the Noire du Velay and Mouton Vendéen sheep breeds. Livest Sci. 2018. 10.1016/j.livsci.2018.07.007.
30. Drouilhet L, Taragnat C, Fontaine J, Duittoz A, Mulsant P, Bodin L, et al. Endocrine characterization of the reproductive axis in highly prolific lacaune sheep homozygous for the FecLL mutation. Biol Reprod. 2010. 10.1095/biolreprod.109.082065.
31. Holm LE, Bendixen C, Eythorsdottir E, Hallsson JH. A frameshift deletion in the GDF9 gene in Icelandic Loa prolific sheep. Anim Genet. 2022. 10.1111/age.13168.
32. Juengel JL, O'Connell AR, French MC, Proctor LE, Wheeler R, Farquhar PA, et al. Identification of a line of sheep carrying a putative autosomal gene increasing ovulation rate in sheep that does not appear to interact with mutations in the Transforming growth factor beta superfamily. Biol Reprod. 2011. 10.1095/biolreprod.110.090514.
33. Demars J, Fabre S, Sarry J, Rossetti R, Gilbert H, Persani L, et al. Genome-wide association studies identify two novel BMP15 mutations responsible for an atypical hyperprolificacy phenotype in sheep. PLOS Genet. 2013. 10.1371/journal.pgen.1003482.
34. Våge DI, Husdal M, Kent MP, Klemetsdal G, Boman IA. A missense mutation in growth differentiation factor 9 (GDF9) is strongly associated with litter size in sheep. BMC Genet. 2013. 10.1186/1471-2156-14-1.
35. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011. 10.1093/bioinformatics/btr330.
36. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. Am J Hum Genet. 2018. 10.1016/j.ajhg.2018.07.015.
37. R Core Team. R: The R Project for Statistical Computing. https://www.r-project.org (2026). Accessed on 20 Sep 2026.
38. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015. 10.18637/jss.v067.i01.

39. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: Tests in linear mixed effects models. J Stat Softw. 2017. 10.18637/jss.v082.i13.
40. Fox J, Weisberg S, An. R companion to applied regression. 2nd ed. California:SAGE; 2011.
41. Searle SR, Speed FM, Milliken GA. Population Marginal Means in the Linear Model: An Alternative to Least Squares Means. Arch de zootecnia. 1996;45(170):201–2.
42. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012. 10.1038/ng.2310.
43. Granato ISC, Galli G, de Oliveira CEG, de Souza MB, Mendonça LF, Fritshe-Neto R. snpReady: a tool to assist breeders in genomic analysis. Mol Breed. 2018. 10.1007/s11032-018-0844-8.
44. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007. 10.1086/519795.
45. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. J Open Source Softw. 2018. 10.21105/joss.00731.
46. Covarrubias-Pazaran G. Renome-assisted prediction of quantitative traits using the R package sommer. Zhang A, editor. PLOS ONE. 2016; 10.1371/journal.pone.0156744
47. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012. 10.1093/bioinformatics/bts606.
48. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018. 10.1093/bioinformatics/bty560.
49. Wingett SW, Andrews S. FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res. 2018; 10.12688/f1000research.15931.2
50. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009. 10.1093/bioinformatics/btp324.
51. Auwera GAVS, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. 1st ed. Sebastopol: O'Reilly Media; 2020.
52. Lex A, Gehlenborg N. Sets and intersections. Nat Methods. 2014. 10.1038/nmeth.3033.
53. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016. 10.1186/s13059-016-0974-4.
54. Fahmy MH. Prolific sheep. 1st ed. Wallingford: CAB International; 1996.
55. Murphy TW, Keele JW, Freking BA. Genetic and nongenetic factors influencing ewe prolificacy and lamb body weight in a closed Romanov flock. J Anim Sci. 2020. 10.1093/jas/skaa283.
56. Pokharel K, Peippo J, Honkatukia M, Seppälä A, Rautiainen J, Ghanem N, et al. Integrated ovarian mRNA and miRNA transcriptome profiling characterizes the genetic basis of prolificacy traits in sheep (Ovis aries). BMC Genomics. 2018. 10.1186/s12864-017-4400-4.

57. McNatty KP, Juengel JL, Pitman JL, Davis GH. Fecundity Genes. Reference module in neuroscience and biobehavioral psychology. Elsevier; 2017. pp. 632–6.
58. Bravo S, Larama G, Paz E, Inostroza K, Montaldo HH, Sepúlveda N. Polymorphism of the GDF9 gene associated with litter size in Araucana creole sheep. Anim Genet. 2016. 10.1111/age.12404.
59. Heaton MP, Smith TPL, Freking BA, Workman AM, Bennett GL, Carnahan JK, et al. Using sheep genomes from diverse U.S. breeds to identify missense variants in genes affecting fecundity. F1000Res. 2017. 10.12688/f1000research.12216.1.
60. Najafabadi HA, Khansefid M, Mahmoud GG, Zhou H, Hickford JGH. Identification of polymorphisms in the oocyte-derived growth differentiation growth factor 9 (GDF9) gene associated with litter size in New Zealand sheep (Ovis aries) breeds. Reprod Domest Anim. 2020. 10.1111/rda.13813.
61. Ji X, Cao Z, Hao Q, He M, Cang M, Yu H, et al. Effects of new mutations in BMPRIB, GDF9, BMP15, LEPR, and B4GALNT2 genes on litter size in sheep. Vet Sci. 2023. 10.3390/vetsci10040258.
62. Wang F, Chu M, Pan L, Wang X, He X, Zhang R, et al. Polymorphism detection of GDF9 gene and its association with litter size in Luzhong mutton sheep (Ovis aries). Animals. 2021. 10.3390/ani11020571.
63. Shakhin AV, Dotsev AV, Deniskova TE, Brem G, Zinovieva NA. PSXII-2 Identification of candidate SNPs associated with high prolificacy in Romanov sheep. J Anim Sci. 2021. 10.1093/jas/skab235.465.
64. Xu H, Akhmet N, Luo Y, Guo Z, Pan C, Song E, et al. Are two beneficial mutations (p.Q249R and 90-bp Indel) within the ovine BMPRIB gene associated with growth traits? Front Vet Sci. 2023. 10.3389/fvets.2023.1280548.
65. You X, Fang Q, Chen C, Cao J, Fu S, Zhang T, et al. A near complete genome assembly of the East Friesian sheep genome. Sci Data. 2024. 10.1038/s41597-024-03581-w.
66. Moghaddar N, Bunter KL, Swan AA, Li L, Gurman PM. The effect of GDF9 on litter size in Australian sheep. In: Association for the Advancement of Animal Breeding and Genetics; 2021.
67. Davis GH, Dodds KG, Wheeler R, Jay NP. Evidence that an imprinted gene on the x chromosome increases ovulation rate in sheep. Biol Reprod. 2001. 10.1095/biolreprod64.1.216.
68. Wang GP, Xu CS. Pancreatic secretory trypsin inhibitor: More than a trypsin inhibitor. World J Gastrointest Pathophysiol. 2010. 10.4291/wjgp.v1.i2.85.
69. Zalazar L, Saez Lancellotti TE, Clementi M, Lombardo C, Lamattina L, De Castro R, et al. SPINK3 modulates mouse sperm physiology through the reduction of nitric oxide level independently of its trypsin inhibitory activity. Reproduction. 2012. 10.1530/REP-11-0107.
70. Huttlin EL, Bruckner RJ, Navarrete-Perea J, Cannon JR, Baltier K, Gebreab F, et al. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell. 2021. 10.1016/j.cell.2021.04.011.
71. Wang D, Eraslan B, Wieland T, Hallström B, Hopf T, Zolg DP, et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Mol Syst Biol. 2019. 10.15252/msb.20188503.

72. Slupe AM, Merrill RA, Strack S. Determinants for substrate specificity of Protein Phosphatase 2A. Enzyme Res. 2011;2011; 10.4061/2011/398751
73. Virant-Klun I, Leicht S, Hughes C, Krijgsveld J. Identification of maturation-specific proteins by single-cell proteomics of human oocytes. Mol Cell Proteom MCP. 2016. 10.1074/mcp.M115.056887.
74. Li Q, Mu L, Yang X, Wang G, Liang J, Wang S, et al. Discovery of oogenesis biomarkers from mouse oocytes using a single-cell proteomics approach. J Proteome Res. 2023. 10.1021/acs.jproteome.3c00157.
75. Yousuf S, Malik WA, Feng H, Liu T, Xie L, Miao X. Integrated analysis of microRNA and mRNA interactions regulating fecundity in the ovaries of two distinct sheep breeds. BMC Genomics. 2025. 10.1186/s12864-025-11408-0.
76. Hu MW, Wang ZB, Jiang ZZ, Qi ST, Huang L, Liang QX, et al. Scaffold subunit Aalpha of PP2A is essential for female meiosis and fertility in mice. Biol Reprod. 2014. 10.1095/biolreprod.114.120220.
77. Tang A, Shi P, Song A, Zou D, Zhou Y, Gu P, et al. PP2A regulates kinetochore-microtubule attachment during meiosis I in oocyte. Cell Cycle. 2016. 10.1080/15384101.2016.1175256.
78. Gu B, Liu H, Xu T, Sui M, Yang Y, Jiang H. Integrated miRNA-mRNA analysis Identifies oar-miR-143 as a potential regulator of follicular dominance in multiparous sheep and validates its target gene CDK2. Anim Reprod Sci. 2026. 10.1016/j.anireprosci.2026.108277.
79. Ohno K, Takeda J, ichi, Masuda A. Rules and tools to predict the splicing effects of exonic and intronic mutations. WIREs RNA. 2018. 10.1002/wrna.1451.

Figures

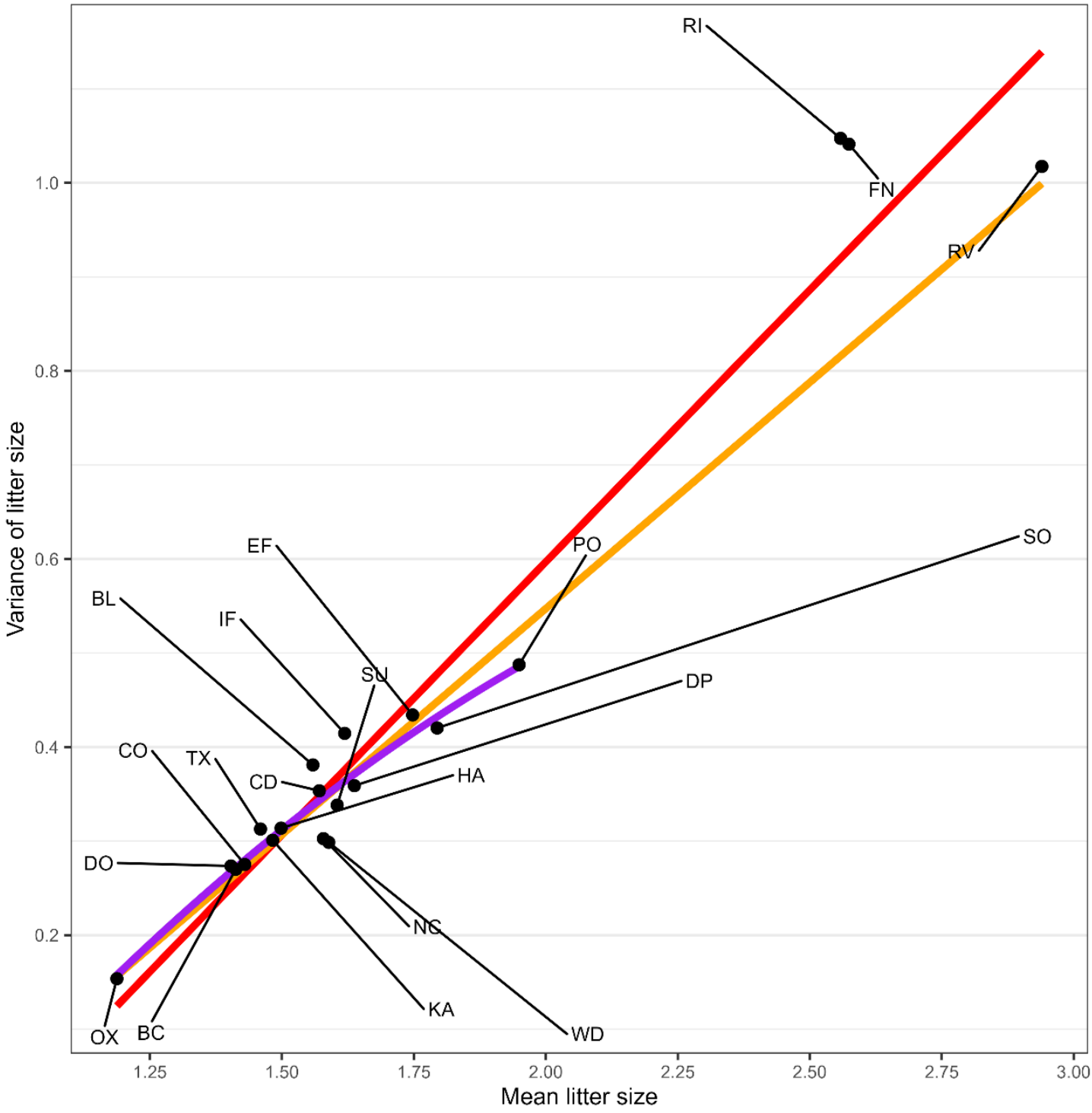


Figure 1

Relationship between litter size variance and average across 20 sheep breeds reared in Canada. Prolific breeds (FN, RI and RV), defined as breeds with a mean litter size greater than two lambs per litter, were clearly separated from the other breeds. The red line represents the quadratic regression fitted across all breeds from the *Pheno dataset*. The orange line represents the regression excluding Finnsheep and Rideau Arcott, whereas the purple line represents the regression including only breeds with a mean litter size below two lambs per litter. Breeds are represented as follow : Border Cheviot (BC), Border Leicester

(BL), Canadian Arcott (CD), Charolais (CO), Dorper (DO), Dorset Polled (DP), East Friesian (EF), Finnsheep (FN), Hampshire (HA), Ile de France (IF), Katahdin (KA), North Country Cheviot (NC), Oxford (OX), Polypay (PO), Rideau Arcott (RI), Romanov (RV), Southdown (SO), Suffolk (SU), Texel (TX), White Dorper (WD).

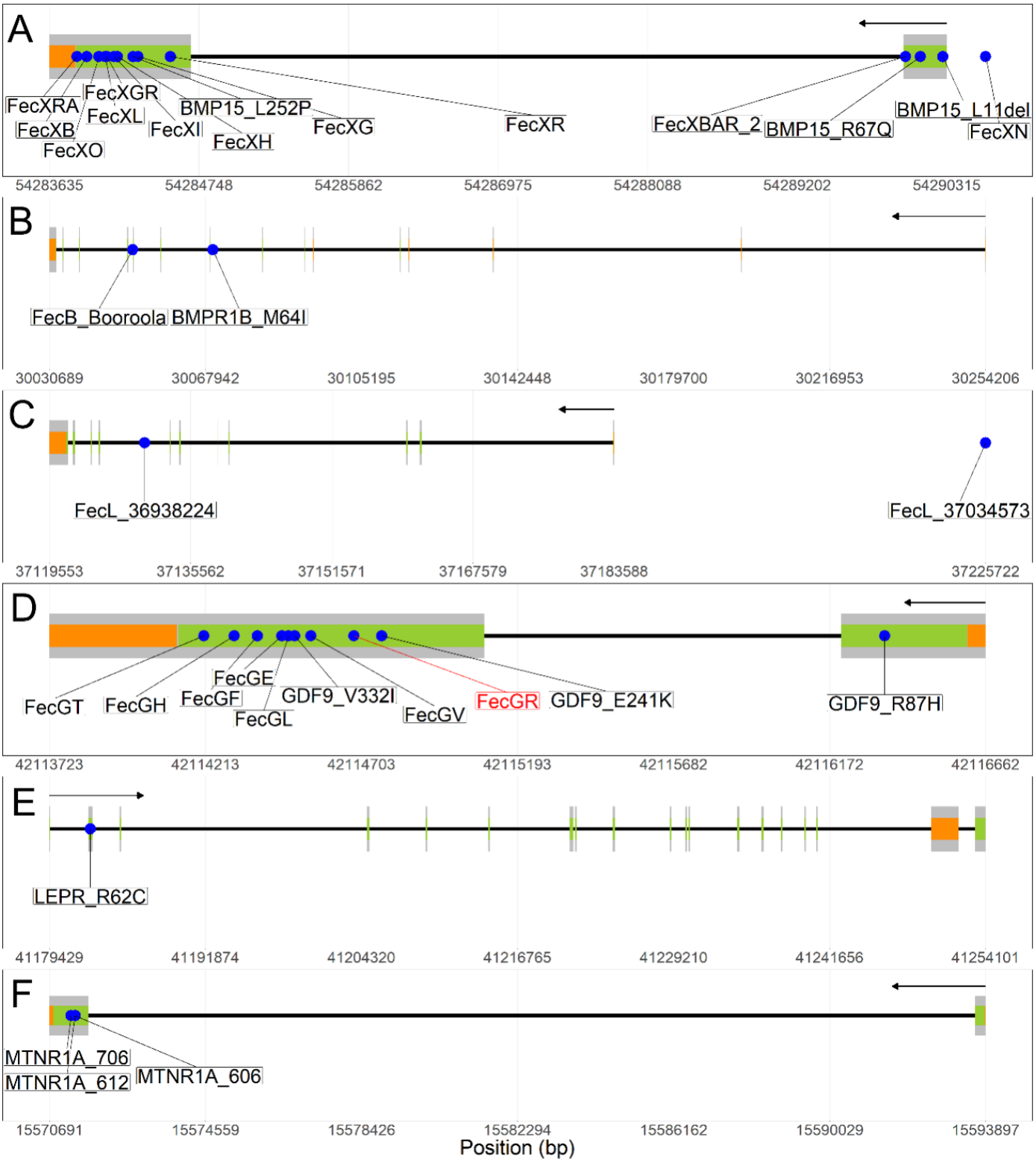


Figure 2

Genomic positions of the markers investigated in this study. All reported elements and their genomic positions were extracted from the Ensembl ARS-UI_Ramb_v3.0.116 annotation file (.GTF) and plotted regardless of transcript ID to represent the overall functional genomic context of each variant. Exons, coding sequences (CDS), and untranslated regions (UTRs) are represented in grey, green, and orange, respectively. Arrows indicate the transcription direction. **A.** *BMP15* gene on chromosome X. The L11 deletion was not analyzed because of poor genotype clustering. **B.** *BMPR1B*gene on chromosome 6. **C.** *B4GALNT2* gene on chromosome 11. **D.***GDF9* gene on chromosome 5. **E.** *LEPR* gene on chromosome 1. **F.***MTNR1A* gene on chromosome 26. *FecGR* (in red) corresponds to the newly identified GDF9 variant.



Figure 3

Alternative allele frequencies of reproduction-related variants showing polymorphism in the *Marker* dataset. The GDF9_R87H (*G1*) and GDF9_E241K (*G4*) variants, as well as MTNR1A_612 and MTNR1A_706, are in complete linkage disequilibrium within the dataset.

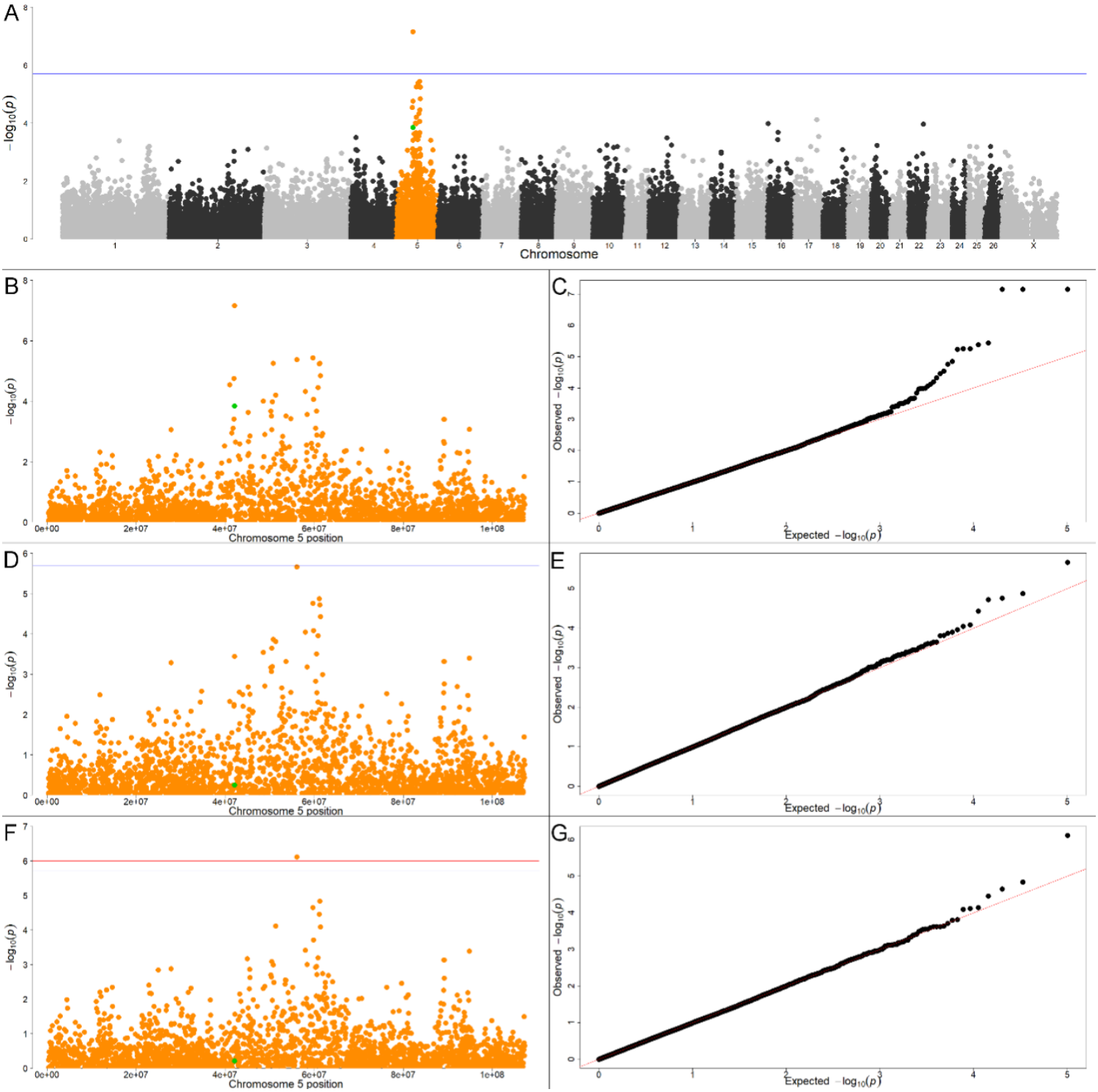


Figure 4

Genome-wide association analysis of litter size in Rideau-Arcott ewes. **A, B and C.** Manhattan and quantile-quantile plots of the GWAS p-values obtained before accounting for the known *FecGF* variant. **D and E**. GWAS results obtained using the same dataset after including *FecGF* as a fixed effect in the model. **F and G.** GWAS results obtained after accounting for *FecGF* and *FecGR*genotypes as fixed effects**.** Red and blue lines represent the Bonferroni-corrected genome-wide significant and suggestive thresholds, respectively. The *FecGF* marker is represented by a green dot in the Manhattan plots.

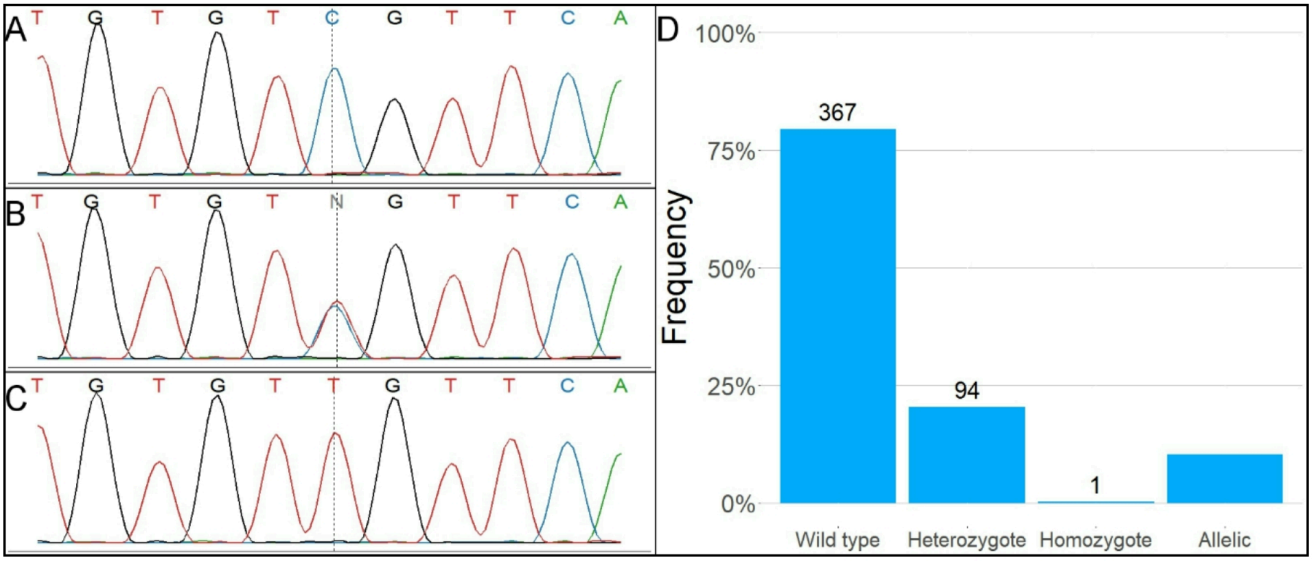


Figure 5

Detection of the **NC_056058.1:g.42114679C>T variant in the GDF9 gene in Rideau Arcott sheep.**Representative Sanger sequencing chromatograms showing the variant site in **A**. wild-type, **B.** heterozygous carrier, and **C.** homozygous mutant animals, respectively. The mutation is positioned in line with the central dashed lines. **D.** Genotype and allele frequencies among the 462 sequenced ewes. Superscripts indicate the number of animals tested for each genotype.

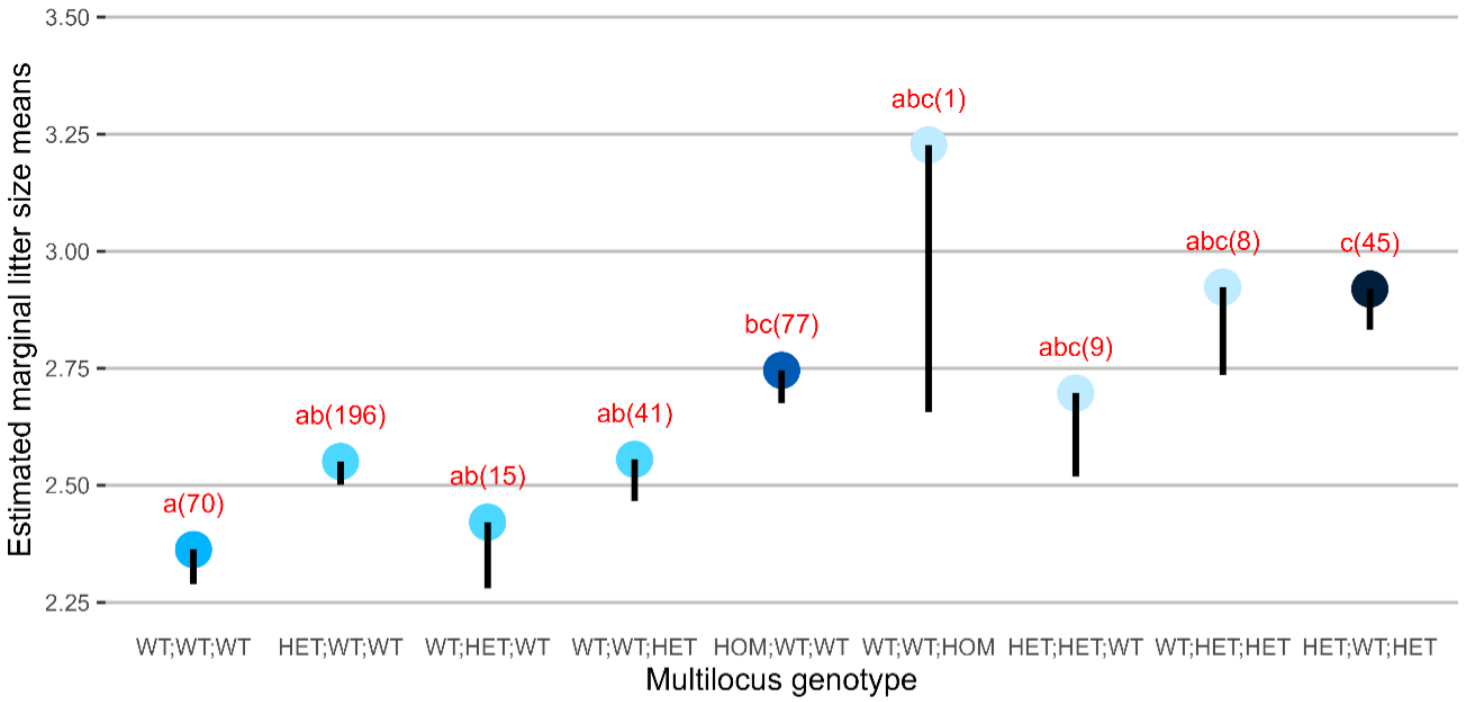


Figure 6

Estimated marginal mean litter size **according to combined genotypes for the non-synonymous *GDF9* mutations.**Multilocus genotypes are presented in the *FecGF*;*G6*;*FecGR*order. Letters indicate significant differences based on the Tukey post hoc test, and the number of ewes for each genotype is provided in parentheses. Dot colors were assigned according to significance letter groups. Downward-oriented black bars represent the standard errors of the estimated marginal means. WT,wild-type; HET, heterozygous carrier; HOM, homozygote mutant animal.

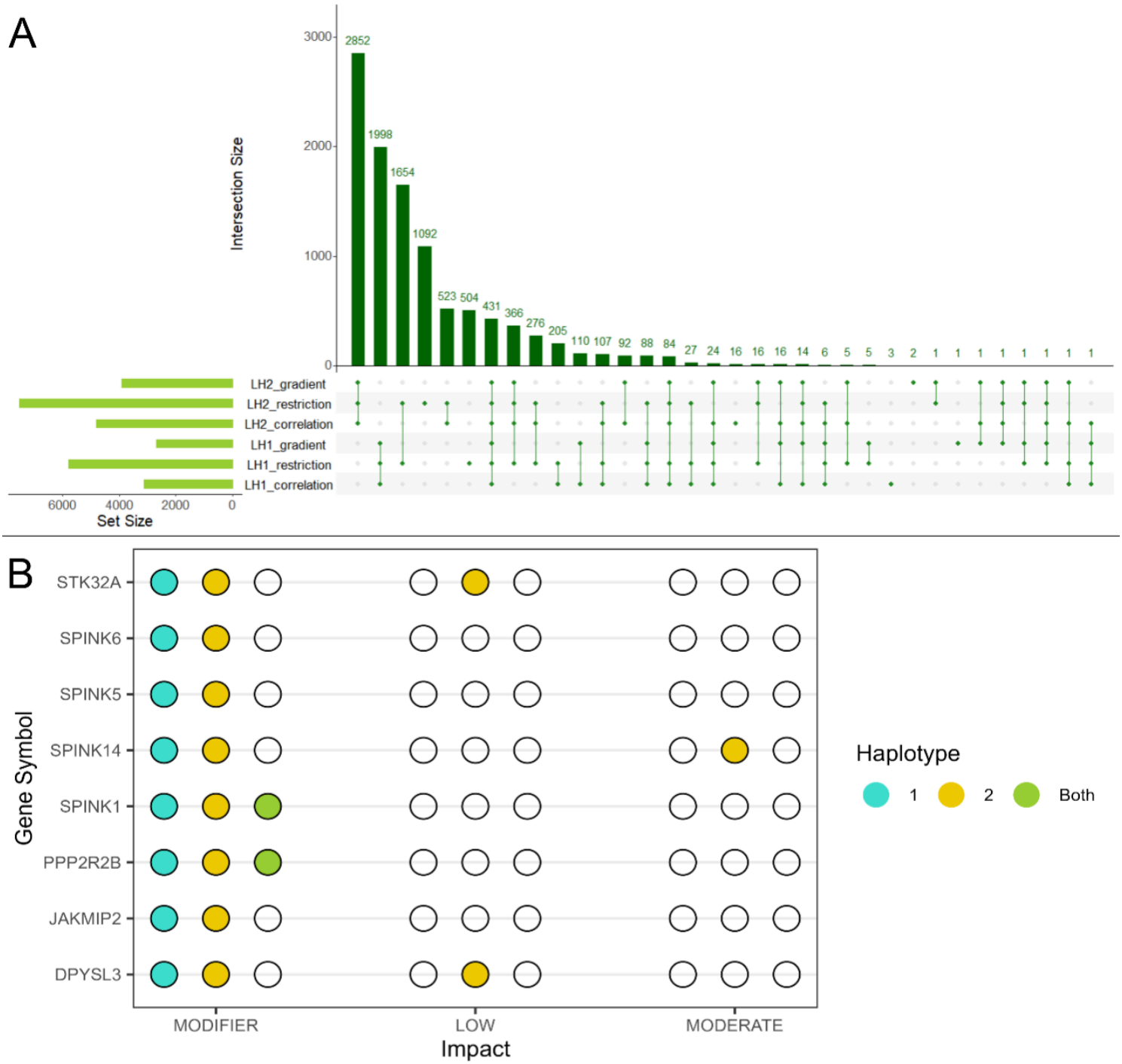


Figure 7

Potential variants and genes **associated with prolificacy in Rideau Arcott sheep. A.**UpSet plot representing the number of variants common to each combination of the three detection metrics (correlation, genotype pattern restriction, and allele-frequency gradient) within LH1, LH2, or both haplotypes. **B.** Plot showing the VEP-predicted impact of variants within each gene that were absent from the Hampshire and Suffolk sequenced animals and fulfilled all three detection metrics in LH1, LH2, or both haplotypes. Only genes with an assigned gene symbol were included. Empty circles indicate the absence of variants with the corresponding predicted impact within the respective gene and haplotype.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile2.xlsx
- AdditionalFile4.xlsx
- AdditionalFile5.xlsx
- AdditionalFile6.xlsx
- AdditionalFile7.xlsx
- AdditionalFile8.xlsx
- AdditionalFile9.xlsx
- AdditionalFile1VF.xlsx
- AdditionalFile3VF.xlsx
- Additionalfileslegends.docx