Genome-wide identification of a regulatory mutation in BMP15 controlling prolificacy in sheep

Louise Chantepie, Loys Bodin, Julien Sarry, Florent Woloszyn, Florence Plisson-Petit, Julien Ruesche, Laurence Drouilhet, Stéphane Fabre

Genome-wide identification of a regulatory mutation in *BMP15* controlling prolificacy in sheep

Short title: Regulatory mutation in BMP15 associated with ovine prolificacy

Louise Chantepie, Loys Bodin, Julien Sarry, Florent Woloszyn, Florence Plisson-Petit, Julien Ruesche, Laurence Drouilhet and Stéphane Fabre*

GenPhySE, Université de Toulouse, INRA, ENVT, Castanet Tolosan, France

*corresponding author

E-mail: stephane.fabre@inra.fr (SF)

Abstract

The search for the genetic determinism of prolificacy variability in sheep has evidenced several major mutations in genes playing a crucial role in the control of ovulation rate. In the Noire du Velay (NV) sheep population, a recent genetic study has evidenced the segregation of such a mutation named *FecL*^{L}. However, based on litter size (LS) records of *FecL*^{L} non-carrier ewes, the segregation of a second prolificacy major mutation was suspected in this population. In order to identify this mutation, we have combined case/control genome-wide association study with ovine 50k SNP chip genotyping, whole genome sequencing and functional analyses. A new single nucleotide polymorphism (OARX:50977717T>A, NC_019484) located on the X chromosome upstream of the *BMP15* gene was evidenced highly associated with the prolificacy variability (P =1.93E^{-11}). The variant allele was called *FecX*^{N} and shown to segregate also in the Blanche du Massif Central (BMC) sheep population. In both NV and BMC, the *FecX*^{N} allele frequency was estimated close to 0.10, and its effect on LS was estimated at +0.20 lamb per lambing at heterozygous state. Homozygous *FecX*^{N} carrier ewes were fertile with increased prolificacy in contrast to numerous mutations affecting *BMP15*. At the molecular level, *FecX*^{N} was shown to decrease BMP15 promoter activity and to impact *BMP15* expression in oocyte. This regulatory action was proposed as the causal mechanism for the *FecX*^{N} mutation to control ovulation rate and prolificacy in sheep.

Author Summary

In the genetic etiology of women infertility syndromes, a focus was done on the oocyte-expressed *BMP15* and *GDF9* genes harboring several mutations associated with

ovarian dysfunctions. In sheep also, mutations in these two genes are known to affect the ovarian function leading to sterility or, on the opposite, increasing ovulation rate and litter size constituting the prolificacy trait genetically selected in this species. Through a genome-wide association study with the prolificacy phenotype conducted in the French Noire du Velay sheep breed, we describe a novel mutation located in the regulatory region upstream of the *BMP15* gene on the X chromosome. This mutation increases litter size by +0.2 lamb per lambing at the heterozygous state, possibly through an inhibition of *BMP15* expression within the oocyte. Our findings suggest a novel kind of *BMP15* variant responsible for high prolificacy, in contrast to all other *BMP15* variants described so far in the coding sequence.

Introduction

There is now an accumulation of evidence that oocyte plays a central role in controlling the ovarian folliculogenesis, from the early stages up to ovulation. Among the local factors produced by the oocyte itself, members of the bone morphogenetic protein/growth and differentiation factor (BMP/GDF) family play an integral role in this control (Persani et al., 2015[1]). Among them, the most important are surely *BMP15* and *GDF9*. Knock-out mice models gave the first evidence of the importance of these two oocyte-derived factors acting individually as homodimers and/or through a synergistic co-operation to control the ovarian function (Elvin et al., 1999, Yan et al., 2001[2,3]).

In human also, a focus was done on *BMP15* and *GDF9* about their implication in various ovarian dysfunctions. Indeed, numerous heterozygous missense mutations have been identified in both genes associated with primary or secondary amenorrhea in different cohorts of women affected by primary ovarian insufficiency (POI) all over

the world. Particularly, the 10-fold higher prevalence of *BMP15* variants among patients with POI compared with the control population supports the causative role of these mutations (Persani et al., 2015[1]). Alteration of *BMP15* and *GDF9* were also searched in association with the polycystic ovary syndrome (PCOS). Here again several missense variants were discovered in both genes, but the pathogenic role of these mutations remains controversial in the etiology of this syndrome. However, several studies have reported an aberrant expression of *BMP15* and *GDF9* in the ovary of PCOS patients (Teixera Filho et al. 2002; Wei et al. 2014 [4,5]). Interestingly, some *BMP15* polymorphisms situated in the 5'UTR are significantly associated with the over response to recombinant FSH applied during assisted reproductive treatment and with the risk to develop an ovarian hyperstimulation syndrome (OHSS, Moron et al. 2006; Hanevik et al. 2011 [6,7]). Finally, polymorphisms in *BMP15* and *GDF9* genes were also searched in association with dizygotic twinning in human. If no convincing results were obtained for *BMP15*, some lost-of-function variants of *GDF9* were observed significantly more frequently in mothers of twins compared to the control population (Palmer et al. 2006; Simpson et al. 2014 [8,9]).

In parallel, the search for the genetic determinism of ovulation rate and prolificacy variability in sheep has also highlighted the crucial role of *BMP15* and *GDF9* by evidencing numerous independent loss-of-function mutations all altering the coding sequence of these two genes (Persani et al. 2015; Abdoli et al. 2016 [1,10]). Depending on the mutation and its hetero- or homozygous state, the phenotype controlled by these mutations in *BMP15* and *GDF9* goes from the early blockade of the folliculogenesis, and subsequent sterility, to an extraordinary increase of the ovulation rate (OR) and thus litter size (LS) of carrier ewes (Galloway et al., 2000; Hanrahan et al., 2004; Sylva et al., 2011; Demars et al., 2013 [11–14]). Thus, sheep exhibiting an extremely high

prolificacy are of great interest for identifying genes and mutations involved in molecular pathways controlling the ovarian function. These animal models have a double interest, in agriculture for the genetic improvement of the prolificacy, and in human clinic for providing valuable candidate genes in the genetic determinism of female infertility or subfertility, as described above.

The Noire du Velay (NV) population is a French local sheep breed mainly reared in the Haute-Loire and Loire departments. Ewes present naturally out-of-season breeding ability, very good maternal characteristics and a quite high prolificacy (mean LS=1.62 lamb per lambing). Large variation in LS has been observed in this breed and a recent genetic study has evidenced the segregation of an autosomal mutation named *FecL*^{L} controlling this trait (Chantepie et al. 2018[15]). This variant located in the intron 7 of the *B4GALNT2* gene and associated with its ectopic ovarian expression, was originally discovered in the Lacaune meat sheep breed, increasing OR and prolificacy (Drouilhet et al. 2013[16]). For the segregation study, more than 2700 NV ewes with LS records were genotyped at the *FecL* locus (Chantepie et al. 2018[15]). Surprisingly, the distribution of LS and the existence of high prolific ewes among the *FecL*^{L} non-carrriers have suggested the possible segregation of a second prolificacy major mutation in this population as already observed in the Lacaune breed carrying both *FecL*^{L} and *FecX*^{L} (Bodin et al. 2007, Drouilhet et al. 2013[16,17]). In order to validate this hypothesis, after specific genotyping excluding all other known mutations affecting OR and LS and segregating in French sheep populations, we have performed a genome-wide association study (GWAS) based on a case/control design. Completed by the whole genome sequencing of two finely chosen animals, we have identified a new regulatory variant called *FecX*^{N} affecting the oocyte-dependent expression of *BMP15* in association with increased prolificacy in sheep.

Results

Genetic association analyses

A first set of genomic DNA from 30 NV ewes without the *FecL*^{L} prolific allele at the *B4GALNT2* locus (LS records ranging from 2.00 to 3.00) was genotyped for already known mutations affecting sheep prolificacy at the 3 other loci, *BMPR1B*, *GDF9* and *BMP15*. Using specific RFLP assay (*BMPR1B*, Wilson et al. 2001[18]) or Sanger sequencing of coding parts (*GDF9* and *BMP15*, Talebi et al, 2018[19]), none of the known mutations were evidenced (data not shown). Thus, to establish the genetic determinism of the remaining LS variation in this population, 80 ewes were genotyped by Illumina Ovine SNP50 Genotyping Beadchip. The allele frequencies of the most highly prolific ewes (cases, n=40, mean LS=2.47) and lowly prolific ewes (controls, n=40, mean LS=1.23) were compared to identify loci associated with LS using GWAS according to the procedures described in the Materials and Methods. Finally, genotype data were obtained from 79 animals (39 cases, 40 controls). Six markers located on OARX were significantly associated with LS variation at the genome-wide level after Bonferroni correction (Fig 1A, Table 1). Importantly, at the chromosome-wide level, a cluster of 26 significant markers encompassed the location of the *BMP15* candidate gene (Fig 1B). In order to better characterize this locus on the X chromosome, we have determined for each individual the most likely linkage phase across 80 markers (10Mb) including the significant region. After haplotype clusterization, a specific segment of 3.5 Mb (50639087–54114793 bp, OARv3.1 genome assembly) was identified to be more frequent in highly prolific cases than in controls (f_{cases}= 0.51 vs. $f_{controls}$=0.37, *P*= 1.92E^{-11}, Chi-square test) (Fig 2). This identified segment contained the *BMP15* gene (50970938-50977454 bp, OARv3.1) well-known to play a crucial role in the ovarian

function and to be a target of numerous mutations in its coding region controlling prolificacy (Persani et al, 2015[1]).

Table 1. Markers significantly associated with litter size

SNP	Chromosome	Position[a]	MAF[b]	P_{Unadj}[c]	P_{Chrom}[d]	P_{Genome}[e]
OARX_51294776.1	OARX	53756339	0.28	$3.52E^{-11}$	$4.18E^{-08}$	$1.67E^{-06}$
s27837.1	OARX	53825247	0.30	$5.83E^{-10}$	$6.92E^{-07}$	$2.77E^{-05}$
s73460.1	OARX	53905939	0.44	$9.59E^{-10}$	$1.14E^{-06}$	$4.55E^{-05}$
s39212.1	OARX	53852735	0.44	$3.46E^{-09}$	$4.11E^{-06}$	$1.64E^{-04}$
OARX_52608221.1	OARX	52367253	0.41	$1.68E^{-08}$	$2.00E^{-05}$	$7.99E^{-04}$
s46003.1	OARX	80222479	0.44	$2.56E^{-07}$	$3.03E^{-04}$	$1.21E^{-02}$
OARX_55032299.1	OARX	48942926	0.30	$1.57E^{-06}$	$1.86E^{-03}$	NS
OARX_72164491.1	OARX	74388397	0.23	$1.71E^{-06}$	$2.03E^{-03}$	NS
OARX_49135019.1	OARX	42475099	0.35	$1.80E^{-06}$	$2.13E^{-03}$	NS
OARX_111306030.1	OARX	92041520	0.27	$2.25E^{-06}$	$2.67E^{-03}$	NS
OARX_102620828.1	OARX	82796975	0.41	$3.02E^{-06}$	$3.59E^{-03}$	NS
s31917.1	OARX	58202482	0.44	$3.12E^{-06}$	$3.70E^{-03}$	NS
OARX_72351736.1	OARX	74590448	0.19	$5.63E^{-06}$	$6.68E^{-03}$	NS
OARX_72263548.1	OARX	74498463	0.17	$6.45E^{-06}$	$7.66E^{-03}$	NS
OARX_49564109.1	OARX	42876169	0.49	$8.16E^{-06}$	$9.69E^{-03}$	NS
DU400878_520.1	OARX	73847207	0.27	$9.96E^{-06}$	$1.18E^{-02}$	NS
s54281.1	OARX	58993959	0.24	$1.20E^{-05}$	$1.42E^{-02}$	NS
s05229.1	OARX	58346644	0.46	$1.71E^{-05}$	$2.03E^{-02}$	NS
s27938.1	OARX	53275559	0.35	$1.77E^{-05}$	$2.10E^{-02}$	NS
OARX_54104393.1	OARX	49870983	0.36	$2.03E^{-05}$	$2.41E^{-02}$	NS
OARX_72236232.1	OARX	74464263	0.18	$2.23E^{-05}$	$2.64E^{-02}$	NS
OARX_111349974.1	OARX	92085118	0.18	$2.23E^{-05}$	$2.65E^{-02}$	NS
OARX_53703822.1	OARX	51193144	0.35	$2.59E^{-05}$	$3.08E^{-02}$	NS
OARX_43227227.1	OARX	36235514	0.32	$3.41E^{-05}$	$4.04E^{-02}$	NS
OARX_51842287.1	OARX	53162079	0.49	$3.43E^{-05}$	$4.08E^{-02}$	NS
OARX_102654502.1	OARX	82837158	0.20	$3.65E^{-05}$	$4.33E^{-02}$	NS

[a] Position of markers are based on the OARv3.1 assembly in bp.
[b] MAF, minor allele frequency.
[c] P_{Unadj} corresponds to exact unadjusted p-value for the Fisher's test.
[d] P_{Chrom} corresponds to p-value after chromosome-wide Bonferroni correction.
[e] P_{Genome} corresponds to p-value after genome-wide Bonferroni correction (NS, non-significant).

Characterization of the mutation

While the *BMP15* gene could be considered as a positional and functional candidate gene, no mutation was evidenced by Sanger sequencing of the *BMP15* coding regions of the most prolific ewes studied. In order to find the potential causal mutation, we sequenced the whole genome of two finely chosen ewes based on the shortest haplotype within the region (homozygous reference vs. homozygous variant) and their opposite extreme phenotypes (LS 1.1 vs. 2.8).

Variant search analysis and annotations through GATK toolkit was limited to the OARX: 50639087-54114793 region. We detected 60 SNPs and 90 small insertions and deletions (INDELs) with quality score >30 (S1 Table). Among them, we particularly focused on the 85 variants located within annotated genes (upstream, exon, intron, splice acceptor or donor, and downstream localization). After filtering these 85 variants for allele sharing with other breeds based on SheepGenome DB (http://sheepgenomesdb.org/) and 68 publicly available domestic sheep genomes (International Sheep Genomics Consortium; http://www.sheephapmap.org/), none of them were removed, all being NV breed specific. Finally, and based on prolificacy gene knowledge, we were particularly interested in one SNP (T>A) identified in the upstream region of the *BMP15* gene at position 50977717 on OARX v3.1. We then developed a RFLP assay to specifically genotype for this polymorphism. Among the 79 animals of the GWAS, 31 ewes were heterozygous and 6 homozygous for the A variant allele. As shown in Table 2, most of the A carrier ewes were in the highly prolific Case group (34 among 39), while only 3 set in the Control group. When associating the LS performance of the 79 ewes to their genotype at the OARX: 50977717T>A SNP, the A non-carriers exhibited a mean LS of 1.36, heterozygous T/A a mean LS of 2.32 and homozygous A/A a mean LS of 2.73 indicating that the A allele of this polymorphism was strongly associated with increased LS in NV (T/A or A/A vs. T/T, $P<1E^{-3}$, one-way ANOVA). Furthermore, this polymorphism appears in total linkage disequilibrium with the six more significant markers from the GWAS analysis (Fig 3). Genotype information at the OARX: 50977717T>A locus was introduced in the GWAS analysis. This SNP appeared as the most significant marker associated to the prolificacy phenotype ($P_{\text{unadjusted}}=1.93E^{-11}$, $P_{\text{Chromosome-wide corrected}}=1.62E^{-14}$ and $P_{\text{Genome-wide corrected}}=9.13E^{-07}$) suggesting that it could be the causal mutation (S3 Figure). In accordance with the *Fec*

167 gene nomenclature, the mutant allele identified upstream of the *BMP15* gene in NV
168 sheep was named *FecX*^{N}.

169

Table 2. Distribution of OARX:50977717T>A SNP genotypes and associated LS in case and control groups

Group		TT	TA	AA
Low prolific control	n=	37	3	
	Raw mean LS	1.22	1.39	
Highly prolific case	n=	5	28	6
	Raw mean LS	2.41	2.43	2.73
Total		**42**	**31**	**6**

170

171 As described for other prolific alleles such as *FecB*^{B}, *FecX*^{G}, *FecG*^{H}, *FecX*^{Gr} and *FecL*^{L},
172 a given mutation can segregate in several sheep populations (Davis et al., 2002;
173 Mullen et al., 2013, Chantepie et al. 2018, Ben Jemaa et al. 2019[15,20–22]). We have
174 tested the *FecX*^{N} allele presence in a diversity of 26 sheep breeds representing 725
175 animals (Rochus et al. 2018[23]). Among the breeds tested, the *FecX*^{N} genotyping has
176 confirmed the segregation of this mutation in NV breed and revealed its presence in
177 the Blanche du Massif Central (BMC) and Lacaune breeds (Table 3). Additionally, the
178 *FecX*^{N} variant was absent from the Ensembl variant database
179 (http://www.ensembl.org) compiling information from i) dbSNP, ii) whole genome
180 sequencing information from the NextGen project (180 animals from various Iranian
181 and Moroccan breeds) and iii) the International Sheep Genome Consortium (551
182 animals from 39 breeds all over the world).

Table 3. *FecX*^{N} genotype distribution from a diversity panel of French ovine breeds

Breed	Total	Genotype		Breed	Total	Genotype	
		+/+ (+/Y)	*N/+ (N/Y)*[a]			*+/+ (+/Y)*	*N/+ (N/Y)*[a]
Berrichon du Cher	29	29		Mourerous	26	26	
Blanche Massif Central	31	27	4	Mouton Vendéen	30	30	
Causse du Lot	32	32		Noire du Velay	28	26	2
Charmoise	31	31		Préalpes du sud	27	27	
Charollais	29	29		Rava	29	29	
Corse	30	30		Romane	29	29	
Ile de France	28	28		Romanov	26	26	
Lacaune (meat)	42	40	2	Rouge de l'Ouest	28	28	
Lacaune (dairy)	40	40		Roussin	30	30	
Limousine	30	30		Suffolk	20	20	
Manech tête rousse	29	29		Tarasconnaise	32	32	
Martinik	22	22		Texel	21	21	
Merinos d'Arles	26	26					
				TOTAL	725	717	8

[a]: *+/+* or *N/+* females, *N/Y* or *N/+* hemizygous males.

183

184 ## *FecX*^{N} genotype frequency and effect on prolificacy

185 Large cohorts of ewes, chosen at random, were genotyped in order to accurately
186 estimate the allele frequencies in the NV and the BMC populations (Table 4). The
187 frequency of the N prolific allele at the *FecX* locus was similar in both populations, 0.11
188 and 0.10, with a distribution of 19.4% and 17.6% heterozygous, 1.5 % and 1%
189 homozygous carriers in NV and BMC, respectively. The genotype frequencies were
190 consistent with the Hardy Weinberg equilibrium (HWE) in both breed (NV *P*= 0.28 and
191 BMC *P*= 0.76).

Table 4. *FecX*^{N} frequencies and effects on LS in NV and BMC breeds.

		Breed						
		NV (n=2323)				BMC (n=2456)		
FecX genotype		*+/+*	*N/+*	*N/N*		*+/+*	*N/+*	*N/N*
Number of animals		1839	450	34		1999	432	25
Frequency (%)		79.2	19.4	1.5		81.3	17.6	1
Raw mean LS		1.66	1.93	2.45		1.56	1.79	1.87
ANOVA solutions [a]								
	+/+	0.00	0.22	0.65	NE	0.00	0.18	0.30
FecL genotype	*L/+*	0.41	0.58	0.56	IE	0.23	0.35	0.17
	L/L	0.72	/	/				

[a.] estimated increased LS compared to *L* and *N* non-carriers from the linear mixed models with interaction between genotypes at the *FecX* and *FecL* loci for the Noire du Velay (NV) breed and between *FecX* genotypes and estrus type (NE= natural estrus; IE = Induced estrus) for the Blanche du Massif Central (BMC) breed.

192

193 Based on the raw mean LS observations, the *FecX*^{N} carrier ewes clearly exhibited
194 increased LS compared to non-carriers in both populations (Table 4). The L prolific
195 allele at the *FecL* locus is also segregating in NV (Chantepie et al. 2018 [15]). Results
196 of the linear mixed model showed that for the NV breed, one copy of the *FecX*^{N} allele
197 significantly increased LS by +0.22 and two copies increased LS by +0.65, while a
198 single copy of the *FecL*^{L} allele increased LS by +0.41 and two copies by +0.72. Based
199 on the 80 ewes genotyped heterozygous at both loci it appeared that the effect of
200 *FecX*^{N} and *FecL*^{L} on LS was not fully additive, the expected LS being significantly
201 slightly reduced by -0.05 (0.58 instead of 0.63) (Fig. 4A). For the BMC population,
202 compared to *FecX*$^{+}$*/FecX*$^{+}$ ewes, *FecX*^{N}*/FecX*$^{+}$ exhibited increased LS by +0.18 and
203 *FecX*^{N}*/FecX*^{N} by +0.30 under natural estrus (Fig. 4B). The use of PMSG for estrus
204 synchronization increased LS significantly among *FecX*$^{+}$*/FecX*$^{+}$ ewes (+0.23) and
205 *FecX*^{N}*/FecX*$^{+}$ ewes (+0.18) while the effect on *FecX*^{N}*/FecX*^{N} ewes was negative (-
206 0.13). The combined effect of the first copy of the *FecX*^{N} allele and the use of PMSG

207 treatment was not fully additive, the interaction being significant although low (0.35
208 instead of 0.41) (Fig. 4B).

209 **Functional effects of the *FecX^N* mutation**

210 As described above, *FecX^N* is located upstream of the coding region of the *BMP15*
211 gene when referencing to the ovine genome v3.1 (ensembl.org) or v4.0
212 (ncbi.nlm.nih.gov). In both versions of the ovine genome, the *BMP15* gene annotation
213 begins at the ATG start site and *FecX^N* is located -290pb upstream, possibly in the
214 5'UTR and/or the proximal promoter region. As a first approach, we took advantage of
215 RNA sequencing data from ovine oocytes publicly available at EMBL-EBI (Bonnet et
216 al., 2013[24]). After reads mapping against the ovine genome (v3.1) using STAR2
217 aligner within the Galaxy pipeline and visualization with Integrative Genome Viewer
218 (IGV), the Fig 5 shows the location of *FecX^N* within the possible 5'UTR of the *BMP15*
219 gene when expressed in the oocyte. Consequently, we have first tested the potential
220 functional impact of *FecX^N* on the *in vitro* stability and translatability of the *BMP15*
221 mRNA. Thus, the reference (T, *FecX$^+$*) and variant (A, *FecX^N*) forms of the ovine
222 BMP15 cDNA (-297, +1183 referring to ATG start codon) were cloned in a pGEM-T
223 vector for subsequent *in vitro* T7 promoter-dependent transcription/translation
224 experiment using reticulocyte lysate solution. As shown in Fig 6, the western blotting
225 of the BMP15 proteins produced from both forms and their chemiluminescent
226 quantification revealed that the *FecX^N* mutation had no significant impact on the overall
227 stability and translatability of the *BMP15* mRNA in this condition.

228 As a second hypothesis, we have tested the *FecX^N* impact on the *BMP15* promoter
229 activity. Two promoter regions were tested ([-743,-11] bp and [-443,-102] bp referring
230 to ATG start codon) cloned in front of the luciferase reporter gene and transiently
231 expressed in CHO cells cultured *in vitro*. As shown by the luciferase assays (Fig 7),

the *FecX*^{N} variant was able to significantly reduce the luciferase activity in the context of both the long or the short BMP15 promoters, indicating the possible inhibitory impact of *FecX*^{N} on *BMP15* gene expression. To go further with this hypothesis, *in vivo* BMP15 gene expression was measured directly on isolated oocytes pools from NV and BMC homozygous ewe carriers and non-carriers of the *FecX*^{N} allele. Real-time qPCR experiments revealed a tendency of the *BMP15* expression to be decreased by 2-fold (*P*=0.17, genotype effect, two-way ANOVA) in the oocytes of *FecX*^{N} carriers despite a large inter-animal variability. In contrast, the expression of the second oocyte-specific prolificacy major gene *GDF9* seemed unaffected (Fig 8).

Discussion

The present study identified the g.50977717T>A variant on the ovine chromosome X upstream of the *BMP15* gene as the most likely causative mutation for the increased prolificacy of the NV ewes. The highly significant genetic association with the extreme LS phenotype, the significant effect of the A variant on increasing prolificacy by +0.2 lamb per lambing in a large set of NV ewes, also found in the BMC genetic background, and the demonstrated action on *BMP15* transcriptional activity all support the causality of this mutation named *FecX*^{N}.

The *BMP15* gene is at the top of the list of candidate genes controlling the ovarian function, ovulation rate and thus prolificacy in the ovine species, with nine independent causal mutations identified out of the sixteen already known. Indeed, 7 SNPs and 2 small INDELs all within the open reading frame were evidenced affecting the BMP15 function. Among these mutations, 2 SNPs and the 2 INDELS impaired the protein production either by generating premature stop codon (*FecX*^{H}, Galloway et al. 2000;

256 *FecX*^{G}, Hanrahan et al. 2004 [11,12]) or by breaking the reading frame (*FecX*^{R},
257 Martinez-Royo et al. 2008; *FecX*^{Bar}, Lassoued et al. 2017[25,26]). The 5 other SNPs
258 generate non-conservative amino acid substitutions all leading to a loss of function of
259 BMP15 ranging from inhibited protein production (*FecX*^{L}, Bodin et al. 2007 [17]),
260 impaired interaction with GDF9 (*FecX*^{I} and *FecX*^{B}, Liao et al. 2004 [27]), to altered cell
261 signalling activity (*FecX*^{Gr} and FecX^{O}, Demars et al. 2013 [14]). In contrast with the 9
262 mutations described above, the *FecX*^{N} variant evidenced in the present study is not
263 located in the open reading frame of *BMP15* and does not alter the protein sequence.
264 However, no other polymorphism genetically linked to *FecX*^{N} was found in the *BMP15*
265 coding sequence when checked by whole genome or local Sanger sequencing of the
266 *BMP15* gene from *FecX*^{N} carrier animals. Of course, this does not rule out the
267 possibility of a polymorphism lying in another gene nearby with a still unknown role in
268 the ovarian function and prolificacy. Nevertheless, we did not find any polymorphism
269 (SNP and INDEL) altering the coding sequence of genes annotated in the significantly
270 LS-associated genetic region of 3.5Mb on OARX (S1 Table), leaving *BMP15* as the
271 most obvious candidate.

272 Whatever the version of the ovine reference genome (Oar_v3.1, Oar_v4.0 or even the
273 last Oar_rambouillet_v1.0) the annotation of the *BMP15* gene always starts at the ATG
274 initiating codon. Using publicly available transcriptome data from ovine oocytes
275 RNAseq analysis, we were able to show that *FecX*^{N} located 290bp upstream of *BMP15*
276 could stand in its 5'UTR region. From our *in vitro* functional analyses, *FecX*^{N} was not
277 demonstrated to influence the translatability of the *BMP15* mRNA, but in the contrary
278 it was shown to decrease the *BMP15* promoter activity. Little is known about
279 transcription factors able to regulate *BMP15* expression. Several regulatory elements
280 were evidenced in the pig *BMP15* promoter hosting consensus binding sites for LHX8,

NOBOX and PITX1 transcription factors. However, only LHX8 was demonstrated as functionally activating the porcine BMP15 promoter activity (Wan et al. 2015 [28]). In human, a regulatory mutation in the 5'UTR of *BMP15* (c.-9C>G) was associated to non-syndromic premature ovarian failure (Dixit et al. 2006 [29]), but also to iatrogenic ovarian hyperstimulation syndrome (Moron et al. 2006 [6]). This mutation was shown to enhance the fixation of the PITX1 factor transactivating the BMP15 promoter (Fonseca et al. 2014 [30]). However, the *FecX*^{N} position does not fit with the syntenic location of porcine LHX8 and human PITX1 binding sites on the ovine *BMP15* promoter. Using the MatInspector promoter analysis tool (Genomatix), we were only able to hypothesize an alteration by *FecX*^{N} of a putative TATA-box like sequence (TTAAATA >TTATATA). Unfortunately, our electromobility shift assay attempts using CHO nuclear extracts failed to demonstrate the binding of any factor at the *FecX*^{N} position, preventing us from defining the precise molecular mechanism by which *FecX*^{N} decreases the *BMP15* promoter activity.

The inhibition of the promoter activity combined with the apparent decreased of *BMP15* mRNA accumulation in homozygous *FecX*^{N}/*FecX*^{N} oocytes seem to confirm the transcriptional regulatory role of *FecX*^{N}. However, the moment we have chosen during the follicular phase of the late folliculogenesis for the comparative analysis between *FecX*$^{+}$ and *FecX*^{N} oocytes from antral follicles could not be optimal to visualize a highly significant differential expression of *BMP15*. The *BMP15* gene expression in ovine oocytes begins during the primary stage of follicular development and its expression increases up to the antral stages (McNatty et al. 2005; Bonnet et al. 2011 [31,32]). Moreover, the streak ovaries phenotype of infertile ewes carrying homozygous mutations in *BMP15* have evidenced its crucial role in controlling the primary to secondary follicle transition (Galloway et al. 2000; Bodin et al. 2007; Lassoued et al.

306 2017 [11,17,26]). Consequently, it would certainly be appropriate to follow the *BMP15*
307 expression in *FecX*^{N} carrier ewes from these early stages of folliculogenesis to better
308 decipher the mutation impact on ovarian physiology. Nevertheless, the fact that *FecX*^{N}
309 inhibits the *BMP15* gene expression fits well with the physiological and molecular
310 models associating BMP system loss-of-function and increased sheep prolificacy
311 (Fabre et al., 2006, Demars et al. 2013[14,33]).

312 One copy of *FecX*^{N} allele significantly increased by +0.30 to +0.50 the raw mean LS of
313 NV ewes. When corrected for different environmental effects and more particularly for
314 the genotype at the *FecL* locus, the estimated effect of *FecX*^{N} on LS was +0.22 lamb
315 per lambing for the first copy and +0.43 for the second copy. This effect was in the
316 range of already known prolific alleles in various sheep breeds (Jansson, 2014 [34]).
317 The effect of *FecX*^{N} on LS seems independent of the genetic background. Indeed, the
318 estimated positive effect of *FecX*^{N} on prolificacy was confirmed in BMC breed with
319 +0.18 lamb per lambing based on natural estrus. Moreover, the same robust effect was
320 observed even in the presence of PMSG for synchronizing the estrus cycles preceding
321 the lambing. The same observation is made for other mutations controlling sheep
322 prolificacy. For instance, the *FecL*^{L} allele exhibited a similar effect on LS in NV (+0.41,
323 present study; +0.42, Chantepie et al. 2018 [15]), Lacaune (+0.47, Martin et al. 2014
324 [35]) and D'man (+0.30, Ben Jemaa et al. 2018 [22]), and this was also observed for
325 the *FecB*^{B} allele introgressed in several populations (Kumar et al. 2008 [36]).

326 By genotyping a diversity panel, we also evidenced the presence of 2 *FecX*^{N} carrier
327 animals in the Lacaune meat strain which will require further genotyping of numerous
328 animals. If this is confirmed, the Lacaune meat breed will be another population, as
329 Belclare, where 3 different natural prolific mutations are segregating (Hanrahan et al.
330 2004; Bodin et al. 2007; Drouilhet et al. 2013 [12,16,17]). The presence of *FecL*^{L} in

331 both NV and Lacaune, and the presence of *FecX*^{N} in NV, BMC and Lacaune, also
332 raises the question of the origin of these mutations. From population structure analysis,
333 it was shown that NV, BMC and Lacaune shared the same origin within the European
334 southern sheep populations that may explain the segregation of the same mutations
335 in these populations (Rochus et al., 2018 [23]).

336 In conclusion, through a case/control GWAS strategy and genome sequencing, we
337 have identified in the NV breed a second prolific mutation named *FecX*^{N} affecting the
338 expression of the *BMP15* gene, a well-known candidate gene controlling OR and LS
339 in sheep. This work confirms the relevance of the whole genome approaches to
340 decipher the genetic determinism of the prolificacy trait. Homozygous *FecX*^{N}/*FecX*^{N}
341 animals were still hyperprolific as already observed for *FecX*^{Gr} and *FecX*^{O}, but in
342 contrast with sterile animals observed for the 7 other *FecX* homozygous variants in
343 *BMP15*. As an upstream regulatory mutation, *FecX*^{N} also contrasts with these 9 other
344 prolific causal mutations all evidenced in the coding part of *BMP15* and altering the
345 protein function. Thanks to this new sheep model, the genetic etiology of ovarian
346 pathologies in women could be improved by searching polymorphisms, not only in the
347 coding region, but also in the regulatory parts driving the *BMP15* expression within the
348 oocyte.

349

350 ## Materials and Methods

351 ### Animals

352 Ewes (*Ovis aries*) from the NV breed (n=2266) were genotyped on blood DNA at the
353 *FecL* locus as already described (Chantepie et al. 2018[15]). In order to test the
354 hypothesis of the segregation of a second major mutation controlling LS in this breed,

a first set of 80 ewes with at least 5 LS records (mean LS=1.84; ranging from 1.00 to 3.50) were selected among the *FecL*$^{+}$ homozygous genotype (n=2151, mean LS=1.58). Subsequently, for NV breed, the effect of the *FecX*^{N} mutation on LS was estimated on 2252 ewes, considering the genotype at the *FecL* locus. The presence of the *FecX*^{N} mutation in other breeds was checked on a diversity panel of 725 animals from 26 French sheep breeds (Rochus et al. 2018[23]; Table 3). For the BMC population, the effect of the *FecX*^{N} mutation on LS was estimated on 2456 ewes. For gene expression analysis, 10 homozygous ewes at the *FecX*^{N} locus (5 carriers and 5 non-carriers of the *N* allele) were bought from private breeders (6 NV and 4 BMC) and reared at INRA experimental facility (agreement number: D3142901). All experimental procedures were approved (approval number 01171.02) by the French Ministry of Teaching and Scientific Research and local ethical committee C2EA-115 (Science and Animal Health) in accordance with the European Union Directive 2010/63/EU on the protection of animals used for scientific purposes.

Biological samples

All blood sampling from the numerous sheep breeds studied were collected from jugular vein (5 ml per animal) by Venoject system with EDTA and directly stored at -20°C for further use. Part of these blood samples (GWAS and diversity panel) was used for extraction of genomic DNA as described (Bodin et al. 2007[17]). All other samples were used for direct genotyping on whole blood without DNA purification (Chantepie et al. 2018[15]).

For ovary collection and oocyte isolation, the estrus cycles of all adult NV and BMC ewes were synchronized with intravaginal sponges impregnated with flugestone acetate (FGA, 30 mg, CEVA) for 14 days. Ovaries were collected at slaughtering during the follicular phase 36h after FGA sponge removal. Cumulus-oocyte complexes (COC)

were immediately recovered from all visible 1-3mm follicles by aspiration using a 1ml syringe with a 26G needle and placed in McCoy's 5A culture medium (Sigma-Aldrich). COC were mechanically dissociated by several pipetting and washing cycles in 150µl drops of McCoy's 5A medium and finally, denuded oocytes devoid of granulosa cells were recovered in 1X PBS. Only intact oocytes with a good homogeneity of the cytoplasm were grouped to obtain two to three pools of 5 oocytes per animal and stored at -80°C before RNA extraction.

Genotyping analyses

The *FecL*^{L} mutation (OAR11:36938224T>A, NC_019468) was genotyped directly on whole blood samples by the KAPA-KASP assay as already described (Chantepie et al. 2018[15]). As a prerequisite before GWAS, a set of 30 high prolific *FecL*$^{+}$*/FecL*$^{+}$ ewes were controlled for the absence of other evidenced major mutations affecting sheep prolificacy in French populations. Using the same KAPA-KASP assay, *FecX*^{L} and *FecX*^{Gr} alleles in *BMP15* were genotyped as described (Chantepie et al. 2018[15]). *FecB*^{B} in the exon 7 of the *BMPR1B* gene (OAR6:29382188A>G, NC_019463.1) was genotyped using forced restriction fragment length polymorphism (RFLP) as described by Wilson et al. (2001)[18].

The whole genome genotyping was performed on 80 ovine genomic DNA using the OvineSNP50 Genotyping Beadchip from Illumina according to the manufacturer's protocol at the Laboratoire d'Analyses Génétiques pour les Espèces Animales, (LABOGENA, Jouy en Josas, France; www.labogena.fr). From the dataset, individuals with a call rate <0.98 were excluded. SNP exclusion thresholds were: call frequency <0.95, and minor allele frequency (MAF) <0.01; or a significant deviation from Hardy-Weinberg equilibrium (HWE) in the controls ($p<1.10^{-6}$). Non-polymorphic SNP positions and markers with no position on the OARv3.1 reference genome map were

405 also discarded. Finally, from the available design of 54241 SNPs available on the
406 Illumina OvineSNP50 Beadchip and 80 selected NV ewes, the final dataset was
407 reduced to 47446 SNPs analyzed in 79 individuals.

408 The *FecX*^{N} mutation (OARX: 50977717T>A, NC_019484) was genotyped by a RFLP
409 analysis using the Mse1 restriction enzyme (New England Biolabs) after a first step of
410 Terra PCR Direct Polymerase Mix amplification (Takara) using one µl sample of total
411 blood. The accuracy of the *FecX*^{N} RFLP genotyping was controlled by Sanger
412 sequencing on few samples using the same amplification primers.

413 PCR amplifications were conducted independently for each locus studied on an
414 ABI2400 thermocycler (Applied Biosystems) with the following conditions: 5min at
415 94°C, 32 cycles of 30s at the specific melting temperature, 30s at 72 °C and 30s at 94
416 °C, followed by 5 min at 72 °C. The primers used in this study are listed in S2 Table.

417 **Whole genome sequencing (WGS) analysis**

418 DNA sequencing libraries were constructed from 1 µg of genomic DNA using TruSeq
419 DNA PCR-free Library Prep kit (Illumina). Sequencing was run on an Illumina HiSeq
420 2500 apparatus using a paired-end read length of 2x150 pb with the Illumina Reagent
421 Kits as already described (Demars et al. 2017[37]). WGS was performed at the
422 Genotoul-GeT core facility (INRA Toulouse, https://get.genotoul.fr). The raw reads of
423 Illumina DNA sequencing were preprocessed by removing adapter sequences. After
424 quality control, the FastQ files and metadata were submitted to the European
425 Nucleotide Archive (ENA) at EMBL-EBI (accession number PRJEB35553). Reads
426 mapping and variants calling were performed using the local instance of Galaxy
427 (https://galaxyproject.org) at the Toulouse Midi-Pyrénées bioinformatics platform
428 (http://sigenae-workbench.toulouse.inra.fr). The cleaned paired reads were combined

and mapped against the ovine genome assembly (Oar_v3.1.86) using BWA-MEM (Galaxy version 0.7.17.1). The resulting BAM files were sorted using Samtools_sort (Galaxy version 1.0.0). Sorted and indexed BAM files were visualized through Integrative Genome Viewer, IGV software version 2.4.10 (Robinson et al. 2011[38]). A GFF3 annotation file was obtained from Ensembl (Ovis_aries.Oar_V3.1.78). We applied GATK version 3.5-0 to performed SNP and InDel discovery and genotyping across the two samples simultaneously using standard filtering parameters according to GATK Best Practices recommendations (DePristo et al. 2011; Van der Auwera et al. 2013[39,40]). Variants effect and annotation were realized by SNPEff version 4.1 and filtering of interesting variants was performed using the SNPSift tool.

In vitro transcription and translation of BMP15

The full-length cDNA of ovine BMP15 (1480bp [-297,+1183] referring to ATG start codon) with or without the *FecX*^{N} mutation was generated from oocyte-derived RNA after a reverse transcription (RT) step (described in RNA extraction and RT paragraph, primers are listed in S2 Table). The resulting PCR products were inserted by TA cloning into pGEM-T Easy plasmid (Promega) possessing T7 and SP6 promoters. The orientation of insertion and exclusion of unexpected PCR-induced mutations were controlled by Sanger sequencing.

In vitro transcription and translation were realized from 500ng of cDNA pGEM-T construct using the TnT T7 Quick Coupled Transcription/Translation kit (Promega) and Transcend Biotin-Lysine-tRNA following the manufacturer protocol. Reactions for each construct were run in duplicate in 6 independent TnT experiments. The resulting BMP15 protein was revealed using Transcend non-radioactive translation detection system with chemiluminescent method (Promega) after reducing SDS-PAGE on a gradient (4-15%) polyacrylamide gel (Promega) and transfer onto nitrocellulose

454 membrane. Chemiluminescent signal was capture by a ChemiDoc MP imaging system
455 and images were analyzed with the Image Lab Software (Bio-Rad).

456 **BMP15 promoter activity**

457 The promoter sequence of the ovine *BMP15* gene was amplified by PCR on genomic
458 DNA from both homozygous *FecX*$^{+}$ and homozygous *FecX*^{N} ewes. Two sizes of
459 fragments were generated for cloning the *BMP15* promoter in front of the luciferase
460 (Luc) reporter gene, a long (lg) form of 732 bp ([-743,-11] bp referring to ATG start
461 codon) and a short (sh) form of 341bp ([-443,-102] bp). The PCR products were
462 engineered for digestion using Kpn1 and Hind3 restriction enzymes (New England
463 Biolabs) and inserted into the pGL4.23 vector (Promega). The four resulting constructs
464 (lgBMP15^{+}-Luc; lgBMP15^{N}-Luc; shBMP15^{+}-Luc; shBMP15^{N}-Luc) were controlled by
465 Sanger sequencing. Primers used to generate these constructs are listed in S2 Table.
466 Twenty-four hours after seeding (3.10^{4} cells/well, 24 wells plate), CHO (Chinese
467 Hamster Ovary) cells were transfected using Lipofectamine 3000 (Invitrogen) with
468 500ng/well of pGL4.23 constructs either empty or containing *BMP15* promoter
469 fragment. Forty-eight hours after transfection, cells were lysed and assayed for
470 luciferase activity (Luciferase reporter assay kit, Promega). Luminescence in relative
471 light units (RLU) was measured by a Glomax microplate reader (Promega). Each
472 construct was assayed in triplicate in 6 independent transfection experiments.

473

474 **RNA extraction, reverse transcription and quantitative PCR**

475 Total RNA from pools of 5 oocytes were extracted using the Nucleospin RNA XS kit
476 according to the manufacturer's protocol (Macherey-Nagel) and including a DNase1
477 treatment. The low quantity of RNA recovered did not allow quantification. So, the

equivalent of 1.25 oocyte was reverse-transcribed using SuperScript II reverse transcriptase (Invitrogen) and anchored oligo(dT)22 primer (1µl at 10 µM). Primer design using Beacon designer 8.20 (Premier Biosoft), SYBR green real-time PCR cycling conditions using QuantStudio 6 Flex Real-Time PCR system (ThermoFisher Scientific) and amplification efficiency calculation ($E=e^{(-1/slope)}$) were as already described in Talebi et al. (2018)[19]. Primer sequences, amplicon length and amplification efficiency are listed in S2 Table. RNA transcript abundance was quantified using the ΔCt method with the mean expression of GAPDH and SDHA as internal references and following the formula $R=[E_{ref}^{Ct\ ref}/E_{target}^{Ct\ target}]$. The two reference genes were validated by the Bestkeeper algorithm (Pfaffl et al., 2004)[41].

Data analysis

Single-marker association analyses were conducted using a Fisher's exact test and a Bonferroni correction has been applied to check for significance levels. The chromosome-wide and genome-wide values have been established as mentioned by Balding et al. 2006 [42] . Statistical analyses were done using PLINK1.9 software under a case/control design [43]. Among the 79 datasets of 47446 SNPs analyzed, the LS trait was considered as case when mean LS ≥ 2.18 (n=39) and control when LS ≤ 1.45 (n=40). Haplotypic association analysis on X chromosome were performed using FastPhase software [44]. Empirical significance levels were calculated using maximum statistic permutation approach (max (T), n=1000).

Allele effect on LS was estimated in NV and BMC breeds on data extracted from the French national database for genetic evaluation and research managed by the Institut de l'Elevage (French Livestock Production Institute) and the CTIG (Centre de Traitement de l'Information Génétique, Jouy-en-Josas, France). Only females born after 2000 were retained (27 754 NV ewes with 122 110 LS records and 110 848 BMC

ewes with 461 405 LS records) with their pedigree over 5 generations. *FecX*^{N} genotype effect on the subset of 79 case/control animals was assessed by one-way ANOVA, follow by Newman-Keuls post-hoc test. For the large animal cohort analyses, the linear mixed models used were as similar as possible to those of the national genetic evaluation system (Poivey et al. 1995 [45]). In the present study, the following fixed effects were considered: i) the genotype at the *FecX* locus, ii) the month of birth (12 levels) iii) a physiological status effect combining parity, age at first lambing, rearing mode and postpartum interval (44 levels) and iv) a combination of the flock year and season effect. Two random effects were added to the model: a permanent environmental effect and an animal additive genetic effect. Moreover, an additional fixed effect of the reproduction type was considered for the BMC breed for which some hormonal treatments are used each year (87% and 13% after natural and induced estrus in the data set). For the NV breed, since the *FecL*^{L} allele is also segregating in the population (Chantepie et al. 2018 [15]), the effect of the genotype at this locus (2252 known and 25502 unknown genotypes) as well as its interaction with the genotypes at the *FecX* locus were considered. All these models were fitted using the ASReml software (Gilmour et al. 2009 [46]).

The comparison between *FecX* alleles for BMP15 protein quantification was analyzed using Student's t-test, using Welch's correction. For reporter luciferase assays, differences between constructs were analyzed by one-way ANOVA followed by Newman-Keuls post-hoc test. QPCR data for *BMP15* and *GDF9* expression in oocytes were analyzed by two-way ANOVA considering genotype and breed effects. $P > 0.05$ was considered as not significant. All these experimental data are presented as means ± SEM and were analyzed using Prism 6 (GraphPad Software Inc.).

Acknowledgments

We thank Claire Chantaduc, Didier Cathalan and Kévin Chile from ROM Sélection managing the NV and BMC populations, for their precious help in the planning of blood sampling. We are grateful to the breeders who made their animals available for this study. LC was supported by a PhD grant co-funded by APIS-GENE through the Proligen project and the European Funds for Regional Development (FEDER) through the Interreg POCTEFA programme in the framework of the PIRINNOVI project (EFA103/15). Part of the NV sampling was supported by the DEGERAM project co-funded by the FEDER Massif Central, the Régions: Aquitaine, Midi- Pyrénées, Limousin and Auvergne; and the French government.

References

1. Persani L, Rossetti R, Di Pasquale E, Cacciatore C, Fabre S. The fundamental role of bone morphogenetic protein 15 in ovarian function and its involvement in female fertility disorders. Hum Reprod Update. 2014;20: 869–883. doi:10.1093/humupd/dmu036

2. Elvin JA, Clark AT, Wang P, Wolfman NM, Matzuk MM. Paracrine Actions Of Growth Differentiation Factor-9 in the Mammalian Ovary. Mol Endocrinol. 1999;13: 1035–1048. doi:10.1210/mend.13.6.0310

3. Yan C, Wang P, DeMayo J, DeMayo FJ, Elvin JA, Carino C, et al. Synergistic roles of bone morphogenetic protein 15 and growth differentiation factor 9 in ovarian function. Mol Endocrinol Baltim Md. 2001;15: 854–866. doi:10.1210/mend.15.6.0662

4. Teixeira Filho FL, Baracat EC, Lee TH, Suh CS, Matsui M, Chang RJ, et al. Aberrant expression of growth differentiation factor-9 in oocytes of women with polycystic ovary syndrome. J Clin Endocrinol Metab. 2002;87: 1337–1344. doi:10.1210/jcem.87.3.8316

5. Wei L-N, Huang R, Li L-L, Fang C, Li Y, Liang X-Y. Reduced and delayed expression of GDF9 and BMP15 in ovarian tissues from women with polycystic ovary syndrome. J Assist Reprod Genet. 2014;31: 1483–1490. doi:10.1007/s10815-014-0319-8

6. Morón FJ, de Castro F, Royo JL, Montoro L, Mira E, Sáez ME, et al. Bone morphogenetic protein 15 (BMP15) alleles predict over-response to recombinant follicle stimulation hormone and iatrogenic ovarian hyperstimulation syndrome (OHSS). Pharmacogenet Genomics. 2006;16: 485–495. doi:10.1097/01.fpc.0000215073.44589.96

7. Hanevik HI, Hilmarsen HT, Skjelbred CF, Tanbo T, Kahn JA. A single nucleotide polymorphism in BMP15 is associated with high response to ovarian stimulation. Reprod Biomed Online. 2011;23: 97–104. doi:10.1016/j.rbmo.2011.02.015

8. Palmer JS, Zhao ZZ, Hoekstra C, Hayward NK, Webb PM, Whiteman DC, et al. Novel variants in growth differentiation factor 9 in mothers of dizygotic twins. J Clin Endocrinol Metab. 2006;91: 4713–4716. doi:10.1210/jc.2006-0970

9. Simpson CM, Robertson DM, Al-Musawi SL, Heath DA, McNatty KP, Ritter LJ, et al. Aberrant GDF9 expression and activation are associated with common human ovarian disorders. J Clin Endocrinol Metab. 2014;99: E615-624. doi:10.1210/jc.2013-3949

10. Abdoli R, Zamani P, Mirhoseini SZ, Ghavi Hossein-Zadeh N, Nadri S. A review on prolificacy genes in sheep. Reprod Domest Anim Zuchthyg. 2016;51: 631–637. doi:10.1111/rda.12733

11. 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;25: 279–283. doi:10.1038/77033

12. 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)1. Biol Reprod. 2004;70: 900–909. doi:10.1095/biolreprod.103.023093

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. 42: 89–92. doi:10.1111/j.1365-2052.2010.02078.x

14. 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;9: e1003482. doi:10.1371/journal.pgen.1003482

15. 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;216: 44–50. doi:10.1016/j.livsci.2018.07.007

16. 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;9: e1003809. doi:10.1371/journal.pgen.1003809

17. 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;148: 393–400. doi:10.1210/en.2006-0764

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;64: 1225–1235.

19. Talebi R, Ahmadi A, Afraz F, Sarry J, Plisson-Petit F, Genêt C, et al. Transcriptome analysis of ovine granulosa cells reveals differences between small antral follicles collected during the

follicular and luteal phases. Theriogenology. 2018;108: 103–117. doi:10.1016/j.theriogenology.2017.11.027

20. Davis GH, Galloway SM, Ross IK, Gregan SM, Ward J, Nimbkar BV, et al. DNA tests in prolific sheep from eight countries provide new evidence on origin of the Booroola (FecB) mutation. Biol Reprod. 2002;66: 1869–1874. doi:10.1095/biolreprod66.6.1869

21. Mullen MP, Hanrahan JP, Howard DJ, Powell R. Investigation of prolific sheep from UK and Ireland for evidence on origin of the mutations in BMP15 (FecX(G), FecX(B)) and GDF9 (FecG(H)) in Belclare and Cambridge sheep. PloS One. 2013;8: e53172. doi:10.1371/journal.pone.0053172

22. Jemaa SB, Ruesche J, Sarry J, Woloszyn F, Lassoued N, Fabre S. The high prolificacy of D'man sheep is associated with the segregation of the FecLL mutation in the B4GALNT2 gene. Reprod Domest Anim. 2018;0. doi:10.1111/rda.13391

23. Rochus CM, Tortereau F, Plisson-Petit F, Restoux G, Moreno-Romieux C, Tosser-Klopp G, et al. Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. BMC Genomics. 2018;19. doi:10.1186/s12864-018-4447-x

24. Bonnet A, Cabau C, Bouchez O, Sarry J, Marsaud N, Foissac S, et al. An overview of gene expression dynamics during early ovarian folliculogenesis: specificity of follicular compartments and bi-directional dialog. BMC Genomics. 2013;14: 904. doi:10.1186/1471-2164-14-904

25. 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;39: 294–297. doi:10.1111/j.1365-2052.2008.01707.x

26. Lassoued N, Benkhlil Z, Woloszyn F, Rejeb A, Aouina M, Rekik M, et al. FecX Bar a Novel BMP15 mutation responsible for prolificacy and female sterility in Tunisian Barbarine Sheep. BMC Genet. 2017;18: 43. doi:10.1186/s12863-017-0510-x

27. Liao WX, Moore RK, Shimasaki S. Functional and molecular characterization of naturally occurring mutations in the oocyte-secreted factors bone morphogenetic protein-15 and growth and differentiation factor-9. J Biol Chem. 2004;279: 17391–17396. doi:10.1074/jbc.M401050200

28. Wan Q, Wang Y, Wang H. Identification and Analysis of Regulatory Elements in Porcine Bone Morphogenetic Protein 15 Gene Promoter. Int J Mol Sci. 2015;16: 25759–25772. doi:10.3390/ijms161025759

29. Dixit H, Rao LK, Padmalatha VV, Kanakavalli M, Deenadayal M, Gupta N, et al. Missense mutations in the BMP15 gene are associated with ovarian failure. Hum Genet. 2006;119: 408–415. doi:10.1007/s00439-006-0150-0

30. Fonseca DJ, Ortega-Recalde O, Esteban-Perez C, Moreno-Ortiz H, Patiño LC, Bermúdez OM, et al. BMP15 c.-9C>G promoter sequence variant may contribute to the cause of non-syndromic premature ovarian failure. Reprod Biomed Online. 2014;29: 627–633. doi:10.1016/j.rbmo.2014.07.018

31. McNatty KP, Galloway SM, Wilson T, Smith P, Hudson NL, O'Connell A, et al. Physiological effects of major genes affecting ovulation rate in sheep. Genet Sel Evol. 2005;37: S25. doi:10.1186/1297-9686-37-S1-S25

638 32. Bonnet A, Bevilacqua C, Benne F, Bodin L, Cotinot C, Liaubet L, et al. Transcriptome profiling of
639 sheep granulosa cells and oocytes during early follicular development obtained by laser capture
640 microdissection. BMC Genomics. 2011;12: 417. doi:10.1186/1471-2164-12-417

641 33. Fabre S, Pierre A, Mulsant P, Bodin L, Di Pasquale E, Persani L, et al. Regulation of ovulation rate
642 in mammals: contribution of sheep genetic models. Reprod Biol Endocrinol RBE. 2006;4: 20.
643 doi:10.1186/1477-7827-4-20

644 34. Jansson T. Genes involved in ovulation rate and litter size in sheep [Internet]. SLU, Departement
645 of Clinical sciences; 2014. Available: https://stud.epsilon.slu.se/6803/

646 35. Martin P, Raoul J, Bodin L. Effects of the FecL major gene in the Lacaune meat sheep
647 population. Genet Sel Evol GSE. 2014;46: 48. doi:10.1186/1297-9686-46-48

648 36. Kumar S, Mishra AK, Kolte AP, Arora AL, Singh D, Singh VK. Effects of the Booroola (FecB)
649 genotypes on growth performance, ewe's productivity efficiency and litter size in
650 Garole×Malpura sheep. Anim Reprod Sci. 2008;105: 319–331.
651 doi:10.1016/j.anireprosci.2007.03.012

652 37. Demars J, Cano M, Drouilhet L, Plisson-Petit F, Bardou P, Fabre S, et al. Genome-Wide
653 Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy
654 Species from Modern Woolly Sheep. Mol Biol Evol. 2017;34: 1722–1729.
655 doi:10.1093/molbev/msx114

656 38. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative
657 genomics viewer. Nat Biotechnol. 2011;29: 24–26. doi:10.1038/nbt.1754

658 39. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for
659 variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet.
660 2011;43: 491–498. doi:10.1038/ng.806

661 40. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From
662 FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline.
663 Curr Protoc Bioinforma Ed Board Andreas Baxevanis Al. 2013;11: 11.10.1-11.10.33.
664 doi:10.1002/0471250953.bi1110s43

665 41. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes,
666 differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using
667 pair-wise correlations. Biotechnol Lett. 2004;26: 509–515.

668 42. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet.
669 2006;7: 781–791. doi:10.1038/nrg1916

670 43. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to
671 the challenge of larger and richer datasets. GigaScience. 2015;4: 7. doi:10.1186/s13742-015-
672 0047-8

673 44. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype
674 data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet.
675 2006;78: 629–644. doi:10.1086/502802

45. Poivey J-P, Tiphine L, Berny B, Julien E. Indexation blup modèle animal chez les ovins allaitants. Rencontres Recherches Ruminants. 1995. pp. 453–456. Available: http://www.journees3r.fr/IMG/pdf/1995_8_actualite_01_poivey.pdf

46. Glimour A, Gogel B, Cullis B, Thompson R. R: ASReml User-Guide Release 3.0. VSN [Internet]. VSN International; 2009. Available: https://asreml.kb.vsni.co.uk/wp-content/uploads/sites/3/2018/02/ASReml-3-User-Guide.pdf

Supporting information captions

S1 Table. List of variants found in the OARX: 50639087-54114793 region. Listing of 60 SNPs and 90 small INDELs with quality score >30.

S2 Table. List of primers used in the study. Locations of primers are based on the OARv3.1 ovine genome assembly available on ensembl.org.

S3 Figure. Genome-wide and chromosome-wide association results integrating the SNP OARX: 50977717T>A. (A) Genome-wide association results for litter size in the NV sheep population. Manhattan plot shows the combined association signals (-$\log_{10}$(p-value)) on the y-axis versus SNPs position in the sheep genome on the x-axis and ordered by chromosome number (assembly OARv3.1). Red line represents the 5% genome-wide threshold. (B) OARX chromosome-wide association results. The curve shows the combined association signals (-$\log_{10}$(p-value)) on the y-axis versus SNPs position on the X chromosome on the x-axis (assembly OARv3.1). Red line represents the 5% chromosome-wide threshold. In both panels, the position of the SNP OARX:50977717T>A is indicated by a red dot. In panel (B), the *BMP15* gene location is indicated by a red arrowhead.

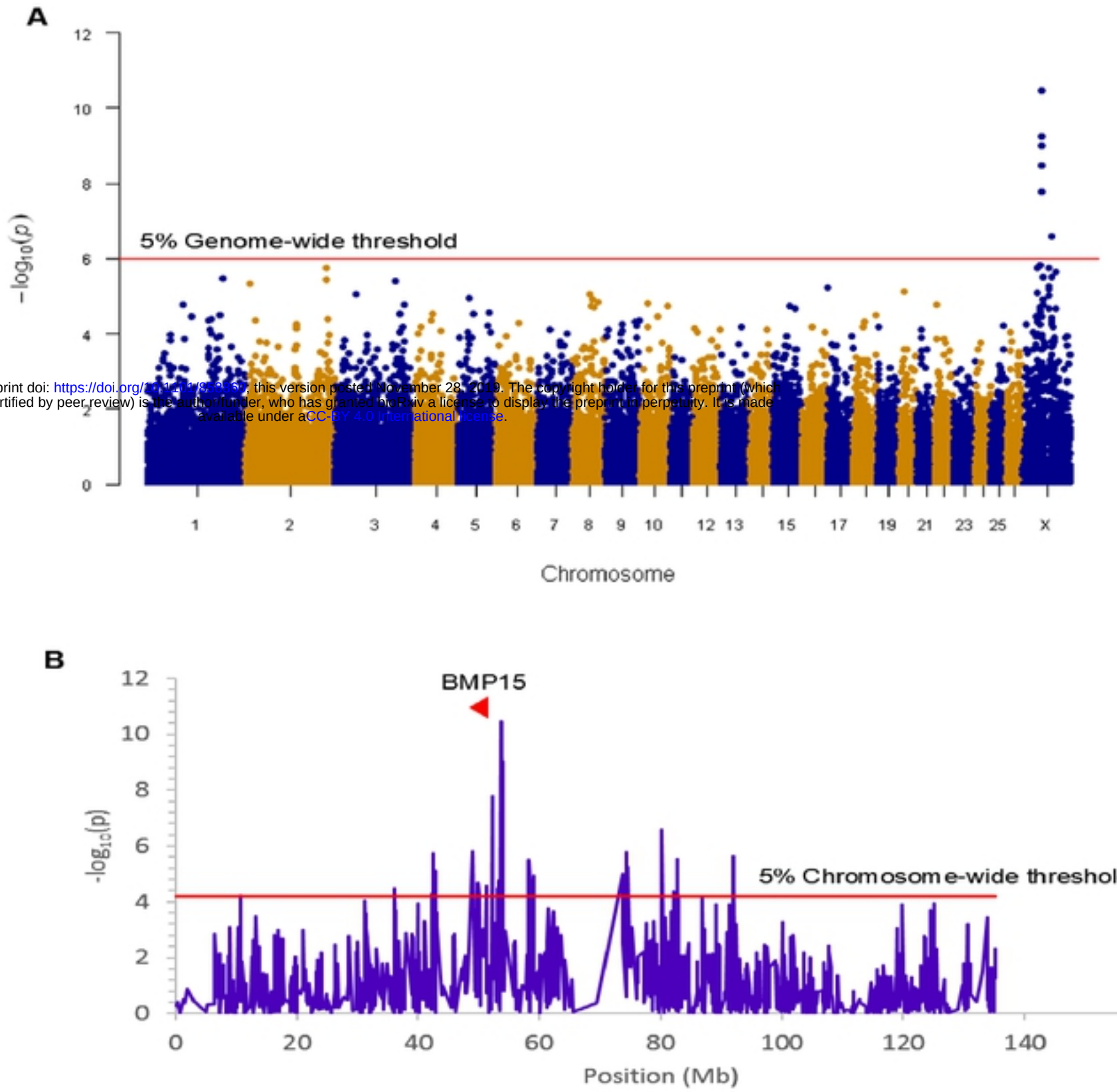

Figure 1. Genome-wide and chromosome-wide association results. (A) Genome-wide association results for litter size in the NV sheep population. Manhattan plot shows the combined association signals ($-\log_{10}$(p-value)) on the y-axis versus SNPs position in the sheep genome on the x-axis and ordered by chromosome number (assembly OARv3.1). Red line represents the 5% genome-wide threshold. (B) OARX chromosome-wide association results. The curve shows the combined association signals ($-\log_{10}$(p-value)) on the y-axis versus SNPs position on the X chromosome on the x-axis (assembly OARv3.1). Red line represents the 5% chromosome-wide threshold. The *BMP15* gene location is indicated by a red arrowhead.

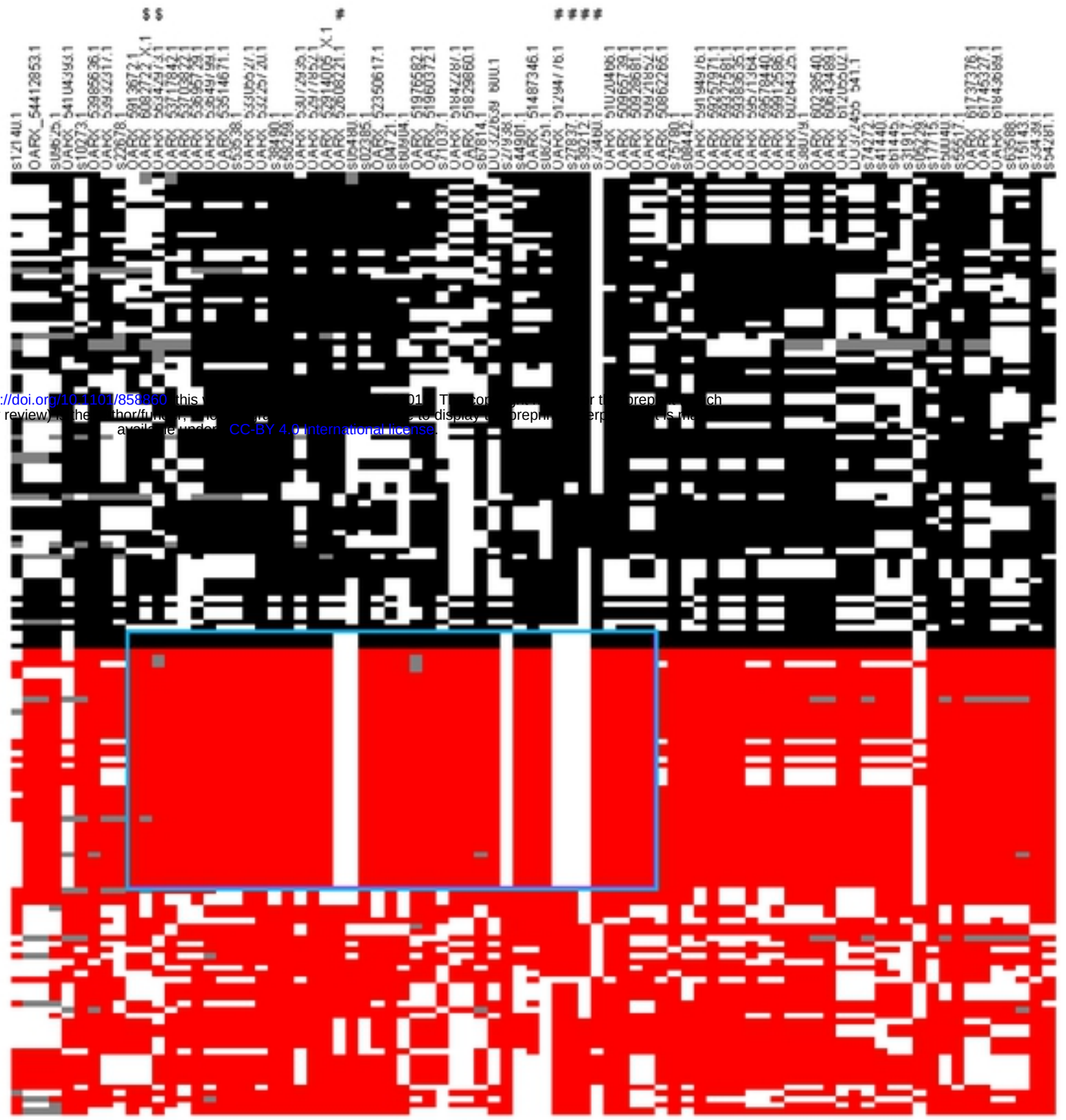

Figure 2. Clusterization of haplotypes reconstructed at the OARX locus. 80 markers encompassing the OARX region of interest (50.6 Mb-54.1 Mb) were selected to construct haplotypes from 39 cases and 40 control animals. Each column represents one SNP and each line represents one haplotype. For one marker (i) allele 1 is in black in controls, or in red in cases, (ii) allele 2 is in white when the phase was unambiguous and (iii) grey color represents unphased SNP. Haplotypes were ordered to distinguish controls versus cases and clusterized to classify similar clades of haplotypes. The # sign flags SNP significantly associated with LS at genome-wide level. The $ sign flags SNP flanking the *BMP15* genes (50970938-50977454 bp). The specific haplotype preferentially selected in highly prolific ewes (cases) is symbolized by the cyan rectangle.

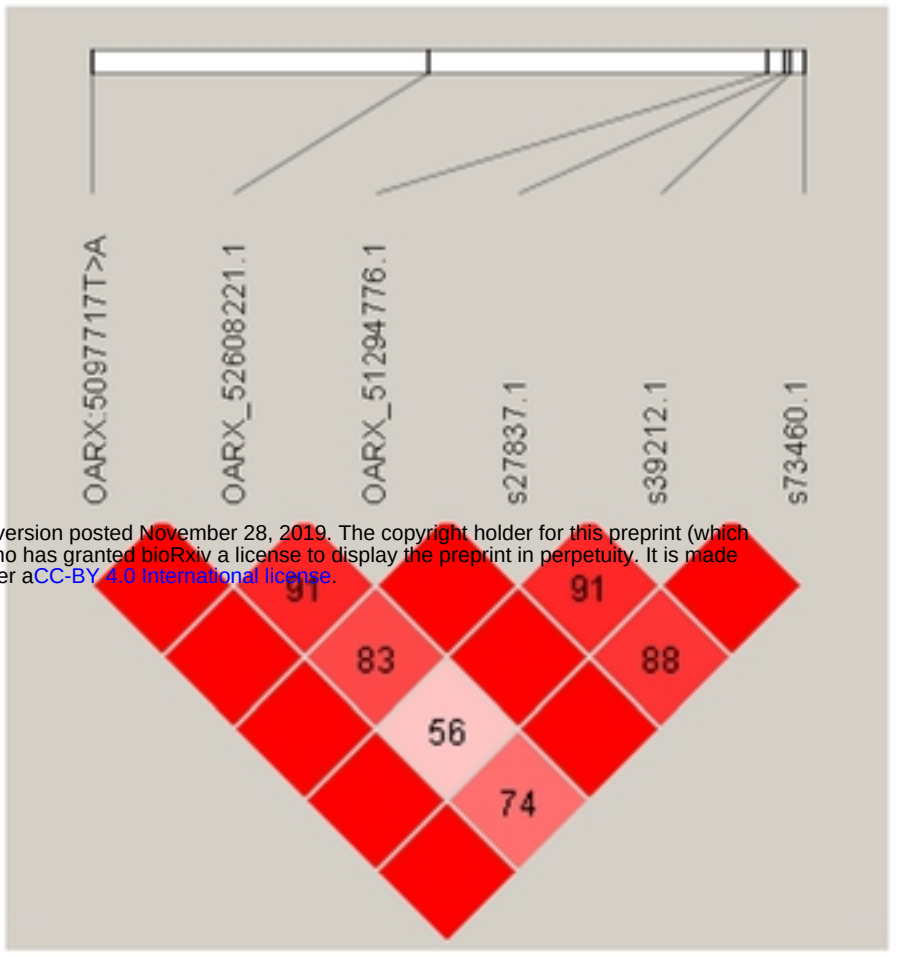

Figure 3. Linkage disequilibrium (LD) plot. Generated by Haploview, pair-wise LD between SNP markers (OARX: 50977717T>A and the five genome-wide significant LS-associated markers) is represented by the D' value. Strong LD is represented in dark red boxes (D'=100) and weaker LD (D'<100) in lighter red boxes.

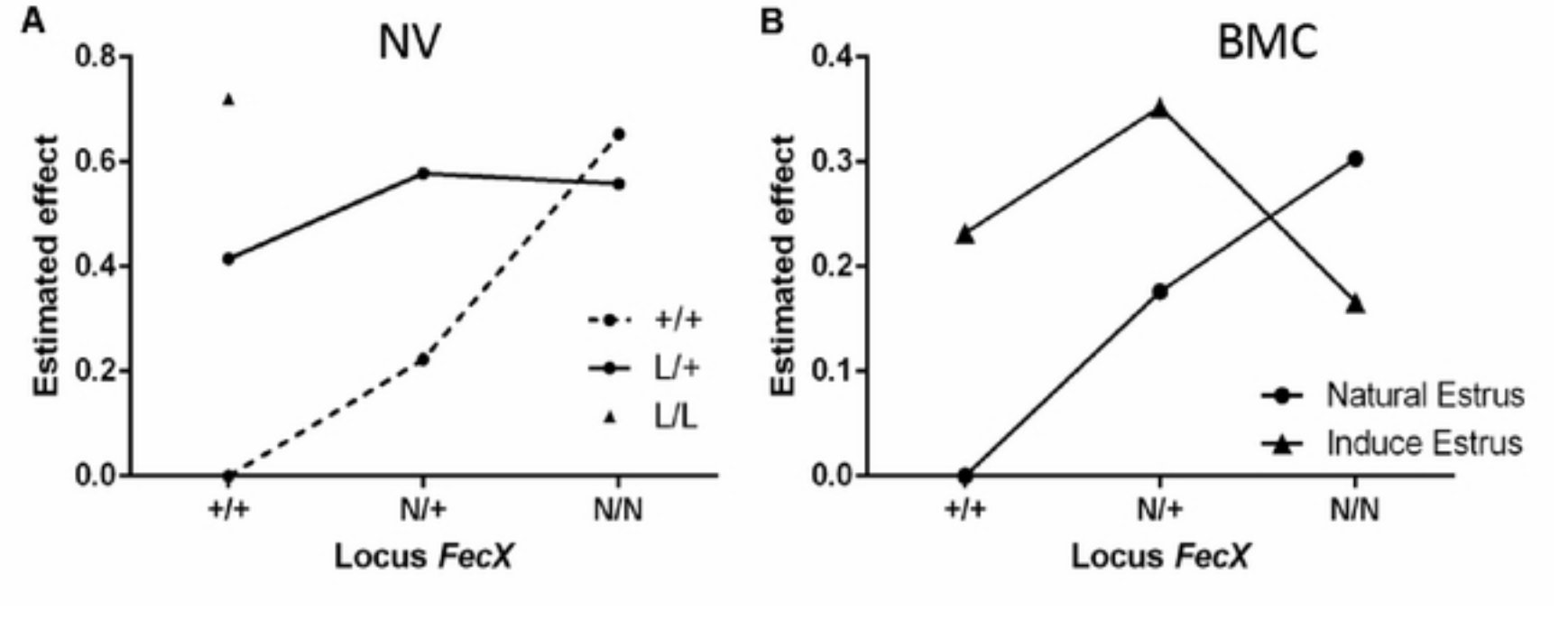

Figure 4. Estimated effect of *FecX^N*. A) Estimated effect of *FecX^N* knowing *FecL^L* genotype in Noire du Velay (NV). B) Estimated effect of *FecX^N* according to induced or natural estrus in Blanche du Massif Central (BMC). Estimated increased LS from the linear mixed models with interaction between genotypes at the *FecX* and *FecL* loci for NV, and between *FecX* genotypes and estrus type for BMC are given relative to *L* and *N* alleles non-carrier ewes (+/+).

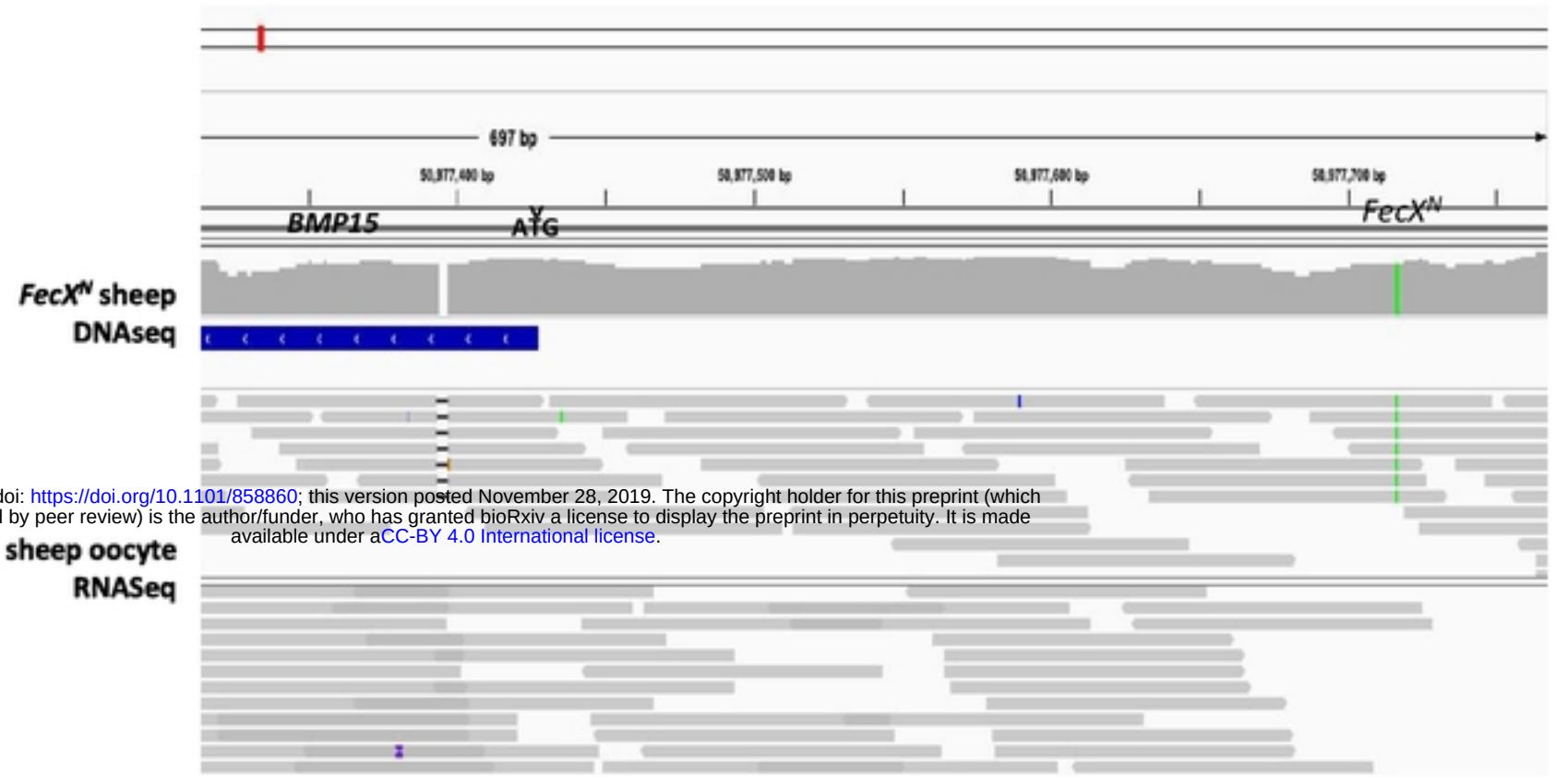

Figure 5. Localization of *FecX^N* in the *BMP15* upstream region. Integrative Genomics Viewer (IGV v2.4.10) snapshot of the ovine *BMP15* gene upstream region alignments (coordinates from OAR v3.1 assembly) with reads from whole genome DNA sequencing (DNASeq) of a homozygous *FecX^N* carrier ewe (green line) and total mRNA sequencing of ovine oocytes from small antral follicles (RNASeq, Bonnet et al. 2013), indicating a possible localization of *FecX^N* in the 5'UTR region of *BMP15* (ATG start codon is indicated by an arrowhead and *BMP15* gene annotation is on the minus strand).

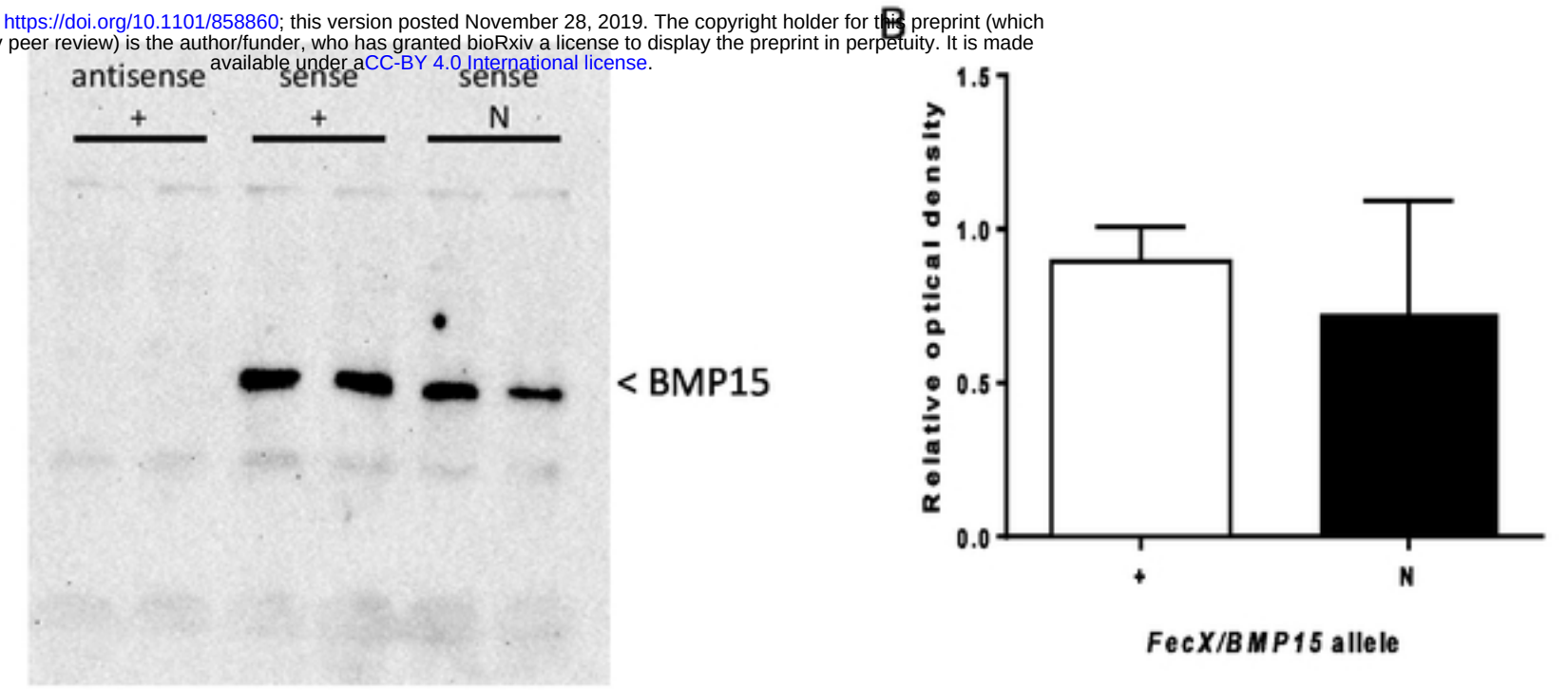

Figure 6. Effect of *FecX*^{N} mutation on the ovine BMP15 protein produced *in vitro*. *In vitro* transcription and translation were realized in duplicate in 6 independent experiments from antisense *BMP15* cDNA carrying the wild-type allele (antisense +) as negative control, or sense *BMP15* cDNA with (sense N) or without *FecX*^{N} (sense +). A) A representative western blot experiment revealing the BMP15 protein. B) Chemiluminescent signal of BMP15 was captured by a ChemiDoc MP imaging system and images were quantified (relative to sense +) and analyzed with the Image Lab Software (Bio-Rad).

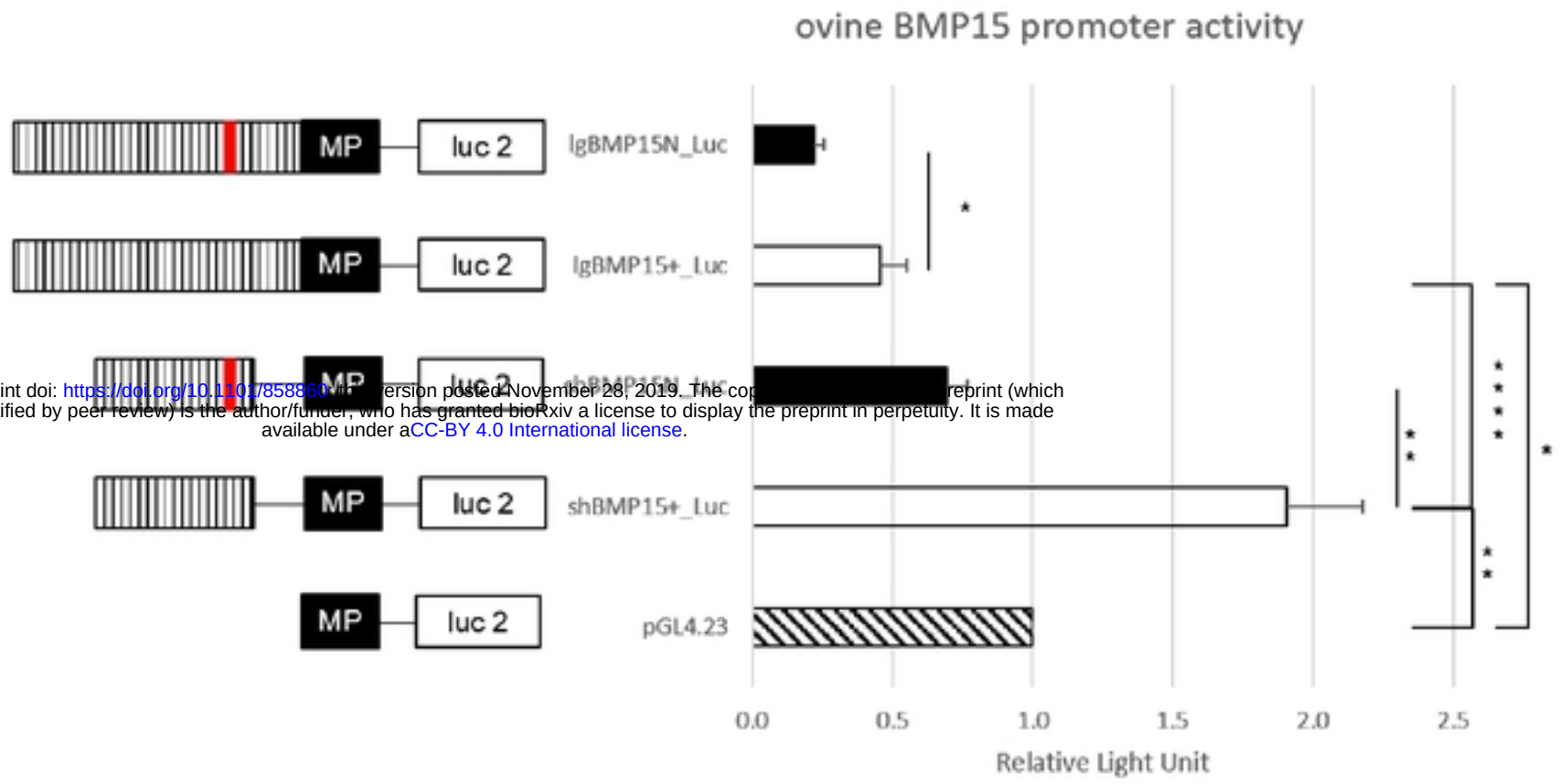

Figure 7. Functional effect of *FecX^N* mutation on the *BMP15* promoter activity. *In vitro* reporter luciferase assay from CHO cells transiently transfected with empty vector or wild-type *BMP15* promoter (BMP15$^+$) or mutant *BMP15* promoter (BMP15^N). Two fragments were generated, a long (lg) form of 732 bp (-743, -11bp referring to ATG start codon) and a short (sh) form of 341bp (-443, -102bp). Results are expressed as means ± SEM of the relative light unit (RLU) from 6 independent transfection experiments in triplicate. Asterisk indicates significant difference *: $p<0.05$; **: $p<0.01$; ****: $p<0.0001$. MP: minimal promotor of pGL4.23, Luc 2: luciferase reporter gene, hatched bar: *BMP15* promotor, red line: *FecX^N* mutation

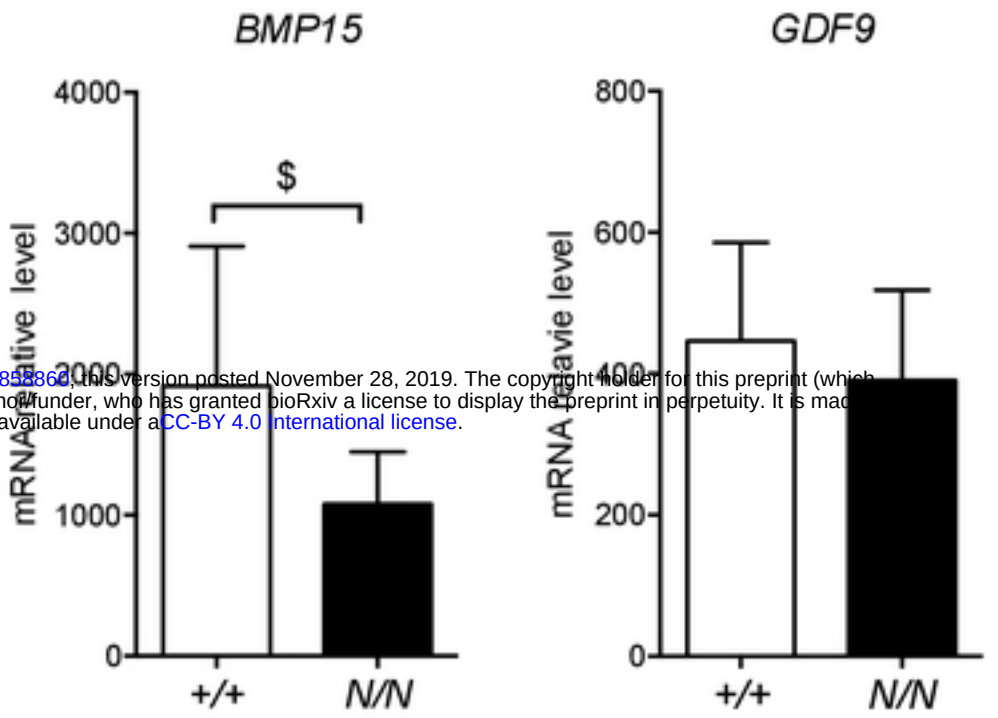

Figure 8. Effect of *FecX^N* mutation on *BMP15* and *GDF9* expression in ovine oocyte. Quantitative real-time PCR results of BMP15 (A) and GDF9 (B) expression in oocyte pools from growing (1-3mm) follicles (n=20 pools of 5 oocytes) from homozygous *FecX^N* carrier (*N/N*, n=5) and non-carrier ewes (+/+, n=5) during the follicular phase of the estrus cycle. Results are expressed as means ± SEM of the mRNA relative level for each genotype, using GAPDH and SDHA as internal references. Raw data were analyzed by two-way ANOVA. Dollar symbol indicates a suggestive difference between genotypes, $: p=0.166.