Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep

Julie Demars,*[,1] Margarita Cano,[†,2] Laurence Drouilhet,[†,1] Florence Plisson-Petit,[1] Philippe Bardou,[1,3] Stéphane Fabre,[1] Bertrand Servin,[1] Julien Sarry,[1] Florent Woloszyn,[1] Philippe Mulsant,[1] Didier Foulquier,[4] Fabien Carrière,[4] Mathias Aletru,[1,5] Nathalie Rodde,[6] Stéphane Cauet,[6] Olivier Bouchez,[1,7] Maarten Pirson,[1,7] Gwenola Tosser-Klopp,[1] and Daniel Allain[1]

[1]GenPhySE, Université de Toulouse, INRA, INPT, ENVT, Castanet-Tolosan, France

[2]Instituto de Genetica, INTA, Buenos Aires, Argentina

[3]SIGENAE, INRA, Castanet-Tolosan, France

[4]Domaine Expérimental de La Fage INRA, Roquefort sur Soulzon, France

[5]Domaine Expérimental de Langlade INRA, Pompertuzat, France

[6]CNRGV INRA, Castanet-Tolosan, France

[7]Get-Plage INRA, Castanet-Tolosan, France

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: julie.demars@inra.fr.

Associate editor: Nadia Singh

Abstract

The composition and structure of fleece variation observed in mammals is a consequence of a strong selective pressure for fiber production after domestication. In sheep, fleece variation discriminates ancestral species carrying a long and hairy fleece from modern domestic sheep (*Ovis aries*) owning a short and woolly fleece. Here, we report that the "woolly" allele results from the insertion of an antisense *EIF2S2* retrogene (called *asEIF2S2*) into the 3′ UTR of the *IRF2BP2* gene leading to an abnormal IRF2BP2 transcript. We provide evidence that this chimeric IRF2BP2/asEIF2S2 messenger 1) targets the genuine sense EIF2S2 RNA and 2) creates a long endogenous double-stranded RNA which alters the expression of both EIF2S2 and IRF2BP2 mRNA. This represents a unique example of a phenotype arising via a RNA-RNA hybrid, itself generated through a retroposition mechanism. Our results bring new insights on the sheep population history thanks to the identification of the molecular origin of an evolutionary phenotypic variation.

Sheep and goats were the first livestock species to be domesticated (Ryder 1981); several domestication events gave rise to domestic sheep (Pedrosa et al. 2005). Initially, sheep were reared for access to meat before human mediated specialization for secondary products such as wool and milk 4,000–5,000 years ago (Chessa et al. 2009). Nevertheless, the exact line of descent between domestic sheep and their wild ancestor remains unclear (Hiendleder et al. 2002). Sheep selected for secondary products like wool characteristics appear to have replaced more primitive populations. Indeed, woolly modern sheep own a single coated fleece whereas wild and ancestral species exhibit a double coated fleece (Piper and Ruvinsky 1997). Nowadays, fleece variation is segregating in the French Romane breed which is a composite breed between the Berrichon du Cher and Romanov (Ricordeau et al. 1992). Due to its parental origins, this breed shows large variability in its fleece type. At birth, a highly variable coat type from a short, woolly fleece (typical to Berrichon du Cher and domestic sheep) up to a hairy, long coat (characteristic of Romanov and primitive sheep) is observed in Romane lambs (Allain et al. 2014) (fig. 1a). Moreover, the birth coat type is a

very highly heritable trait within this breed (Allain et al. 2014). Here, we used 1) the Romane breed to map, identify and functionally validated the mutation responsible of fleece variation and 2) various sheep populations representative of both ancestral species and modern sheep breeds to genetically confirm this singular causal mutation.

The Fleece Variation Locus Encompasses the Unique IRF2BP2 Gene on Chromosome 25

To establish the genetic determinism of fleece variation in sheep, experimental half-sib families were organized in the Romane breed. The trial involved a total of two thousands lambs produced from ten unrelated rams (supplementary table S1, Supplementary Material online). Approximately 57% of the lambs presented a hairy and long coat similar to the rustic Romanov breed (fig. 1b and c) and 43% were comparable to Berrichon du Cher individuals with a short and woolly fleece (fig. 1d and e). Based on the phenotypic segregation in ram families, the hypothesis of a single locus was rejected since 1) only two phenotypes were observed excluding an additive inheritance model and 2) no sire family produced an unique
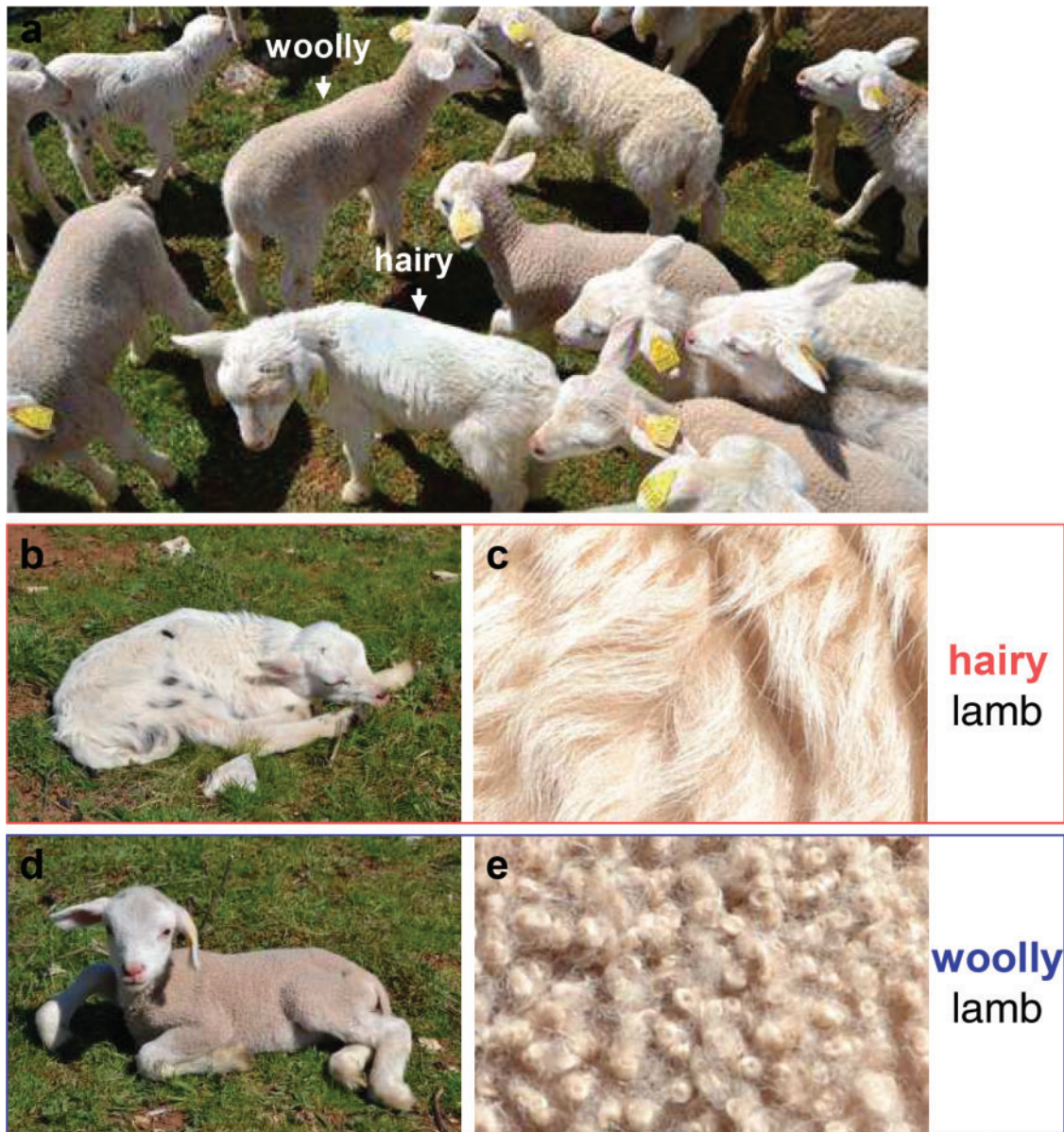
Open Access

Letter

Fig. 1. Fleece variation observed in lambs of the Romane breed. (*a*) Segregation of the birth coat type in 1-month-old Romane animals. (*b, c*) Romane lamb carrying a long and hairy coat similar to primitive and ancestral sheep species. Hairy breeds have a double coated fleece with a coarse outer coat made of hair fibers and a fine inner coat composed of woolly fibers. (*d, e*) Romane lamb owning a short and woolly coat typical to domestic modern sheep. Woolly breeds have more woolly fibers and carry a single coated fleece with all fibers nearly similar in dimensions (Hiendleder et al. 2002).

birth coat type ruling out a dominance mechanism (supplementary table S1, Supplementary Material online). Moreover, when considering progeny phenotypes based on dam traits, our data likely suggest a genetic model involving a few major genes with dominance/recessive effects (supplementary table S1, Supplementary Material online).

To gain insights into the molecular basis of fleece variation, a quantitative trait locus (QTL) detection for the birth coat type was carried out on a subset of the Romane half-sib protocol ($n = 759$) using the OvineSNP50 Genotyping BeadChip (supplementary table S2, Supplementary Material online). In parallel, a sampling of eight QTL rams and few offspring per family was also genotyped with the Ovine

Infinium HD SNP BeadChip ($n = 135$) (supplementary Materials and Methods and supplementary table S2, Supplementary Material online). Thus, we took advantage of this data set for imputing, with the Beagle program (Browning and Browning 2009), ungenotyped SNPs within the rest of the half-sib experimental design. A genome-wide association study (GWAS) was performed using mixed models implemented into the GEMMA software (Zhou and Stephens 2012; Zhou et al. 2013). Two loci, located on ovine chromosomes 25 (QTL$_{Oar25}$) and 13 (QTL$_{Oar13}$), underlie fleece variation in the Romane breed (fig. 2*a* and supplementary fig. S1, Supplementary Material online) confirming the oligogenic determinism hypothesis. The QTL$_{Oar25}$ showed
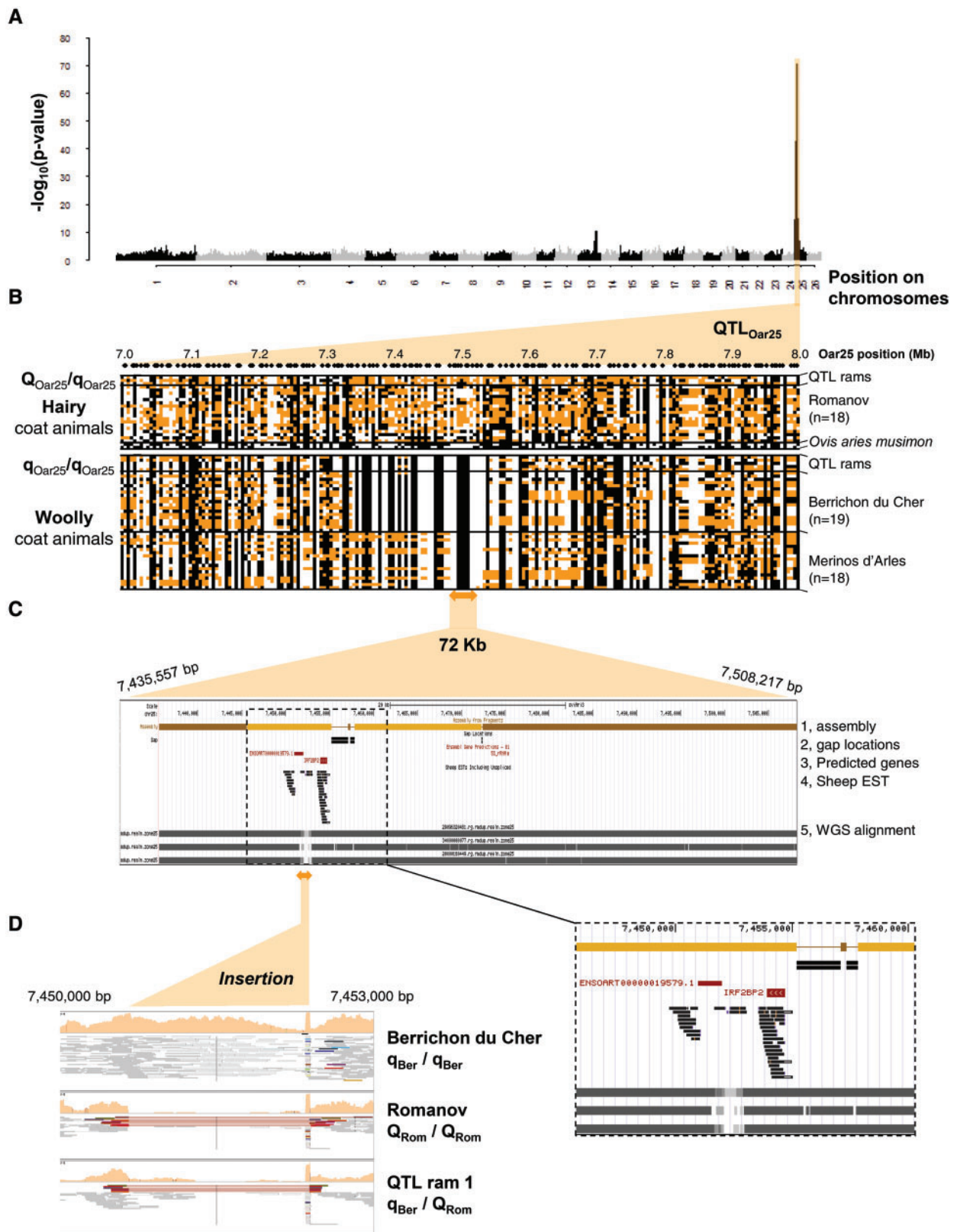
FIG. 2. Positional cloning of the QTL$_{Oar25}$. (*a*) Manhattan plot shows on the *y*-axis the significance versus the chromosomal position (Oarv3.1) on the *x*-axis. (*b*) Fine mapping of the QTL$_{Oar25}$. Genotypes of 65 sheep for the QTL$_{Oar25}$ segment centered around the most significant GWAS peak, and encompassing 208 SNPs. The positions of the SNPs are mentioned on the scale at the top. Each column represents one SNP and each line represents one animal. Homozygous genotypes are shown in black or white, heterozygous genotypes in orange. (*c*) Window of the QTL$_{Oar25}$ interval (Oarv3.1) extracted from the UCSC genome browser (https://genome-euro.ucsc.edu) and a zoom of the region around the *IRF2BP2* gene. (*d*) Screen capture of the suggested insertion extracted from IGV. Orange represents the depth coverage. The red lines correspond to reads which align to two positions separated of 1500 bp. EST, expression sequence tag and WGS, whole genome sequencing.
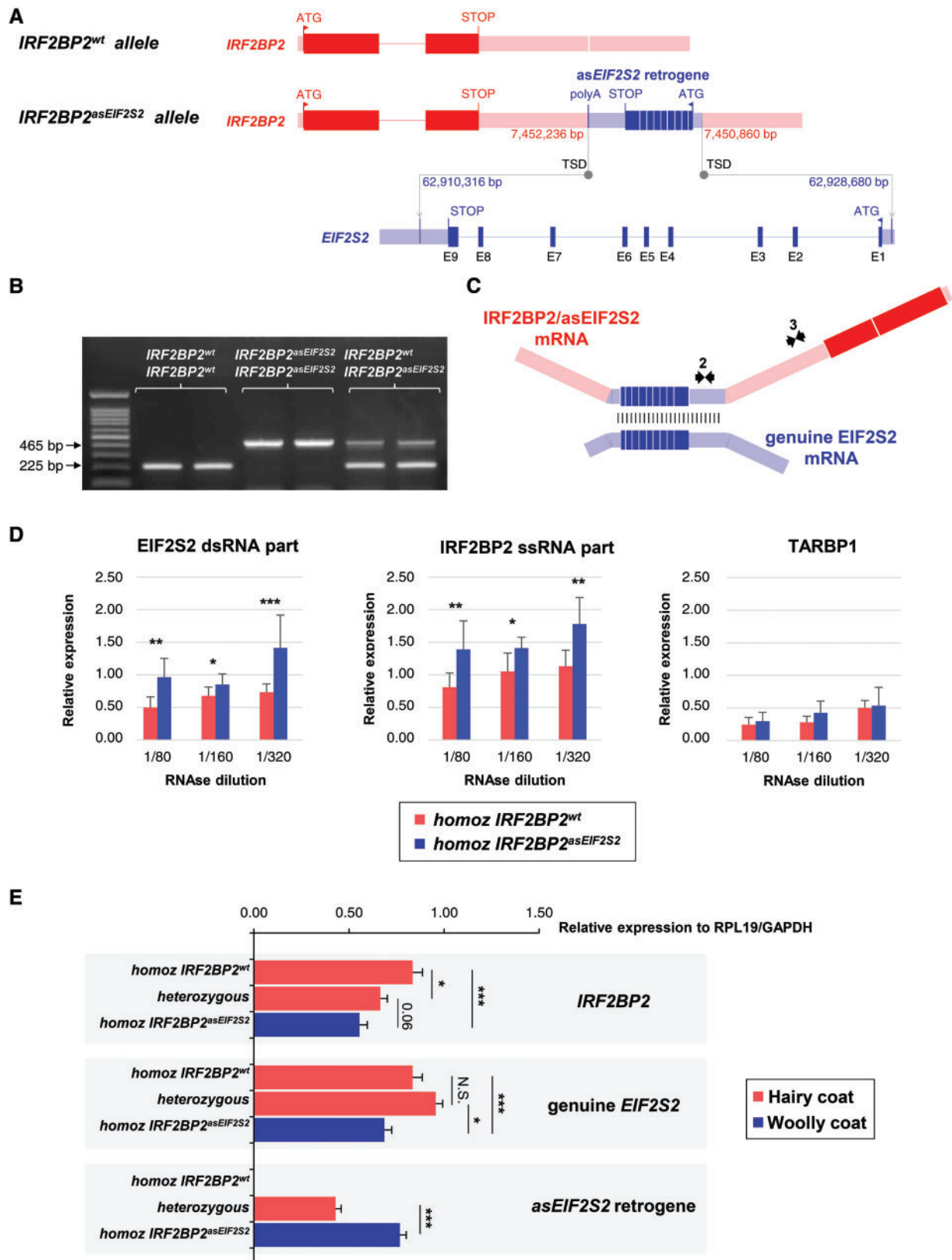
Fig. 3. Characteristics and molecular effects of the QTL$_{Oar25}$ mutation. (*a*) The exact structure of the genuine *EIF2S2* gene is conserved except for 5′ and 3′ UTR sequences which are missing. Sequences corresponding to *IRF2BP2* and *EIF2S2* are, respectively, in red and blue. The positions of the breakpoints on the Oarv3.1 genome assembly are also mentioned. (*b*) Genotyping of the QTL$_{Oar25}$ mutation in cDNA extracted from skin samples. The lower band corresponds to the wild type allele (*IRF2BP2^{wt}*) and the upper band corresponds to the insertion (*IRF2BP2^{asEIF2S2}* allele). (*c*) Putative EIF2S2 RNA–RNA complex in lambs carrying the *IRF2BP2^{asEIF2S2}* allele. Arrows represent pairs of primers used in the EIF2S2 RNAse A protection assay (*d*) specific of either EIF2S2 dsRNA part (2) or IRF2BP2 ssRNA part (3). The numbers of pairs of primers referred to the supplementary table S3

both the strongest signal (s67158.1 located 7,727,709 bp, q-value $= 2.60E^{-65}$) and the largest effect (0.62 $\pm$ 0.017 phenotypic standard deviation) and had already been identified in several populations (Allain et al. 2006; Vitezica et al. 2007). The QTL_{Oar13} also had both significant signal (s27858.1 located 62,539,468 bp, q-value $= 1.96E^{-07}$) and effect (0.39 $\pm$ 0.024 phenotypic standard deviation) but seemed specific to the Romane breed. Interestingly, the QTL_{Oar25} region overlapped with a selective signature in ancestral populations (Kijas et al. 2012; Fariello et al. 2014) implying that fleece variation phenotype has been under selection over the period of domestication.

To fine map the major locus and identify the shortest woolly haplotype, we combined genotyping results from experimental and production populations. A linkage analysis within each ram family, performed with the QTLMap software (Elsen et al. 1999), showed that three out of eight sires were heterozygous for QTL_{Oar25} (supplementary fig. S2a, Supplementary Material online). After phasing haplotypes for the region of interest, we used the software Plink (Chang et al. 2015) to identify segments that were identical by state, narrowing the likely QTL_{Oar25} region to a 198 kb haplotype (7,299,164–7,510,378 bp) (supplementary fig. S2b, Supplementary Material online). In an independent effort, purebred animals at the origin of the Romane breed (Berrichon du Cher and Romanov) as well as primitive double coated sheep (Ovis musimon) and single coated animals (Merinos d'Arles) were genotyped using the Ovine Infinium HD SNP BeadChip. We then performed homozygosity mapping for all individuals known as woolly modern sheep. The QTL_{Oar25} region was significantly decreased to 72 kb (7,435,557–7,508,217 bp) encompassing the unique IRF2BP2 gene which was not previously known to be involved in fleece variation (fig. 2b and c).

The Fleece Variation Mutation Corresponds to the Insertion of an EIF2S2 Retrogene

To fulfill various gaps existing in the reference sheep genome for the QTL_{Oar25} region, the CH243-319B1 Bacterial Artificial Chromosome (BAC) spanning the Oar25 locus (7,303,632–7,523,874 bp) was sequenced via the PacBio technology sequencing (see Supplementary Materials and Methods, Supplementary Material online). The BAC sequence was then integrated into the Oarv3.1 genome assembly since both ovine BAC library (Dalrymple et al. 2007) and reference genome (Jiang et al. 2014) were produced from a Texel sheep. Therefore, this perfect sequence covering the whole QTL_{Oar25} interval was then considered as a novel reference for further analyses. To identify the causal mutation of QTL_{Oar25}, we

sequenced on an Illumina HiSeq 3000 system the whole genome of seven selected individuals carrying either the "Q" ancestral hairy allele (Romanov, $n = 2$), the "q" modern woolly allele (Berrichon du Cher, $n = 3$) or both (heterozygous QTL sires, $n = 2$). After sequences alignment and variant filtering, a total of 18 variants, including only intergenic or intron polymorphisms, remained putative causal mutations. Visual inspection of sequences using the Integrative Genomics Viewer (IGV) software (https://www.broadinstitute.org/igv/) pinpointed an insertion of approximately 1500 bp and located within the IRF2BP2 gene. As expected and shown on figure 2d, Berrichon du Cher individuals were homozygous for the insertion as the reference Texel sequence whereas Romanov animals did not carry the insertion and QTL rams were heterozygous. Interestingly, this insertion encompassed the ENSOARG00000019579.1 processed pseudogene (fig. 2c).

To characterize this singular variant, Sanger sequencing was performed for two animals (Berrichon du Cher vs. Romanov). We identified within the Berrichon du Cher individual, using the BLAST program (https://www.ncbi.nlm.nih.gov/), an insertion of an antisense EIF2S2 retrogene (asEIF2S2) into the 3' UTR of the IRF2BP2 gene (called $IRF2BP2^{asEIF2S2}$ allele) (fig. 3a and supplementary fig. S3a, Supplementary Material online). Thus, the ancestral allele ($IRF2BP2^{wt}$) of the QTL_{Oar25} present in rustic Romanov sheep corresponds to the absence of asEIF2S2 whereas single coated modern sheep, as Berrichon du Cher and Texel being the reference genome, carried the mutated woolly $IRF2BP2^{asEIF2S2}$ allele. Accordingly, the inserted retrogene originated from the ancestral copy of EIF2S2 since it carries the C allele at rs162016183 fixed in primitive animals (supplementary fig. S3b, Supplementary Material online). Three EIF2S2 genes are annotated in the Oarv3.1 assembly including the genuine EIF2S2 located on Oar13 and two EIF2S2 pseudogenes located on Oar7 and Oar25.

Remarkably, the genuine EIF2S2 gene (62,909,286–62,928,755 bp) is located nearby the most significant marker of the QTL_{Oar13} (s27858.1, 62,539,468 bp). We fine-mapped the QTL_{Oar13} by taking advantage of both linkage analyses and populations haplotypic information as previously. The two likely remaining genetic intervals span 129 Kb (62,475,621–62,604,749 bp) and 69 kb (62,817,543–62,886,830 bp) and therefore excluded the genuine EIF2S2 gene as a candidate gene for the QTL_{Oar13} (supplementary fig. S4, Supplementary Material online). Nonetheless, Sanger sequencing was performed for three animals (Berrichon du Cher, Romanov and a QTL sire) and we highlighted three polymorphisms including one in the 3' UTR, one synonymous (rs160604021), and one nonsynonymous

FIG. 3 Continued
and figure S9, Supplementary Material online. (d) Messengers of animals homozygous $IRF2BP2^{wt}$ or $IRF2BP2^{asEIF2S2}$ were treated with several doses of RNAse A, reverse-transcribed and submitted subsequently to real-time quantitative PCR. In addition to primers located along the RNA–RNA complex, the TARBP1 gene located close to the QTL_{Oar25} (7,259,194–7,324,294 bp) was used as control of ssRNA. The relative expression corresponds to the comparison of the target gene to the mean of internal housekeeping genes and then normalized with the expression of the same nontreated samples. Data are mean $\pm$ s.d. $^*P < 5E^{-02}$, $^{**}P < 1E^{-02}$, and $^{***}P < 1E^{-03}$ (two-tailed unpaired Student's t-test). e, Quantification of messengers in skin biopsies of lambs homozygous for the $IRF2BP2^{wt}$ ($n = 15$), heterozygous ($n = 30$) and exhibiting a double coated fleece or homozygous for the asEIF2S2 retrogene insertion ($n = 24$) and displaying a single coated fleece. Data are mean $\pm$ s.d. $^*P < 5E^{-02}$, $^{**}P < 5E^{-03}$, $^{***}P < 5E^{-04}$ (nonparametric test). TSD, target site duplication, Homoz, homozygous and Heteroz, heterozygous.

Table 1. Segregation of the QTL_{Oar25} Mutation in Sheep Displaying Either a Double- or Single-Coated Fleece According to Sheep Breed Standards.

Breed	Genotyped Animals (n)	Phenotype	$IRF2BP2^{wt}$ $IRF2BP2^{wt}$	$IRF2BP2^{wt}$ $IRF2BP2^{asEIF2S2}$	$IRF2BP2^{asEIF2S2}$ $IRF2BP2^{asEIF2S2}$
Ovis orientalis (NextGen project)	19	Long and hairy	19	0	0
Ovis aries musimon (NextGen project)	20	Long and hairy	20	0	0
Ovis orientalis	15	Long and hairy	15	0	0
Ovis aries musimon	18	Long and hairy	18	0	0
Ovis canadensis	3	Long and hairy	3	0	0
Ovis dalli	2	Long and hairy	2	0	0
Causses du Lot	20	Long and hairy	15	4	1
Corse	16	Long and hairy	16	0	0
Limousine	18	Long and hairy	16	2	0
Manech tête rousse	25	Long and hairy	25	0	0
Rava	19	Long and hairy	16	3	0
Romanov	18	Long and hairy	16	2	0
	193				
Berrichon du Cher	35	Short and woolly	0	0	35
Blanche du Massif Central	20	Short and woolly	0	0	20
Charmoise	22	Short and woolly	0	0	22
Charollais	22	Short and woolly	0	0	22
Île-de-France	23	Short and woolly	0	0	23
Lacaune (lait)	35	Short and woolly	0	0	35
Lacaune (viande)	34	Short and woolly	0	0	34
Mérino d'Arles	18	Short and woolly	0	0	18
Mérino de Rambouillet	27	Short and woolly	0	0	27
Mourerous	16	Short and woolly	1	0	15
Ouessant	18	Short and woolly	0	0	18
Préalpes du Sud	17	Short and woolly	0	1	16
Rouge de l'Ouest	16	Short and woolly	0	0	16
Roussin de la Hague	21	Short and woolly	0	0	21
Suffolk	18	Short and woolly	0	0	18
Tarasconnaise	15	Short and woolly	0	0	15
Texel	24	Short and woolly	1	0	23
Vendéen	21	Short and woolly	0	0	21
	402				
Noire du Velay	19	Both	1	6	12
Romane	17	Both	1	2	14

NOTE.—A total of four ancestral species and 24 French ovine breeds were genotyped with an average of 21.3 ($\pm$5.7) animals per breed. We also exploited data from the international NextGen project which are publicly available (http://nextgen.epfl.ch/).

(rs162016183). The genotype of the missense SNP, obtained *via* the Ovine Infinium HD SNP BeadChip, was confirmed for the whole QTL protocol. The significant association obtained (q-value = 2.02E^{-05}), due to the linkage disequilibrium structure, was much lower than the best signal (q-value = 1.96E^{-07}). Altogether, these results ruled out a functional mutation within the genuine *EIF2S2* gene as the causal mutation of the QTL_{Oar13} but *cis*-regulatory effects affecting this gene may not be discarded. To understand the genetic architecture of the birth coat trait, the segregation of both QTL_{Oar25} and QTL_{Oar13} was analyzed. We highlighted that almost all lambs carrying at least one "Q" haplotype at the QTL_{Oar25} are hairy animals indicating a dominant effect of the ancestral wild type allele (supplementary fig. S5, Supplementary Material online). However, many homozygous animals for "q" haplotype bear a long and hairy coat rather than a woolly coat as expected (supplementary fig. S5, Supplementary Material online). Although the QTL_{Oar13} showed a significant effect within this genotypic group, itself as well as phenotypic errors may not explain these inconsistencies. These data suggest either incomplete penetrance or the implication of a third locus into the genetic determinism of the fleece variation in the Romane breed.

Given the large impact of the QTL_{Oar25} on the birth coat phenotype, we then genotyped the whole half-sib Romane protocol for the *asEIF2S2* insertion. The mutation segregated perfectly with the QTL_{Oar25} status for the 8 QTL rams since heterozygous sires were heterozygous for the insertion and homozygous rams were homozygous for the mutated $IRF2BP2^{asEIF2S2}$ allele (supplementary fig. S6, Supplementary Material online). The addition of genotypes as fixed effect into the GWAS model showed that 1) the initial significant signal disappeared emphasizing the potential causal role of this variant (supplementary fig. S7, Supplementary Material online) and 2) this polymorphism explained 65% of the fleece phenotypic variance. Moreover, we report, *via* the genotyping of this mutation in double coated sheep ($n = 193$) including ancestral species as Asiatic and Mediterranean Mouflon (*Ovis orientalis* and *Ovis musimon*) and in various single coated sheep breeds ($n = 402$), a perfect discrimination between both groups. The mutated allele was almost fixed in all woolly modern sheep (frequency [$IRF2BP2^{asEIF2S2}$] = 0.99) and *vice versa* for primitive hairy sheep (frequency [$IRF2BP2^{wt}$] = 0.96) (table 1). The few exceptions are more likely atypical phenotypes. Indeed, the classification of populations relies on the standard of the breed and not on

phenotypic measurements as we have done for the Romane breed. Altogether, these results demonstrated that the mutated $IRF2BP2^{asEIF2S2}$ allele is recessive to the wild type ancestral allele and is associated with the woolly coat phenotype in modern domestic breed sheep.

Insertion of EIF2S2 into IRF2BP2 Creates a dsRNA and Modulates Both Messengers

To dissect the molecular mechanism responsible for fleece variation, functional experiments were performed on skin samples of Romane lambs displaying distinct phenotypes and genotypes at the QTL_{Oar25} mutation. We firstly showed that the inserted retrogene was transcribed into a chimeric IRF2BP2/asEIF2S2 transcript in individuals carrying the $IRF2BP2^{asEIF2S2}$ allele (fig. 3b). This result suggested a potential role of this aberrant RNA to target the genuine sense EIF2S2 messenger thereby creating a double-stranded EIF2S2 RNA (EIF2S2 dsRNA) (fig. 3c). Therefore, we performed a EIF2S2 RNAse A protection assay by using this ribonuclease that cleaves only single stranded RNA (ssRNA) molecules. Messengers from both types of homozygous animals were RNAse A treated and we demonstrated the protection of a long endogenous EIF2S2 dsRNA in lambs exhibiting a single coated fleece and carrying the woolly $IRF2BP2^{asEIF2S2}$ allele (fig. 3d). Indeed, a significant difference is observed between both homozygous groups ($IRF2BP2^{wt}$ vs. $IRF2BP2^{asEIF2S2}$) for all primers pairs tested along the RNA–RNA hybrid suggesting that the whole complex seems protected by the enzymatic treatment likely due to its structural conformation (fig. 3d and supplementary fig. S8, Supplementary Material online). The formation of double-stranded RNA can trigger various mechanisms modulating gene expression. The three best supported ones are RNA masking, establishment of chromatin marks and RNA interference via the production of small interfering RNA (siRNA) as shown in mouse (Watanabe et al. 2008). Accordingly, we expected that the naturally formed EIF2S2 dsRNA might regulate especially genuine EIF2S2 transcripts which would produce the functional EIF2S2 protein. We tested this assumption by quantifying both EIF2S2 and IRF2BP2 mRNA accumulation and stability using real-time qPCR. Indeed, lambs homozygous for the mutated $IRF2BP2^{asEIF2S2}$ allele and owning a short and woolly coat showed a significant reduction of both the genuine EIF2S2 and IRF2BP2 messengers (fig. 3e).

Our analysis of fleece variation reveals a novel retroposition gene mechanism generating RNA–RNA hybrids. Retroposition seems a common process since a number of cases have been described in various organisms (Kaessmann et al. 2009) as shown for the TRIM5-CypA fusion gene in owl monkeys (Sayah et al. 2004) or the chimeric sphinx gene responsible of courtship behaviors in Drosophilia (Dai et al. 2008). Nevertheless, the contribution of retrogenes to cellular and organismal phenotypes remains poorly understood although recent studies have suggested that mammalian retrogenes would encode siRNAs which are important for the control of their own parental source genes (Tam et al. 2008; Watanabe et al. 2008). Thus, retrogenes do not necessarily represent evolutionary dead-ends, but might provide the raw material for functionally important evolutionary innovations like wool production. Indeed, such trait was strongly selected after domestication and became an important economical trait in the sheep industry corresponding to ancestral selection signature for genes related to wool properties (Ciani et al. 2015). Even though it is complex to estimate when this selective sweep occurred in the sheep populations history (Chessa et al. 2009), this singular mutation accounts for most if not all types of birth coat in sheep and discriminates perfectly wild species from woolly modern breeds. Furthermore, these results will contribute to a sustainable sheep production system since lambs carrying a hairy fleece are more adapted and robust to environmental harsh conditions (Allain et al. 2014). In addition, our findings pinpointed EIF2S2 and IRF2BP2 as novel master genes in the mammalian hair development. Keratins and Keratine Associated Proteins (KAPs) have been associated with most hair defaults so far in human disorders (Shimomura et al. 2010) as well as in sheep fiber anomalies (Li et al. 2009). However, results from Li et al. (2009) as others (Shimomura et al. 2010) suggested the involvment of a common regulatory process since several KAPs were underexpressed or absent in Merino sheep displaying a particular fleece phenotype. The functions of both interest genes likely support this assumption since EIF2S2 (Eukaryotic Translation Initiation Factor 2 Subunit Beta) encodes a translation factor and IRF2BP2 (Interferon Regulatory Factor 2 Binding Protein 2) acts as a transcriptional coregulator. Although further analyses need to be done to elucidate how EIF2S2 and IRF2BP2 regulate fleece formation, we believe that these new genes provide intriguing general insights into molecular basis of hair folliculogenesis and might open up new fields of research.

Materials and Methods

All experimental and analytical procedures are described in the Supplementary Materials and Methods file, Supplementary Material online.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

References

Allain D, Foulquie D, Autran P, Francois D, Bouix J. 2014. Importance of birthcoat for lamb survival and growth in the Romane sheep breed extensively managed on rangelands. *J Anim Sci.* 92:54–63.

Allain D, Schibler L, Mura L, Barillet F, Sechi T, Rupp R, Casu S, Cribiu E, Carta a. 2006. QTL detection with DNA markers for wool traits in a sheep backcross Sarda * Lacaune resource population. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production; 2006 Aug 13–18; Belo Horizonte, Minas Gerais, Brazil: WCGALP publisher. p. 5–7.

Browning BL, Browning SR. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet.* 84:210–223.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7.

Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, Kao RR, Pemberton JM, Beraldi D, Stear MJ, et al. 2009. Revealing the history of sheep domestication using retrovirus integrations. *Science* 324:532–536.

Ciani E, Lasagna E, D'Andrea M, Alloggio I, Marroni F, Ceccobelli S, Delgado Bermejo JV, Sarti FM, Kijas J, Lenstra JA, et al. 2015. Merino and Merino-derived sheep breeds: a genome-wide intercontinental study. *Genet Sel Evol.* 47:64.

Dai H, Chen Y, Chen S, Mao Q, Kennedy D, Landback P, Eyre-Walker A, Du W, Long M. 2008. The evolution of courtship behaviors through the origination of a new gene in Drosophila. *Proc Natl Acad Sci U S A.* 105:7478–7483.

Dalrymple BP, Kirkness EF, Nefedov M, McWilliam S, Ratnakumar A, Barris W, Zhao S, Shetty J, Maddox JF, O'Grady M, et al. 2007. Using comparative genomics to reorder the human genome sequence into a virtual sheep genome. *Genome Biol.* 8:R152.

Elsen J-M, Mangin B, Goffinet B, Boichard D, Le Roy P. 1999. Alternative models for QTL detection in livestock. I. General introduction. *Genet Sel Evol.* 31:213.

Fariello M-I, Servin B, Tosser-Klopp G, Rupp R, Moreno C, Cristobal MS, Boitard S. 2014. Selection signatures in worldwide sheep populations. *PLoS One* 9:e103813.

Hiendleder S, Kaupe B, Wassmuth R, Janke A. 2002. Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proc Biol Sci.* 269:893–904.

Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W, et al. 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344:1168–1173.

Kaessmann H, Vinckenbosch N, Long M. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet.* 10:19–31.

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, et al. 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* 10:e1001258.

Li SW, Ouyang HS, Rogers GE, Bawden CS. 2009. Characterization of the structural and molecular defects in fibres and follicles of the merino felting lustre mutant. *Exp Dermatol.* 18:134–142.

Pedrosa S, Uzun M, Arranz J-J, Gutiérrez-Gil B, San Primitivo F, Bayón Y. 2005. Evidence of three maternal lineages in Near Eastern sheep supporting multiple domestication events. *Proc Biol Sci.* 272:2211–2217.

Piper LR, Ruvinsky A. 1997. The genetics of sheep. Oxon, UK: CAB International. p. 471–504.

Ricordeau G, Tchamitchian L, Brunel JC, Nguyen TC, François D. 1992. La race ovine INRA 401: un exemple de souche synthétique. *INRA Prod Anim Hors-Série* 255–262.

Ryder ML. 1981. A survey of European primitive breeds of sheep.pdf. *Ann Génét Sél Anim.* 13:381–418.

Sayah DM, Sokolskaja E, Berthoux L, Luban J. 2004. Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. *Nature* 430:569–573.

Shimomura Y, Wajid M, Petukhova L, Kurban M, Christiano AM. 2010. Autosomal-dominant woolly hair resulting from disruption of keratin 74 (KRT74), a potential determinant of human hair texture. *Am J Hum Genet.* 86:632–638.

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, et al. 2008. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453:534–538.

Vitezica ZG, Moreno CR, Lantier F, Lantier I, Schibler L, Roig A, François D, Bouix J, Allain D, Brunel J-C, et al. 2007. Quantitative trait loci linked to PRNP gene controlling health and production traits in INRA 401 sheep. *Genet Sel Evol.* 39:421–430.

Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, et al. 2008. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453:539–543.

Zhou X, Carbonetto P, Stephens M. 2013. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet.* 9:e1003264.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet.* 44:821–824.