



HAL
open science

Détection de mutations homozygotes létales chez les petits ruminants

Maxime Ben Braiek

► **To cite this version:**

Maxime Ben Braiek. Détection de mutations homozygotes létales chez les petits ruminants. Biologie animale. Institut National Polytechnique de Toulouse - INPT, 2022. Français. NNT : 2022INPT0102 . tel-04155209

HAL Id: tel-04155209

<https://hal.inrae.fr/tel-04155209v1>

Submitted on 7 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Institut National Polytechnique de Toulouse

Présentée et soutenue par
Maxime BEN BRAIEK

Le 16 décembre 2022

Détection de mutations homozygotes létales chez les petits ruminants.

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingenieries**

Spécialité : **Infectiologie, Physiopathologie, Toxicologie, Génétique et Nutrition**

Unité de recherche :
GenPhySE- Unité Génétique, Physiologie et Systèmes d'Elevage

Thèse dirigée par
Stéphane FABRE et Carole MORENO

Jury

M. Didier BOICHARD, Rapporteur
M. Xavier MONTAGUTELLI, Rapporteur
Mme Juliette RIQUET, Examinatrice
M. Gwendal RESTOUX, Examineur
Mme Véronique BLANQUET, Examinatrice
M. Stéphane FABRE, Directeur de thèse



Détection de mutations homozygotes létales chez les petits ruminants

2019-2022

BEN BRAIEK Maxime



*A la mémoire de
Carole Moreno-Romieux*

Remerciements

Je tiens à remercier vraiment l'ensemble des personnes qui ont rendu ce travail possible. La recherche c'est avant tout un travail d'équipe et de collaboration.

Tout d'abord, je tiens à remercier la région Occitanie et APIS-GENE pour le financement de cette thèse (projet HOMLET).

Je tiens à remercier les membres du jury : Didier Boichard et Xavier Montagutelli pour avoir accepté d'évaluer mon travail de thèse en tant que rapporteurs, ainsi que les examinateurs Juliette Riquet, Gwendal Restoux et Véronique Blanquet.

Un grand merci à Dominique Pantalacci pour le suivi des doctorants et votre réactivité, ainsi qu'à l'ensemble de la direction de l'école doctorale SEVAB.

Une thèse ne peut se faire sans ses encadrants : Carole Moreno-Romieux et Stéphane Fabre. Merci à vous d'avoir cru en moi pendant cette thèse. Vous étiez à la croisée des mondes entre la génétique quantitative et la génétique moléculaire. Au cours de cette thèse, j'ai pu montrer que ces deux disciplines pouvaient facilement dialoguer ! Même si Carole, tu disais qu'on faisait des choses bizarres les "biologistes humides". Carole, c'est avec plein de tristesse que tu nous as quittés si rapidement, laissant un grand vide pour l'ensemble des membres du labo. . . Merci pour toutes les connaissances que tu m'as transmises. Même si tu n'as pas vu l'achèvement de cette thèse, j'espère que tu es très fière de ce qu'on a pu réaliser tous ensemble ! Stéphane, un grand merci, tu m'as appris énormément de choses pendant cette thèse. Tu es toujours disponible quand j'ai besoin de ton aide. Merci pour toutes les connaissances que tu m'as transmises. Merci pour les corrections du manuscrit. Je pense qu'au travers de cette thèse on est devenu de vrais chasseurs de mutations (même si tu avais déjà cette réputation) ! La rédaction de la thèse a été semée d'embûches mais on a réussi tous ensemble à mettre sur le papier ces beaux résultats acquis pendant ce travail de thèse. Je vais pouvoir me reposer un peu maintenant !

J'aimerais également remercier les membres de mon comité de thèse : Chris Hozé, Aurélien Capitan, Alain Ducos, Carine Capel, Isabelle Palhière et Jean-Michel Astruc. Mention spéciale pour Aurélien, qui nous a embarqués dans un beau projet pour valoriser nos résultats sur le

cornage/cryptorchidie. Merci Julie et Aurélie de participer à ce travail. Ce serait vraiment génial d'avoir une jolie carrière comme toi Aurélien sur les anomalies génétiques. J'espère qu'on aura de nouvelles occasions de pouvoir collaborer de près ou de loin sur ce sujet si passionnant ! Merci à Sébastien et Chris de m'avoir mis dans les mains le programme de déficit pendant mon stage de fin d'études, je crois que j'en ai fait un très bon usage.

Merci également à l'ensemble de l'équipe pédagogique de la spécialisation Filières Animales Durables de Bordeaux Sciences Agro : Marie-Pierre Ellies, Hervé Jacob, Guillaume Durand et Alain Ducos, de m'avoir toujours encouragé dans mes choix de faire de la recherche. Petite pensée pour Guillaume, j'ai largement utilisé Inskape pour faire les schémas de mon manuscrit. Alain, qui m'a suivi tout au long de ma "petite carrière" dans la recherche : tuteur de mon premier stage de 2A d'école d'ingénieur agronome, stage 3A et membre de mon comité de thèse.

Merci aux membres du projet PRESAGE, j'espère vraiment que ce projet va aboutir à créer cet observatoire en petits ruminants. Je tiens également à remercier Maria et Marcel (CRAG, Barcelone) pour leur collaboration dans le cadre du projet VarGoats. Merci à Cord Drögemüller (VetSuisse Faculty, Berne) pour sa collaboration sur les mutations létales. Merci à l'ensemble des professionnels des filières, sans vous ce travail de recherche n'aurait pas avancé aussi vite. Je tiens à remercier tout particulièrement Céline et Francis (CDEO) pour leur confiance, c'était un vrai plaisir de travailler ensemble. C'est une immense fierté d'avoir résolu le problème de la cryptorchidie qui vous posait tant de problèmes depuis plus de 50 ans en race Manech Tête Rousse. Un grand merci aux éleveurs qui ont accepté de réaliser les croisements à risque chez eux.

Un grand merci à l'équipe GenROC, vous êtes une équipe au top ! C'était vraiment un bonheur de venir au labo tous les matins. Merci à Julien, Florence et Florent de m'avoir appris tellement de choses à la paillasse et m'avoir aidé à réaliser les manip. Merci à Gwen, Carine, Laurence pour leur soutien et connaissances transmises. Mention spéciale pour Laurence. C'est avec toi que j'ai réalisé mon premier stage en 2A d'ingénieur. A l'époque je n'avais jamais tenu une pipette dans mes mains et puis maintenant me voilà en train de finir la rédaction de ma thèse. C'est grâce à toi que j'ai découvert ce monde de la recherche et j'espère pouvoir continuer le plus longtemps possible. Merci pour ta bienveillance, ton soutien et ton gros travail de relecture de ce manuscrit. J'espère vous avoir régalés avec mes pâtisseries !! Je tiens également à remercier les membres de l'équipe GeSPR qui m'ont aidé tout au long de ma thèse. Petite pensée pour toi Jean-Louis, c'était un plaisir de partager ton bureau pendant mon stage et le début de ma thèse. Merci à l'équipe de Langlade : Fred, Arnaud, Némuel et François d'avoir pris soins de mes brebis et agneaux ! Merci aux personnes avec lesquelles j'ai pu interagir tout au long de cette thèse : Diane Buisson, Philippe Bardou, Marjorie Chassier, Virginie Clément, Hélène Larroque, Rachel Rupp, Jérôme Raoul, Isabelle Palhière, Hélène Gilbert ...

Un énorme merci à tous les membres de l'unité GenPhySE. Merci aux membres du labo de m'avoir accompagné pendant cette période : Julie, Sophie, Pitou, Chloé, Juliette, Martine,

Guillaume, Katia, Laure, Lisa, Nathalie I., Nathalie M., Maguy, Jean-Noël, Patrick, Vincent, le personnel administratif de GenPhySE. . . En espérant n'oublier personne.

Merci à Soline Szymczak d'avoir contribué aux résultats de ma thèse. C'était vraiment agréable de t'accueillir dans notre équipe. Au travers de ton stage de M2, j'ai pu voir qu'encadrer quelqu'un n'est pas si simple que ça !

Merci à l'ensemble des thésards et stagiaires de l'unité GenPhySE pour la très bonne ambiance au sein du laboratoire pendant ces années. Merci aux "vieux" thésards, CDD et stagiaires : Sophie, Audrey, Vanille, Tiphaine, Claire, Estelle T., Louise, Emilie, Marc, Harmonie, Océane, Estelle J., Quentin, Marie et mon super compère de stage Amaury, ainsi que l'ensemble des autres stagiaires pour ces moments inoubliables !! Ainsi qu'aux doctorants en fin de thèse comme moi : Vincent et Guillermo. Courage aux autres doctorants, le dernier coup de collier pour la rédaction du manuscrit n'est pas simple mais ça en vaut la peine !!

Merci à ma famille de m'avoir toujours soutenu dans cette épreuve. Merci à mes parents et ma sœur Justine d'être toujours présents quand j'ai besoin de vous ! Même si la distance n'est pas toujours simple pour vous voir. Une pensée pour mon grand-père qui nous a quitté il y a bientôt deux ans, j'espère que de là-haut tu es fier de ce que j'ai pu accomplir ! Et si j'ai réussi avec succès ce travail de thèse, c'est peut-être parce que "tu tiens ça de ton grand-père" comme tu me le disais si souvent.

Merci à Chloé d'être rentrée dans ma vie il y a maintenant bientôt deux ans. Tu es une personne exceptionnelle, tu as changé ma vie, jamais je n'aurai imaginé pouvoir rencontrer une fille comme toi. Merci pour ton aide sur le manuscrit et ton soutien pendant cette période. Dans un an ça sera à ton tour de soutenir ta thèse, courage ! Je serai toujours là pour te soutenir comme tu l'as fait pour moi. Un grand merci à ma belle-famille de m'avoir accueilli et pour leur soutien tout au long de cette thèse. Et puis, je ne t'oublie pas Lilou, toujours là pour jouer à la balle et demander des caresses, cela m'a permis de me distraire pendant l'écriture de ce manuscrit.

Résumé

Les anomalies génétiques, définies comme des phénotypes plus ou moins délétères et déviant par rapport à une population, résultent de mutations naturelles uniques dans l'ADN, rares et héréditaires. On estime que chaque individu est porteur de 1 à 5 mutations hautement délétères dans son génome. Pour identifier une anomalie génétique, l'approche classique repose sur l'étude des phénotypes délétères et de l'ADN des individus atteints. Les différents outils de génotypage, l'analyse d'association et le séquençage de génomes complets permettent de mettre en évidence les mutations causales associées. Cependant, lorsque la mutation est létale, notamment au stade embryonnaire, les échantillons d'ADN ne sont pas disponibles. Ainsi pour les animaux d'élevage, d'autres approches dites de génétique inverse, permettent de tirer parti des nombreux génotypes obtenus dans le cadre de la sélection génomique. Des analyses statistiques dédiées de ces génotypes permettent de rechercher des régions du génome en déficit d'animaux homozygotes, supposées porter des mutations récessives létales. En effet, le déficit d'homozygotes s'expliquerait par la mort précoce des animaux avant leur génotypage.

Le premier objectif de cette thèse a été d'étudier la faisabilité d'une telle approche dans les populations de petits ruminants engagées dans un programme de sélection génomique. Cette méthode a permis de mettre en évidence 13 régions (ou haplotypes) indépendantes uniquement en races ovines Lacaune (LDHH) et Manech Tête Rousse (MTRDHH), mais pas dans les autres races laitières ovines ou caprines. Le deuxième objectif a été d'étudier leurs effets sur les caractères de reproduction et de production. L'analyse des caractères de fertilité et de mortinatalité permet de confirmer que certains des haplotypes détectés sont associés à de la létalité au cours du développement embryonnaire (LDHH1, 2 et 9 ; MTRDHH2), ou autour de la naissance (LDHH3 et 6, et MTRDHH1). Les autres haplotypes sont supposés conduire à de la mortalité plus tardive (MTRDHH3) ou être associés à des défauts morphologiques contre-sélectionnés. Avec des fréquences de porteurs variant de 3% à 16%, certains haplotypes offrent un potentiel avantage sélectif des béliers porteurs hétérozygotes, notamment pour la production laitière augmentée de leurs filles (LDHH, 3, 9 et 11 ; MTRDHH2). Le troisième objectif consistait à identifier les mutations causales sous-jacentes de ces régions en déficit par l'analyse des séquences de génomes entiers des hétérozygotes. Parmi les 13 haplotypes, des mutations candidates "perte de fonction" ont été identifiées dans 9 gènes, *ORC5* pour LDHH1, *IDI1* pour LDHH2, *PREB* et *GPN1* pour LDHH3, *CCDC65* pour LDHH6, *EDC3* pour LDHH9, *MMUT* pour MTRDHH1, *SLC33A1* pour MTRDHH2 et *FCGR1A* pour MTRDHH3. De façon à prouver le caractère récessif léthal, des

accouplements à risque ont été réalisés entre animaux hétérozygotes pour générer des individus homozygotes. La mortalité chez le jeune agneau homozygote de quelques jours a pu être prouvée pour les mutations dans *CCDC65* et *MMUT*. Quant à la mutation dans *SLC33A1*, elle entraînerait une perte embryonnaire dans le premier mois de gestation. Par ailleurs, trois haplotypes semblent être associés à des mutations de gènes déjà connus chez les ovins en lien avec la sensibilité aux mammites (*SOCS2* dans LDHH11), la couleur de toison noire (*ASIP* dans MTRDHH5) et l'absence de cornes elle-même liée à de la cryptorchidie (*RXFP2* dans MTRDHH4).

L'ensemble de ce travail de thèse apporte de nouvelles connaissances sur les anomalies génétiques existantes dans les populations ovines. La gestion de ces anomalies par la connaissance des phénotypes affectés et le génotypage systématique des mutations causales permettra d'améliorer la fertilité globale, de limiter la mortalité des agneaux, et plus généralement d'améliorer la santé et le bien-être des animaux de ces populations.

Mots-clés : Petits ruminants, génomique, anomalies génétiques, mutations délétères, génétique inverse.

Abstract

Genetic defects, defined as phenotypes more or less deleterious and deviating from a population, result from unique natural mutations in DNA. They are rare and hereditary. It is estimated that each individual carries 1 to 5 highly deleterious mutations in its genome. To identify a genetic defect, the conventional approach is based on the study of the deleterious phenotypes and DNA of affected individuals. The various genotyping tools, association analysis and sequencing of whole genomes make it possible to highlight the associated causal mutations. However, when the mutation is lethal, especially in the embryonic stage, DNA samples are not available. For example, in livestock, other approaches known as reverse genetics take advantage of the numerous genotyping data obtained in the context of genomic selection. Dedicated statistical analyses of these genotypes make it possible to search for regions of the genome lacking homozygous animals, thus supposed to carry lethal recessive mutations. Indeed, the lack of homozygous is explained by the early death of animals before their genotyping.

The first objective of this thesis was to study the feasibility of such an approach in small ruminant populations engaged in a genomic selection program. This method has highlighted 13 independent regions (or haplotypes) only in Lacaune (LDHH) and Manech Tête Rouse (MTRDHH) sheep breeds but not in other dairy sheep or goats. The second objective was to study their effects on reproductive and production traits. Analysis of fertility and stillbirth rate characteristics confirms that some of the haplotypes detected are associated with lethality during embryonic development (LDHH1, 2 and 9; MTRDHH2), or around birth (LDHH3 and 6; MTRDHH1). The other haplotypes are assumed to lead to later mortality (MTRDHH3) or to be associated with counter-selected morphological defects. With carrier frequencies ranging from 3% to 16%, some haplotypes offer a selective advantage of heterozygous carrier rams, especially for the increased milk production of their daughters (LDHH, 3, 9 and 11; MTRDHH2). The third objective was to identify the underlying causal mutations in these homozygous deficient regions by analysis of whole genome sequences of heterozygous. Of the 13 haplotypes, candidate “loss of function” mutations were identified in 9 genes, *ORC5* for LDHH1, *IDI1* for LDHH2, *PREB* and *GPN1* for LDHH3, *CCDC65* for LDHH6, *EDC3* for LDHH9, *MMUT* for MTRDHH1, *SLC33A1* for MTRDHH2 and *FCGR1A* for MTRDHH3. In order to prove the lethal recessive characteristic of these mutations, at-risk mating was made by crossing heterozygous animals to generate homozygous individuals. Lethality of newborn homozygous lambs has been proven for mutations in *CCDC65* and *MMUT*. Concerning the mutation in *SLC33A1*, it would cause embryonic loss in

the first month of gestation. In addition, three haplotypes appear to be associated with mutations of genes already known in sheep in relation to mastitis sensitivity (*SOCS2* in LDHH11), black fleece (*ASIP* in MTRDHH5) and the absence of horns itself related to cryptorchidism (*RXFP2* in MTRDHH4).

Overall, this thesis work brings new knowledge on the genetic defects existing in the ovine populations. Managing these defects through knowledge of affected phenotypes and systematic genotyping of causal mutations will improve global fertility, limit lamb mortality, and more generally will improve the health and animal welfare of these populations under selection.

Keywords : Small ruminants, genomics, genetic defects, deleterious mutations, reverse genetics.

Table des matières

I	Introduction bibliographique : les anomalies génétiques en petits ruminants	
1	Comprendre l'émergence des anomalies génétiques dans les populations de petits ruminants	25
1.1	Données générales sur les filières de petits ruminants	25
1.1.1	Chiffres clés sur les filières de petits ruminants en France	25
1.1.2	Description des cheptels ovins laitiers	26
1.1.3	Description des cheptels caprins laitiers	28
1.2	Organisation de la sélection génétique	29
1.2.1	Les acteurs de la sélection	29
1.2.2	Le contrôle de performances et les caractères en sélection	30
1.2.3	De la sélection classique à la sélection génomique	33
1.2.3.1	<i>Généralités</i>	33
1.2.3.2	<i>Les schémas ovins laitiers à l'ère de la génomique</i>	35
1.2.3.3	<i>Les schémas caprins laitiers à l'ère de la génomique</i>	38
1.3	Des structures de populations favorables à l'émergence d'anomalies génétiques	40
1.3.1	Les populations d'élevage ont des effectifs génétiques limités	40
1.3.2	Dérive génétique et consanguinité	42
2	Les anomalies génétiques : définitions, modes de transmission et d'action, évolution des fréquences	43
2.1	L'ADN, source de variabilité génétique	43
2.1.1	L'ADN, support de l'information génétique	43
2.1.2	Les différents polymorphismes de l'ADN et leurs impacts biologiques	47
2.1.2.1	<i>Les variants de une à quelques bases</i>	47
2.1.2.2	<i>Les variants structuraux</i>	49

2.1.2.3	Occurrence des différents types de variants en lien avec des anomalies . . .	49
2.2	Les anomalies en petits ruminants et leurs modes de transmission	50
2.2.1	Les anomalies génétiques identifiées en petits ruminants	51
2.2.2	Anomalies autosomiques récessives et co-dominantes	61
2.2.3	Anomalies autosomiques dominantes	62
2.2.4	Anomalies liées aux chromosomes sexuels	63
2.2.5	Anomalies mitochondriales	64
2.2.6	Anomalies chromosomiques	65
2.2.7	Anomalies avec déterminisme génétique complexe	67
2.3	Les mécanismes influençant la fréquence des variants	71
2.3.1	Effet de la dérive génétique	71
2.3.2	Sélection équilibrante : avantage des hétérozygotes	71
3	Un contexte favorable à l'étude des anomalies génétiques dans les populations d'élevage	73
3.1	Une meilleure connaissance des génomes et le développement des technologies haut-débit facilitent l'étude des anomalies	73
3.1.1	Les différentes technologies pour séquencer un génome	73
3.1.1.1	<i>Le séquençage de première génération : le séquençage Sanger</i>	<i>73</i>
3.1.1.2	<i>Le séquençage de deuxième génération : le séquençage en lectures courtes</i>	<i>74</i>
3.1.1.3	<i>Le séquençage de troisième génération : le séquençage en lectures longues</i>	<i>75</i>
3.1.2	Construction d'une séquence de référence <i>de novo</i>	77
3.1.3	Traitements bioinformatiques des données de séquence	81
3.1.3.1	<i>Alignements des lectures et détection de variants</i>	<i>81</i>
3.1.3.2	<i>Prédiction de l'effet des variants</i>	<i>82</i>
3.1.4	Le génotypage	83
3.1.4.1	<i>Principe et fonctionnement des puces de génotypages</i>	<i>83</i>
3.1.4.2	<i>Les différentes puces disponibles en petits ruminants</i>	<i>84</i>
3.1.4.3	<i>Utilisation des puces dans le cadre de la sélection génomique</i>	<i>86</i>
3.2	Des dispositifs d'hérédo-surveillance facilite la remontée de cas	87
4	Les différentes méthodes pour identifier les anomalies génétiques . . .	89
4.1	Est-ce que l'anomalie est d'origine génétique ?	89
4.2	Du phénotype au génotype : méthodes de génétique classique	90
4.2.1	Approche gènes candidats	90
4.2.2	Détection de QTL : Analyses de liaison/association tout génome	90
4.2.3	Cartographie par homozygotie	92
4.3	Du génotype au phénotype : la génétique inverse	94
4.3.1	Exploitation des données de séquence pour identifier des variants délétères rares	94
4.3.2	Identification de régions du génome présentant des déficits d'animaux homozygotes à partir de données de génotypage	94

5	Situation du sujet et objectifs de la thèse	103
----------	--	------------

II

Identification d'haplotypes en déficit d'animaux homozygotes à partir des données de génotypages

6	Méthodes	111
6.1	Description des animaux génotypés	111
6.2	Contrôle qualité, imputation et phasage des génotypages	111
6.3	Identification des haplotypes en déficit d'animaux homozygotes	113
7	Résultats : Identification de DHH en petits ruminants	117

III

Effets des haplotypes sur les caractères

8	Fertilité et mortinatalité	129
8.1	Méthodes	129
8.2	Résultats : confirmation de la létalité	131
9	Caractères en sélection	135
9.1	Méthodes	135
9.1.1	Analyse des Daughter Yield Deviation	135
9.1.2	Analyse de l'Index de Synthèse Ovins Laitiers en Manech Tête Rousse	136
9.2	Résultats : potentiel avantage sélectif des animaux hétérozygotes aux DHH136	
10	Évolution des fréquences des haplotypes	141
10.1	Évolution de la fréquence des LDHH en Lacaune	141
10.2	Évolution de la fréquence des MTRDHH en Manech Tête Rousse	143

IV

Recherche des gènes candidats et des mutations pour chaque haplotype

11	Recherche de gènes candidats	149
11.1	Méthodes	149
11.2	Résultats : identification de gènes candidats fonctionnels	149
12	Recherche des mutations candidates	155
12.1	Méthodes	155
12.1.1	Description des données de séquençage	155
12.1.2	Détection de variants et annotation	156
12.1.3	Identification des mutations candidates	156
12.2	Résultats : identification des mutations candidates localisées dans les DHH	159
12.2.1	Description des mutations identifiées dans les régions associées à de la mortalité embryonnaire	163
12.2.2	Description des mutations identifiées dans les régions associées à de la létalité pendant toute la gestation	166
12.2.3	Description des mutations identifiées dans les régions associées à de la mortalité périnatale	170
12.2.4	Description des mutations identifiées dans les régions associées à de la mortalité juvénile ou à des défauts morphologiques	176
12.3	Développement de tests de géotypages spécifiques aux mutations candidates	184
12.3.1	Développement d'un test spécifique de géotypage	184
12.3.2	Validation à grande échelle des mutations candidates	184
12.3.2.1	<i>Haplotypes en lien avec des mutations supposées létales</i>	184
12.3.2.2	<i>Haplotypes en lien avec des mutations connues en ovins associées à des caractères de morphologie</i>	184

V

Validation de la causalité des mutations candidates

13	Anomalies génétiques induisant de la mortalité pendant la gestation	193
13.1	Croisements <i>ORC5</i> et <i>ID11</i> en Lacaune	193
13.1.1	Accouplements à risque et diagnostic de gestation par échographie	193
13.1.2	Choix d'un test moléculaire de gestation précoce	194
13.1.2.1	<i>Dosage de la progestérone</i>	195

13.1.2.2	Expression par qPCR des gènes <i>MX1</i> et <i>STAT1</i>	195
13.1.3	Analyse du pouvoir de diagnostic du test moléculaire par qPCR sur l'état de gestation	197
13.1.3.1	Analyses statistiques	197
13.1.3.2	Résultats	197
13.2	Croisements <i>SLC33A1</i> en Manech Tête Rousse	200
13.2.1	Dispositif expérimental d'accouplements des porteurs <i>SLC33A1_dupG</i>	200
13.2.2	Réussite à l'IA et mortalité périnatale	200
13.2.3	Pertes fœtales	203
14	Anomalies génétiques induisant de la mortalité périnatale	207
14.1	Croisements <i>MMUT</i> en Manech Tête Rousse	207
14.1.1	Dispositif expérimental d'accouplements des porteurs du SNV <i>MMUT</i>	207
14.1.2	Mortalité périnatale	207
14.1.3	Expression et activité de <i>MMUT</i>	211
14.2	Croisements <i>CCDC65</i> en Lacaune	213
14.2.1	Dispositif expérimental d'accouplements des porteurs du SNV <i>CCDC65</i>	213
14.2.2	Mortalité périnatale	213
15	Anomalies génétiques en lien avec des défauts morphologiques	217
15.1	<i>SOCS2</i> : stature et sensibilité aux mammites	217
15.2	<i>ASIP</i> : coloration de la toison	220
15.3	<i>RXFP2</i> : cornage et cryptorchidie	224
15.3.1	Ségrégation de la mutation cornage en Manech Tête Rousse	225
15.3.2	Etude de l'expression de <i>RXFP2</i>	230
15.3.3	Identification d'une mutation causale de la cryptorchidie	232
15.3.4	Lien entre <i>MTRDHH4</i> et cryptorchidie	239

VI

Discussion générale et perspectives

16	Génétique inverse pour la recherche de déficit d'animaux homozygotes chez les petits ruminants	249
16.1	Existence d'haplotypes en déficit d'animaux homozygotes dans les races ovines laitières	249
16.2	Réflexion sur la méthode de recherche de déficit	250
16.3	Quelles sont les opportunités futures ?	251

17	Les mutations dans les haplotypes en déficit	253
17.1	Découverte de nombreuses mutations récessives candidates	253
17.2	Limites de la méthode utilisée pour rechercher les mutations	255
18	Causalité des mutations	257
18.1	Mutations létales embryonnaires	257
18.2	Mutations létales périnatales	259
18.2.1	Mutation dans <i>MMUT</i>	259
18.2.2	Mutation dans <i>CCDC65</i>	260
18.3	Mutations en lien avec la morphologie	261
18.4	Les mutations non étudiées	263
18.5	Pertinence des mutations identifiées en tant que modèles animaux de pathologies	264
19	Gestion dans les schémas de sélection	267
19.1	Évolution de la fréquence des mutations	267
19.2	Les mutations identifiées sont-elles spécifiques de races?	268
19.3	Comment gérer les anomalies génétiques dans les schémas de sélection?	270

VII

Productions scientifiques

20	Articles scientifiques	275
21	Communication lors de congrès	377

VIII

Annexes

IX

Références bibliographiques

Introduction bibliographique : les anomalies génétiques en petits ruminants

1	Comprendre l'émergence des anomalies génétiques dans les populations de petits ruminants	25
1.1	Données générales sur les filières de petits ruminants	
1.2	Organisation de la sélection génétique	
1.3	Des structures de populations favorables à l'émergence d'anomalies génétiques	
2	Les anomalies génétiques : définitions, modes de transmission et d'action, évolution des fréquences	43
2.1	L'ADN, source de variabilité génétique	
2.2	Les anomalies en petits ruminants et leurs modes de transmission	
2.3	Les mécanismes influençant la fréquence des variants	
3	Un contexte favorable à l'étude des anomalies génétiques dans les populations d'élevage	73
3.1	Une meilleure connaissance des génomes et le développement des technologies haut-débit facilitent l'étude des anomalies	
3.2	Des dispositifs d'héredo-surveillance facilite la remontée de cas	
4	Les différentes méthodes pour identifier les anomalies génétiques	89
4.1	Est-ce que l'anomalie est d'origine génétique ?	
4.2	Du phénotype au génotype : méthodes de génétique classique	
4.3	Du génotype au phénotype : la génétique inverse	
5	Situation du sujet et objectifs de la thèse	103

1. Comprendre l'émergence des anomalies génétiques dans les populations de petits ruminants

Les objectifs de ce chapitre sont de décrire les différentes races françaises de petits ruminants, de comprendre comment est organisée la sélection et pourquoi ces populations sont favorables à l'émergence d'anomalies génétiques.

1.1 Données générales sur les filières de petits ruminants

L'émergence de l'agriculture a été initiée par la domestication d'espèces animales et végétales sauvages. Les petits ruminants ont été domestiqués au cours de la période néolithique dans le croissant fertile entre 11 000 - 9 000 ans avant Jésus-Christ (McHugo, Dover, & MacHugh, 2019). Les moutons (*Ovis aries*) et chèvres (*Capra hircus*) modernes sont originaires du mouflon asiatique (*Ovis orientalis*) et de la chèvre sauvage bezoar ibex (*Capra aegagrus*). Au cours de l'histoire, l'homme a sélectionné progressivement les animaux pour répondre à ses besoins pour l'alimentation (production de lait, viande), l'habillement (toison, cuir) et le travail (docilité, force motrice) (Alberto et al., 2018; Chessa et al., 2009; Deng et al., 2020; Zheng et al., 2020). L'évolution des phénotypes contribue à ce jour en une grande biodiversité (plus de 200 races reconnues dans le monde entier pour chacune des espèces) et les animaux sont répartis sur l'ensemble du globe, adaptés à de nombreux environnements très variés allant des zones montagneuses froides et humides du nord de l'Europe aux régions arides et chaudes d'Afrique et d'Asie. On dénombre à ce jour 1,2 milliards d'ovins et 1 milliard de caprins dans le monde (FAO, <https://www.fao.org/livestock-systems/global-distributions/en/>).

1.1.1 Chiffres clés sur les filières de petits ruminants en France

En France, on observe une grande diversité des systèmes de production associée à un grand nombre de races de petits ruminants. Sur notre territoire, 57 races ovines (52 allaitantes et 5 laitières) et 14 races caprines (laitières) sont officiellement reconnues (Prache et al., 2018).

En 2020, le cheptel ovin français se classe au 4ème rang européen avec 7,3 millions d'animaux répartis dans 39 milliers d'exploitations, loin derrière l'Espagne avec plus de 15 millions d'ovins (Agreste, 2021, 2022). Au niveau national, ce cheptel est composé de 30% d'animaux laitiers et 70% d'animaux allaitants. En 2021, la production ovine française pour la consommation de viande se classe au 5ème rang européen avec 4,3 millions de têtes (85% ovins allaitants et 15% ovins laitiers de réforme) ~ 81 000 tonnes équivalent carcasse (recul de -12,4% en dix ans) (IDELE & CNE, 2021b; Agreste, 2022). Pour la production laitière, la France se classe au 4ème rang européen avec 297 millions de litres produits (Agreste, 2021). La production est exclusivement localisée dans les régions Occitanie, Nouvelle-Aquitaine et Corse, et est en grande partie transformée en

fromages par l'industrie laitière (Roquefort et pâtes pressées non cuites principalement).

Le cheptel caprin français quant à lui est principalement destiné à la production laitière (très peu de consommation de viande de chevreaux en Europe contrairement en Asie ou Afrique). En 2020, le cheptel est composé de 1,4 million de têtes réparties dans 10,8 milliers d'exploitations (4ème rang européen) (Agreste, 2021). Deux pays se classent largement en tête : la Grèce (3,5 millions de têtes) et l'Espagne (2,6 millions). Cependant, la France se classe 1ère des pays européens pour la production laitière avec 508 millions de litres de lait collectés. Les élevages sont répartis globalement sur l'ensemble du territoire même si 35% des caprins sont élevés dans la région Nouvelle-Aquitaine (Agreste, 2020). Le lait est fortement valorisé puisque 25% du lait produit est directement transformé à la ferme (valorisation en AOP, AOC, IGP ...) tandis que 75% sont livrés aux industriels (IDELE & CNE, 2021a).

Les exploitations laitières de petits ruminants ont des effectifs importants, plus de 75% des exploitations ont plus de 200 animaux en production. Par la suite, un intérêt particulier sera donné aux races laitières.

1.1.2 Description des cheptels ovins laitiers

En France, le cheptel ovins laitier se compose de 5 races (IDELE, 2022b) : la Lacaune (60,5% de l'effectif national), la Manech Tête Rousse (23,0%), la Basco-Béarnaise (8,3%), la Manech Tête Noire (5,6%) et la Corse (2,6%), réparties dans trois grands bassins de production (Figure 1.1 ; Table 1.1).

Dans le rayon de Roquefort, on retrouve la Lacaune avec plus de 800 000 brebis. Une grande partie du lait est valorisée pour être transformée en AOP *Roquefort*. La singularité de la race Lacaune repose sur l'existence de deux rameaux : laitier (800 000 brebis) et allaitant (300 000 brebis) à partir des années 1970. La Lacaune est originaire du Massif-Central et est caractérisée par une tête longue et fine, une toison avec des poils fins et par une absence de cornes.

Dans le bassin des Pyrénées-Atlantiques, on retrouve les trois races ovines laitières des Pyrénées (ROLP) : la Manech Tête Rousse, la Manech Tête Noire et la Basco-Béarnaise. Le lait sert à la production de fromage AOC *Ossau-Iraty*. La plupart des exploitations de Basco-Béarnaises sont fromagères avec une production directe à la ferme. Il existe également une valorisation des agneaux de lait par le label "agneau de lait des Pyrénées Label Rouge" depuis 1996, le reste de la production est exporté en Espagne pour engraissement. Ces races pyrénéennes sont adaptées à vivre dans des zones difficiles. On retrouve la Manech Tête Rousse dans les coteaux tandis que la Manech Tête Noire est localisée en altitude dans la montagne basque. La Basco-Béarnaise est localisée dans la partie béarnaise des Pyrénées et en Haute-Saoule. Ces races présentent une longue toison ancestrale jarreuse. Les races Manech Tête Noire et Basco-Béarnaise présentent des cornes enroulées tandis que chez la Manech Tête Rousse, on retrouve des animaux sans cornes, avec cornillons ou cornus. La race Basco-Béarnaise est caractérisée par une tête busquée. Toutes ces races sont adaptées à la transhumance l'été. Les Manechs ont été importées du pays basque espagnol dans les années 1970 et sont très apparentées génétiquement aux races Latxa espagnoles (Legarra et al., 2014).

Dans le bassin Corse, la race Corse est caractérisée par son adaptation à des conditions

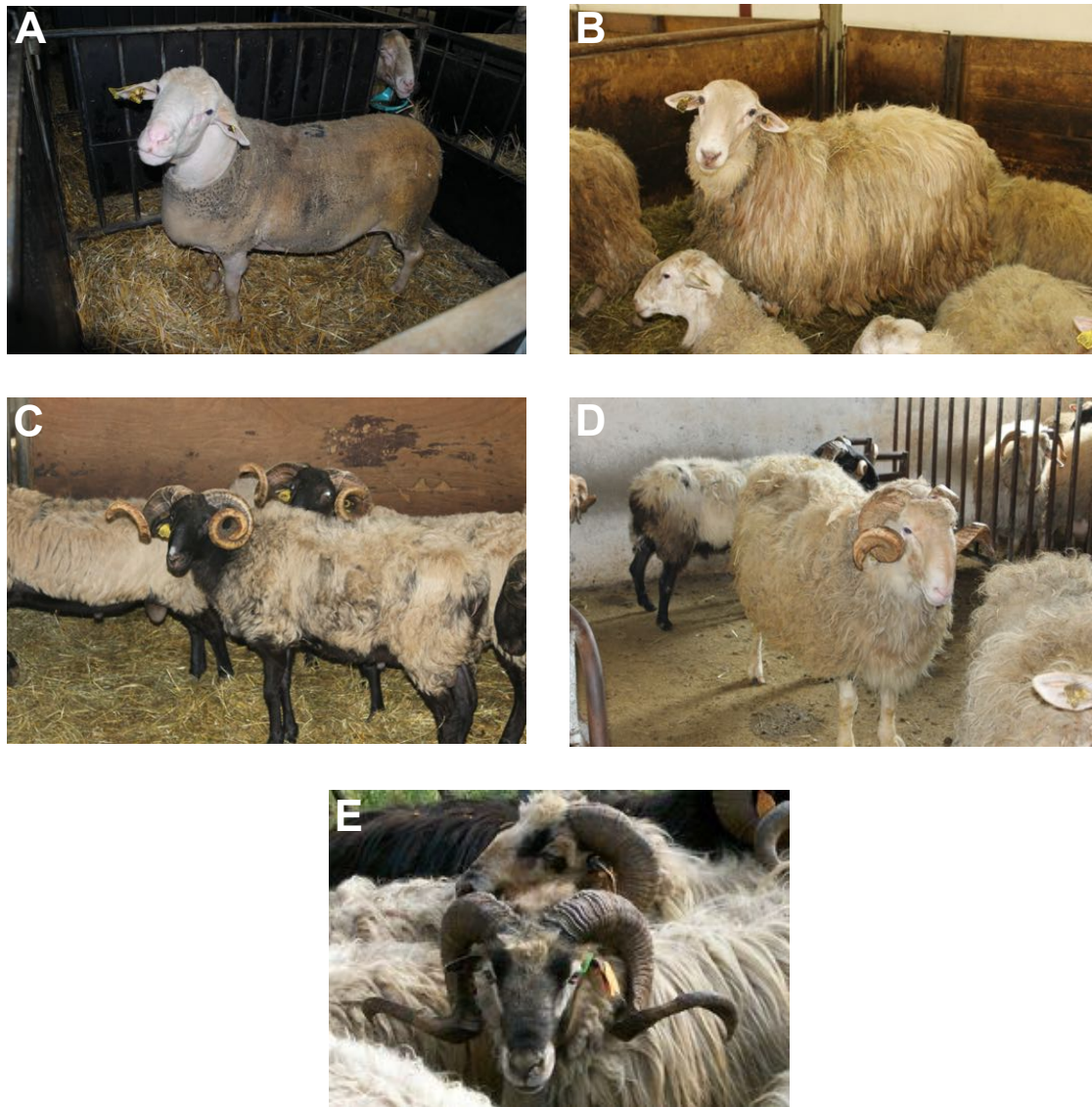


FIGURE 1.1 – **Races ovines laitières françaises.** (A) Bélière Lacaune; (B) Bélière Manech Tête Rousse; (C) Bélière Manech Tête Noire; (D) Bélière Basco-Béarnaise; (E) Bélière Corse. *Crédits photos : D. Buisson (IDELE), OS Corse.*

climatiques difficiles et son lait est valorisé en AOC *Brocciu*. Elle présente également une toison jarreuse et des cornes.

TABLE 1.1 – Principales races ovines laitières françaises. CLO = Contrôle Laitier Officiel ; *Total des deux entreprises de sélection en race Lacaune lait (Lacaune Ovitest et Service Elevage de la Confédération de Roquefort) (IDELE, 2022b)

Bassin	Roquefort	Pyrénées-Atlantiques			Corse
Race	Lacaune	Manech Tête Rousse	Manech Tête Noire	Basco-Béarnaise	Corse
Effectif total des brebis	890 000*	270 000	120 000	80 000	100 000
Effectif de brebis en CLO	172 495*	65 676	7 523	23 743	15 841
Nombre de troupeaux en CLO	372	327	67	98	51
Production moyenne (durée lactation)	346 litres (176 jours)	251 litres (164 jours)	184 litres (149 jours)	224 litres (148 jours)	153 litres (188 jours)

1.1.3 Description des cheptels caprins laitiers



FIGURE 1.2 – Principales races caprines laitières. (A) Chèvres de race Alpine ; (B) Chèvres de race Saanen. *Crédits photos : Capgènes*

En France, le cheptel caprin laitier se compose de 2 races principales : l'Alpine (62,3% de l'effectif national) et la Saanen (34,2%) (Figure 1.2; Table 1.3). La race Alpine est originaire de Savoie et se retrouve majoritairement dans la vallée de la Loire, de la Saône, du Rhône et dans le Poitou-Charente. Il s'agit de la race la plus présente sur le territoire français. L'Alpine est une race rustique capable de pâturer dans les zones montagneuses. Les animaux ont des poils ras et une couleur chamoisée (Capgènes, 2018a). La race Saanen est la race la plus représentée

dans le monde. En France, on la retrouve principalement dans le Sud-Est, le Centre et l'Ouest de la France. Elle est caractérisée par une toison blanche composée de poils courts et denses (Capgènes, 2018b). Les principaux fromages AOP produits sont le Sainte-Maure de Touraine (26%), le Rocamadour (18%), le Selles-sur-Cher (15%) et le Chavignol (13%) (IDELE & CNE, 2021a).

TABLE 1.3 – **Principales races caprines laitières françaises**. Remarque : En caprins, la production laitière est exprimée à 250 jours (durée de référence) pour compenser l'hétérogénéité des durées de lactation (beaucoup de lactations longues). CLO = Contrôle Laitier Officiel (IDELE, 2022a)

Race	Alpine	Saanen
Effectif total des chèvres	450 000	350 000
Effectif de chèvres en CLO	146 854	80 750
Nombre de troupeaux en CLO	372	327
Production sur 250 jours	822 kg	818 kg
Production moyenne (durée de lactation)	978 kg (317 jours)	1 034 kg (329 jours)
Taux protéique (TP)/ Taux butyreux (TB) en g/kg	34,0/38,3	32,9/36,7

1.2 Organisation de la sélection génétique

Chaque race laitière citée précédemment fait l'objet d'une sélection orchestrée par un programme de sélection spécifique. Initialement basée sur les performances phénotypiques des futurs reproducteurs et des apparentés, la sélection s'est fortement améliorée avec le développement de la génomique.

1.2.1 Les acteurs de la sélection

En race Lacaune, les deux entreprises de sélection (ES) Ovitest et le service élevage de la Confédération de Roquefort en lien avec l'organisme de sélection (OS) UPRA Lacaune coordonnent les schémas de sélection. En ROLP, il s'agit de l'OS ROLP associé à l'entreprise de sélection CDEO (Centre Départemental de l'Élevage Ovin) qui sont en charge de la gestion des schémas des trois races laitières des Pyrénées. En Corse, le schéma est géré par l'entreprise de sélection CORSIA en lien avec l'OS Brebis Corse. Le contrôle de performances est assuré par les deux ES Lacaune, CDEO en ROLP et par la chambre d'agriculture en Corse (SUAE Corse du Sud et SUAE Haute-Corse). L'ensemble de ces organismes sont membres du Comité National Brebis Laitières (CNBL). En caprins laitiers, l'organisme et entreprise de sélection (OES) Capgènes gère les deux races laitières et la race Angora (élevée pour la production de mohair). L'ensemble de ces entreprises ont un fort lien avec l'UMT GPR (Unité Mixte Technologique - Génétique des

Petits Ruminants) qui regroupe l'Institut de l'Élevage (IDELE) et INRAE.

1.2.2 Le contrôle de performances et les caractères en sélection

Depuis septembre 1997, tous les ovins/caprins nés et élevés en France sont identifiés de façon individuelle (Ministère de l'Agriculture et de l'Alimentation, 2018b) sous la responsabilité de l'EDE (Établissement de l'Élevage). Depuis 2006, la réglementation est encadrée à l'échelle européenne et l'identifiant animal comprend le code pays et un numéro à 11 chiffres (indicatif de marquage à 6 chiffres qui est lié au N° de l'exploitation à 8 chiffres + Numéro d'ordre à 5 chiffres).

Dans les filières laitières, les critères de sélection ont évolué au fil du temps, adaptés en fonction des choix des gestionnaires des races et de la disponibilité des phénotypes (Figure 1.3). L'enregistrement de l'ensemble des caractères est effectué en routine lors du contrôle de performances des brebis/chèvres soumis au contrôle laitier officiel (CLO) ou simplifié (CLS), ce qui représente 20-30% de l'effectif d'une race. Le CLO est réservé aux éleveurs sélectionneurs.

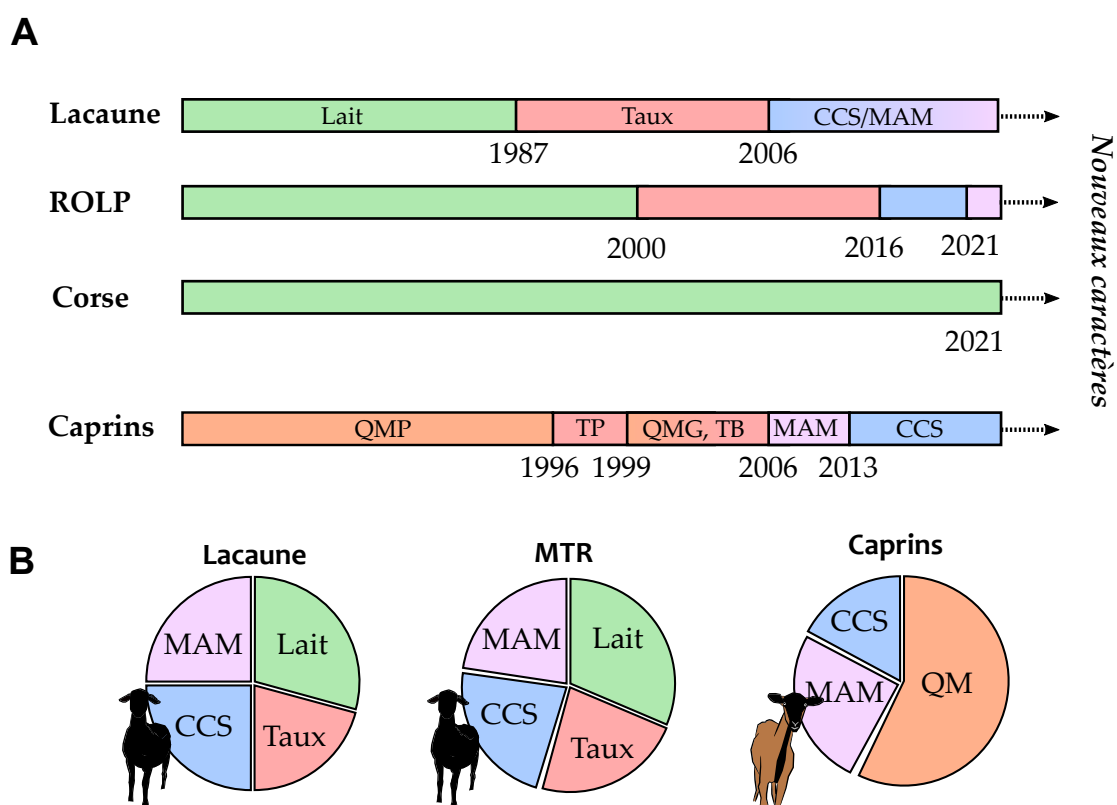


FIGURE 1.3 – Évolution des caractères pris en compte dans les schémas en ovins et caprins laitiers. (A) Temporalité des critères de sélection ; (B) Pondération de l'objectif de sélection dans les différentes races en 2021. QMP : Quantité de Matière Protéique, TP : Taux Protéique, QMG : Quantité de Matière Grasse, TB : Taux butyreux, CCS : Comptage Cellules somatiques dans le lait, MAM : Morphologie Mammaire, ROLP : Races Ovines Laitières des Pyrénées, MTR : Manech Tête Rousse. *Adapté d'après* Astruc et Buisson (2018d), V. Clément (IDELE, comm. pers.).

En ovins, le CLO est un contrôle de type AC (Nomenclature ICAR : "le lait du jour est calculé en multipliant le lait au contrôle par un coefficient au contrôle x troupeau appelé coefficient AC") correspondant à un contrôle mensuel d'une seule des 2 traites journalières et a pour vocation à enregistrer les performances des béliers pour l'évaluation génétique (Ministère de l'Agriculture et

de l'Alimentation, 2018b). Le CLO est un contrôle quantitatif (Production laitière : 5 contrôles/brebis/campagne laitière) et qualitatif (TB, TP et CCS : 2 à 3 contrôles/brebis/campagne laitière). Dans le protocole AC, les événements de reproduction doivent être enregistrés (Date de mise-bas, anomalie de reproduction, mode de reproduction, nombre d'agneaux nés et sexe). Le CLS concerne les éleveurs utilisateurs (contrôle de type D). Le contrôle est plus souple et il n'a pas vocation à être utilisé dans l'indexation. Il s'agit d'un contrôle qualitatif (1 à 2 mesures de la production laitière sur la campagne/brebis). En caprins, il existe sept protocoles de contrôles laitiers (A, AY, CY, AZ, CZ, AC et AT) plus ou moins contraignants (Ministère de l'Agriculture et de l'Alimentation, 2018a). Les contrôles qualitatifs et quantitatifs de type A (contrôle de toutes les traites quotidiennes) et AT (contrôle d'une seule des deux traites quotidiennes) sont les plus fréquents.

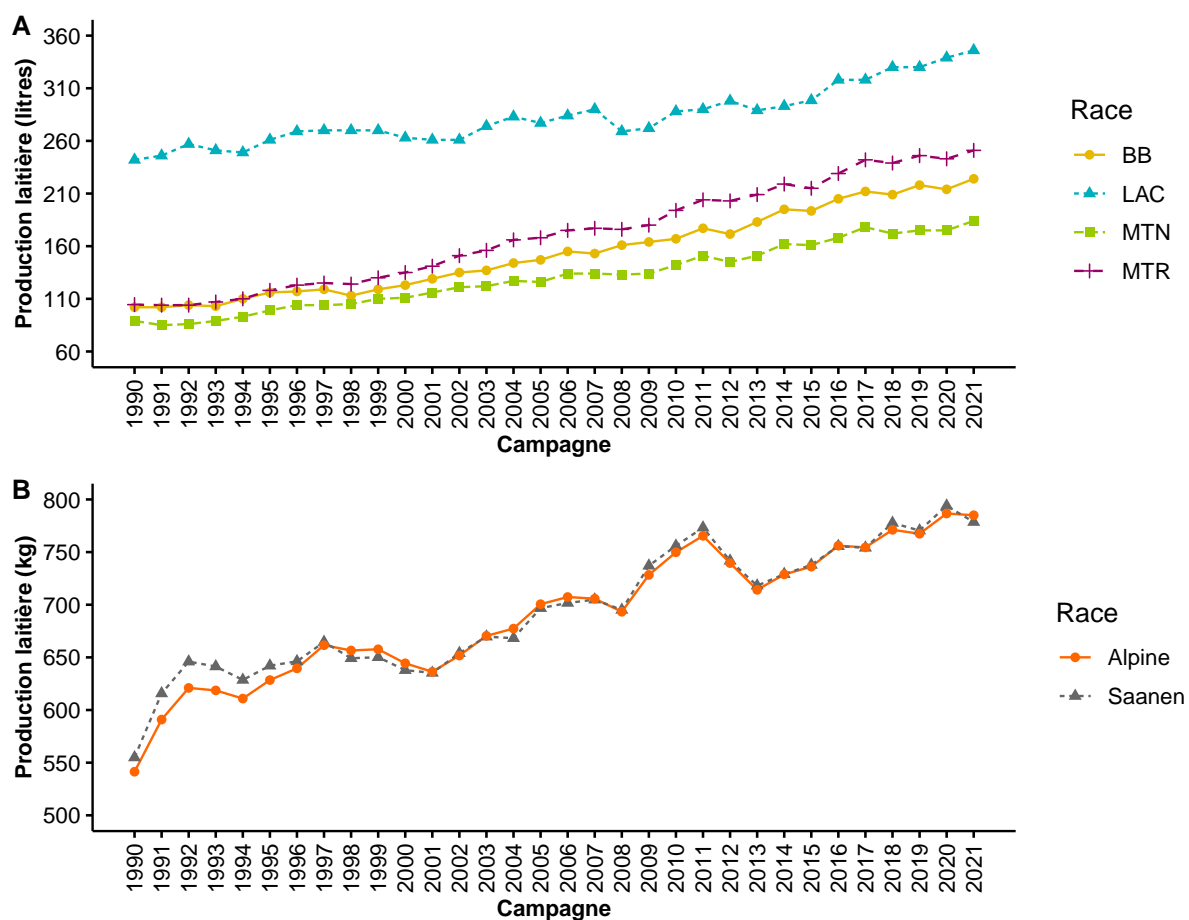


FIGURE 1.4 – Évolution de la production laitière entre 1990 et 2021 dans les races ovines et caprines laitières. (A) Production laitière en races ovines Basco-Béarnaise (BB), Lacaune (LAC), Manech Tête Noire (MTN) et Manech Tête Rousse (MTR). (B) Quantité de lait à 250 jours en races caprines Alpine et Saanen.

Pendant de longues années, les schémas de sélection ont œuvré exclusivement pour augmenter la production laitière (Figure 1.3 A). En ovins lait, la quantité de lait a doublé en quarante ans (Figure 1.4). Par exemple, en race MTR la production est passée de 102 litres (en 128j) en 1990 à 224 litres (148j) en 2021. A noter qu'à ce jour, la quantité de lait est l'unique critère de sélection de la race Corse. En caprins, la sélection n'a jamais porté directement sur la quantité de lait mais sur la quantité de matière protéique qui est fortement corrélée (0,8) à la production laitière

(Teissier et al., 2019). Les schémas ont ensuite progressivement introduit les taux protéiques (TP) et butyreux (TB) dans les critères de sélection (1987 en Lacaune, 2000 en ROLP et 1996-99 en caprins). Dès 2006, des critères fonctionnels ont pris place en lien avec la santé (comptage des cellules somatiques, indicateur de l'inflammation de la mamelle) et conformation de la mamelle (MAM) (cellules et morphologie en 2006 en Lacaune, cellules en 2016 et morphologie en 2021 en ROLP, morphologie dès 2006 et cellules en 2013 en caprins). En ovins, les pointages de la mamelle sont mesurés sur la brebis en première lactation et portent sur 5 postes : angle du trayon (ANG), sillon entre les 2 héli-mamelles (SIL), distance entre le plancher de la mamelle et le jarret (PLJ) et le déséquilibre entre les 2 héli-mamelles (DES). En caprins, 11 postes sont mesurés : profil de la mamelle (PRM), hauteur du plancher (PLA), largeur de l'attache-arrière (AAR), avant-pis (AVP), orientation des trayons (ORT), tour de poitrine (TPO), ouverture des pieds (OUP), longueur des trayons (LOT), forme des trayons (FOT), inclinaison des trayons (INT) et la forme de l'arrière-pis (FAP).

L'ensemble de ces critères sont regroupés en index de synthèse : ISOL (Indice de Synthèse Ovins Laitier) en ovins lait et ICC (Index Combiné Caprin) en caprins (Figure 1.3 B). Les pondérations des caractères dans l'objectif de sélection sont variables en fonction des races et espèces. En race Lacaune, on observe un équilibre entre les caractères de production et fonctionnels. En ROLP (MTR et BB), l'objectif de sélection est composé de 60% de caractères de production et 40% de caractères fonctionnels. Ces races ont intégré les cellules en 2016 et très récemment, la morphologie de la mamelle afin de sélectionner des brebis avec la mamelle bien tendue mieux adaptée à la traite. En caprins, l'ICC combine deux index : l'index de production caprine (IPC) et l'index morphologique caprine (IMC). En Alpine, l'ICC = IPC + 0,5 IMC alors qu'en Saanen un poids plus important est donné à la morphologie, ICC = IPC + 0,6 IMC. L'IPC (= Index MP + 0,4 Index TP + 0,2 Index MG + 0,1 Index TB) et l'IMC (= 0,2 Avant pis + 0,2 Profil + 0,2 Plancher + 0,2 Orientation des trayons + 0,2 Largeur attache arrière) sont identiques entre les deux races. A noter que les cellules ne figurent pas en tant que telles dans l'ICC mais font l'objet d'un index fourni à part.

Les années 1990-2000 ont été marquées par un épisode de tremblante ovine. Il s'agit d'une maladie à prion (Encéphalopathie Spongiforme Subaiguë Transmissible, ESST) affectant le système nerveux central et conduisant à la mort rapide des individus atteints. Les recherches menées ont permis de mettre en évidence des marqueurs génétiques associés à la résistance à la tremblante (Goldmann et al., 1990; Laplanche et al., 1993). Un Programme National d'Amélioration Génétique pour la Résistance à la Tremblante (PNAGRT) a été mis en place en 2001 pour éliminer les animaux de génotype ARQ (forme ancestrale, sensible) et VRQ (Brochard et al., 2006). Depuis 2003, la gestion de la tremblante ovine par le génotypage du gène *PRNP* responsable de la tremblante ovine n'est pas intégrée en tant que telle dans l'ISOL (Figure 1.5). La sensibilité à la tremblante est essentiellement liée à l'allèle Q. De ce fait, aujourd'hui seuls les béliers ARR/ARR entrent en centre d'élevage et diffusent leur génétique dans les schémas de sélection.

L'essor de la sélection génétique permis par le développement de la génomique (voir [paragraphe 1.2.3](#)), permet d'avoir des index plus fiables et d'améliorer les modèles d'indexation. Ceci conduit

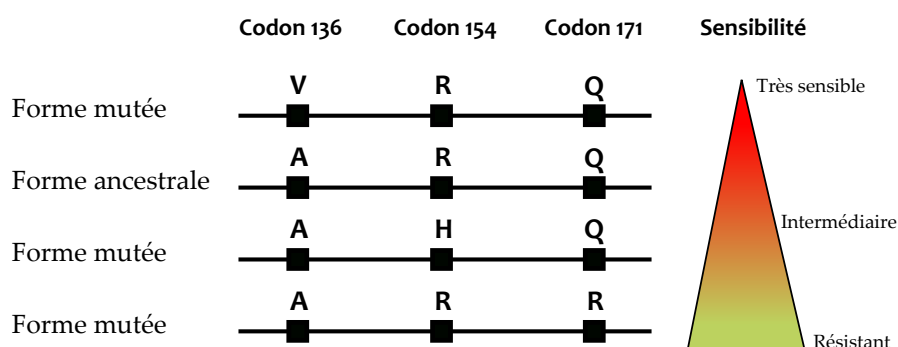


FIGURE 1.5 – Lien entre les différents haplotypes utilisés en sélection au gène *PRNP* (OAR13*) et le degré de sensibilité à la tremblante. Les analyses génétiques ont permis d'identifier une dizaine de polymorphismes dans le gène *PRNP*. Les trois plus importants utilisés en sélection sont les polymorphismes présents dans les codons : codon 136 : Alanine (A) > Valine (V), codon 154 : Arginine (A) > Histidine (H) et codon 171 : Glutamine (Q) > Arginine (R). *Adapté de Tortereau (2016)*. OAR est la nomenclature des chromosomes dans l'espèce ovine (*Ovis aries*).

aujourd'hui, à la prise en compte de nouveaux caractères dans les schémas de sélection. A ce jour, différents projets visent à inclure des caractères liés à la fertilité mâle (qualité de la semence) et femelle, la longévité fonctionnelle des femelles, les aptitudes fonctionnelles des béliers (pointage de la morphologie des mâles), la résistance au parasitisme, la résilience et l'efficacité des animaux, la qualité du lait et le comportement à la traite (Astruc & Buisson, 2018d).

1.2.3 De la sélection classique à la sélection génomique

1.2.3.1 Généralités

Pour pouvoir être sélectionné, un caractère doit être transmis à la descendance (voir [Encadré 1. Principe de l'amélioration génétique](#)). Des pré-requis comme l'identification des animaux (pedigree et état civil des animaux : numéro national unique, race, date de naissance et mort) et la création d'une base de sélection (ensemble de femelles inscrites aux dispositifs officiels de contrôle de performances) sont essentiels. Sélectionner dans une population, c'est orienter les accouplements entre les meilleurs femelles et mâles pour obtenir des descendants avec les meilleures valeurs génétiques possibles. Les meilleurs animaux sont ceux qui ont les meilleures valeurs génétiques additives pour les caractères de l'objectif de sélection. Les gestionnaires des races doivent conjointement diffuser le progrès génétique et gérer la variabilité génétique dans la race.

Encadré 1. Principe de l'amélioration génétique (Minvielle, 2010)

■ Héritabilité d'un caractère

Les performances d'un individu (notées P) pour un caractère donné est le reflet du potentiel génétique de l'animal (noté G) et de l'environnement (noté E).

$$(1) \quad P = G + E$$

La composante G combine l'effet additif des gènes sur le caractère (noté A) et des effets de dominance (D) et d'épistasie (noté I). Ainsi l'équation (1) devient :

$$(2) \quad P = A + D + I + E$$

En sélection génétique, l'objectif est de choisir les animaux avec les meilleures valeurs génétiques additives. En génétique quantitative, la sélection sur un caractère repose sur la notion d'héritabilité h^2 . Au sens étroit, elle correspond à la part de la composante génétique qui sera transmise à la descendance, définie par le rapport entre la variance génétique additive et la variance phénotypique :

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$

L'héritabilité est une valeur comprise entre 0 et 1. Plus cette valeur s'approche de 1, plus il est facile de faire une sélection sur ce caractère. On qualifie l'héritabilité de faible ($h^2 < 0,2$), modérée ($0,2 < h^2 < 0,4$) ou élevée ($h^2 > 0,4$).

■ Le progrès génétique annuel

L'efficacité de la sélection repose sur la notion de progrès génétique annuel (noté ΔG) :

$$\Delta G = \frac{i \times \rho \times \sigma_A}{t}$$

Avec i : intensité de sélection ; ρ : précision de la valeur des index (correspond à la racine carrée du coefficient de détermination (CD)) ; σ_A : écart-type de la valeur génétique additive de la population et t : intervalle de génération, il correspond au temps qui s'écoule entre la naissance et la mise en reproduction des reproducteurs.

Afin de sélectionner les animaux, il est nécessaire d'estimer (modèles de prédiction) les valeurs génétiques additives. Un individu transmet en espérance la moitié de la valeur génétique additive à ses descendants. La sélection génétique classique utilise le BLUP (*Best Linear Unbiased Predictor*) - modèle animal et repose sur l'enregistrement des performances de l'individu et de ses apparentés. La précision de l'index s'accompagne d'un coefficient de détermination (CD). Il correspond à la précision de l'estimation qui dépend du nombre d'enregistrements de phénotypes et de l'héritabilité des caractères. Plus le CD est grand, plus la valeur génétique additive estimée est précise. Les caractères de production ont des héritabilités modérées à élevées (0,30 à 0,50 en ovins/caprins). Pour les caractères fonctionnels, le comptage des cellules somatiques est faiblement héritable ($h^2=0,13$ en ovins et $h^2=0,19$ en caprins) et les héritabilités des pointages mamelles sont faibles à élevées ($h^2=0,19-0,33$ en ovins et $h^2=0,12-0,50$ en caprins) (Barillet et al., 2016 ; Teissier et al., 2019).

La sélection se fait au niveau individuel (choix des femelles de renouvellement en élevage) et collectif (sélection des mâles d'IA). L'organisation de la sélection est pyramidale. Chez les éleveurs sélectionneurs du noyau de sélection, des accouplements programmés entre les femelles et mâles avec les meilleures valeurs génétiques sont réalisés. Les descendants issus de ces accouplements

permettent d'une part d'avoir des jeunes mâles reproducteurs (à évaluer) et d'autre part d'évaluer les performances des filles des mâles reproducteurs du CIA (centre d'insémination animale), c'est le testage sur descendance.

Le processus de sélection classique est long et coûteux. L'optimisation du progrès génétique repose notamment sur l'utilisation des biotechnologies de la reproduction, des méthodes d'induction-synchronisation des chaleurs et de la révolution génomique. Dans les schémas laitiers, le progrès génétique passe principalement par la voie mâle et repose sur la pratique de l'insémination animale (IA). En petits ruminants, la saisonnalité et l'utilisation de semence fraîche imposent d'utiliser un nombre de reproducteurs importants. Néanmoins, en caprins, l'insémination en semence congelée est réalisée. L'évolution de la fertilité sur IA en petits ruminants ([Figure 1.6](#)) est relativement stable au cours du temps, avec une tendance à l'augmentation ces dernières années en ovins. Le déclin de la fertilité à l'IA chez les bovins (Pryce et al., 2004) associé à une augmentation importante de la production laitière, n'est pas observé dans nos filières de petits ruminants. En Lacaune, 95% des agnelles de renouvellement du noyau de sélection sont issues d'IA, suggérant une sélection implicite sur la fertilité. En ROLP, cette évolution est plus limitée, que l'on explique par un taux d'IA plus faible dans le schéma (45-60%) avec seulement 70% d'agnelles de renouvellement du noyau de sélection issues d'IA. De plus, les conditions dans les Pyrénées sont plus exigeantes : herbe, état des brebis au moment de la lutte... Chez les caprins, on remarque que les chevrettes (primipares) ont une moins bonne fertilité que les agnelles, expliqué par une mise à la reproduction trop précoce. Il est également important de noter qu'une seule IA est pratiquée en petits ruminants alors qu'en bovins jusqu'à 3 IA consécutives par campagne peuvent être réalisées si la vache n'est pas gestante.

L'avènement de la génomique en petits ruminants (mise en place de la sélection génomique en 2015 en Lacaune, 2017 en ROLP et 2018 en caprins) a permis d'améliorer le processus de sélection (Astruc et al., 2016b ; Clement et al., 2020). La sélection génomique permet de prédire la valeur génétique additive d'un candidat reproducteur en analysant un grand nombre de marqueurs SNP (*Single Nucleotide Polymorphism*, variant d'une paire de bases) répartis sur l'ensemble de son génome (voir [paragraphe 2.1.2.1](#)). La prédiction de cette valeur repose sur l'utilisation d'une population de référence correspondant à un ensemble d'individus de la même race qui sont génotypés et phénotypés (Georges et al., 2019). En petits ruminants, cette population est exclusivement composée de mâles d'IA testés sur descendance (Astruc & Buisson, 2018e). Ainsi, avec l'apparentement du candidat reproducteur avec la population de référence, il est possible dès la naissance de connaître sa valeur génétique estimée. La méthode d'indexation retenue est le single-step Genomic BLUP (ssGBLUP) (Astruc & Buisson, 2018f).

L'utilisation de la sélection génomique a principalement permis d'augmenter l'intensité de sélection conduisant à augmenter le progrès génétique de l'ordre de 15-20% par an (Astruc & Buisson, 2018a). Les schémas de sélection génomique sont similaires entre ovins et caprins laitiers avec quelques spécificités détaillées dans les [Figures 1.7 et 1.8](#).

1.2.3.2 Les schémas ovins laitiers à l'ère de la génomique

Le fonctionnement des schémas de sélection ovins laitiers en Lacaune et en ROLP est décrit dans la [Figure 1.7](#).

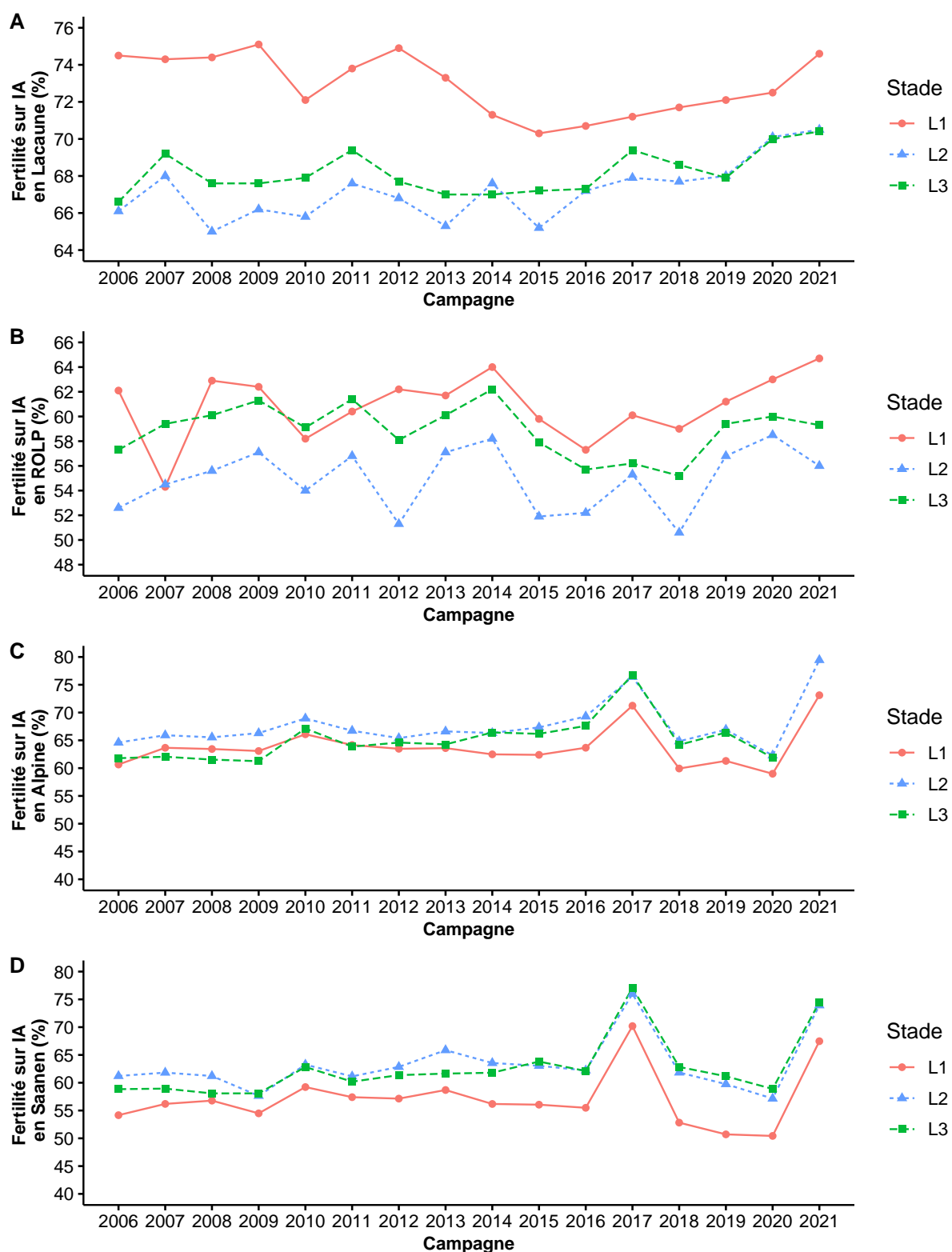


FIGURE 1.6 – Évolution de la fertilité à l'insémination animale dans les races ovines et caprines laitières entre 2006 et 2021. L1= lactation 1, L2= lactation 2 et L3= lactation 3 et plus. Les données brutes sont extraites des bases de données nationales SIEOL (Système d'Information en Elevage Ovin Laitier) (J-M, Astruc, comm. pers.) et SIECL (Système d'Information en Elevage Caprin Laitier) (V. Clément, comm. pers.)

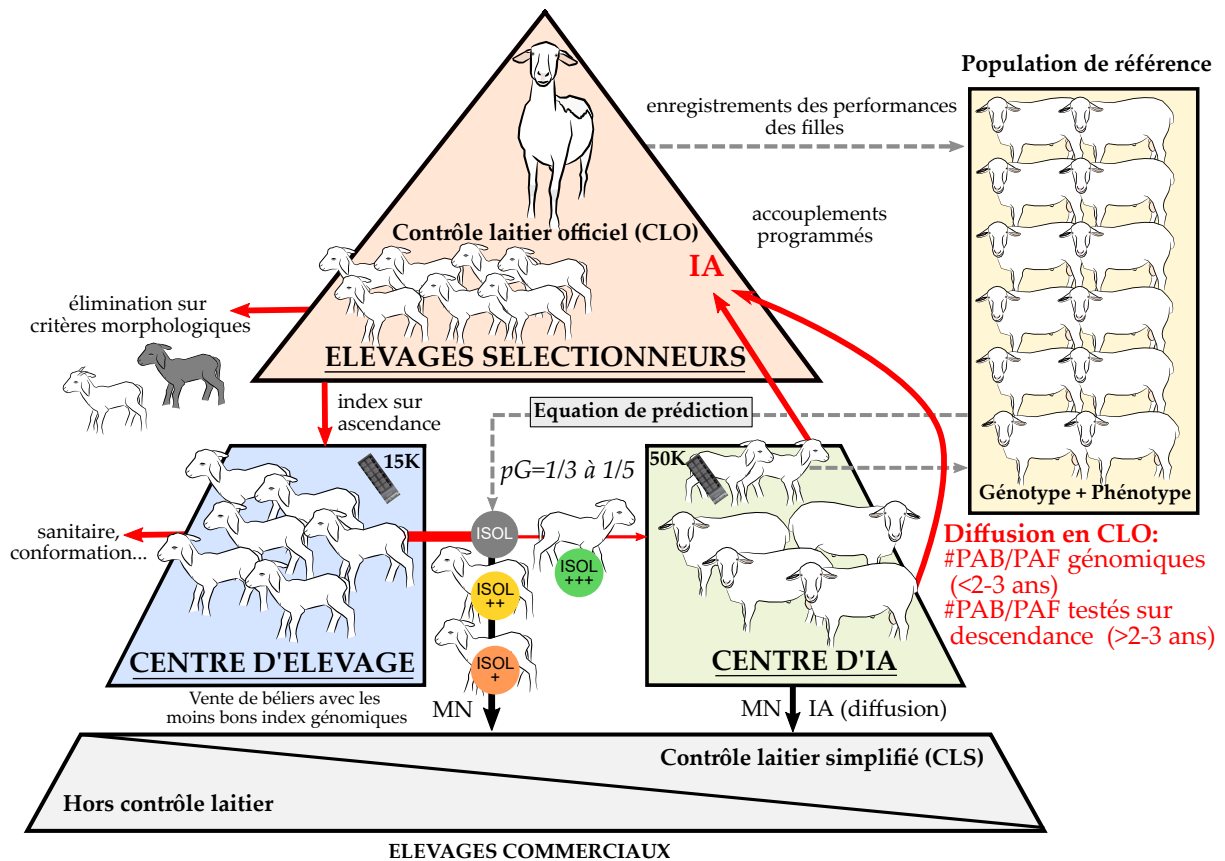


FIGURE 1.7 – Fonctionnement des schémas de sélection ovins laitiers en Lacaune et ROLP. PAB= Père à béliers, PAF= Père à filles, MN= monte naturelle, IA= insémination animale.

La diffusion de la génétique dans les schémas repose sur la pratique de l'IA en semence fraîche. Ceci implique d'avoir un grand nombre de béliers en CIA (1400 béliers en Lacaune et 600 béliers en MTR) (Buisson et al., 2018). En effet, le taux d'IA dans les schémas varie de 85% en Lacaune et 45-60% en ROLP. Les agneaux nés d'accouplements programmés entre meilleurs reproducteurs, avec les meilleurs index sur ascendance et conformes aux standards de la race, sont génotypés dès le premier mois après la naissance (Astruc et al., 2016a; Astruc & Buisson, 2018c, 2018b). Les défauts morphologiques peuvent être liés suivant la race à la présence de tâches, la couleur de toison, la présence ou non de cornes, la forme de la tête/oreilles/mâchoire, le type de la laine ou de la stature. En Lacaune, les candidats à la sélection rentrent ensuite en centres d'élevage (Le Bourguet ou La Glène). En ROLP, la capacité d'accueil ne permet pas d'accueillir l'ensemble des agneaux candidats, de ce fait les candidats restent chez les éleveurs sélectionneurs jusqu'à la diffusion de l'index génomique. Durant cette période, de nombreux agneaux sont éliminés sur critères sanitaires ou défauts de conformation. Par exemple, l'UPRA Lacaune décline les animaux s'ils présentent un dos cassé, épaules étroites, aplombs tordus, paturons "assis", jarrets serrés et/ou un manque de développement. Le taux de réforme est de 50% en Lacaune et de 10-15% en ROLP. Vers 4-5 mois, la diffusion des index génomiques permet de sélectionner les meilleurs candidats (index de synthèse ovins laitiers génomique que l'on nomme ISOLg). Cette prédiction est permise par l'équation de prédiction ssGBLUP et repose sur la population de référence (=ensemble des béliers du CIA). La pression de sélection génomique exercée (pG)

est importante puisqu'elle varie de 1/3 à 1/5 (Astruc & Buisson, 2018a). L'ISOLg permet de classer les agneaux en quatre catégories : agneaux avec un ISOL négatif (index négatif ou erreur paternité), les agneaux avec un ISOL intermédiaire (classés en +, ++ ou ++++) et les béliers pour le CIA avec les meilleurs ISOL. Les agneaux candidats intègrent par la suite le CIA et vont exclusivement diffuser (à partir de 1,5 ans en Lacaune et 2,5 ans en ROLP) dans les troupeaux en CLO pour faire au moins 120 IA afin d'obtenir les premiers index sur descendance. Une partie des jeunes béliers est également réformée lors de l'entraînement à la production de semence. Une pression de sélection sur descendance contribue à éliminer 15-30% des béliers génomiques. Après la première campagne, les animaux sont évalués sur descendance et diffusent leur génétique jusqu'à 4-5 ans. Les agneaux non retenus pour rentrer en CIA (ISOL +, ++ et ++++) sont vendus dans les élevages en CLS et hors contrôle laitier pour améliorer la génétique dans toute la race.

1.2.3.3 Les schémas caprins laitiers à l'ère de la génomique

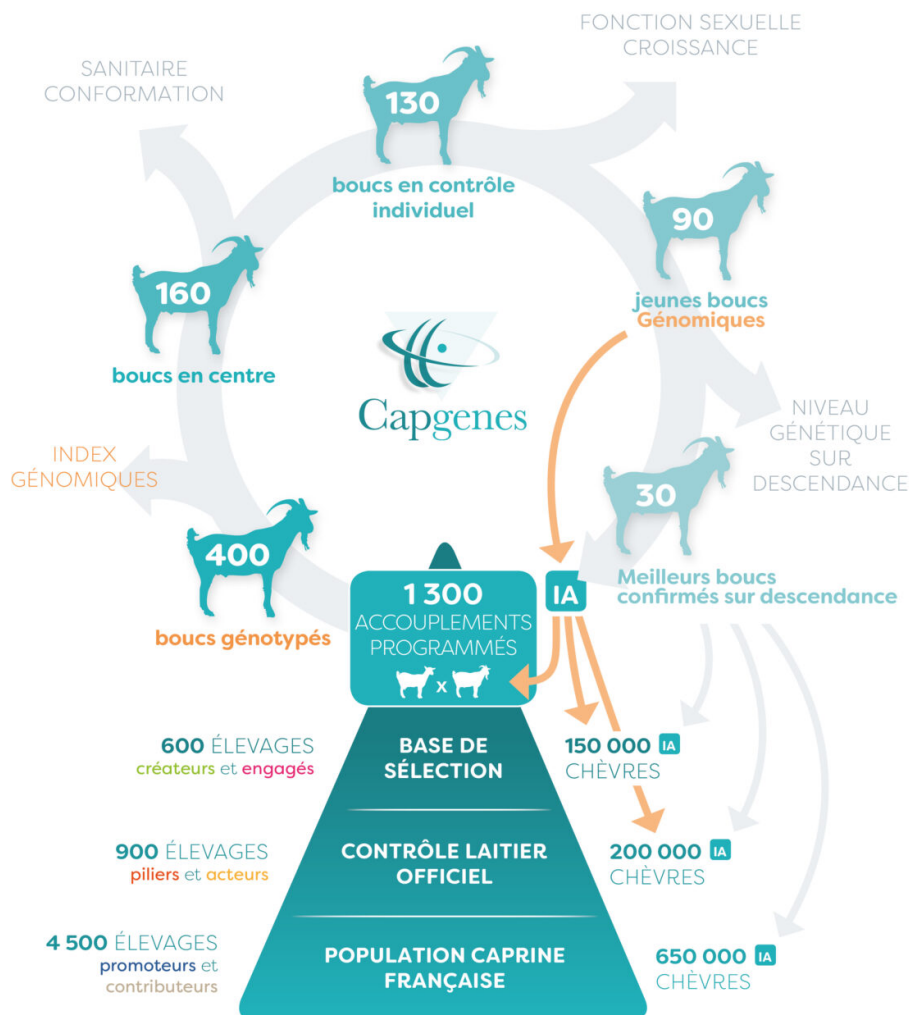


FIGURE 1.8 – **Fonctionnement des schémas de sélection caprins laitiers.** Les effectifs indiqués dans la figure sont pour les deux races Alpine et Saanen avec un ratio 65% et 35%, respectivement. *Source* : Capgènes.

Les schémas caprins laitiers s'organisent de la même manière que les ovins laitiers. Les schémas

sont décrits dans la [Figure 1.8](#). A chaque campagne, 1300 accouplements élités sont programmés (845 en Alpine et 455 en Saanen). Les boucs candidats sont génotypés vers 1 mois et leurs index génomiques sont disponibles vers 3 mois. Suivant leur valeur génétique, les meilleurs boucs ($pG=0,4$) entrent à Capgènes lorsqu'ils ont entre 3 et 5 mois et commencent à produire de la semence entre 6 et 8 mois. Les boucs seront ensuite collectés pendant un peu moins de 2 ans avant d'être abattus. Comme chez les ovins, des réformes liées au sanitaire, conformation ou standard de race et défauts de production de semence contribuent à réduire le nombre des jeunes boucs avant de devenir reproducteur (taux de réforme de 60%). En caprins, la diffusion du progrès génétique passe par l'IA en semence congelée.

1.3 Des structures de populations favorables à l'émergence d'anomalies génétiques

Les espèces d'élevage ont des effectifs génétiques limités, ce qui contribue à amplifier le phénomène de dérive génétique et augmenter la consanguinité.

1.3.1 Les populations d'élevage ont des effectifs génétiques limités

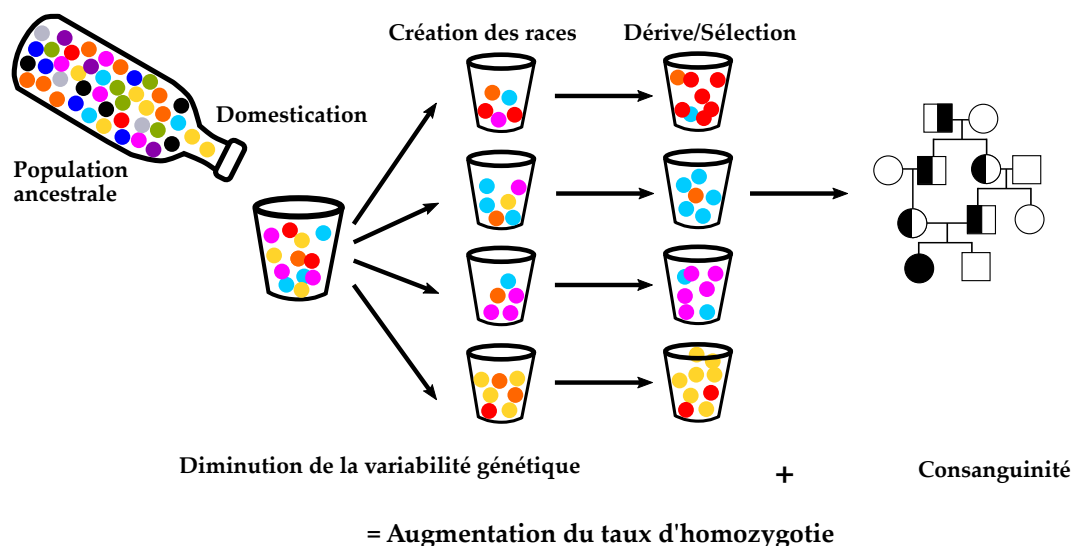


FIGURE 1.9 – Trajectoire évolutive des allèles d'une population en sélection

Depuis la domestication, l'homme a sélectionné des animaux adaptés à ses besoins (production laitière et viande, docilité ...). La diversité des allèles initialement présents dans la population ancestrale s'est fortement réduite lors de la création des races (définies comme un ensemble d'individus qui se ressemblent morphologiquement) (Liu et al., 2019 ; Albalat & Cañestro, 2016) (Figure 1.9). Même si les effectifs au sein des races sont parfois très élevés, le patrimoine génétique des animaux de rente s'est appauvri par l'utilisation d'un faible nombre d'animaux reproducteurs élités, notamment par la pratique de l'IA. La gestion de la diversité génétique est un enjeu majeur pour les gestionnaires des races.

La variabilité génétique d'une population peut être caractérisée par trois indicateurs (Figure 1.10). Ces derniers sont calculés dans le cadre de l'observatoire de variabilité génétique des ruminants (VARUME) obtenus par l'analyse des données généalogiques (Danchin-Burge et al., 2017). L'indicateur A_e correspond au nombre d'ancêtres efficaces, c'est-à-dire combien d'individus sont à l'origine des allèles portés par les individus de la race. Il permet de mettre en évidence des goulots d'étranglement. Par rapport aux bovins laitiers, le A_e est plus élevé en petits ruminants. L'indicateur N_e correspond à la taille efficace de la population (par exemple si $N_e = 100$, cela signifie que la variabilité génétique de la population correspond à l'accouplement au hasard de 50 mâles et 50 femelles non apparentés). Son calcul est basé sur l'évolution de la consanguinité (définit un lien génétique de parenté entre reproducteurs) (voir Encadré 2. Taux de consanguinité et taille efficace d'une population). La taille efficace intra-race est importante en races ovines (Lacaune et MTR) et caprines (Alpine et Saanen) laitières ($N_e=151-336$) alors qu'elle est plus faible en races ovines MTN et BB (83 à 95). La sélection des animaux ARR/ARR résistants à la tremblante ovine a conduit dans les années 2000 à un goulot d'étranglement. L'indicateur N_{gen}

correspond à la profondeur des généalogies. En ovins et caprins laitiers, les qualités des généalogies sont très satisfaisantes puisque l'on est capable de remonter dans le pedigree jusqu'à 10 générations.

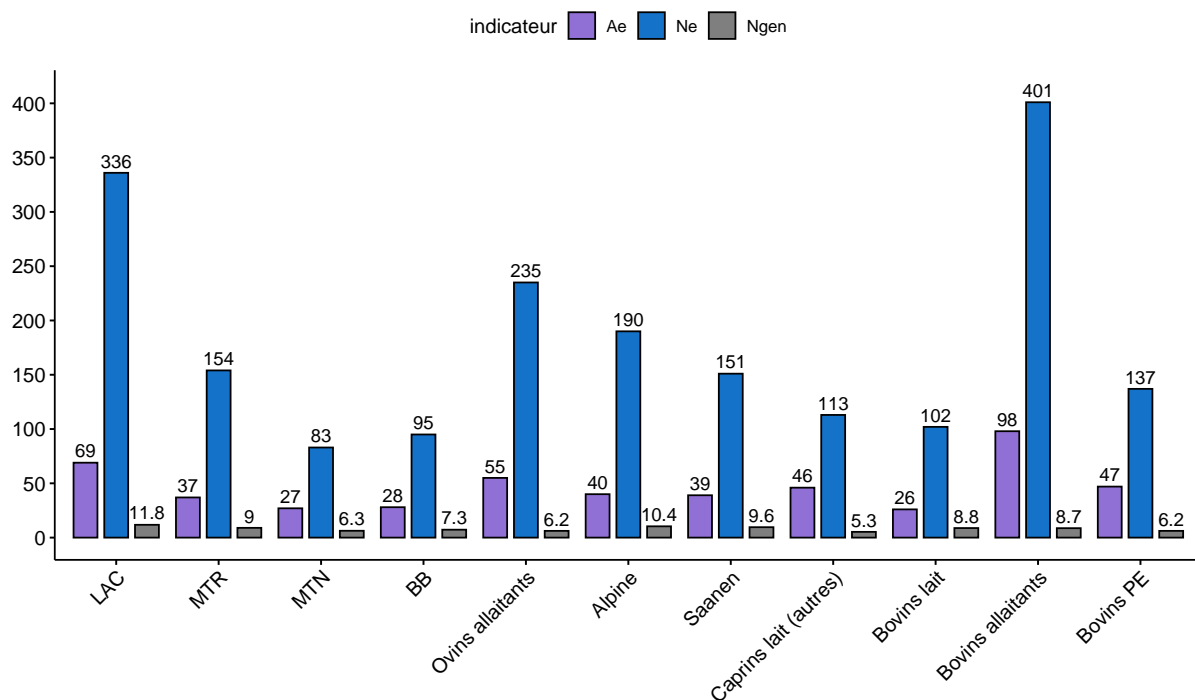


FIGURE 1.10 – Principaux indicateurs de variabilité génétique pour les races de ruminants en France. Avec les probabilités d'origine des gènes (Ae), la taille efficace (Ne) et le nombre de générations connus (Ngen). LAC= Lacaune, MTR= Manech Tête Rousse, MTN= Manech Tête Noire, BB= Basco-Béarnaise, Bovins PE= comprend les races à petits effectifs (PE) et les races étrangères. *Source* : VARUME.

Encadré 2. Taux de consanguinité et taille efficace d'une population

$$\Delta F = F_t - F_{t-1} = \frac{1}{2N_e}(1 - F_{t-1})$$

Connaissant le taux de consanguinité (F) entre la génération $t - 1$ et t , on peut calculer un effectif génétique (N_e) réalisé :

$$N_e = \frac{(1 - F_{t-1})}{2\Delta F}$$

L'état de l'art en ovins et caprins montre une meilleure gestion de la diversité génétique comparée aux bovins, expliquée par la saisonnalité et l'utilisation de semences fraîches qui imposent un nombre plus important de reproducteurs. Un bélier Lacaune réalise au maximum 1 200 doses alors que chez l'espèce bovine où l'utilisation de semence congelée est possible, un taureau produit plusieurs dizaines de milliers de doses/an.

Avec l'avènement des outils génomiques, l'étude des ROH (run of homozygosity) se montre pertinente pour caractériser la diversité génétique d'une race. Ces derniers sont des régions contiguës du génome où un individu est homozygote, hérités de ses deux parents, eux-mêmes issus d'un même ancêtre commun (Ceballos et al., 2018). Plus ces régions sont grandes et nombreuses

le long du génome, plus elles témoignent d'une consanguinité proche.

1.3.2 Dérive génétique et consanguinité

La réduction du nombre d'ancêtres contributeurs A_e et de la taille efficace N_e contribue d'une part à augmenter la consanguinité au cours du temps et d'autre part à réduire la variabilité génétique des animaux. Le taux de consanguinité (=accouplement entre reproducteurs apparentés) est relativement faible en ovins laitiers en race Lacaune ($\Delta F/\text{génération} = +0,29\%$) et MTR ($\Delta F/\text{génération} = +0,42\%$) sur la période 2000-2009 (Buisson et al., 2018). Les coefficients de consanguinité dans les deux espèces sont similaires ($F \sim 2.5-3\%$). Dans les autres races ovines (BB et MTN), l'augmentation du taux de consanguinité est proche des bovins laitiers (1% par génération) (Gutiérrez-Reinoso et al., 2022). L'augmentation de la consanguinité conduit inévitablement à augmenter le taux d'homozygotie des génomes à l'origine de l'émergence d'anomalies génétiques. En effet, chaque individu est porteur de 1 à 5 mutations hautement délétères dans son génome (notion de fardeau génétique), responsables d'anomalies à l'état homozygote (Charlier et al., 2016 ; Georges et al., 2019).

Les phénomènes évolutifs (sélection et dérive génétique) contribuent également à impacter la variabilité génétique des populations en modifiant les fréquences des allèles au cours des générations (voir [paragraphe 2.3](#)).

A retenir La sélection des petits ruminants est structurée (structure pyramidale des schémas de sélection), organisée (gestion par les ES et OS, contrôle de performances) et améliorée avec l'avènement de la sélection génomique dans les populations laitières. La sélection génomique permet de génotyper tous les candidats à la sélection dès la naissance pour prédire les valeurs génétiques des jeunes candidats. La pratique de la sélection contribue à créer un biais de représentativité entre les animaux de la population et les reproducteurs des CIA. En effet, seuls les meilleurs individus avec les meilleures valeurs génétiques et conformes aux standards de la race sont gardés. Ceci conduit inévitablement à réduire la diversité allélique d'une race. Les populations de petits ruminants ont des effectifs génétiques modérés conduisant à l'émergence d'anomalies génétiques du fait de la dérive génétique (fluctuation aléatoire des allèles) et de la consanguinité (accouplements entre reproducteurs apparentés). Un compromis entre variabilité génétique et pression de sélection est désormais possible par la sélection génomique afin d'assurer des schémas de sélection efficaces. ■

2. Les anomalies génétiques : définitions, modes de transmission et d'action, évolution des fréquences

Une anomalie génétique est définie comme un phénotype déviant dans une population et résulte d'une mutation aléatoire de l'ADN (Boichard, Grohs, Danchin-Burge, & Capitan, 2016). Ces anomalies peuvent être délétères (voir létales) ou neutres (par exemple : couleur de toison) et sont liées pour la plupart à une seule région génomique (monogénique). Elles sont qualifiées de congénitales lorsque le phénotype est visible dès la naissance. Nous verrons par la suite qu'il existe différents déterminismes génétiques des anomalies. Au cours de l'évolution, ces variants, pourtant délétères, peuvent se maintenir dans les populations sous l'effet de forces évolutives (dérive et sélection).

2.1 L'ADN, source de variabilité génétique

2.1.1 L'ADN, support de l'information génétique

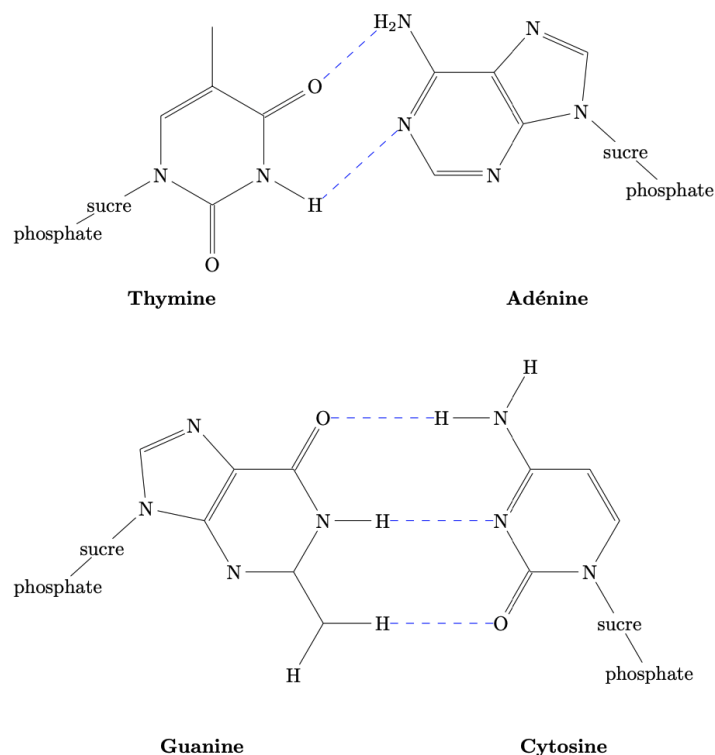


FIGURE 2.1 – Structure moléculaire de l'ADN. Les liaisons hydrogène sont représentées par des tirets bleus.

L'Acide DésoxyriboNucléique (ADN) est le support universel de l'information génétique. L'ADN est composé de deux chaînes orientées antiparallèles et complémentaires de polymères de

nucléotides formant une double hélice (Watson and Crick, 1953). Les monomères sont reliés entre eux par des liaisons covalentes sur un même brin de type liaison phosphodiester (liaison entre l'extrémité 3'-OH du nucléotide n et l'extrémité 5'-Phosphate du nucléotide n+1). Un nucléotide est composé d'une base nucléique, d'un sucre (pentose) et d'un groupement phosphate. Les bases nucléiques sont la Thymine (T), l'Adénine (A), la Guanine (G) et la Cytosine (C) (Figure 2.1). La Guanine et l'Adénine sont des bases purines tandis que la Thymine et la Cytosine sont des pyrimidines. Ces bases s'apparient entre elles par des liaisons hydrogènes (trois liaisons entre G et C et une double liaison entre A et T). Une richesse en paires A-T, permet à l'ADN de s'ouvrir plus facilement. L'enchaînement des nucléotides forme la séquence et définit le code génétique de chaque individu.

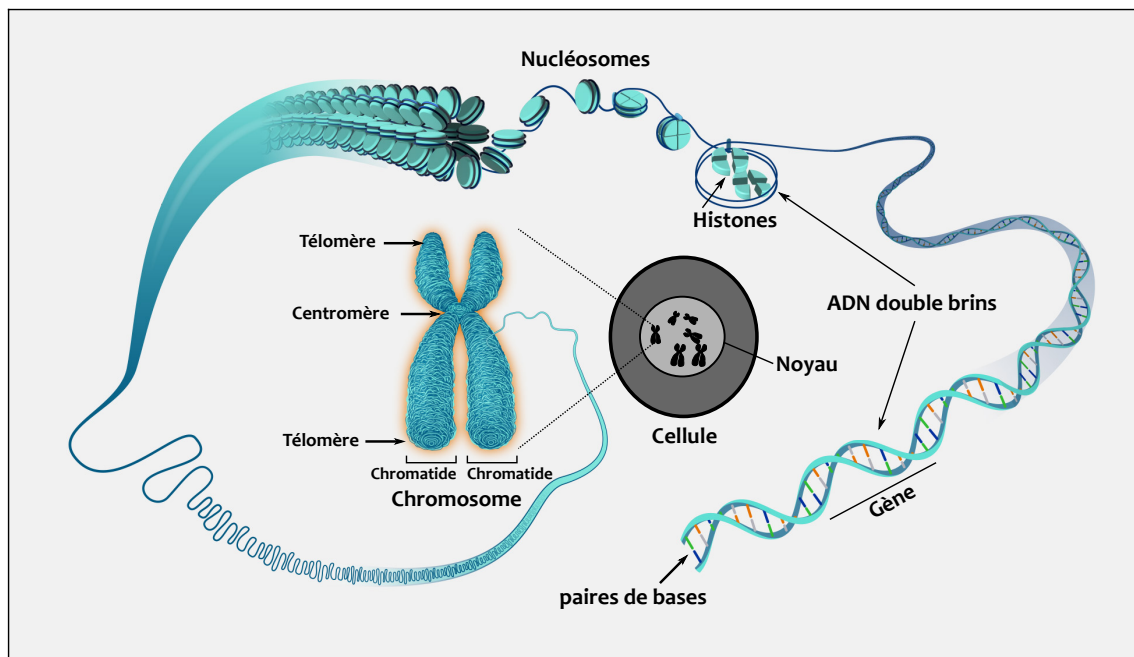


FIGURE 2.2 – De la cellule à l'ADN. Adapté d'après <https://www.genome.gov/genetics-glossary/Chromosome>.

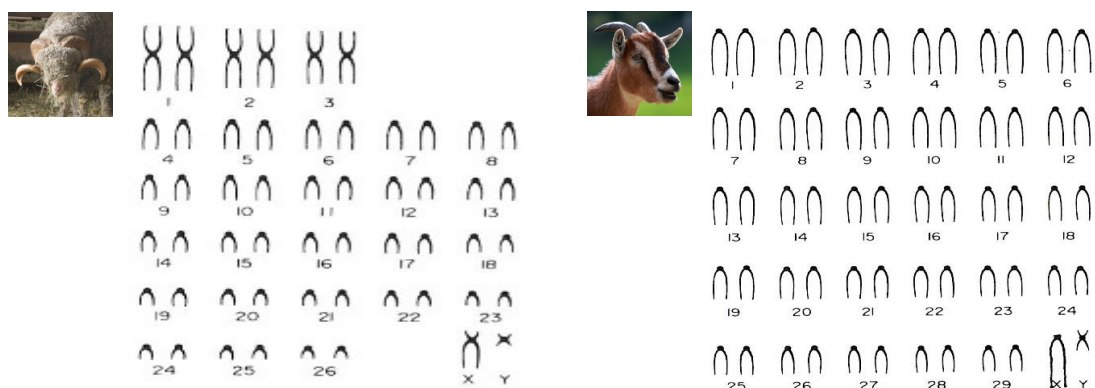


FIGURE 2.3 – Caryotypes des ovins et caprins

Les gènes, supports de l'information génétique, sont portés par les chromosomes (Figure 2.2). Ils sont localisés dans le noyau des cellules chez les eucaryotes. Les chromosomes résultent de la condensation de la chromatine et l'ADN est enroulé autour d'octamères d'histones pour former

les nucléosomes. Les génomes ovins et caprins (Figure 2.3) sont composés de deux chromosomes sexuels (XX pour la femelle et XY pour le mâle) et 26 paires d'autosomes ($2n=54$) chez le mouton, 29 paires d'autosomes chez la chèvre ($2n=58$). Le caryotype de la chèvre diffère du mouton puisque tous les chromosomes sont acrocentriques, c'est-à-dire qu'ils n'ont qu'un seul bras.

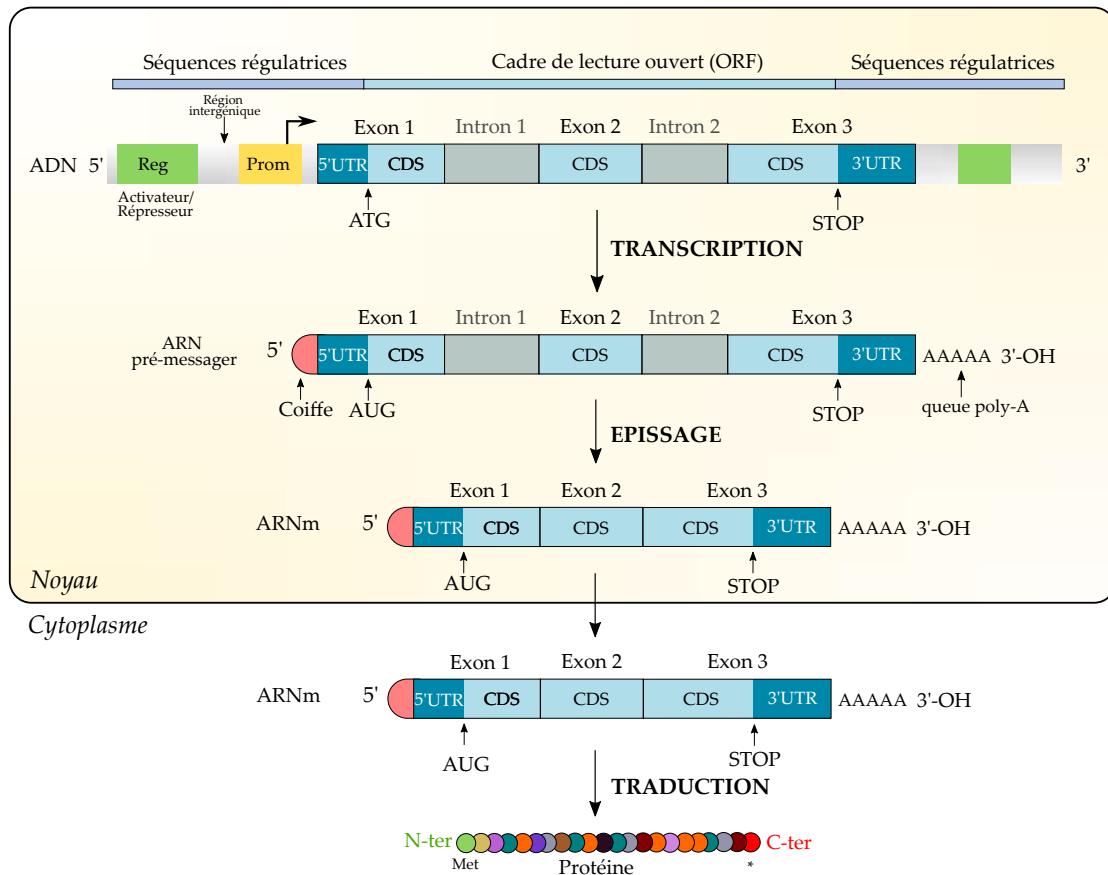


FIGURE 2.4 – **Du gène à la protéine.** Seul le brin 5' → 3' est représenté. Reg= Séquences régulatrices, Prom= Promoteur, ORF= *Open Reading Frame*/Cadre ouvert de lecture, CDS= *Coding sequence*/séquence codante, UTR= *Untranslated region*/ Région non traduite.

Un gène correspond à une séquence d'ADN double brin (chez les eucaryotes, seulement 2% de l'ADN est codant). Un gène est composé de séquences régulatrices (rôle dans la régulation de l'expression du gène) et d'un cadre ouvert de lecture composé de séquences codantes (exons) et non-codantes (introns). L'ensemble des étapes permettant de passer du gène à la protéine est illustré dans la Figure 2.4. Lors de la transcription, l'ARN polymérase II utilise le brin 5' → 3' comme matrice pour synthétiser une molécule d'ARN pré-messager simple brin. La synthèse est initiée à la lecture du codon d'initiation de la transcription et se termine à la lecture du site de fin de transcription. Dans l'ARN, la thymine (T) est remplacée en Uracile (U). L'ARN pré-messager présente une coiffe en 5'-P et une queue poly-A en 3'-OH. Lors de l'épissage, les introns sont éliminés avant d'être exportés dans le cytoplasme. Ensuite, les ribosomes (complexes protéiques intervenant dans la traduction) vont venir se fixer sur l'ARNm et grâce aux ARN de transfert (ARNt), l'ARNm est alors traduit en protéine. Il s'agit d'une longue chaîne d'acides aminés reliés entre eux par des liaisons peptidiques. La traduction démarre à la lecture du codon d'initiation de la traduction (AUG, Méthionine) et se termine à la lecture d'un codon STOP

(TAA, TAG ou TGA). La séquence des acides aminés de la protéine s'obtient par la lecture d'une fenêtre non-chevauchante de 3 nucléotides, appelés codon. Les 20 acides aminés constituant le code génétique sont décrits dans la [Table 2.1](#).

TABLE 2.1 – **Liste des 20 acides aminés constituant le code génétique.** Les acides aminés sont classés en 4 catégories : les acides aminés apolaires et hydrophobes (Alanine, Valine, Leucine, Isoleucine, Proline, Phénylalanine, Tryptophane, Méthionine et Glycine), les acides aminés polaires non chargés (Cystéine, Sérine, Thréonine, Tyrosine, Asparagine et Glutamine), les acides aminés chargés - (Acide aspartique et Acide glutamique) et les acides aminés chargés + (Lysine, Arginine et Histidine).

		Deuxième base									
		U		C		A		G			
Première base	U	UUU	Phe (F)	UCU	Ser (S)	UAU	Tyr (Y)	UGU	Cys (C)	U	Troisième base
		UUC	Phe (F)	UCC	Ser (S)	UAC	Tyr (Y)	UGC	Cys (C)	C	
		UUA	Leu (L)	UCA	Ser (S)	UAA	Stop (*)	UGA	Stop (*)	A	
		UUG	Leu (L)	UCG	Ser (S)	UAG	Stop (*)	UGG	Trp (W)	G	
	C	CUU	Leu (L)	CCU	Pro (P)	CAU	His (H)	CGU	Arg (R)	U	
		CUC	Leu (L)	CCC	Pro (P)	CAC	His (H)	CGC	Arg (R)	C	
		CUA	Leu (L)	CCA	Pro (P)	CAA	Gln (Q)	CGA	Arg (R)	A	
		CUG	Leu (L)	CGC	Pro (P)	CAG	Gln (Q)	CGG	Arg (R)	G	
	A	AUU	Ile (I)	ACU	Thr (T)	AAU	Asn (N)	AGU	Ser (S)	U	
		AUC	Ile (I)	ACC	Thr (T)	AAC	Asn (N)	AGC	Ser (S)	C	
		AUA	Ile (I)	ACA	Thr (T)	AAA	Lys (K)	AGA	Arg (R)	A	
		AUG	Met (M)	ACG	Thr (T)	AAG	Lys (K)	AGG	Arg (R)	G	
	G	GUU	Val (V)	GCU	Ala (A)	GAU	Asp (D)	GGU	Gly (G)	U	
		GUC	Val (V)	GCC	Ala (A)	GAC	Asp (D)	GGC	Gly (G)	C	
		GUA	Val (V)	GCA	Ala (A)	GAA	Glu (E)	GGA	Gly (G)	A	
		GUG	Val (V)	GCG	Ala (A)	GAG	Glu (E)	GGG	Gly (G)	G	

Polaires Polaires - Polaires + Apolaires

L'expression des gènes peut être modulée par des facteurs de transcription qui vont agir sur les séquences régulatrices ou par des phénomènes épigénétiques. L'épigénétique est définie comme l'ensemble des marques présentes sur le génome induisant des changements dans l'expression des gènes (et donc des phénotypes) sans altérer la séquence d'ADN à la différence des différents polymorphismes (Berger et al., 2009). Les principaux phénomènes épigénétiques sont la méthylation de l'ADN, les modifications d'histones (méthylation/acétylation) et l'action des ARN non-codants (Beaujean et al., 2020). Ces phénomènes épigénétiques varient au cours de la vie d'un individu en fonction de l'environnement. Le long de la molécule d'ADN, une cytosine suivie de guanine peut être méthylée par une enzyme, l'ADN méthyltransférase (ajout d'un groupe méthyle (CH_3)). Lorsque ces sites CpG sont localisés au niveau des promoteurs, la méthylation de l'ADN peut réprimer l'expression des gènes (Schübeler, 2015). Les modifications post-traductionnelles des histones affectent l'accessibilité de l'ADN lors de la transcription et participent donc à activer/inhiber l'expression des gènes en fonction de l'état de condensation de la chromatine

(Kouzarides, 2007). Les ARN non-codants sont des molécules simples brins et sont classés en deux catégories : les microARN (composés d'une vingtaine de nucléotides) ou les ARNlnc (ARN long non-codants, composés de plus de 200 nucléotides). Ces ARN ne possèdent pas de cadre de lecture, ce qui les différencie des ARNm. Les microARN sont capables de réguler les activités post-transcriptionnelles en s'appariant avec des ARNm cibles aboutissant à leur dégradation (Huntzinger & Izaurralde, 2011). Les ARNlnc sont depuis longtemps connus pour leur rôle dans la régulation des gènes par des modifications de la condensation de la chromatine et l'altération de la stabilité et la traduction des ARNm impactant les voies de signalisation (Statello, Guo, Chen, & Huarte, 2021). Par exemple, il existe un mécanisme d'interférence par ARN (RNAi) permettant la dégradation des ARNm par un complexe protéique RISC (*RNA-induced silencing complex*) basé sur l'action de miARN (exemple avec le phénotype callipyge décrit dans le [paragraphe 2.2.7](#)).

2.1.2 Les différents polymorphismes de l'ADN et leurs impacts biologiques

Les mutations de l'ADN peuvent avoir des origines environnementales (rayonnements, carences) ou génétiques (Ducos et al., 2002). Lors de la réplication de l'ADN, des nouvelles mutations (néo-mutations) peuvent se produire, ce taux de mutation est de l'ordre de 10^{-8} par nucléotide et par génération chez l'homme. Ces mutations sont à l'origine de la diversité génétique. Si ces mutations affectent les cellules germinales (transmises à la descendance), elles peuvent être à l'origine d'anomalies génétiques héréditaires (Boichard, Grohs, Danchin-Burge, & Capitan, 2016). La taille des génomes ovins et caprins est d'environ 2,6 Gb, ce qui correspond à l'apparition d'environ 30 nouveaux (*de novo*) variants par gamète (et donc le double par individu). Bien que la plupart d'entre eux soient localisés dans des régions non codantes, chaque individu est porteur de 1 à 5 mutations hautement délétères dans son génome (Georges et al., 2019). Ces variants peuvent correspondre à la modification d'une seule base jusqu'à plusieurs kilobases. L'impact d'un variant dépend également de sa localisation sur la séquence d'ADN. La détection des polymorphismes d'un individu se fait par comparaison avec une séquence de référence (Voir [paragraphe 3.1.2](#)).

2.1.2.1 Les variants de une à quelques bases

Séquence référence	AGTTCATGCTGGACCTG T CAGAAGCCGAGG
SNP	AGTTCATGCTGGACCTG C CAGAAGCCGAGG
Séquence référence	AGTTCATGCTGGACCTGTCAGAAGCCGAGG
Insertion	AGTTCATGCTGGACCTG A TCAGAAGCCGAGG
Séquence référence	AGTTCATGCTGGACCTG T CAGAAGCCGAGG
Délétion	AGTTCATGCTGGACCTG - CAGAAGCCGAGG
Séquence référence	AGTTCATGCTGGACCTG TC AGAAGCCGAGG
Microsatellite	AGTTCATGCTGGACCTG TCTCTC CAGAAGCCGAGG

TABLE 2.2 – Les différents types de variants de petites tailles. *Adapté de* Vignal (2011).

La séquence du génome est légèrement variable entre individus appartenant à une même espèce. Ces variations (appelées polymorphismes) impliquent des changements d'une à quelques

bases (Table 2.2) (Vignal, 2011).

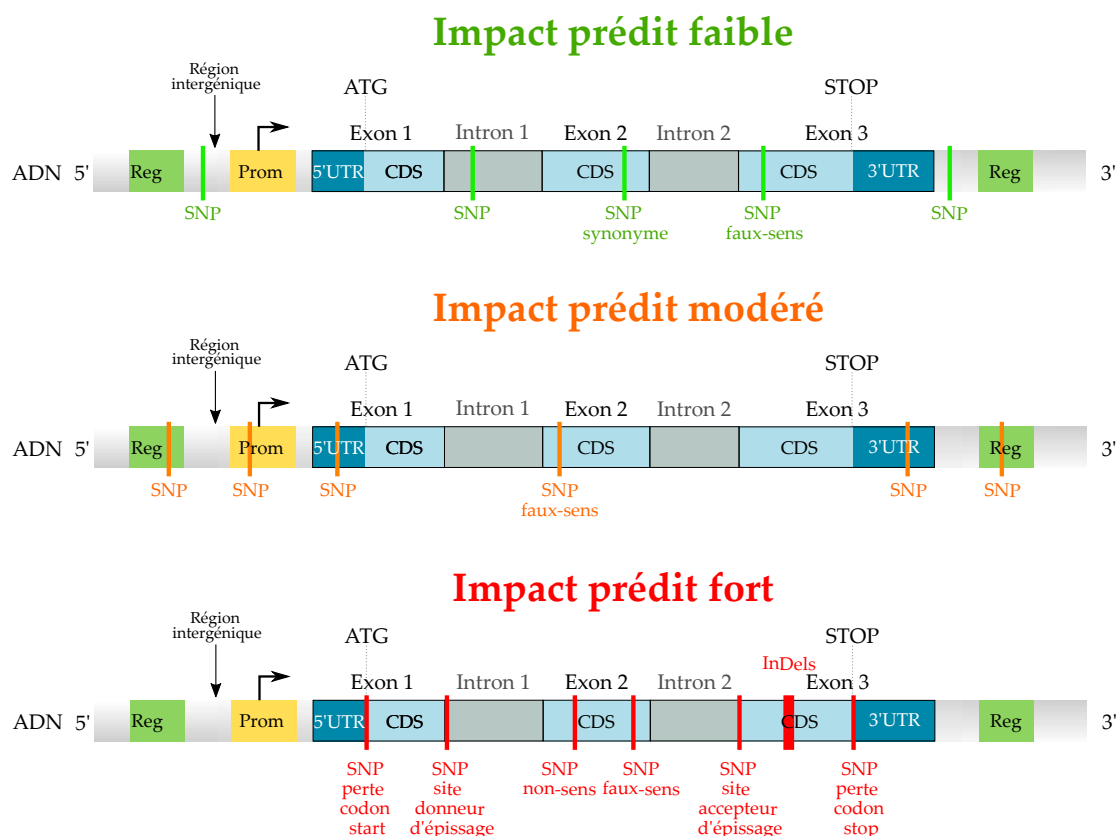


FIGURE 2.5 – **Impacts biologiques des variants de type SNP.** Seul le brin 5' → 3' est représenté. Reg= séquences régulatrices, Prom= Promoteur, CDS= Coding sequence/ séquence codante, UTR= Untranslated region/ Région non traduite, SNP= Single Nucleotide Polymorphism/ variant d'une paire de bases. La présence de mutations de type Insertion/Délétion (InDel) peut également entraîner un décalage du cadre de lecture dans la séquence codante.

Les polymorphismes de type SNP (*Single Nucleotide Polymorphism*) correspondent au remplacement d'une base par une autre. Ces mutations ponctuelles d'une seule paire de bases de l'ADN peuvent affecter des régions diverses au sein du génome (Figure 2.5). Parmi les SNP, les transitions (changement purine-purine ou pyrimidine-pyrimidine $A \leftrightarrow G$ ou $C \leftrightarrow T$) sont plus fréquentes que les transversions (changement purine-pyrimidine : $A \leftrightarrow C$ ou T ; et $G \leftrightarrow C$ ou T) (Hodgkinson and Eyre-Walker, 2010). Lorsque ces SNP sont localisés dans des régions intergéniques, non traduites ou introniques, ils n'ont dans la plupart du temps aucun impact. Cependant, lorsque ces mutations affectent les régions codantes (exons des gènes), l'impact peut être faible à fort suivant le type de mutation. Trois types de mutation peuvent se produire dans le CDS : (i) la mutation même-sens/silencieuse : le changement de nucléotide n'a pas d'effet sur la protéine, plusieurs codons codent pour le même acide aminé (exemple : $TGT \rightarrow TGC$ > Cystéine) ; (ii) la mutation faux-sens (*missense*) : le changement de nucléotide a un effet sur la séquence protéique, le nouveau codon code pour un autre acide aminé (exemple : $TGT \rightarrow TGG$ > Cystéine → Tryptophane) et (iii) la mutation non-sens (*stop-gain*) : le changement de nucléotide conduit à l'apparition d'un codon stop (exemple : $TTA \rightarrow TGA$ > Leucine → Stop).

Les SNP localisés dans les sites de régulation ont des impacts modérés et ont un rôle sur la régulation de l'expression du gène en empêchant ou permettant la fixation de facteur de transcription ou en modulant les impacts de la méthylation. Les SNP présents dans les régions UTR (régions non traduites de l'ARNm) ont également des impacts modérés et peuvent contribuer à rendre moins stable l'ARNm en modifiant sa durée de vie. Les mutations localisées dans les sites d'épissage (donneurs/accepteurs) situés à proximité des exons, peuvent conduire à créer ou supprimer des exons et auront donc des conséquences fortes sur l'expression du gène et les séquences protéiques produites. Les SNP induisant une perte de codon start ou stop ont également des effets forts puisque la protéine ne peut pas être traduite ou être plus longue.

Les polymorphismes de type petites insertions ou délétions (Table 2.2) peuvent être de tailles variables allant d'une paire de bases à plusieurs dizaines de paires de bases. Si ces mutations sont localisées dans les régions codantes, elles entraînent un décalage du cadre de lecture (*frameshift*) conduisant à l'apparition d'un codon stop prématuré, la protéine produite est alors tronquée et/ou modifiée. Si l'insertion/délétion conduit à la suppression/ajout d'un multiple de 3 nucléotides (mutation *inframe deletion* ou *insertion*), cela conduit à la perte/ajout d'un ou plusieurs acides aminés.

Des séquences répétées de quelques paires de bases (2 à 4 pb) (Table 2.2), appelées microsatellites, peuvent elles aussi avoir des impacts forts quand elles rompent le cadre de lecture. Ces derniers ont été à l'origine des premières analyses de diversité de l'ADN avant l'avènement de l'utilisation des SNP.

Les variants qui ont un impact biologique fort sur la protéine (non-sens, site accepteur/donneur, perte codon start/stop, faux-sens, frameshift et Indels) sont appelés mutations pertes de fonction (*Loss-of-function*, LoF) et sont les principaux variants responsables d'anomalies génétiques (MacArthur et al., 2012).

2.1.2.2 Les variants structuraux

Les variants de structure sont des mutations et correspondent à de grandes portions d'ADN (souvent >50pb) (Figure 2.6) (Alkan et al., 2011). Ce terme inclut des grandes insertions/délétions, des duplications, des inversions et translocations. Les variants de structure qui résultent d'un changement du nombre de copies (délétion, insertion et duplication) dans la séquence d'ADN sont appelés CNV (*Copy Number Variation*). L'étude des variants de structure est très récente et rendue possible par le développement des nouvelles techniques de séquençage en lectures longues (voir paragraphe 3.1.1.3). Compte tenu de la taille des variants structuraux (>50pb), de nombreux gènes sont compris dans ces portions de génome ce qui peut conduire à des impacts biologiques forts.

2.1.2.3 Occurrence des différents types de variants en lien avec des anomalies

Comme le montre la Figure 2.7, les polymorphismes de type SNP (mutations faux-sens et non-sens) représentent plus de 50% des anomalies identifiées. On retrouve, au second rang, les mutations de type insertions/délétions (~ 20-30%). Les variants structuraux de type CNV sont peu fréquents même si le développement des nouvelles techniques de séquençage permet de les

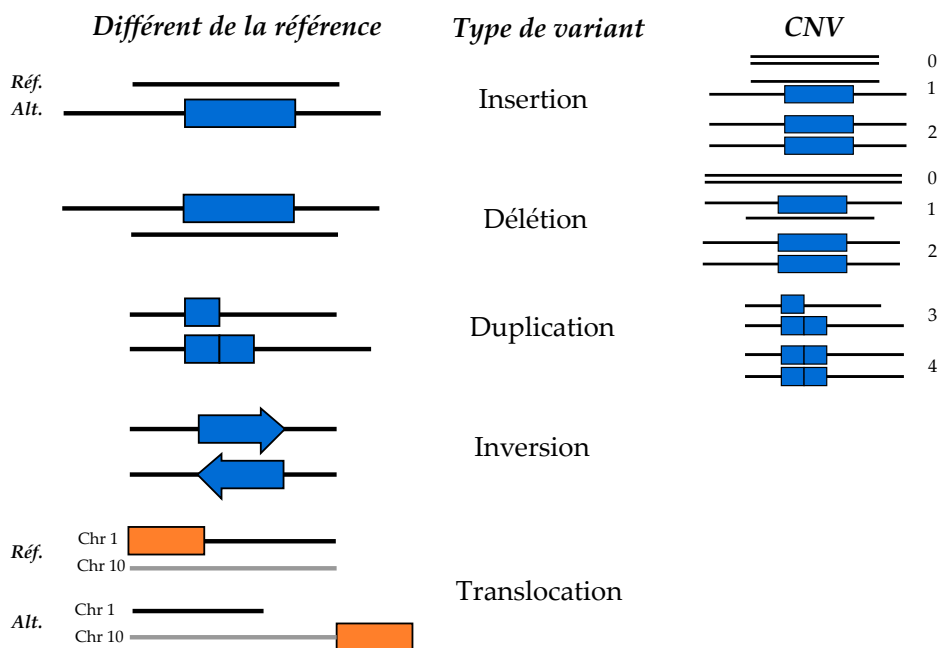


FIGURE 2.6 – Les différents types de variants structuraux. CNV = Copy Number Variation. Adapté de Hollox et al. (2022).

étudier. Ce nombre devrait donc évoluer très rapidement dans la prochaine décennie. Les variants les moins fréquents sont les variants structuraux de type inversion ou translocation. L'analyse des translocations passe principalement par la construction d'un caryotype (coûteux).

Nous venons de voir que les différentes familles de polymorphismes présentent des caractéristiques très distinctes (fréquence, variabilité) et sont adaptées à des utilisations spécifiques. En effet, la facilité de génotypage est essentielle pour les étudier à grande échelle, ce qui est le cas dans le cadre de la sélection génomique où l'on est capable de génotyper plusieurs milliers de SNP pour chaque individu.

2.2 Les anomalies en petits ruminants et leurs modes de transmission

Historiquement, l'étude des anomalies génétiques portait exclusivement sur la description clinique des individus atteints sans en connaître le déterminisme moléculaire. L'amélioration de la connaissance des génomes (voir paragraphe 2.1.2.1) et des technologies (voir paragraphe 3.1) ont contribué à résoudre le déterminisme génétique de nombreuses anomalies génétiques en identifiant les mutations causales. Depuis longtemps, les scientifiques ont la volonté de synthétiser les découvertes à l'échelle internationale. L'ensemble de ces résultats est disponible dans des bases de données publiques chez l'homme et la souris. La base *Online Mendelian Inheritance in Man* (OMIM, <https://www.omim.org/>) regroupe à ce jour 7 233 phénotypes avec déterminisme génétique connu chez l'homme. La même base existe chez la souris (MGI, *Mouse Genome Informatics*, <http://www.informatics.jax.org>). Ce travail a également été réalisé chez les animaux (hors espèces de laboratoire) dans la base de données *Online Mendelian Inheritance in Animals* (OMIA, <https://www.omia.org/home>). Cette base internationale regroupe l'ensemble des anomalies héréditaires et d'autres caractères pour 346 espèces (hors espèces de laboratoire)

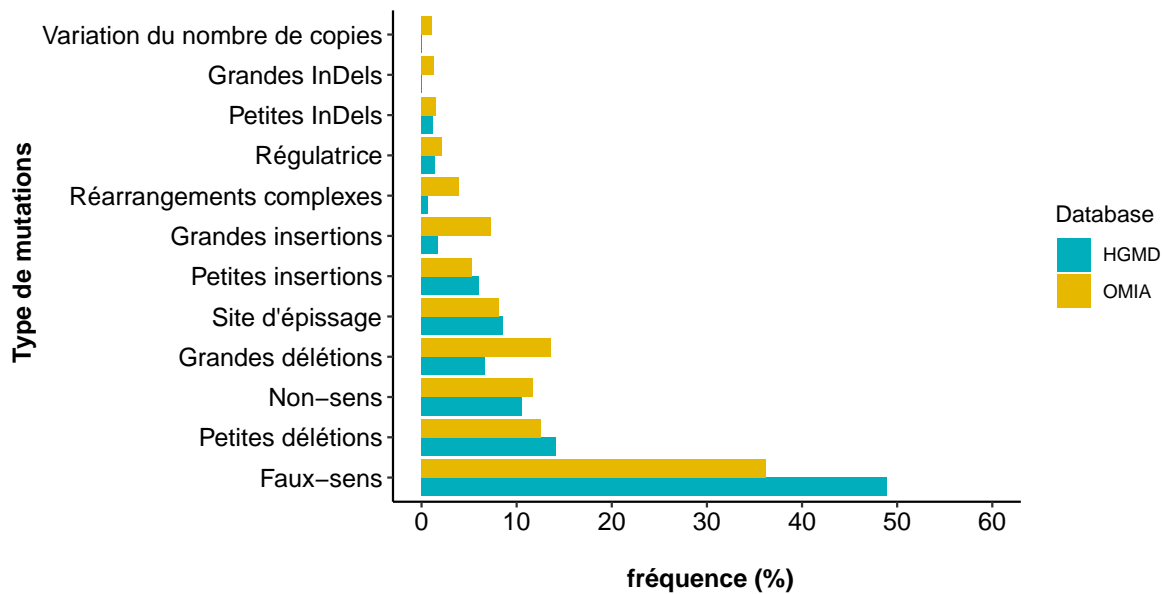


FIGURE 2.7 – Fréquences des différents types de mutations causales identifiées chez l'homme et les espèces animales (hors espèces de laboratoire ~ 346 espèces). Ces données sont extraites des bases de données HGMD (Human Gene Mutation Database, <https://www.hgmd.cf.ac.uk>, consulté le 22/07/2022) et OMIA (Online Mendelian Inheritance in Animals, <https://www.omia.org>, consulté le 22/07/2022). A cette date, 352 731 et 1354 mutations causales sont répertoriées chez l'homme et les espèces animales (hors espèces de laboratoire), respectivement.

pour lesquels les phénotypes, gènes et variants associés sont décrits (Nicholas & Hobbs, 2014). Elle est gérée par Frank Nicholas et Imke Tammen de l'Université de Sydney (Australie). L'évolution du nombre de mutations causales identifiées chez les espèces animales ne cesse d'augmenter ces dernières années (Figure 2.8). La vache et le chien sont les deux espèces pour lesquelles le nombre de mutations causales identifiées est le plus important. A ce jour en petits ruminants, le nombre d'anomalies identifiées reste limité.

2.2.1 Les anomalies génétiques identifiées en petits ruminants

Chaque individu diploïde possède deux copies de chaque gène, l'une reçue de son père et l'autre de sa mère. On note "A" l'allèle normal et "a" l'allèle muté responsable de l'anomalie. Ainsi, chaque descendant peut être de génotype AA, Aa ou aa. La fréquence de ces génotypes dépend du mode de transmission. Ces modes sont décrits dans les paragraphes suivants. A l'aide de la base de données OMIA, j'ai pu réaliser un inventaire des anomalies génétiques avec un déterminisme génétique connu chez les ovins et caprins (Table 2.3). A la fin de l'année 2021, 110 mutations causales ont été identifiées et répertoriées dans la base. Chez le mouton, 80 mutations causales ont été identifiées (41 sont délétères) que l'on associe à 47 phénotypes distincts. Dans la plupart des cas, un phénotype est associé exclusivement à un gène. Pour les caprins, seulement 30 mutations causales identifiées ont été rapportées (11 sont délétères) que l'on associe à 16 phénotypes distincts. A titre comparatif, chez les vaches, c'est 247 mutations causales rapportées dans la base (209 sont délétères) et elles sont associées à 181 phénotypes distincts.

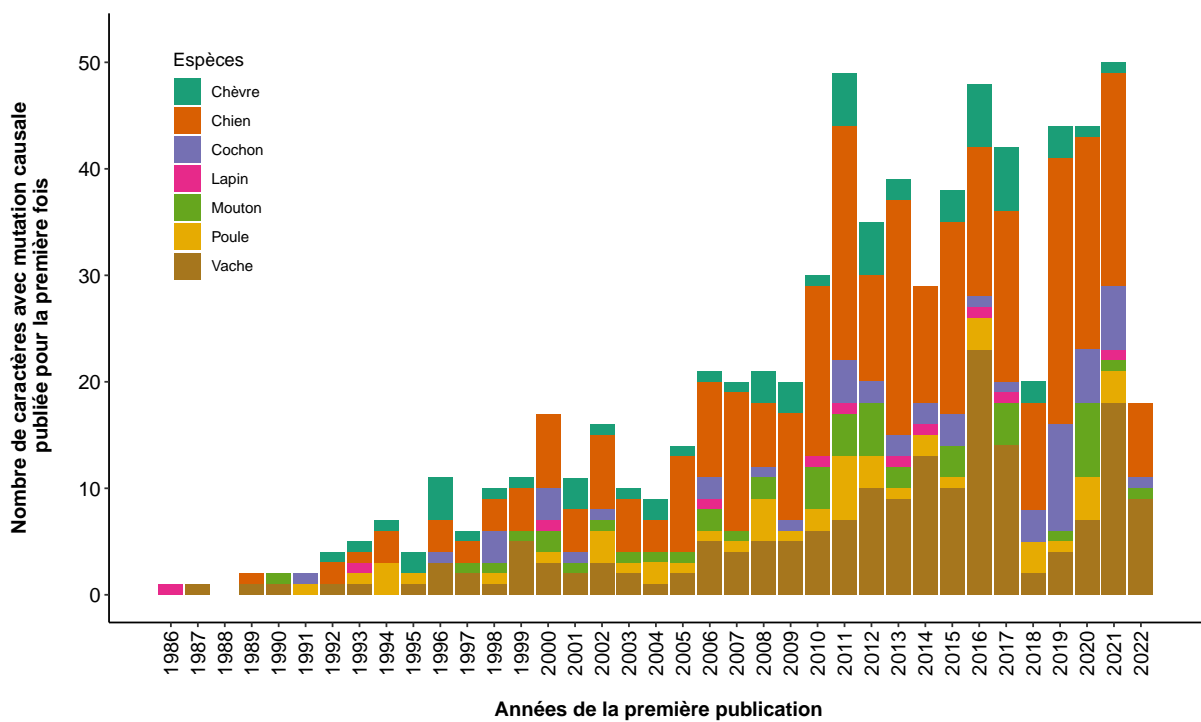


FIGURE 2.8 – Nombre de mutations causales identifiées pour des caractères mendéliens au cours du temps chez sept espèces animales (hors espèces de laboratoire). Extrait de la base de données OMIA le 22/07/2022.

TABLE 2.3 – Liste des anomalies génétiques avec mutations causales identifiées chez les petits ruminants. Extrait de OMIA le 17/12/2021. (... + ...) = association de plusieurs mutations, as = antisense, CNV = Variation du nombre de copies, Co-dom = co-dominance, DEL = délétion, DELINS = délétion et insertion, Dom = dominant, DUP = duplication, ERV = rétrovirus endogène, FS = frameshift (décalage cadre de lecture), IG = intergénique, INS = insertion, Intersex. = intersexualité, MS = missense (faux-sens), NS = nonsense (non-sens, codon-stop), rec = récessif, SD = surdominance polaire, SS = splice site (site d'épissage), UTR = région non traduite, VS = variant de structure.

Espèce	Gène	Phénotype	Région codante	Région non-codante	Région régulatrice	Transmission	Références
Tissu adipeux							
OV	<i>BCO2</i>	Graisse jaune	NS	VS : INS 7,9kb (ERV)		rec	(Vège & Boman, 2010 ; Kent et al., 2021)
Glande endocrine, exocrine et mammaire							
OV	<i>SOC2</i>	Mammites (comptage des cellules somatiques important dans le lait)	MS			rec	(Rupp et al., 2015)
CAP	<i>CSN1S1</i>	Réduction de caséine (Alpha S1)			3'UTR INS	-	(Pérez et al., 1994)
CAP	<i>CSN1S2</i>	Absence de caséine (Alpha S2)	NS			-	(Ramunno et al., 2001)
CAP	<i>CSN2</i>	Absence de caséine (β)	FS (DEL 1pb)		SNP	-	(Cosenza et al., 2007 ; Persuy et al., 1999)
CAP	<i>TG</i>	Goître (hypertrophie de la thyroïde)	NS			rec	(Veenboer & de Vijlder, 1993)

Homéostasie/Métabolisme						
OV	F8	Hémophilie A (hémorragies spontanées)	FS	liée à l'X	(Porada et al., 2010)	
OV	UROD	Porphyrie cutanée tardive (déficit activité hépatique)	MS	rec	(Nezamzadeh et al., 2005)	
Système immunitaire						
OV	LOC105603932/ ZNF389, TMEM154	Résistance/Susceptibilité aux lentivirus	MS1-6, FS1-2	5'UTR DEL 1pb	Multigénique	(White et al., 2014; Heaton et al., 2012)
Stockage lysosome						
OV	CLN5, CLN6, GBA, LOC443060	Céroïde lipofuscinose neuronale (épilepsie, perte de vision)	SS SNP, MS1-3, DEL 83pb	rec	(Bond et al., 2013; Frugier et al., 2008; Tammen et al., 2006; Tynnela, 2000; H. Zhou et al., 2017)	
OV	GLB1, HEXA	Gangliosidose à GM1 ou GM2/maladie de Tay-Sachs (déficit en hexosaminidase)	MS, SS SNP	rec	(Torres et al., 2010; Walker et al., 2012)	
OV	PYGM	Maladie de stockage du glycogène (Syndrome McArdle)	SS SNP	rec	(Tan et al., 1997)	
CAP	GNS	Mucopolysaccharidose	NS	rec	(Cavanagh et al., 1995)	

CAP	MANBA	Béta-mannosidose (surcharge lysosomale, perte auditive)	FS	rec	(Leipprandt et al., 1996)
Proliféricité					
OV	<i>B4GALNT2</i>	Proliféricité	(SNP1+2)	co-dom	(Drouilhet et al., 2013)
OV	<i>BMP15</i>	Proliféricité (stérilité chez les homozygotes)	MS1-6, NS1-2, DEL 17pb (FS), DELINS (FS)	liée à l'X	(Bodin et al., 2007; Calvo et al., 2020; Chantepie et al., 2020; Demars et al., 2013; Galloway et al., 2000; Hamrahan et al., 2004; Lassoued et al., 2017; Martinez-Royo et al., 2008)
OV	<i>BMPR1B</i>	Proliféricité	MS	1/2 dom	(Wilson et al., 2001)
OV	<i>GDF9</i>	Proliféricité (stérilité chez les homozygotes)	DEL 4pb (FS), MS1-5	1/2 dom	(Hamrahan et al., 2004; Holm et al., 2022; Nicol et al., 2009; Silva et al., 2011; Souza et al., 2014; Våge et al., 2013)
Tégument (Peau/toison)					
OV	<i>ADAMTS2</i>	Dermatosparaxis - Syndrome d'Ehlers-Danlos (défaut production collagène conduisant à des déchirures)	NS, MS	rec	(H. Zhou et al., 2012; Monteagudo, Ferrer, et al., 2015)
OV	<i>HR</i>	Hypotrichose (arrêt développement pileux)	NS	rec	(Finocchiaro et al., 2003)

OV	<i>IRF2BP2</i>	Toison laineuse (courte et frisée)		3'UTR INS asEIF2S2	rec	(Demars et al., 2017)
OV	<i>ITGB4, LAMC2</i>	Epidermolyse bulleuse jonctionnelle (absence de peau aux extrémités des membres et muqueuses)	DEL 4pb, DEL 2pb, NS		rec	(Mömke et al., 2011 ; Suárez-Vega et al., 2015 ; Fabre et al., 2020)
Anomalies musculaires						
OV	<i>CLCN1</i>	Myotonie (perte de tonus musculaire)	MS		rec	(Monteagudo, Tejedor, et al., 2015)
OV	<i>DLK1 + PEG11</i>	Hypertrophie musculaire (Callipyge)		IG SNP	surdominance polaire	(Freking et al., 2002 ; Georges et al., 2003)
OV	<i>MSTN</i>	Hypertrophie musculaire		3'UTR SNP		(Clöp et al., 2006)
OV	<i>TNNI1</i>	Dystrophie musculaire progressive congénitale	SS DEL 1pb		rec	(Clayton et al., 2020)
CAP	<i>CLCN1</i>	Myotonie (perte de tonus musculaire)	MS		dom	(Beck et al., 1996)
Anomalies du système nerveux central et/ou se traduisant par des troubles neurologiques						
OV	<i>AGTPBP1</i>	Maladie neurone moteur	MS		rec	(Zhao, Onteru, Dittmer, et al., 2012)

OV	<i>FGF14</i>	Ataxie spino-cérébelleuse	NS	dom (pénétrance incomplète)	(Dittmer et al., 2017)
OV	<i>PLA2G6</i>	Dystrophie neuro-axonale	(SS SNP + NS)	rec ?	(Letko et al., 2021)
OV	<i>PRNP</i>	Résistance/Sensibilité à la tremblante (Encéphalopathie spongiforme)	MS1-4	-	(Goldmann et al., 1990 ; Lancher et al., 1993 ; Moum et al., 2005)
OV	<i>RELN</i>	Lissencéphalie et hypoplasie cérébelleuse	DEL 31pb	rec	(Suárez-Vega et al., 2013)
CAP	<i>PRNP</i>	Résistance/Sensibilité à la tremblante (Encéphalopathie spongiforme)	NS, MSI-8	-	(Barillet et al., 2009 ; Benestad et al., 2012 ; Billinis et al., 2002 ; Goldmann et al., 2011 ; Papasavva-Stylianou et al., 2007 ; Vaccari et al., 2006)
Rein/système urinaire					
OV	<i>AGXT</i>	Hyperoxalurie (déficit enzymatique hépatique péroxysomal)	MS	rec	(Letko et al., 2020)
Anomalies du système digestif					
OV	<i>EDNRB</i>	Syndrôme de Shah-Waardenburg (hypopigmentation, obstruction intestinale)	SV (DEL 110kb)	rec	(Lühken et al., 2012)

OV	<i>OBSL1</i>	Syndrome brachygnathie, cardiomégalie et hypoplasie rénale (encéphalocèle occipital, kystes rénaux, fibrose hépatique et polydactylie)	DEL 1pb (FS)	rec	(Woolley et al., 2020)
OV	<i>TMEM67</i>	Syndrome de dysplasie fibrokystique hépato-rénale de type Meckel	(MS1+MS2)	rec	(Stayner et al., 2017)
Couleur de toison					
OV	<i>ASIP</i>	Toison blanche	CNV (DUP 190kb)	dom	(Norris & Whan, 2008)
OV	<i>ASIP</i>	Toison noire	DEL 9b, DEL 5pb (FS), MS	rec	(Norris & Whan, 2008 ; Smit et al., 2002)
OV	<i>MC1R</i>	Toison sombre (surproduction d'eumelanine)	MS1-3	dom	(Rochus et al., 2019 ; Våge et al., 1999)
OV	<i>MLPH</i>	Dilution de la couleur (tâches grises à la place de tâches noires)	NS	rec	(Posbergh et al., 2020)
OV	<i>TYRP1</i>	Toison fauve (marron clair)/noire	MS	dom/rec	(Gratten et al., 2007)
OV	<i>TYRP1</i>	Toison brune	DEL 2pb (FS), NS, MS	rec	(Paris et al., 2019 ; Raadsma et al., 2013)
CAP	<i>ASIP</i>	Différentes couleurs de toison	CNV 1-5	-	(Fontanesi et al., 2011 ; Henkel et al., 2019)

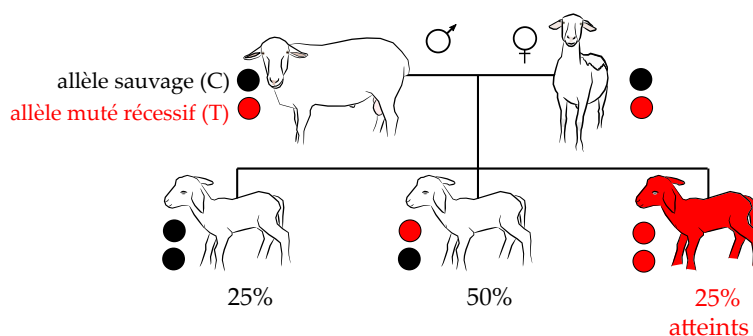
CAP	<i>EDNRA</i>	Toison avec tâches blanches	(MS)	+CNV)	1/2 dom	(Menzi et al., 2016)
CAP	<i>KIT</i>	Présence de tâches blanches (absence mélanocytes)		CNV	-	(Henkel et al., 2019)
CAP	<i>KIT</i>	Toison blanche		CNV	-	(Henkel et al., 2019)
CAP	<i>MC1R</i>	Toison noire	MS		dom	(Fontanesi et al., 2009)
CAP	<i>MC1R</i>	Rouge	NS		rec	(Fontanesi et al., 2009)
CAP	<i>TYPR1</i>	Toison brune	MS		dom	(Becker et al., 2014)
Malformations du squelette (tête, corps et membres) et articulations						
OV	<i>DMP1</i>	Rachitisme hypophosphatémique (défaut minéralisation tissu osseux)	NS		rec	(Zhao et al., 2011)
OV	<i>FGFR3</i>	Chondrodysplasie (agneau araignée)	MS		rec	(Beever et al., 2006)
OV	<i>SLC13A1</i>	Chondrodysplasie (nanisme)		DEL 1pb (FS)	rec	(Zhao, Onteru, Piripi, et al., 2012)
Malformations oculaires et troubles de la vision						
OV	<i>CNGA3</i>	Achromatopsie (cécité diurne)	NS, MS		rec	(Gootwine et al., 2017; Reichner et al., 2010)
OV	<i>PITX3</i>	Microphthalmie (absence d'œil/œil plus petit)	MS		rec	(Becker et al., 2010)

Anatomie cranio-faciale						
OV	<i>HMX1</i>	Microtie (défaut de développement du pavillon de l'oreille)	VS : DUP 76pb (downstream <i>HMX1</i>)	dom	(He et al., 2020)	
OV	<i>HOXD1</i>	Polycérate (=cornes multiples) (hétérozygote) et mortalité embryonnaire/fœtale (homozygote)	SD (DEL 4pb)	co-dom	(Allais-Bonnet et al., 2021)	
OV	<i>OTX2</i>	Otocéphalie (non migration (durant l'embryogenèse) des oreilles en position crâniale))	NS	dom	(Paris et al., 2020)	
OV	<i>RXFP2</i>	Sans cornes	3'UTR INS 1,8kb asEEF1A1	rec	(Wiedemar & Drögemüller, 2015)	
CAP	<i>FOXL2, ERG</i>	Intersexualité associée à l'absence de cornes (PIS)	VS (DEL 11,7kb + DUP INS 480kb)	dom (sans cornes) / rec (intersex.)	(Pailloux et al., 2001 ; Simon et al., 2020)	
CAP	<i>HOXD1</i>	Polycérate (=cornes multiples) (hétérozygote) et mortalité embryonnaire/fœtale (homozygote)	DELINS 177 kb	dom	(Allais-Bonnet et al., 2021)	

2.2.2 Anomalies autosomiques récessives et co-dominantes

Une anomalie est autosomique récessive lorsque la mutation est localisée sur un autosome et que l'allèle muté à l'état homozygote impacte le phénotype (l'individu possède les deux copies de cet allèle, l'une héritée de son père, l'autre de sa mère). A l'état hétérozygote, cette mutation n'a pas d'impact sur le phénotype (Figure 2.9 A). Lorsque deux parents hétérozygotes s'accouplent, 75% des descendants sont non atteints (25% d'homozygotes non porteurs et 50% d'hétérozygotes) et 25% sont atteints (homozygotes mutés). Par exemple, en ovins plusieurs mutations localisées dans le gène *ITGB4* (*Integrin Subunit Beta 4*) sont associées au syndrome d'épidermolyse bulleuse jonctionnelle (OMIA 001948-9940) (Table 2.3). Les deux variants identifiés sont un non-sens OAR11 (Oar_rambouillet_v1.0) : g.7,412,626C>T (p.Arg885*) (Figure 2.9 A) et une délétion de 4pb conduisant à une rupture du cadre de lecture OAR11 (Oar_rambouillet_v1.0) : g.7,425,460_7,425,463del, affectant le même gène et conduisant au même phénotype (Fabre et al., 2020 ; Suárez-Vega et al., 2015). Les individus homozygotes atteints sont caractérisés par des fragilités épithéliales au niveau des extrémités des membres et muqueuses provoquant la mort périnatale de l'agneau. Plusieurs variants chez l'homme conduisant à ce phénotype ont également été identifiés (OMIM 147557) dans le même gène.

A Mutation non-sens OAR11:g.7,412,626C>T dans *ITGB4*



B Mutations régulatrices OAR11:g.36,938,224T>A et OAR11:g.37034573A>G (*FecL^L*) dans *B4GALNT2*

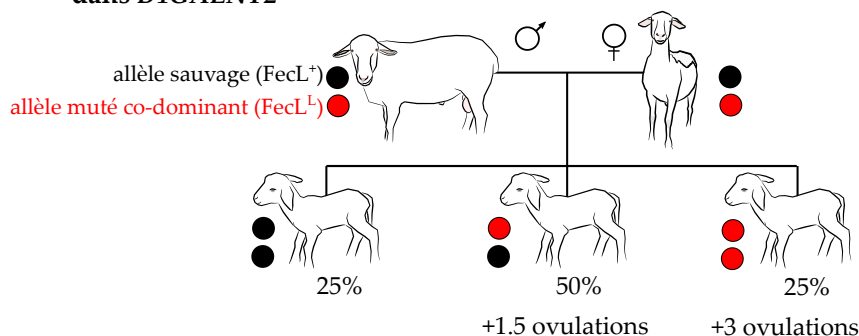


FIGURE 2.9 – Mode de transmission autosomique récessive et co-dominance. (A) Situation de récessivité avec la mutation non-sens OAR11 (Oar_rambouillet_v1.0) : g.7,412,626C>T, p.Arg885*. (B) Cas de co-dominance avec les mutations régulatrices OAR11 (Oar_rambouillet_v1.0) : g.36,938,224T>A et g.37034573A>G (les deux allèles mutés sont regroupés sous le nom de mutation *FecL^L*).

De plus, l'allèle muté peut également être co-dominant, c'est-à-dire que chaque allèle participe en commun à la détermination d'un phénotype "intermédiaire" entre les deux phénotypes

gérés par chacun des allèles (Figure 2.9 B). Ce phénomène inclut souvent un effet additif (Boichard, Grohs, Danchin-Burge, & Capitan, 2016). Par exemple en ovins, deux mutations OAR11 (Oar_rambouillet_v1.0) : g.36,938,224T>A (intron 7 de *B4GALNT2* (*Beta-1,4-N-Acetyl-Galactosaminyltransferase 2*)) et g.37034573A>G (région intergénique entre *B4GALNT2* et *IGF2BP1* (*Insulin like growth factor 2 mRNA binding protein 1*)) en parfait déséquilibre de liaison, sont associées à une augmentation du nombre d'ovulations (hyperprolificité) variable en fonction du génotype des individus (OMIA 001885-9940) (Drouilhet et al., 2013, 2009). L'effet additif de l'allèle permet aux brebis hétérozygotes d'avoir un nombre d'ovulations par cycle augmenté de +1,5 et de +3 pour les brebis homozygotes par rapport aux brebis non porteuses de la mutation (nombre d'ovulations de 1,5 pour les +/+). La mutation entraîne une expression ectopique très importante de *B4GALNT2* dans l'ovaire. Chez le mâle, aucun effet de la mutation *FecL^L* n'a été démontré.

2.2.3 Anomalies autosomiques dominantes

Le cas de dominance correspond à la situation où une seule copie de l'allèle est nécessaire pour exprimer le phénotype. Les individus hétérozygotes et homozygotes ont des phénotypes identiques (Figure 2.10).

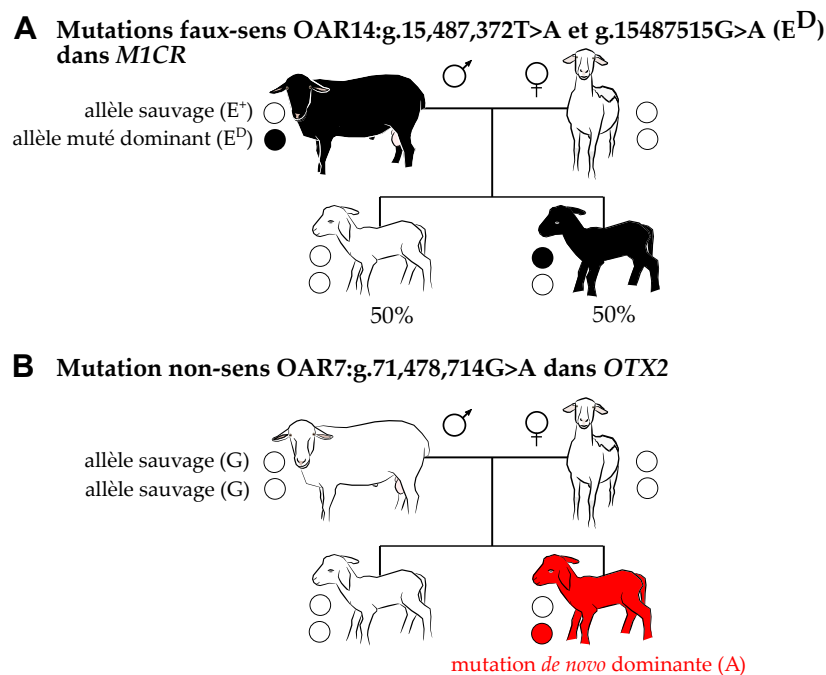


FIGURE 2.10 – Anomalie autosomique dominante. (A) Cas de dominance avec l'exemple des mutations faux-sens OAR14 (Oar_rambouillet_v1.0) : g.15,487,372T>A (p.Met73Lys) et g.15487515G>A (p.Asp121Asn) (regroupées sous le nom de mutation E^D) responsable de la couleur de toison noire. (B) Cas de dominance avec une mutation apparue *de novo* chez un descendant avec l'exemple de la mutation non-sens OAR7 (Oar_rambouillet_v1.0) : g.71478714G>A (p.Arg89*) localisée dans le gène *OTX2*.

Dans les populations d'élevage, les anomalies génétiques à déterminisme dominant sont souvent impliquées dans des variations phénotypiques liées à la morphologie telles que la couleur de toison, sans impacter la viabilité des individus porteurs de la mutation (et est donc transmissible à la descendance). De nombreux exemples sont cités dans la Table 2.3, affectant les gènes *MC1R*,

ASIP et *TYRP1*. Par exemple, deux mutations faux-sens OAR14 (Oar_rambouillet_v1.0) : g.15,487,372T>A (p.Met73Lys) et g.15,487,515G>A (p.Asp121Asn) (regroupées sous le nom de mutation E^D) sont en co-ségrégation dans le gène *MC1R* (*MelanoCortin 1 Receptor*) associées à la couleur de toison noire (OMIA 001199-9940) (Rochus et al., 2019; Våge et al., 1999). Ces mutations entraînent une surproduction d'eumélanine et une seule copie de l'allèle E^D est nécessaire pour obtenir un animal noir (Figure 2.10 A).

Cependant, lorsque la mutation à déterminisme dominant a un impact délétère, les individus atteints sont dans la plupart du temps porteur d'une mutation *de novo*. Par exemple, un variant *de novo* non-sens a été mis en évidence dans le gène *OTX2* (*Orthodenticle Homeobox 2*) en ovins OAR7 (Oar_rambouillet_v1.0) : g.71,478,714G>A (p.Arg89*) conduisant à une otocéphalie (non migration des oreilles pendant le développement embryonnaire) accompagnée de nombreuses malformations congénitales létales pour l'agneau (OMIA 002227-9940, Figure 2.10 B) (Paris et al., 2020).

Il est important de noter que dans la plupart des cas de mutations dominantes délétères, le premier individu porteur de cette mutation est souvent mosaïque c'est-à-dire que la mutation n'a pas lieu au stade une cellule. De ce fait, une partie des cellules constituant l'individu est porteuse de la mutation. Si le mosaïcisme n'atteint pas la lignée germinale alors la mutation n'est pas diffusée. Si la lignée germinale est entièrement atteinte, la mutation diffusera selon la ségrégation à partir d'un individu hétérozygote, soit la moitié des descendants porteurs. En fonction du niveau de mosaïcisme de la lignée germinale, une plus faible proportion de porteurs (<1/2) est attendue (Spinner & Conlin, 2014).

2.2.4 Anomalies liées aux chromosomes sexuels

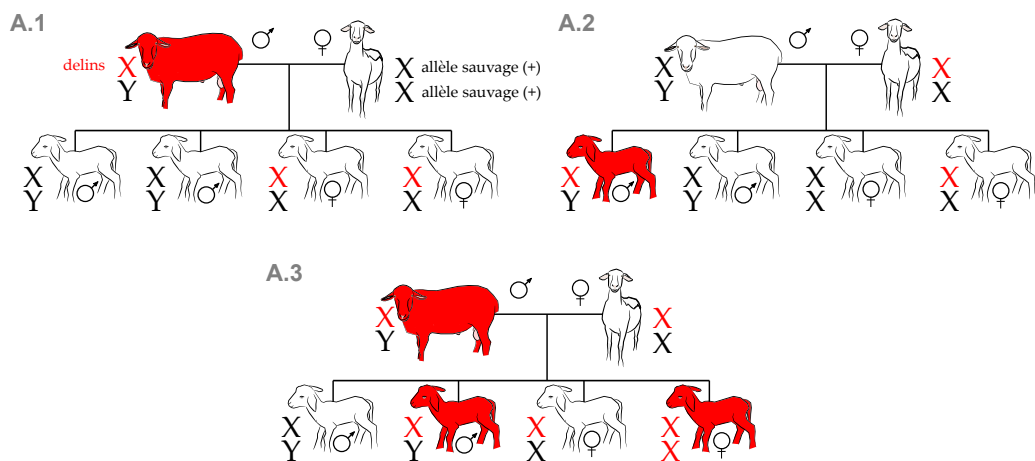
Les modes de transmission décrits précédemment sont également applicables aux chromosomes sexuels (X ou Y). Les mâles ont un chromosome X et un chromosome Y tandis que les femelles ont deux chromosomes X. Dans la littérature, les anomalies liées aux chromosomes sexuels concernent principalement le chromosome X. En effet, peu de gènes sont présents sur le chromosome Y et sont principalement impliqués dans la formation des testicules et la production du sperme (Lau, 2020). Les mâles porteurs de ce type de mutation sont stériles et ne peuvent donc pas la transmettre aux descendants. Pour les anomalies localisées sur le chromosome X, le déterminisme peut être dominant ou récessif (Figure 2.11).

A titre d'exemple, une mutation récessive InDel résultant en un décalage du cadre de lecture OARX (Oar_rambouillet_v1.0) : g.86,301,507_86,301,516delinsTAATTAATACC dans le gène *F8* (*Coagulation Factor VIII*) est responsable d'hémophilie de type A identifiée en ovins (OMIA 000437-9940, Figure 2.11 A) (Porada et al., 2010). Une seule (et unique) copie de la mutation est nécessaire pour qu'un mâle (Figure 2.11 A.1 et A.2) exprime le phénotype (hémarthrose des genoux) tandis que pour la femelle, les deux copies de la mutation sont nécessaires pour exprimer le phénotype (Figure 2.11 A.3).

De nombreuses mutations identifiées sur les chromosomes sexuels sont associées à de la stérilité ou à une faible fertilité, notamment chez les femelles (mutations dans les gènes *BMP15* (*Bone Morphogenetic Protein 15*) et *GDF9* (*Growth Differentiation Factor 9*), Table 2.3). Par exemple,

la mutation faux-sens $FecX^L$ OARX (Oar_rambouillet_v1.0) :g.56,594,981C>T (p.Cys321Tyr) (OMIA 002306-9940) est dominante chez la femelle pour le taux d'ovulation (+1,5) (Figure 2.11 B) et récessive pour le phénotype de stérilité (Figure 2.11 B.3) (Bodin et al., 2007). A l'état hétérozygote, la mutation dans *BMP15* réduit l'activité du système de signalisation des BMP, ce qui conduit à diminuer la prolifération et la sensibilité à la FSH (hormone de stimulation folliculaire) des cellules de la granulosa. La stérilité à l'état homozygote est caractérisée par des ovaires non développés, bloqués au stade primaire de la folliculogenèse liée à une quantité de BMP trop faible (Monniaux et al., 2009). Chez le mâle, aucun effet de la mutation n'a été identifié.

A Mutation OARX:g.86,301,507_86,301,516delinsTAATTAATACC dans *F8*



B Mutation faux-sens OARX:g.56,594,981C>T ($FecX^L$) dans *BMP15*

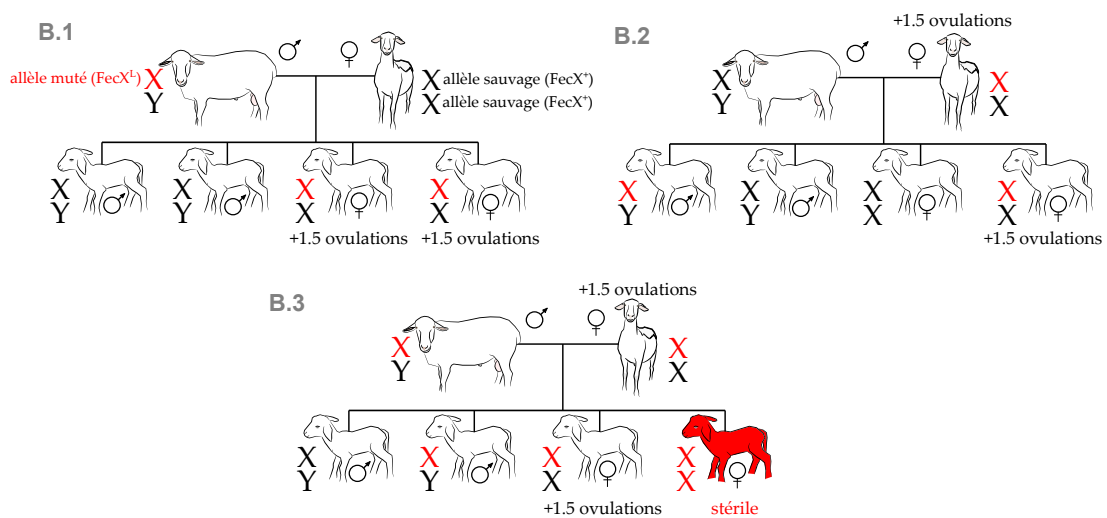


FIGURE 2.11 – Anomalies liées à l’X. (A) Cas d’une mutation récessive liée à l’X OARX :g.86,301,507_86,301,516delinsTAATTAATACC dans *F8*. (B) Cas de la mutation faux-sens OARX :g.56,594,981C>T (appelée $FecX^L$) dans *BMP15* qui est dominante pour le phénotype de prolificité et récessive pour le phénotype de stérilité.

2.2.5 Anomalies mitochondriales

Le nombre de gènes localisés dans la mitochondrie est assez faible : 37 gènes en ovins (NC_001941.1) et 37 en caprins (NC_005044.2). Cependant, le nombre de mutations affectant

ces gènes est important. Pour chaque individu, l'ensemble des mitochondries sont héritées de sa mère via le contenu cytoplasmique de l'ovocyte fécondé. De ce fait, les mâles atteints ne transmettent jamais la mutation aux descendants tandis que les femelles vont transmettre la mutation à l'ensemble des descendants (Figure 2.12). Suivant la mutation et le gène impacté, l'effet sur le phénotype peut être plus ou moins délétère.

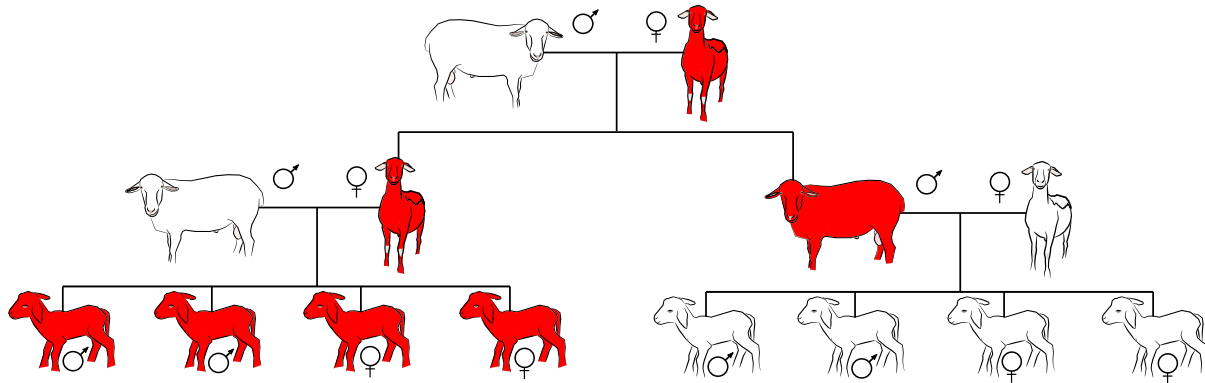


FIGURE 2.12 – **Transmission des anomalies mitochondriales.** Les animaux représentés en rouge expriment le phénotype.

Aucune mutation mitochondriale n'est remontée dans la base de données OMIA. Néanmoins, plusieurs études ont décrit de telles mutations. Par exemple, deux substitutions non synonymes (Phe33Leu et Asp171Asn) et une mutation de type InDel ont été observées dans le gène mitochondrial du cytochrome B (*CYTB*), résultant en une protéine tronquée (Pal et al., 2019). Les agneaux porteurs des mutations présentent une anémie et un dysfonctionnement de la plupart des organes vitaux (foie et reins).

2.2.6 Anomalies chromosomiques

Des anomalies liées à un nombre anormal de chromosomes (ou portions) peuvent avoir lieu, c'est l'aneuploïdie. Comme chez l'Homme, des anomalies chromosomiques sont identifiées dans les espèces animales d'élevage pouvant avoir un impact sur le nombre et/ou la structure des chromosomes (Table 2.4). Elles peuvent être responsables d'anomalies congénitales, de mortalité embryonnaire, d'infertilité ou de cancers. Lors de la fécondation, un gamète déséquilibré peut fusionner avec un gamète normal. Cette anomalie génétique résulte d'un défaut de ségrégation des chromosomes lors de la méiose. Un gamète équilibré est composé d'une seule copie de l'ensemble des chromosomes tandis qu'un gamète déséquilibré possède un morceau de chromosome en moins ou en plus. Dans la plupart des cas, l'animal à l'origine de ces gamètes déséquilibrés est porteur à l'état hétérozygote d'une translocation chromosomique (Figure 2.13). Un animal transloqué équilibré hétérozygote est phénotypiquement normal, mais il est capable de transmettre l'anomalie à ses descendants. On distingue deux types de translocations : réciproque et robertsonienne.

Dans le cas d'une translocation réciproque, des paires de chromosomes ont échangé réciproquement des portions de chromosomes (Figure 2.13 A). Dans ce cas, un animal transloqué hétérozygote produit trois types de gamètes (25% normal, 25% équilibré et 50% déséquilibré). Après fécondation, 50% des zygotes sont viables (normal ou transloqué équilibré hétérozygote). Dans le cas d'un déséquilibre, l'issue est souvent délétère avec des mortalités embryonnaires

TABLE 2.4 – Anomalies chromosomiques identifiées en petits ruminants. *Adapté de Iannuzzi et al. (2021)*

Anomalie chromosomique	Conséquence	Fréq.	Nombre d'anomalies en ovins	Nombre d'anomalies en caprins	Références
Anomalies du nombre de chromosomes					
Monosomie du X (X0)	Infertilité chez la femelle	+	3	1	(Zartman et al., 1981 ; Baylis et al., 1984 ; Berry et al., 2018 ; Bhatia & Shanker, 1990)
Syndrome XXY	Hypoplasie testiculaire chez le mâle	+	2	2	(Bruere et al., 1969 ; Moraes et al., 1980 ; Takebayashi & Jorge, 1986 ; Bhatia & Shanker, 1992)
Inversion du sexe	Stérilité	+	2	2	(Ferrer et al., 2009 ; Albarella et al., 2019 ; Pailhoux et al., 1994 ; Vaiman et al., 1996)
Anomalies de structure des chromosomes (translocation)					
Translocation réciproque (rcp)	Fertilité réduite	++	6	0	(Ducos et al., 2008 ; Ananthawat-Jonsson et al., 1992 ; C. P. Popescu & Tixier, 1984 ; C. P. Popescu, 1990 ; Iannuzzi et al., 2013, 2014)
Translocation robertsonienne (rb)		+++	7	5	(Switoński & Stranzinger, 1998 ; Bruère & Mills, 1971 ; Pearce et al., 1994 ; Chaves et al., 2003 ; C. Popescu & Boscher, 1972 ; Evans et al., 1973 ; Gonçalves et al., 1992 ; Burguete et al., 1987)

précoces ou des malformations congénitales souvent létales. A ce jour, 7 translocations réciproques ont été identifiées en ovins (Iannuzzi et al., 2021). Ce type de translocation aboutit souvent à une fertilité réduite ou à une stérilité chez les animaux transloqués hétérozygotes. Ce qui constitue souvent une alerte pour détecter de telles anomalies chez les espèces domestiques.

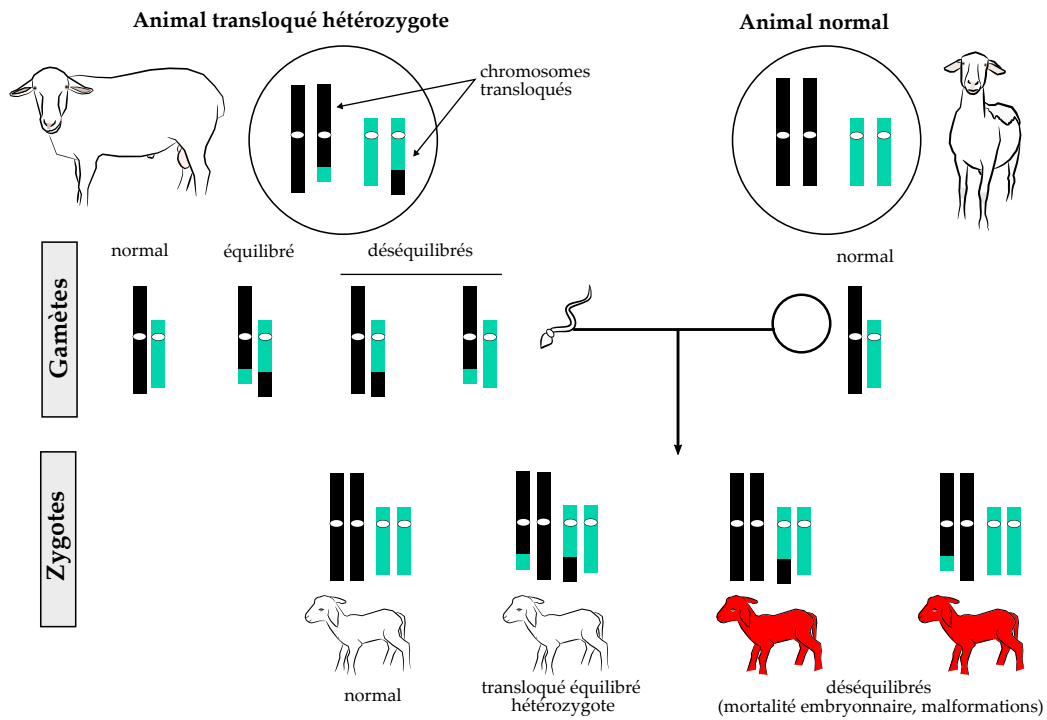
Dans le cas d'une translocation robertsonienne, des chromosomes acrocentriques de paires différentes fusionnent entièrement entre eux par leur centromère (Figure 2.13 B). Dans ce cas, un animal transloqué hétérozygote produit trois types de gamètes (20% normal, 20% équilibré et 60% déséquilibré). Après fécondation, les zygotes peuvent être normaux (16,7% des descendants), viables mais transloqués équilibrés hétérozygotes (16,7%), atteints de trisomie (33,3%) ou de monosomie (33,3%). Dans le cas de trisomie, les individus peuvent avoir des symptômes mais être viables, alors que la monosomie conduit inévitablement à la mort de l'individu. A ce jour, 7 translocations robertsoniennes ont été identifiées en ovins et 5 en caprins (Iannuzzi et al., 2021).

Le nombre d'anomalies chromosomiques rapporté est très faible comparé aux anomalies de une à quelques paires de bases. Ceci s'explique par le fait que ce type de mutation a un impact très fort sur le phénotype et est lié à un reproducteur en particulier. Les contrôles chromosomiques sont réalisés en France principalement dans les espèces porcines et bovines. La plateforme de contrôle chromosomique UMR GenPhySE a déjà réalisé plus de 1000 et 3000 caryotypes dans les espèces bovines et porcines respectivement. En bovins, des décrets imposent le contrôle systématique des mâles reproducteurs pour la recherche de la translocation robertsonienne rb(1;29). Elle a été identifiée dans plusieurs races dans les années 1970, et elle est responsable d'une baisse importante de fertilité (Darré et al., 1972). En porc, un contrôle chromosomique des verras terminaux des CIA est réalisé avant leur diffusion. Avec l'accumulation des analyses, la prévalence de translocations est de 1/200 et elles sont dans 90% des cas des translocations réciproques associées à une baisse moyenne de 40% de la prolificité (A. Pinton, comm. pers.).

2.2.7 Anomalies avec déterminisme génétique complexe

L'ensemble des anomalies décrites ci-dessous ont une transmission mendélienne. Cependant, des phénomènes tels que l'empreinte parentale (*Genomic imprinting*) peut conduire à une transmission non-mendélienne du caractère. Cela signifie que pour un individu, la copie héritée de sa mère et celle de son père ne s'expriment pas de la même manière. Ce mécanisme repose sur la présence de marques épigénétiques dans les cellules germinales des parents pouvant activer ou réprimer l'expression d'un gène. Un exemple bien connu chez les petits ruminants est le phénotype callipyge (hypertrophie musculaire) lié à une mutation intergénique OAR18 (Oar_rambouillet_v1.0) : g.66,187,430A>G située dans un élément de contrôle dans un domaine soumis à empreinte (OMIA 001354-9940) (Cockett et al., 1996 ; Georges et al., 2003). Le mode de transmission est qualifié de "surdominance polaire" par les auteurs, c'est-à-dire qu'un individu doit hériter d'une seule version de l'allèle muté provenant de son père pour présenter cette hypertrophie, l'allèle hérité de sa mère doit être sauvage. A la naissance, les agneaux sont normaux, ce n'est qu'entre 30 à 80 jours après la naissance qu'ils vont développer le phénotype (masse musculaire augmentée de 30 à 50%). Ce mécanisme assez complexe repose sur l'expression ectopique dans le muscle squelettique (stade postnatal) de deux gènes d'expression paternelle *DLK1* (*Delta Like Non-Canonical Notch Ligand*

A. Translocation réciproque



B. Translocation robertsonienne

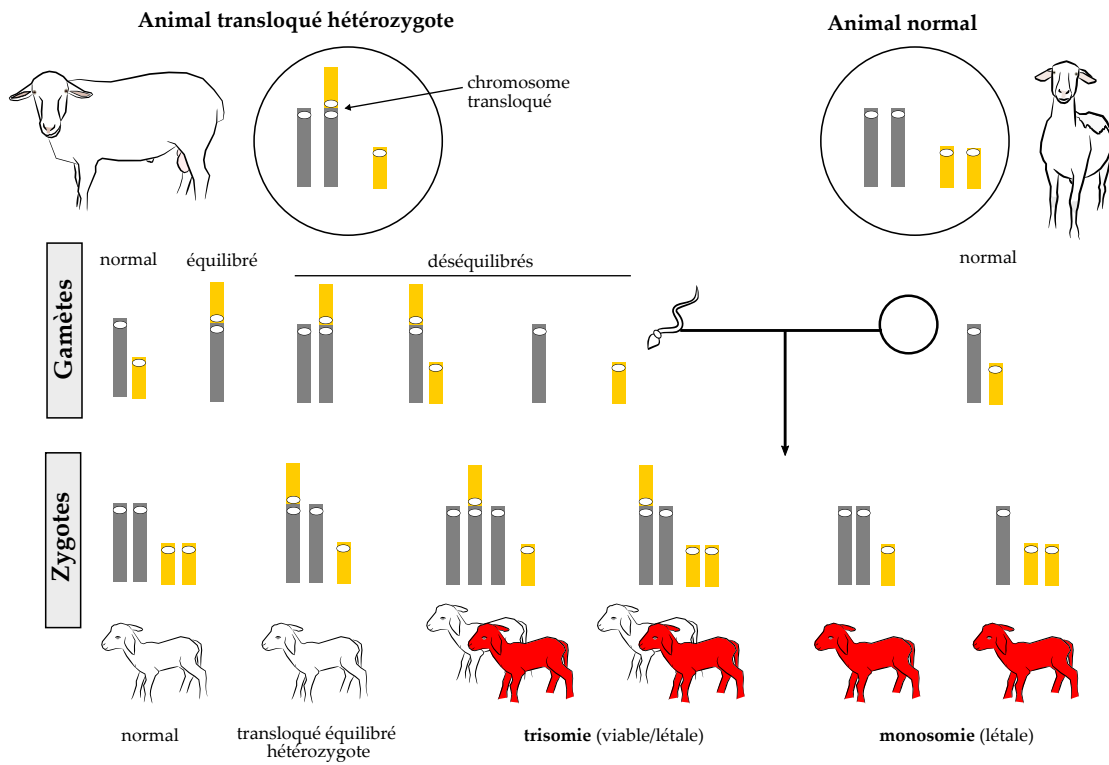


FIGURE 2.13 – Les anomalies génétiques à l'échelle des chromosomes.

1) et *PEG11* (*Retrotransposon Gag Like 1*) (Georges et al., 2003; Xu et al., 2015; Jammes & Renard, 2010; O'Doherty et al., 2015; Lewis & Redrup, 2005). Le mécanisme moléculaire est décrit dans la [Figure 2.14](#). La découverte de ce type de mécanisme ouvre la voie pour identifier des anomalies complexes chez l'homme.

Initialement décrit comme monogénique, le déterminisme génétique d'une anomalie génétique peut être plus compliqué. Une mutation peut être à pénétrance incomplète c'est-à-dire que les individus possèdent le génotype à risque mais ne présentent pas le phénotype délétère attendu, cela peut également être vu comme de l'expressivité incomplète du phénotype. Par exemple, en ovins une mutation non-sens OAR10 (Oar_rambouillet_v1.0) : g.88095843G>A (p.Gln16*) dans le gène *FGF14* (*Fibroblast Growth Factor 14*) est responsable d'ataxie spino-cérébelleuse (OMIA 001723-9940) avec un mode de transmission autosomal dominant avec pénétrance incomplète (Dittmer et al., 2017). L'expressivité du phénotype est dépendant de l'expression d'autres gènes (mécanismes compensatoires) et de l'environnement, mais le mécanisme moléculaire précis n'est pas encore connu. Néanmoins, une très belle illustration existe en race bovine laitière Montbéliarde avec l'identification d'une mutation de type délétion BTA14 (ARS-UCD1.2) : g.26,402,250_26,402,254del (p.Lys594Alafs*29) induisant un décalage du cadre de lecture et l'apparition prématurée d'un codon stop dans le gène *CHD7* (*Cadherin 7*) responsable du syndrome CHARGE ("C pour colobome, H pour *heart* : malformations du cœur, A pour atrésie choanale : les fentes nasales peuvent être bouchées, R pour retard de croissance, G pour défauts de l'appareil génito-urinaire, et E pour *ear* : troubles de l'oreille") (ONAB, 2021). Cette mutation a été identifiée chez un taureau d'insémination ("Etsar") qui a diffusé massivement avant que l'anomalie ne soit identifiée. La grande variabilité des phénotypes et l'âge d'apparition des symptômes ont rendu son identification plus difficile. L'étude fine des descendants du taureau a pu mettre en évidence six régions modulatrices qui interviennent dans l'expressivité du phénotype (Bourneuf et al., 2017; Capitan et al., 2014; Boichard, Grohs, Danchin-Burge, & Capitan, 2016).

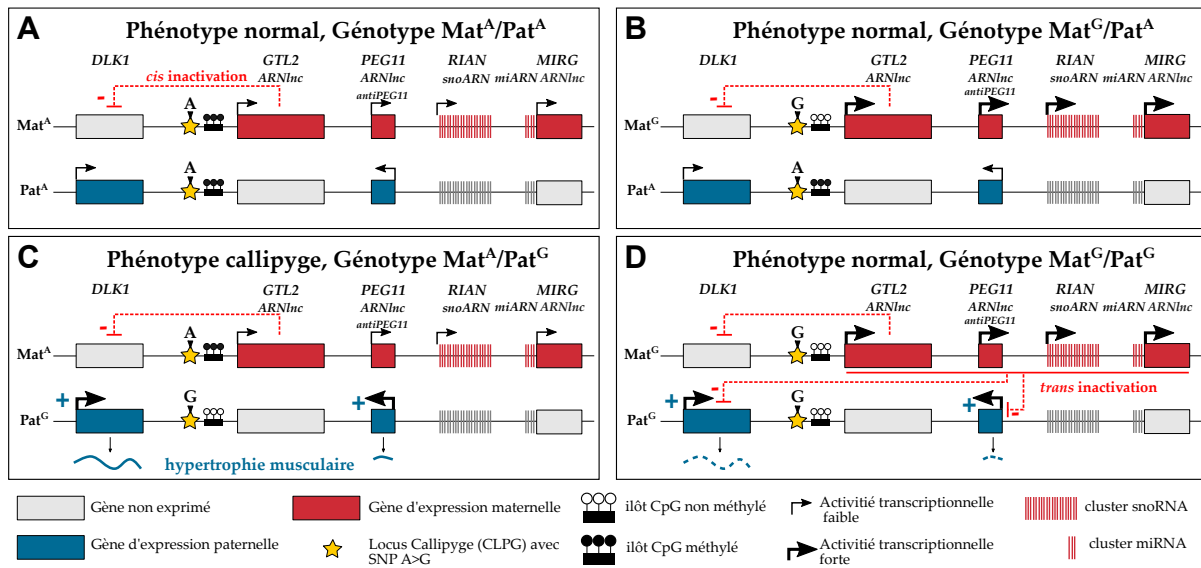


FIGURE 2.14 – Modèle de régulation de l'expression des gènes impliqués dans le phénotype callipyge chez le mouton lié à la mutation OAR18 :g.66,187,430A>G (CLPG). La mutation est localisée dans un élément de contrôle longue distance (LRCE : *long-range control element*), symbolisée par l'étoile. Les gènes d'expression paternelle *DLK1* et *PEG11* codent pour des protéines. Les gènes d'expression maternelle *GLT2*, *antiPEG11* et *MIRG* codent pour des ARNlnc. Le gène *RIAN* correspond à un cluster de petits ARN nucléolaires (snoARN). Mat= chromosome maternel, Pat=chromosome paternel. (A) **Phénotype normal, Génotype Mat^A/Pat^A** : en amont du gène *GLT2*, les îlots CpG sont méthylés réprimant la transcription des gènes en aval. La faible expression maternelle de *GLT2* (ARNlnc) inhibe en *cis* l'expression maternelle de *DLK1*. *DLK1* et *PEG11* sont exprimés très faiblement à partir de l'allèle paternel. (B) **Phénotype normal, Génotype Mat^G/Pat^A** : lorsque la mutation est portée par le chromosome maternel, les expressions de *GLT2*, *antiPEG11* et *MIRG* augmentent fortement et de nombreux miARN et snoARN sont produits. Les îlots CpG sont déméthylés facilitant la transcription des gènes en aval. (C) **Phénotype callipyge, Génotype Mat^A/Pat^G** : quand la mutation est portée par l'allèle paternel, les animaux sont de phénotype callipyge expliqué par une surexpression des gènes *DLK1* et *PEG11* permis par une hypométhylation des îlots CpG. (D) **Phénotype normal, Génotype Mat^G/Pat^G** : lorsque la mutation est portée par les chromosomes paternel et maternel, les individus ont un phénotype normal. La présence de la mutation CLPG sur le chromosome maternel contribue à fortement exprimer les ARNlnc maternelles tandis que sur le chromosome paternel, elle entraîne une surexpression de *DLK1* et *PEG11*. Les ARNlnc, snoRNA et miRNA maternels vont alors inactiver en *trans* l'expression de *DLK1* et *PEG11*. Il a été démontré que la *trans*-inactivation de ces gènes chez les animaux CLPG/CLPG repose sur un mécanisme de clivage (ARN interférent) des ARNm par un complexe protéique RISC (*RNA-induced silencing complex*) rendu possible par l'action de plusieurs miARN parfaitement complémentaires aux ARNm de *DLK1* et *PEG11*. Ainsi, les protéines *DLK1* et *PEG11* sont produites en très faibles quantités. *Adapté de* (Georges et al., 2019; Jammes & Renard, 2010; O'Doherty et al., 2015; Lawson et al., 2013; Xu et al., 2015; Lewis & Redrup, 2005)

2.3 Les mécanismes influençant la fréquence des variants

L'efficacité reproductive des reproducteurs favorise la transmission d'anomalies génétiques. Leur utilisation massive (notamment le cas en monte naturelle en petits ruminants) peut conduire à diffuser des anomalies génétiques dominantes. Dans ce cas, les conséquences sont limitées à l'échelle de la race mais graves à l'échelle d'un élevage. Les individus porteurs de mutations dominantes (exprimant le phénotype) délétères sont rapidement éliminés parfois en une seule génération. Cependant, si une mutation récessive délétère ségrège dans une population, elle existe principalement à l'état hétérozygote (les animaux sont phénotypiquement normaux). Ce n'est qu'après plusieurs générations (5-10), que la mutation peut se retrouver à l'état homozygote. Il est important de rappeler que les mutations responsables d'anomalies génétiques récessives sont des variants rares et peuvent être éliminées au cours des générations par hasard (dérive génétique) ([paragraphe 2.3.1](#)) ou sélectionnées lorsqu'elles présentent un avantage sélectif à l'état hétérozygote ([paragraphe 2.3.2](#)) (Old et al., 1993 ; Hedrick, 2015).

2.3.1 Effet de la dérive génétique

Les variants responsables d'anomalies génétiques ont des fréquences faibles à modérées (1-10%) et peuvent rapidement être éliminés en quelques générations (Georges et al., 2019). Comme nous l'avons vu dans le [paragraphe 1.3.1](#), la création des races a contribué à diminuer la diversité allélique (effectif génétique réduit) (Liu et al., 2019). De ce fait, cette perte de diversité s'est accompagnée d'une augmentation inévitable de la consanguinité. Afin d'étudier l'évolution de la fréquence d'une mutation délétère dans les populations d'élevage, des simulations sur 50 générations ont été réalisées prenant en compte une taille efficace de la population fixée à $N_e=100$ et une fréquence initiale de la mutation fixée à 1% (Derks & Steensma, 2021). Ainsi, par dérive génétique, on observe qu'un variant délétère (sans avantage sélectif) peut se maintenir à une fréquence très faible dans la population pendant de nombreuses générations avant d'être éliminé.

2.3.2 Sélection équilibrante : avantage des hétérozygotes

L'autre phénomène pouvant expliquer le maintien de variant délétère dans les populations est la sélection. En effet, certaines mutations donnent un avantage phénotypique aux individus à l'état hétérozygote. Avec une meilleure fitness (capacité reproductive), les reproducteurs vont plus facilement diffuser dans les schémas de sélection. Cela explique qu'une mutation délétère peut se retrouver à une fréquence stable (entre 5-15%) dans une population (Derks & Steensma, 2021). A ce jour en ovins, des variants identifiés dans cinq gènes présentent un avantage sélectif à l'état hétérozygote. La majorité des mutations de prolificité identifiées dans les gènes *BMP15* et *GDF9* est associée à un meilleur taux d'ovulation et une meilleure prolificité à l'état hétérozygote tandis qu'à l'état homozygote, elles sont responsables de stérilité ([Table 2.3](#)). Les animaux porteurs de la mutation faux-sens OAR6 (Oar_rambouillet_v1.0) : g.128,784,747A>T (p.Pro700Glu) dans le gène *FGFR3* (*Fibroblast Growth Factor Receptor 3*) présentent une meilleure conformation tandis qu'à l'état hétérozygote elle conduit à de la létalité (syndrome de l'agneau araignée). La mutation faux-sens OAR3 (Oar_rambouillet_v1.0) : g.139,302,270C>T (p.Arg96Cys) identifiée dans le gène *SOCS2* (*Suppressor Of Cytokine Signaling 2*) est responsable d'un comptage somatique

élevé dans le lait (susceptibilité aux mammites) à l'état homozygote alors qu'à l'état hétérozygote elle permet une augmentation de la production laitière et les animaux ont un poids plus élevé. En race de mouton sauvage Soay, la mutation responsable de l'absence de cornes (rétroinsertion de 1,8kb dans le 3'UTR du gène *RXFP2* (*Relaxin Family Peptide Receptor 2*) à l'état hétérozygote (+/P) est associée à une meilleure fitness expliquée par le compromis entre meilleure capacité reproductive et survie. Les animaux cornus (+/+) ont une forte capacité reproductive mais une survie courte tandis que que les animaux sans cornes (P/P) survivent longtemps mais ne reproduisent que très rarement (Johnston et al., 2013). Enfin, les animaux porteurs de la mutation callipyge présentent une hypertrophie musculaire (décrit précédemment dans le [paragraphe 2.2.7](#)). De nombreux exemples d'avantages sélectifs sont décrits dans la littérature chez les espèces d'élevage (Derks & Steensma, 2021). Un exemple bien connu chez le bovin est la délétion de 660kb (OAR12 (UMD3.1) : g.2,010,0648_20,763,119del) qui est létale à l'état homozygote alors qu'elle est associée à une meilleure production laitière chez la Nordic Red Cattle (Kadri et al., 2014).

La sélection de l'allèle responsable de l'anomalie peut également être indirecte. Il s'agit de l'effet "auto-stop", c'est-à-dire que l'allèle muté est en déséquilibre de liaison avec un locus d'intérêt en sélection (Boichard, Grohs, Danchin-Burge, & Capitan, 2016). En monogastrique, Derks, Lopes, et al. (2018) ont mis en évidence une délétion de 212kb (contient deux gènes : *BBS9* et *BMPEP*) qui est létale à l'état homozygote (*BBS9*) mais est associée à une meilleure croissance à l'état hétérozygote (*BMPEP*). La sélection sur la croissance a donc tiré le variant délétère par "auto-stop".

A retenir Les anomalies génétiques sont inévitables et résultent de mutations spontanées de l'ADN, source de variabilité génétique. Les mutations ont souvent des effets biologiques neutres, parfois forts (hautement délétères) sur la protéine et conduisent à des anomalies. Le déterminisme génétique d'une anomalie est dans la plupart des cas monogénique et est associé à une mutation perte de fonction (non-sens, site accepteur/donneur, perte codon start/stop, faux-sens avec impact prédit fort, frameshift et Indels). Il existe une diversité de modes de transmission et d'actions des anomalies. Les plus étudiées sont les mutations récessives et dominantes. Suivant le mode de transmission, l'émergence d'une anomalie peut être plus ou moins rapide. De plus, la fréquence des variants génétiques (supposés rares) peut fluctuer sous l'effet de la dérive génétique (élimination progressive au cours des générations) et de la sélection (maintien du variant à une fréquence stable lié à un avantage sélectif pour des caractères d'intérêt des hétérozygotes). ■

3. Un contexte favorable à l'étude des anomalies génétiques dans les populations d'élevage

Le développement des nouvelles technologies haut-débit et la mise en place de système de surveillance des anomalies contribuent fortement à connaître et intégrer les anomalies génétiques dans le processus de sélection afin d'éviter leur diffusion.

3.1 Une meilleure connaissance des génomes et le développement des technologies haut-débit facilitent l'étude des anomalies

Depuis les années 2000, l'avènement des nouvelles technologies a fortement contribué à l'amélioration de la connaissance des génomes telles que le séquençage de génome entier et l'utilisation en routine de puces de génotypes SNP (plusieurs milliers de marqueurs).

3.1.1 Les différentes technologies pour séquencer un génome

Au cours des années, trois grandes révolutions technologiques de séquençage se sont successivement développées, mais restent complémentaires. Séquencer un génome entier (plusieurs Mégabases à Gigabases), c'est connaître l'enchaînement des bases tout au long du génome. Les approches de biologies moléculaires et informatiques ont évolué conjointement pour gérer la grande quantité de données produites. Les techniques de séquençage utilisées diffèrent selon les objectifs (construction d'un assemblage *de novo* ou re-séquençage de nouveaux individus d'une espèce pour de la détection de variants).

3.1.1.1 Le séquençage de première génération : le séquençage Sanger

En 1977, une révolution biologique a eu lieu avec le développement de la méthode de séquençage Sanger (Sanger et al., 1977). Son principe est illustré dans la [Figure 3.1](#). Cette technique repose sur l'hybridation d'une amorce oligonucléique complémentaire au brin d'ADN simple brin et de la synthèse du brin complémentaire (même principe qu'une PCR classique). A partir de l'extrémité 3'-OH de l'amorce, l'ADN polymérase vient ajouter un à un des désoxyribonucléotides (dNTP) complémentaires au brin d'ADN matrice. Les dNTP présents dans le mélange réactionnel sont de deux types : (i) dNTP classiques et (ii) dNTP modifiés et marqués (fluorescents ou radioactifs), appelés didésoxyribonucléotides (ddNTP), qui ne possèdent pas de groupe 3'-OH empêchant la formation de la liaison phosphodiester avec le nucléotide suivant. La synthèse du brin est donc arrêtée et une grande quantité de fragments de tailles différentes est générée. En pratique, le séquençage se fait dans quatre tubes présentant chacun un ddNTP différent. La lecture de la séquence se fait soit par migration sur gel d'acrylamide (Méthode 1) soit par la migration, excitation et lecture sur un séquenceur capillaire (Méthode 2). La connaissance de cartes génétique et physique était essentielle pour ordonner la séquence (Tagu et al., 2018). La méthode Sanger a

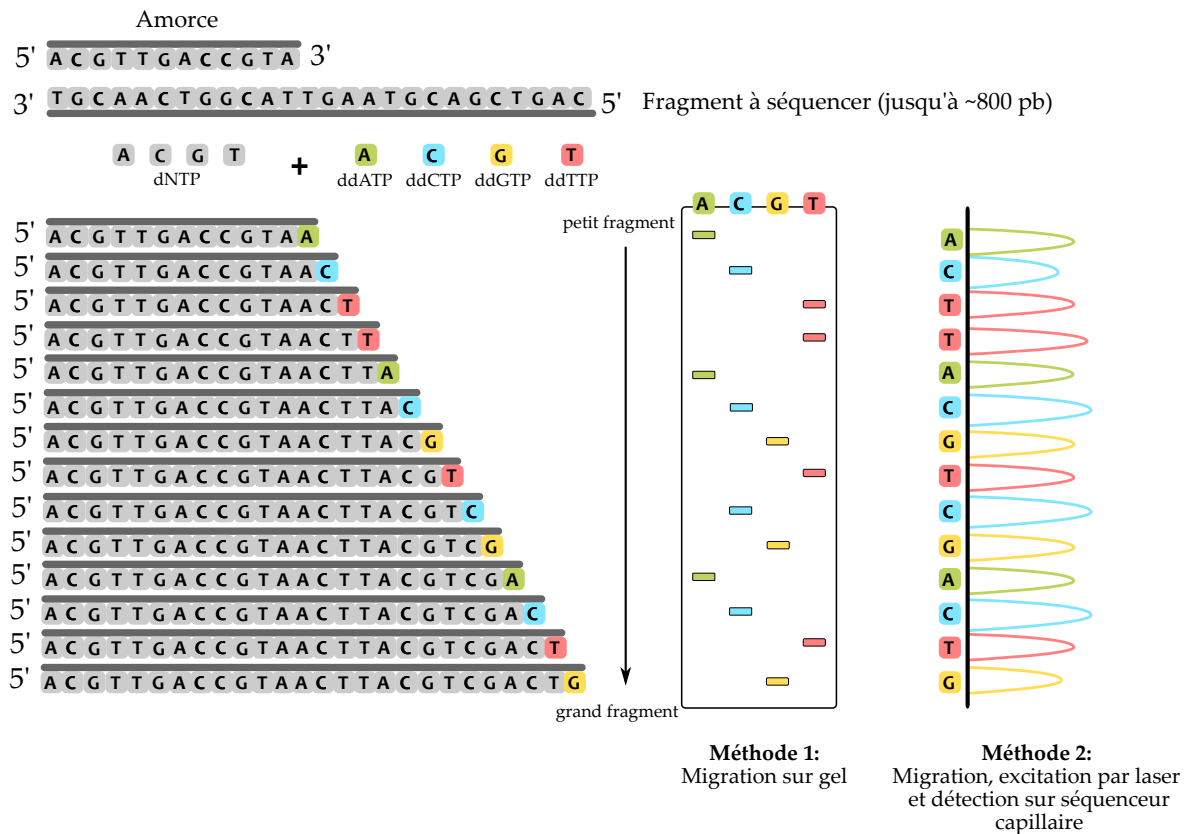


FIGURE 3.1 – Principe du séquençage Sanger

permis de séquencer pour la première fois le génome humain, publié en 2001 par le consortium Human Genome Project (Venter et al., 2001 ; Lander et al., 2001). Les données de séquence sont obtenues par une stratégie *shotgun* (fragmentation aléatoire de l'ADN) et par l'utilisation de BAC (*Bacterial Artificial Chromosomes*). Le séquençage s'est achevé au bout de 13 ans et a coûté 3 milliards de dollars. Bien que révolutionnaire, cette méthode a rapidement montré des limites puisqu'elle permet seulement de séquencer de courts fragments d'ADN (jusqu'à 800 pb) avec un coût élevé et un débit réduit. Son application aujourd'hui se limite à séquencer des petites portions du génome.

3.1.1.2 Le séquençage de deuxième génération : le séquençage en lectures courtes

La recherche de variants issus d'un individu n'est possible que par son séquençage. Cependant, la méthode Sanger ne permet pas de séquencer massivement des génomes entiers. Les développements technologiques (*next generation sequencing*, NGS) ont rapidement permis de rendre plus accessible le séquençage lié à des coûts et une durée de séquençage réduits (Figure 3.2).

Les techniques actuelles de séquençage haut-débit reposent sur le principe du séquençage Sanger mais il est miniaturisé et parallélisé. Différentes technologies de séquençage existent : Illumina, Roche 454, Ion Torrent, SOLiD (Tagu et al., 2018). L'ensemble de ces technologies passe par une étape de préparation de banques d'ADN. Les séquenceurs HiSeq avec la technologie Illumina sont les plus répandus dans le monde et reposent sur le principe de séquençage en

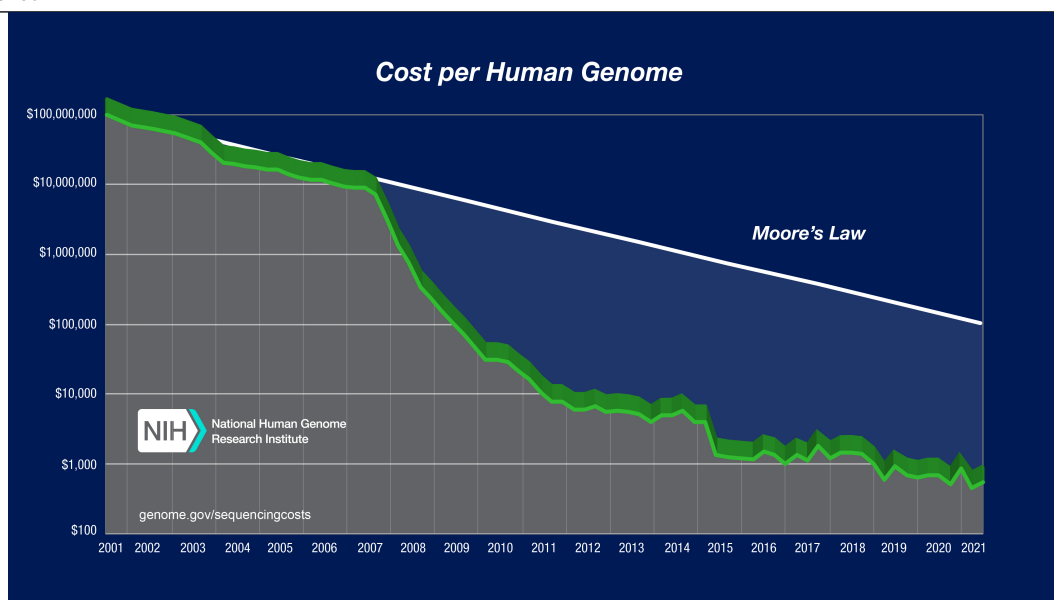
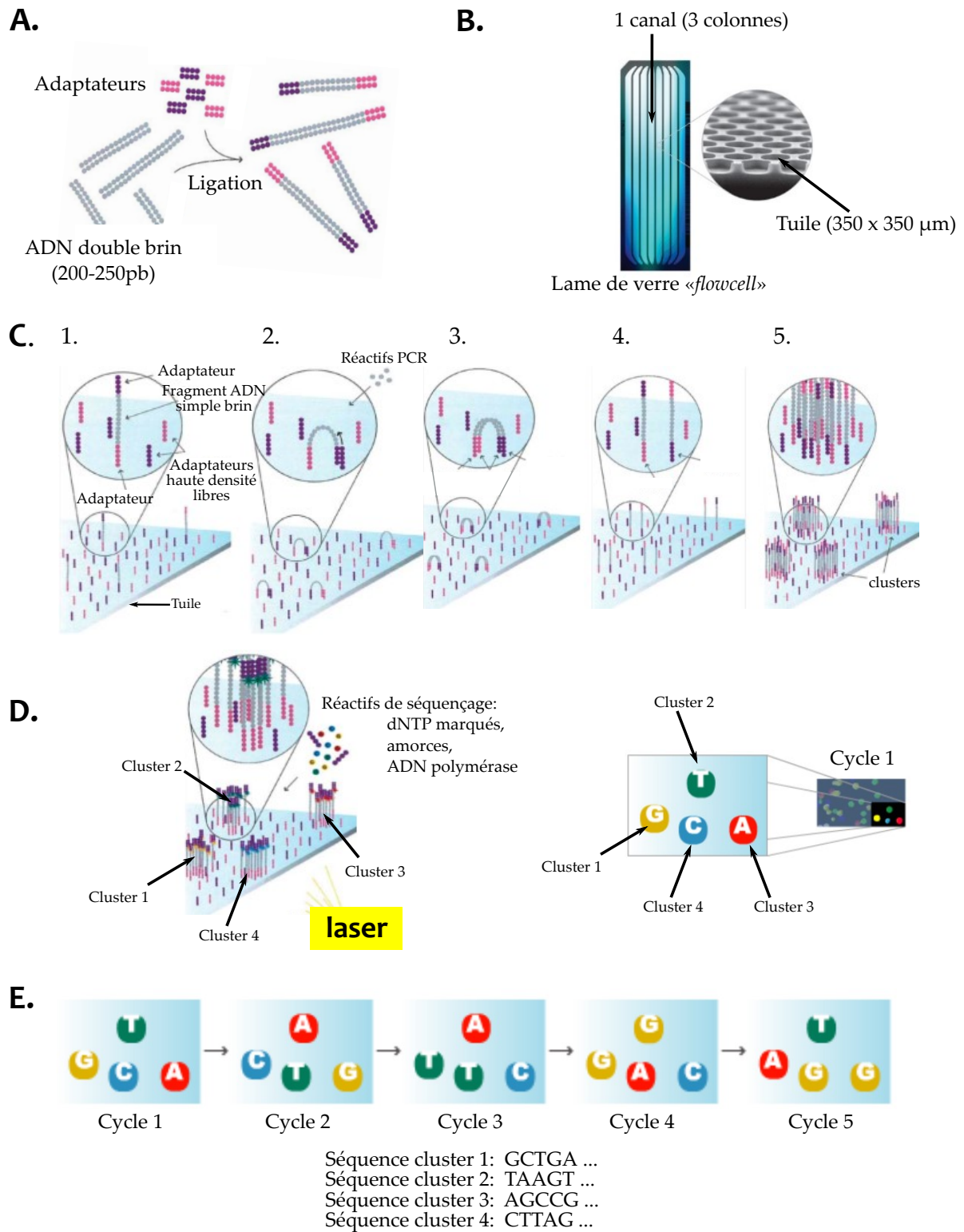


FIGURE 3.2 – Évolution des coûts de séquençage d'un génome complet. Source : NCBI (<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>)

lectures courtes (*short reads*) par synthèse d'acides nucléiques (*Sequencing By Synthesis, SBS*). Son principe est illustré dans la Figure 3.3. L'ADN est découpé en fragments de 200 à 250 pb. Des adaptateurs spécifiques vont ensuite se lier sur chacun des fragments (Figure 3.3 A). La banque de fragments ainsi constituée va être déposée sur une lame de verre (*flowcell*). Elle est constituée de 8 canaux constitués chacun de 3 colonnes subdivisées chacune en une centaine de tuiles. Chaque tuile possède entre 20 et 30 000 amorces complémentaires des adaptateurs. Une étape préliminaire d'amplification PCR clonale est réalisée. Chaque molécule d'ADN simple brin va venir s'hybrider avec l'amorce complémentaire présente sur la tuile (Figure 3.3 C1). L'adaptateur libre de la molécule d'ADN va établir un pont pour venir s'hybrider avec une amorce complémentaire à proximité (Figure 3.3 C2). L'ajout des réactifs PCR va permettre de synthétiser le brin complémentaire (Figure 3.3 C3). L'ADN double brin est ensuite déshybridé pour obtenir deux molécules d'ADN simple brin (Figure 3.3 C4). Seul le brin identique à la molécule d'ADN est gardé. Les étapes précédentes sont répétées pour former des clusters de molécules identiques (Figure 3.3 C5). La réaction de séquençage repose sur l'incorporation de quatre dNTP marqués avec un fluophore différent (A : rouge, C : bleu, G : jaune et T : vert) (Figure 3.3 D). Les brins simple brin obtenus par PCR clonale sont utilisés comme matrice et l'ADN polymérase ajoute un à un les nucléotides marqués. A l'ajout de chaque nucléotide, la réaction s'arrête et le fluophore est excité par un laser, une caméra va ensuite lire la fluorescence. L'ADN polymérase présente une activité exonucléase et élimine le fluophore permettant l'ajout du nucléotide marqué suivant.

3.1.1.3 Le séquençage de troisième génération : le séquençage en lectures longues

La technique de séquençage en lectures courtes présente des limites dès que les régions à analyser sont complexes (régions répétées par exemple). Le développement du séquençage en lectures longues (*Third Generation Sequencing, TGS*) permet de séquencer des fragments de 10 à 80 kb pouvant même atteindre 1 Mb (van Dijk et al., 2018). Trois techniques de séquençage en

FIGURE 3.3 – Séquençage de l'ADN par la technique Illumina. *Adapté de Illumina (2022c)*

lectures longues existent sur le marché (van Dijk et al., 2018) : (i) séquençage en temps réel d'une molécule (*Single Molecule Real Time*, SMRT) développé par la société PacBio, (ii) le séquençage par nanopore développé par la société Oxford Nanopore Technologies (ONT) et (iii) séquençage à longues lectures synthétiques (*Synthetic Long Read*, SLR) développé par Illumina et la société 10X genomics. Dans la technologie SMRT de PacBio, l'ADN polymérase est fixée au fond de plusieurs milliers de puits présents sur une plaque de verre. L'ADN simple brin (10-20kb) à séquencer va alors se fixer au fond du puit et la polymérase réalise la synthèse du brin complémentaire en ajoutant un à un des dNTP marqués par fluophore. La fluorescence est mesurée en temps réel contrairement à la technologie Illumina où la synthèse est discontinue (fixation du dNTP, lecture de la fluorescence, élimination du fluophore, ajout d'un nouveau dNTP...). Dans la technologie Nanopore, aucune amplification en amont n'est nécessaire. Les fragments d'ADN à haut poids moléculaire (20-30 kb) sont ligués aux deux extrémités à des adaptateurs doubles brins comportant des enzymes motrices. Les puces sont composées de 512 pores protéiques (nanopores) intégrés sur une membrane en polymère synthétique. Un champ électrique est exercé pour attirer l'ADN vers le nanopore. L'adaptateur va venir se fixer à proximité du pore et l'enzyme motrice ouvre l'ADN double brin permettant ainsi à l'ADN simple brin de passer dans le nanopore. Le passage des différentes bases dans le pore génère des variations électriques qui sont reconnues et traduites en bases. Enfin, la technologie SLR permet de contourner les limites des lectures courtes et passe par une préparation particulière des échantillons avant d'être séquencés par la technologie Illumina. Un fragment d'ADN de 50-100kb est placé dans une micelle contenant une bille de gel renfermant les réactifs pour une PCR, l'ADN polymérase et plusieurs adaptateurs avec barcode unique. Lors de l'amplification, de nombreux fragments courts sont synthétisés. Après séquençage, le traitement bioinformatique permet de regrouper les fragments courts présentant le même barcode et donc de reconstruire en partie la molécule d'ADN contenue dans la micelle. Cependant, cette méthode ne permet pas d'avoir accès à l'intégralité de la séquence du fragment d'ADN dans le cas de régions répétées.

Le champ d'application des lectures longues permet d'avoir des génomes de meilleure qualité, de construire des génomes de référence *de novo* (voir [paragraphe 3.1.2](#)), de séquencer directement les molécules d'ARN (sans passer par une étape d'obtention des ADNc), de faire des analyses épigénétiques sur l'ensemble du génome (méthylation de l'ADN). Ces techniques sont également efficaces pour la détection de variants structuraux, qui étaient auparavant difficilement détectables par lectures courtes. Cependant, le coût d'un séquençage en lectures longues (un génome *de novo* en 20X coûte 3 000€, durée de run de 4h sur du PacBio) reste plus élevé qu'un séquençage en lectures courtes (un génome de re-séquencage en 15X coûte 1 000€, durée de run de 48h sur du NovaSeq).

3.1.2 Construction d'une séquence de référence *de novo*

La connaissance de la séquence génomique d'une espèce est nécessaire pour connaître l'organisation et la position des gènes le long des chromosomes. Toute analyse génétique nécessite la connaissance d'un génome de référence (Tagu et al., 2018). Si aucun génome de référence n'existe, un assemblage *de novo* est réalisable à partir des données brutes obtenues par séquençage

(Figure 3.4). Les lectures chevauchantes sont regroupées en contigs. Grâce aux paires de lectures chevauchantes (*pair-end reads*) de tailles connues, l'enchaînement des contigs peut être déterminé. Un scaffold regroupe un ensemble de contigs. Cependant, un scaffold peut présenter des trous de séquences appelés GAP, matérialisés par des nucléotides indéterminés (notés "N") dans l'assemblage final. A ce jour, la construction d'un assemblage *de novo* combine un séquençage à courtes et longues lectures afin d'avoir un génome de meilleure qualité et d'être le plus informatif possible. L'apport des lectures longues facilite l'assemblage en permettant de remplacer les nucléotides indéterminés par de vrais nucléotides. Un assemblage exclusivement constitué de lectures longues reste limité pour le moment lié à un taux d'erreurs de séquençage trop important ($\sim 10\%$ suivant les technologies). La qualité d'un assemblage s'apprécie par des indicateurs tels que : la taille de l'assemblage (taille attendue déterminée par spectrométrie de flux), le pourcentage de nucléotides indéterminés (N), le nombre de contigs et scaffolds (indicateur de la fragmentation de l'assemblage), le N50 de scaffolds/contigs (50% de la taille totale de l'assemblage se trouve dans les scaffolds/contigs plus longs que cette valeur) et le L50 de scaffolds/contigs (nombre scaffolds/contigs supérieurs ou égaux à la longueur N50 et comprenant donc la moitié des bases de l'assemblage). Un génome de référence est stocké dans un format *fasta*.

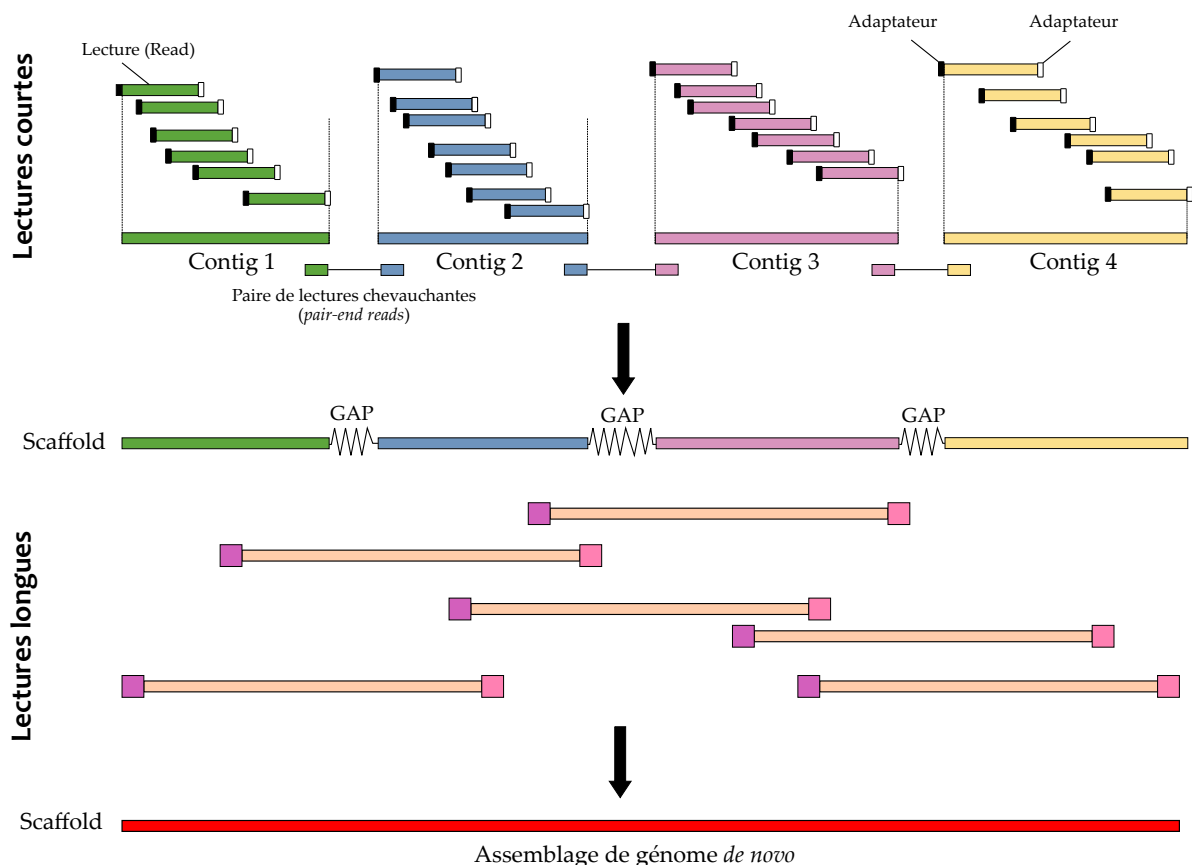


FIGURE 3.4 – Construction d'un génome de référence *de novo*

Chez les petits ruminants, plusieurs séquences de référence ont été réalisées en lien avec l'évolution des différentes technologies (Table 3.4). Aujourd'hui, les séquences de référence les plus récentes sont la version 2 de Rambouillet (ARS-UI_Ramb_v2.0) et la version ARS1 pour les ovins et caprins, respectivement. Ces assemblages sont de très bonne qualité lorsque l'on regarde les

différents indicateurs. L'assemblage ARS-UI_Ramb_v2.0 est composé de seulement 142 scaffolds soit une baisse de 2498 par rapport à la version précédente Oar_rambouillet_v1.0 et possède un contig N50 de 43,2 Mb. A ce jour en ovins, les versions Oar_v3.1 et Oar_rambouillet_v1.0 restent encore très utilisées. En effet, l'utilisation d'un nouvel assemblage nécessite un réaligement complet de séquences des individus déjà séquencés (processus long). Le développement de ces séquences de référence repose sur la collaboration de scientifiques regroupés au sein des consortia internationaux ovins (*International Sheep Genomics Consortium*, ISGC) et caprins (*International Goat Genome Consortium*, IGGC). Cela a également permis le développement des puces de génotypages SNP (voir [paragraphe 3.1.4](#)).

TABLE 3.1 – Principaux génomes de référence utilisés en ovins et caprins. OV= Ovins, CAP= Caprins. *Source* : NCBI

Génome de référence		Oar_v3.1 (OV) 2012	Oar_v4.0 (OV) 2015	Oar_rambouillet_v1.0 (OV) 2017	ARS-UI_Ramb_v2.0 (OV) 2021	CHIR_1.0 (CAP) 2013	ARS1.2 (CAP) 2016
Race (sexe)	Texel (F+M)	Texel	Rambouillet (F)	Rambouillet (F)	Yunnan black (F)	San Clemente (M)	
Couverture du génome	150x	166x	126x	55x	86x	50x	
Technologie de séquençage	Illumina GAII ; 454	Illumina GAII ; 454 ; PacBio RSII	HiSeq X Ten ; PacBio RSII	Oxford Nanopore PromethION ; PacBio RSII ; Illumina HiSeq	Illumina GAIIx	PacBio	
GenBank (GCA)	000298735.1	000298735.2	002742125.1	016772045.1	000317765.1	001704415.2	
Longueur (Gb)	2,6	2,6	2,9	2,6	2,6	2,9	
Nombre de scaffolds	5 697	5 465	2 640	141	77 431	29 885	
Scaffold N50	100,1 Mb	100,0 Mb	107,7 Mb	101,3 Mb	14,4 Mb	87,3 Mb	
Scaffold L50	8	8	8	8	56	13	
Nombre de contigs	130 764	48 481	7 485	225	337 494	30 377	
Contig N50	40,4 kb	150,5 kb	2,6 Mb	43,2 Mb	18 934	26,2 Mb	
Contig L50	18 404	5 008	313	24	38 700	32	
Nombre chromosomes	27	27	27	27	30	30	

3.1.3 Traitements bioinformatiques des données de séquence

Les données brutes sorties des séquenceurs ne permettent pas de lire directement le génome des individus séquencés. Il est nécessaire d'utiliser des outils bioinformatiques pour aligner les petits fragments d'ADN et de les comparer à la séquence de référence pour mettre en évidence des variations génétiques.

3.1.3.1 Alignements des lectures et détection de variants

Lorsqu'un assemblage de génome est disponible, des re-séquençages de génome de nouveaux individus sont réalisés pour rechercher des nouveaux variants, c'est ce qui est fait lorsque l'on recherche le déterminisme génétique d'une anomalie. Il s'agit principalement d'un séquençage obtenu par lectures courtes moins coûteux. Avec l'essor des NGS, plusieurs milliards de lectures de séquence sont générées. Il devient alors nécessaire de reconstruire le puzzle par des traitements bioinformatiques afin de les ordonner à leur position le long des chromosomes. Il devient aisé de reconstruire le génome d'un individu en réalisant un alignement sur la séquence de référence et d'identifier les variants (*calling*) en comparant les deux séquences. Le principe du traitement bioinformatique dans le cas d'un re-séquençage de génome d'un nouvel individu est représenté dans la [Figure 3.5](#).

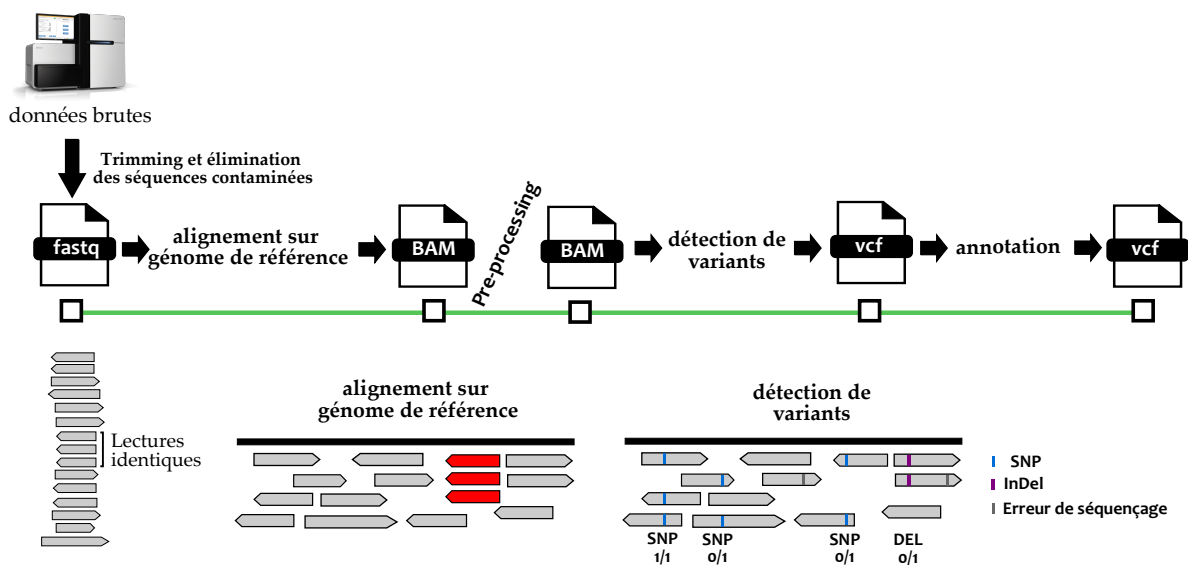


FIGURE 3.5 – Traitements bioinformatiques des données de séquence issues d'un re-séquençage de génome

Les données brutes sorties des séquenceurs sont composées de plusieurs milliers de fichiers stockés sous le format *fastq* correspondant chacun à l'information d'une lecture associée à sa qualité. Au préalable, les amorces des séquences sont éliminées (*trimming*) et les séquences présentant des contaminations sont éliminées (Blast contre les génomes *E. coli*, de phages ou levures par exemple). L'ensemble de ces fichiers va être lu par un logiciel d'alignement tel que BWA (Li & Durbin, 2010) afin d'aligner l'ensemble des lectures contre le génome de référence. Les lectures alignées sont stockées dans un fichier sous un format binaire BAM. Une étape de pré-traitement (*pre-processing*) permet (i) d'éliminer les duplicats de lecture pour éviter les risques de propagation d'erreurs lors de la détection de variants (*MarkDuplicates*) et (ii) une

recalibration des bases (*Base Quality Recalibration Score*, BQRS) est effectuée, permettant de détecter les erreurs de séquençage et d'estimer la précision du *calling* de chaque base et (iii) de réaliser des réalignements locaux. Les logiciels de *calling* tels que GATK (McKenna et al., 2010), SAMtools (Li, 2013) permettent d'identifier les petits variants de type SNP et InDels. Le génotypage à chaque variant est exprimé par rapport à la référence : 0/0 si toutes lectures portent l'allèle identique à la référence, 1/1 si toutes les lectures portent l'allèle alternatif et 0/1 si les deux types lectures sont observés. Un SNP peut également être multiallélique (>2 allèles), les différents allèles sont alors codés 2, 3, 4 etc. La profondeur des lectures est importante pour attribuer correctement le génotype d'un variant. Des méthodes spécifiques sont utilisées pour détecter les variants de structure car ils sont complexes à mettre en évidence par l'approche présentée ci-dessus.

3.1.3.2 Prédiction de l'effet des variants

Après avoir construit un assemblage *de novo* (format fasta), il est nécessaire de connaître la position et l'organisation des éléments structuraux qui composent le génome. On distingue deux types d'annotation : (i) l'annotation structurale et (ii) l'annotation fonctionnelle. L'annotation structurale correspond à la position des éléments fonctionnels tels que la position des gènes (régions UTR, introns, exons) et des régions régulatrices. L'annotation fonctionnelle consiste à proposer une information biologique de ces éléments. En général, l'annotation dans les régions codantes est souvent précise alors que dans les régions intergéniques et intragéniques l'information est souvent pauvre. Des outils bio-informatiques comme *SnpEFF* permettent de prédire ces effets en se basant sur des bases de données existantes (Ensembl, RefSeq, UCSC) pour lesquelles l'annotation est déjà réalisée (fichier d'annotation de type *gff*). Ce type de fichier est réalisé par des logiciels capables de prédire (*ab initio*) la position des gènes par des modèles statistiques prenant en compte des données de RNA-seq (déterminer les limites exons/introns) et les homologies avec d'autres espèces annotées. Les algorithmes détectent la structure d'un gène défini par un codon d'initiation (ATG) et un codon stop (TAA, TAG ou TGA). Cependant, lorsque le gène prédit est inconnu, un identifiant lui est attribué sans être associé à une fonction biologique. L'impact biologique d'un variant (nul à fort) est dépendant de sa localisation sur le génome de référence (voir [paragraphe 2.1.2.1](#)). D'autres outils comme *SIFT* ou *Polyphen2* permettent de prédire l'impact des substitutions non conservatives sur la protéine. A ce jour, l'annotation ovine ARS-UI_Ramb_v2.0 (GCF_016772045.1) est composée de 33 801 gènes (dont 21 257 gènes codant pour des protéines) et l'annotation caprine ARS1 (GCF_001704415.2) est composée de 28 911 gènes (dont 20 572 gènes codant pour des protéines). A l'instar du projet d'annotation du génome humain ENCODE, le projet FAANG (*Functional Annotation of ANimal Genomes*), visant à améliorer l'annotation des génomes des animaux de rente, est basé sur l'étude de données de transcriptome (RNA-seq), d'accessibilité de la chromatine (ATAC-seq) et de topologie 3D du génome (Hi-C) (Clark et al., 2020). Comme dans de nombreuses espèces, des projets 1 000 génomes ont vu le jour en petits ruminants. Dans le cadre du consortium ovins (ISGC) et caprins (IGGC), 935 séquences ovines (21 pays, 69 races) et 1159 séquences caprines (35 pays, 103 races) sont annotées (Kijas et al., 2009; Denoyelle et al., 2021).

3.1.4 Le génotypage

L'accumulation des données de séquence dans une espèce permet d'identifier des millions de polymorphismes issus de races/populations différentes permettant de capturer un maximum de diversité génétique (Denoyelle et al., 2021; Rupp et al., 2016; Groenen et al., 2012; Kijas et al., 2009; Tosser-Klopp et al., 2014; Ka-Shu Wong et al., 2004). L'accumulation du nombre de variants SNP détectés a permis le développement de puces SNP offrant la possibilité de génotyper plusieurs milliers de marqueurs très rapidement. Ces puces sont utilisées pour conduire une analyse d'association tout génome (i.e. déterminer les régions du génome associées à un caractère) et dans le cadre de la sélection génomique (indexation génomique). L'ensemble des SNP est regroupé dans des bases de données telles que dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>).

3.1.4.1 Principe et fonctionnement des puces de génotypages

Les marqueurs présents sur les puces ont été choisis pour qu'ils soient partagés par un maximum de races pour être les plus informatifs possibles (notion d'universalité par espèce) avec en général une fréquence de l'allèle le plus rare (*Minimum Allele Frequency*, MAF) > 5% et un taux de réussite de génotypage (Call Rate du SNP) de 95%. Les marqueurs sont répartis sur l'ensemble du génome et respectent l'équilibre de Hardy-Weinberg (voir [Encadré 3. Principe de l'équilibre de Hardy-Weinberg](#)). Cependant, une puce par espèce est développée.

Encadré 3. Principe de l'équilibre de Hardy-Weinberg (HWE) (Minvielle, 2010)

On considère un SNP bi-allélique A et B , de fréquences alléliques $f(A)$ et $f(B)$. Dans une population de grande taille, sans migration (régime fermé), non soumise à la sélection, sans mutations, les fréquences alléliques $p = f(A)$ et $q = f(B)$ restent constantes d'une génération à l'autre. De plus, s'il y a panmixie, on est alors capable de déduire à partir des fréquences alléliques, les fréquences génotypiques (et donc le nombre d'individus attendus pour chacun des génotypes) :

Génotype	AA	AB	BB
Nombre d'individus observés	N_{AA}	N_{AB}	N_{BB}
Nombre d'individus attendus (théorique)	Np^2	$N2pq$	Nq^2

Test du χ^2 (ddl=1) :

$$\chi^2 = \sum \frac{(Obs - Att)^2}{Att} = \frac{(N_{AA} - Np^2)^2}{Np^2} + \frac{(N_{AB} - N2pq)^2}{N2pq} + \frac{(N_{BB} - Nq^2)^2}{Nq^2}$$

Rmq : Le seuil de significativité à 5% du χ^2 , avec ddl=1, est à 3,84.

Si le nombre d'individus observés correspond au nombre d'individus attendus pour chacun des génotypes (valeur de test du $\chi^2 < 3,84$), on dit que le marqueur SNP est en équilibre d'HWE. Si ce n'est pas le cas, le marqueur n'est plus à l'équilibre (écart à la loi de HWE). Dans la plupart des cas, un SNP d'une puce en déséquilibre de HWE résulte d'une erreur technique de génotypage.

Ces puces reposent sur l'utilisation de technologies qui sont miniaturisées et parallélisées. A ce jour, deux plateformes dominent le marché : Illumina et Affymetrix. Elles utilisent une chimie différente mais le principe reste le même. L'ADN génomique est amplifié puis fragmenté et est déposé sur une matrice contenant plusieurs milliers de sondes. Les fragments d'ADN vont alors s'hybrider à la sonde produisant un signal fluorescent spécifique de chaque allèle du SNP. L'accumulation de la fluorescence est lue et est interprétée pour chaque SNP pour en déduire le génotype.

Le mode opératoire de la plateforme Illumina est décrit dans la [Figure 3.6](#) (Gunderson et al., 2005). L'ADN génomique est amplifié ($\times 1000$) de manière non-spécifique par une réaction isotherme sans avoir recours à une PCR classique qui peut induire des biais dans les zones enrichies en GC. L'ADN est ensuite fragmenté en molécules de 300-600pb et est précipité et resuspendu ([Figure 3.6 A](#)). L'échantillon est alors chargé sur la plaque de verre ([Figure 3.6 B](#)). La puce est une lame de verre composée de plusieurs millions de micropuits contenant chacun une bille (Illumina, 2022b). Chaque bille est spécifique d'un SNP et présente à sa surface plusieurs centaines de milliers de sondes oligonucléotides identiques (~ 50 pb). Une bille est présente en une vingtaine d'exemplaires sur la puce qui sont répartis de manière aléatoire. Lors de l'étape d'hybridation, les fragments d'ADN simples brins vont venir se fixer sur les sondes possédant la séquence complémentaire ([Figure 3.6 C](#)). Lors de l'incubation, l'ADN polymérase ajoute un nucléotide marqué ce qui arrête l'élongation. Les bases A et T sont marquées au dinitriphényl (DNP) et les bases C et G sont marquées à la biotine. L'ADN génomique est ensuite éliminé. Afin d'acquérir un signal fluorescent, des anticorps anti-DNP marqués par fluorescence rouge se fixent sur le nucléotide marqué au DNP et sont contre-colorés avec des anticorps secondaires conjugués au DNT ([Figure 3.6 D](#)). De même, la biotine est directement colorée avec la streptavidine marquée par fluorescence verte qui est contre-colorée avec des anticorps anti-streptavidine biotinylés (Illumina, 2022a). L'ensemble de ce processus permet d'amplifier le signal fluorescent. La mesure de la fluorescence est réalisée par un scanner (Illumina *iScan*) qui utilise des lasers rouges et verts. L'intensité de la fluorescence est ensuite interprétée par un logiciel (*Genome Studio*) permettant ainsi de déterminer le génotype pour l'ensemble des individus à chaque SNP (Illumina, 2010).

Le génotypage par puce SNP permet également de génotyper des petites insertions/délétions et des variants structuraux. Dans ces cas de figure, un design différent de sonde est réalisé.

3.1.4.2 Les différentes puces disponibles en petits ruminants

A ce jour, plusieurs puces de génotypage SNP existent sur le marché avec des technologies (Illumina ou Affymetrix) et des densités différentes. En ovins, trois principales puces sont couramment utilisées :

- La puce **ISGC SheepLD** (coût de 16€ par échantillon) est de faible densité (environ 15 000 à 18 000 marqueurs suivant la version). Elle présente 11 342 marqueurs en commun avec la puce moyenne densité **OvineSNP50 BeadChip**. Cette puce permet également d'ajouter à la demande des SNP d'intérêts lors des nouvelles mises à jour.
- La puce **OvineSNP50 BeadChip** (coût de 14€) est de moyenne densité et a été industrialisée par Illumina avec la collaboration du consortium ovin (ISGC) lors du projet HAPMAP. Elle est composée de 54 241 SNP avec un espacement moyen entre marqueur de 50,9kb

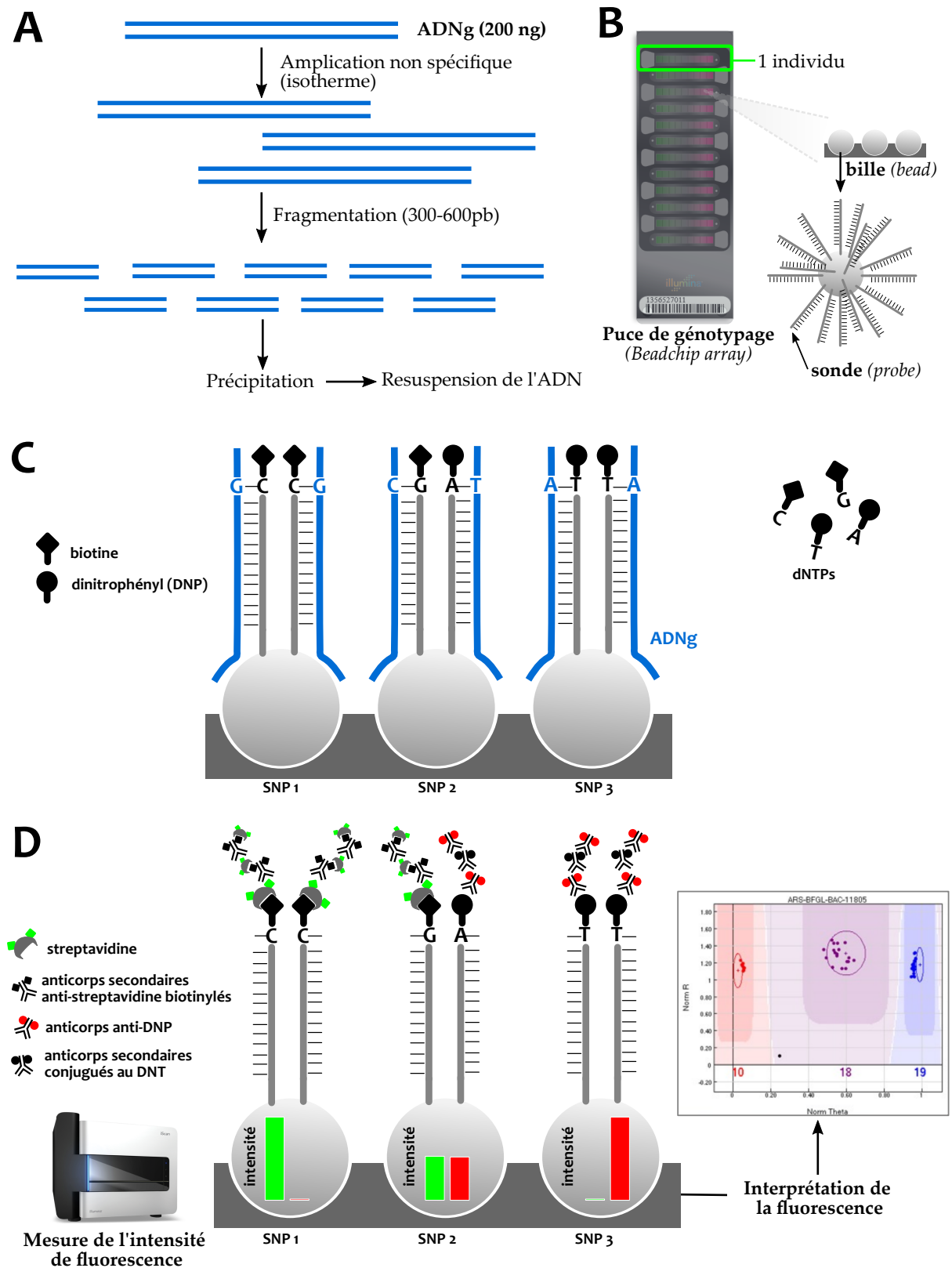


FIGURE 3.6 – Fonctionnement du génotypage haut-débit d'un marqueur SNP sur puce Illumina utilisant la chimie *Infinium UltraHD*. Adapté de (Illumina, 2009) (A) Préparation de l'ADN génomique. (B) Puce de génotypage. (C) Hybridation spécifique de l'ADN aux sondes oligonucléiques et extension d'une seule base. Par simplicité, seulement deux sondes sont représentées sur chacune des billes. (D) Marquage fluorescent des dNTPs, mesure et interprétation de la fluorescence.

(Kijas et al., 2009; Illumina, 2015). Une puce moyenne densité est également disponible chez ThermoFisher (Axiom™ Ovine Genotyping Array 50K) avec la technologie Affymetrix.

- La puce **OvineHD BeadChip** (AgResearch's SheepHD) est une puce cliente (support offert par le fournisseur pour y placer les marqueurs de son choix) de haute densité avec environ 600 000 SNP (coût de 90€).

A noter que depuis 2022, une nouvelle puce cliente **AgResearch 60K** a vu le jour et résulte de la fusion entre les puces ISGC SheepLD et OvineSNP50 Genotyping BeadChip. Cette puce a un coût équivalent à la puce ISGC SheepLD.

En caprins, une seule puce de moyenne densité existe sur le marché (GoatSNP50 Illumina chip) depuis 2011 et a récemment été mise à jour en 2021 (Goat_IGGC_65K_v2) (Tosser-Klopp et al., 2014). Elle est composée de 59 727 SNP (IGGC, 2021) et a un coût de 19€. Une puce moyenne densité ThermoFisher est également en cours de développement.

3.1.4.3 Utilisation des puces dans le cadre de la sélection génomique

L'ensemble des puces citées précédemment est utilisé en recherche pour effectuer des analyses génétiques (analyse d'association, analyse de diversité ...). Le recours à la puce 600K permet de gagner en puissance puisque qu'elle permet de mettre en évidence des régions d'intérêt plus petites.

Les puces basses et moyennes densités sont utilisées dans le cadre de la sélection génomique (voir [paragraphe 1.2.3](#)). L'information moléculaire portée par les marqueurs SNP permet d'estimer plus précisément les valeurs génétiques des animaux (possible dès la naissance) sans collecter de phénotype. L'utilisation des puces basses densités (15 000 SNP) permet pour un coût moindre d'avoir accès à l'information moléculaire sur les 50 000 marqueurs de la puce de moyenne densité par les processus de phasage et d'imputation. Le phasage permet à partir des génotypes des parents, de déduire l'origine des allèles (paternel et maternel) du descendant génotypé. L'imputation de génotype est un processus d'inférence de génotypes non observés qui repose sur un échantillon d'individus déjà génotypés. Cela permet de prédire les génotypes des marqueurs non typés sur la puce basse densité présents sur la puce moyenne densité (35 000 autres marqueurs) et de corriger les génotypes manquants des marqueurs (erreurs de génotypage) (Marchini & Howie, 2010) ([Table 3.2](#)). Pour cela, on prend en compte les génotypes des animaux de la population de référence et du pedigree. Cette méthode repose sur le déséquilibre de liaison entre marqueurs.

L'imputation en ovins est fiable : Call Rate par animal >98%, Call Rate par SNP >98% et un r^2 par SNP >93% pour les races laitières Lacaune et Manech Tête Rousse (Larroque et al., 2017). Dans les schémas de sélection, les candidats à la sélection sont génotypés sur la puce basse densité puis seuls les candidats retenus (30%) sont regénotypés mais cette fois sur la puce moyenne densité pour entretenir la population de référence (précision de l'imputation et phasage des générations futures). La nouvelle puce ovine 60K permettra de réaliser dès la campagne 2022 un seul génotypage. A ce jour, dans le cadre de la sélection génomique, c'est 50 000 et 7 000 génotypages disponibles en ovins et caprins laitiers, respectivement. A titre de comparaison, en bovins laitiers français, le nombre de génotypages a largement augmenté cette dernière décennie puisque l'on passe de 8 500 génotypages en 2009 à 975 000 génotypages en 2019 (VanRaden, 2020).

TABLE 3.2 – **Principe de l'imputation de la puce basse densité vers la moyenne densité.** Les SNP imputés sont représentés en bleu et les erreurs d'imputation en rouge. Les phases 1 et 2 correspondent aux phases paternelles et maternelles, respectivement. *Reproduit de Larroque et Astruc (2017).*

Densité	Phase	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10	SNP11	SNP12	SNP13	SNP14	SNP15	SNP16	SNP17	SNP18
50K	1	A	C	G	A	C	T	G	C	A	C	G	T	G	G	C	T	G	C
	2	T	C	A	A	C	T	C	T	A	G	G	T	G	C	T	T	G	T
LD	1	A	.	G	.	.	.	G	C	.	C	.	.	.	G	C	.	.	C
	2	T	.	A	.	.	.	C	T	.	G	.	.	.	C	T	.	.	T
LD → 50K	1	A	C	G	A	C	T	G	C	A	C	G	T	G	G	C	T	G	C
	2	T	C	A	A	T	T	C	T	A	G	G	T	G	C	T	A	G	T

3.2 Des dispositifs d'hérédo-surveillance facilite la remontée de cas

L'apparition des anomalies génétiques est inévitable. Sans gestion, l'anomalie peut avoir des conséquences importantes dans une filière. Agir tôt lors de l'émergence d'une anomalie est essentiel pour limiter sa propagation. Plusieurs pays ou institutions ont créé des dispositifs d'hérédo-surveillance. Le Danemark est le premier pays à avoir créé un tel observatoire dès 1989 avec le *Danish Bovine Genetic Disease Program* (Agerholm et al., 1993). En 2002, c'est au tour de la France de créer son observatoire national des anomalies bovines "ONAB" (ONAB, 2021 ; Grohs et al., 2016). La création de cet observatoire a été motivée par les crises liées aux anomalies *Complex vertebral malformation, CVM* (Thomsen et al., 2006) et *Bovine Leucocyte Adhesion Deficiency, BLAD* (Shuster et al., 1992) pour lesquelles la semence de taureaux porteurs hétérozygotes des anomalies a été fortement diffusée dans la population Prim'Holstein avant d'être prise en charge. A ce jour, l'ONAB compte plus de 30 anomalies avec mutations causales identifiées. Le rôle des observatoires est multiple : (i) identifier l'émergence de nouvelles anomalies (fiche descriptive remplie par les éleveurs, les techniciens d'élevage, les vétérinaires), (ii) comprendre l'anomalie et (iii) gérer son élimination progressive. Dans d'autres espèces, des observatoires ont été créés en porc (Riquet et al., 2016) et en équin (IFCE, 2008). Il est important de préciser que l'analyse génétique ne peut se faire sans échantillons biologiques. C'est pourquoi les observatoires ont un lieu de stockage centralisé des échantillons de type Centre de Ressources Biologiques (CRB, <https://www.crb-anim.fr/>).

La plupart du temps, la remontée des anomalies se fait via des réseaux établis chez les espèces domestiques. Par exemple, une grande base de données (<http://www.maladieshereditairesduchien.com>)

regroupe toutes les maladies génétiques identifiées chez le chien, qui constitue une espèce modèle pour les maladies génétiques humaines (van Steenbeek et al., 2016). L'existence de ces réseaux repose notamment sur les initiatives des facultés vétérinaires. Par exemple, le professeur C. Drögemüller (université de Berne, Suisse) est très sollicité pour identifier le déterminisme génétique des anomalies génétiques chez les animaux de rente.

A ce jour, aucun observatoire n'existe en petits ruminants. La remontée de cas reste occasionnelle et est principalement expliquée par la faible valeur économique d'un agneau ou chevreau (<50€). Même si, comme dans d'autres espèces, un certain nombre d'anomalies génétiques ségrège dans ces populations.

Néanmoins, la volonté des professionnels d'avoir leur propre observatoire des anomalies génétiques pour détecter précocement l'émergence d'anomalies génétiques et de les gérer le plus rapidement est rendu possible par le projet Casdar PRESAGE 2021-2024 (<https://idele.fr/presage/>). Tout comme pour l'ONAB, la création d'un tel observatoire a été motivée par l'émergence d'une anomalie avec de fortes conséquences dans la filière. Il s'agit de l'anomalie *epidermolyse bulleuse* identifiée en race allaitante mouton Vendéen (Fabre et al., 2020). L'objectif majeur de ce projet est d'évaluer les conditions nécessaires pour la création d'un tel observatoire. Ce projet repose sur 5 axes : (1) communiquer et sensibiliser les différents acteurs du projet aux anomalies génétiques, (2) organiser le futur observatoire, (3) faciliter la remontée de cas observés sur le terrain, permise par le développement d'une application mobile, (4) définir les méthodes pour identifier les anomalies génétiques par l'étude de cas pilotes et (5) proposer des modèles de gestion des anomalies dans les filières (Buisson, 2021). L'ensemble de mes travaux de thèse s'inscrit dans ce projet.

A retenir L'étude des anomalies génétiques est possible par (i) des avancées technologiques et une (ii) facilité de remontée des cas permise par les systèmes d'hérédosurveillance. L'évolution des technologies de séquençage (Sanger, séquençage haut-débit lectures courtes ou longues) a permis d'avoir des génomes de référence dans chaque espèce et de mieux connaître les génomes (annotations structurale et fonctionnelle). La disposition d'un grand nombre d'individus séquencés a permis de construire des puces SNP permettant de connaître l'information moléculaire de nombreux individus pour un coût raisonnable. De plus, les mutations causales identifiées peuvent être mises sur ces puces afin de pouvoir gérer les accouplements dans les schémas de sélection. Le point limitant pour identifier les anomalies génétiques repose sur une remontée précoce des premiers cas et de leur description clinique fine. Différents observatoires dédiés ont ainsi été créés afin de centraliser la remontée des cas et de faciliter l'analyse génétique. ■

4. Les différentes méthodes pour identifier les anomalies génétiques

L'identification du déterminisme génétique d'une anomalie impose d'avoir accès à une grande quantité d'informations diverses : phénotypes, pedigree, données de génotypage et de séquençage de génomes entiers. L'annotation fonctionnelle du génome (voir [paragraphe 3.1.3.2](#)) s'avère ensuite indispensable pour localiser les régions/variants candidats. Différentes méthodes permettent de mettre en évidence le déterminisme génétique d'une anomalie. Les approches classiques de génétique reposent sur l'enregistrement de phénotypes et l'analyse de l'ADN des individus phénotypés (approche du phénotype au génotype). Avec l'avènement des données moléculaires, une autre approche a été développée en tirant parti uniquement des données de génotypage/séquençage disponibles pour identifier des variants supposés responsables d'anomalies génétiques sans connaître préalablement le phénotype. Dans ce chapitre, les méthodes décrites concernent principalement des anomalies récessives car elles sont beaucoup plus nombreuses (plus diffusées dans les populations) et sont plus facilement identifiées.

4.1 Est-ce que l'anomalie est d'origine génétique ?

Avant de lancer toute analyse génétique, il est nécessaire de s'interroger si l'anomalie à l'étude a une cause génétique. Par exemple, l'anomalie de freemartinisme (coexistence d'ovaires et de testicules chez un même individu qui se traduit le plus souvent par une apparence ambiguë des organes génitaux externes à la naissance) peut être identifiée comme une anomalie supposée d'ordre génétique. Chez le bovin, il existe un cas de freemartinisme avec des causes génétiques (Caryotype XX et présence de *SRY* (*Sex-determining Region of Y chromosome*)) (Bresciani et al., 2015). Toutefois dans la plupart des cas, le freemartinisme est un défaut congénital non génétique. Par exemple, dans l'espèce ovine, un certain nombre d'agnelles peut se retrouver avec un appareil génital externe (vulve et clitoris visibles) et des testicules en position inguinale (Chaban et al., 2015). Cette anomalie est souvent identifiée par les inséminateurs lors de la première mise à la reproduction. Or, les freemartins apparaissent lorsque des connexions vasculaires se forment entre les placentas de fœtus jumeaux de sexe différent. Il en résulte une masculinisation de l'appareil reproducteur femelle par une réponse aux hormones produits par le fœtus mâle (AMH et testostérone) à des degrés variables (Cabianca et al., 2007).

Ainsi, une étude préliminaire du pedigree (individus atteints provenant du même père ou spécifiques d'un élevage) vont permettre de mettre en évidence un potentiel lien génétique. Lorsque l'anomalie est supposée d'ordre génétique, il est essentiel d'avoir une description clinique précise. Un phénotype approximatif va entraîner un biais dans l'analyse génétique. Des animaux peuvent être considérés comme atteints de la même anomalie (phénotypes proches) mais les déterminismes

génétiques sont différents. Par exemple, l'épidermolyse bulleuse est associée à 4 groupes de maladies génétiques : épidermolyse bulleuse simplex, épidermolyse bulleuse jonctionnelle, épidermolyse bulleuse dystrophique et épidermolyse bulleuse Kindler. Chacune de ces catégories est liée à des mutations causales localisées dans des gènes différents (Has et al., 2020).

4.2 Du phénotype au génotype : méthodes de génétique classique

L'ensemble des méthodes décrites ci-dessous permettent de mettre en évidence des régions du génome susceptibles d'abriter la mutation causale. Une fois la région détectée, des re-séquençages de génomes complets permettront d'identifier dans un second temps la mutation causale.

4.2.1 Approche gènes candidats

Lorsque le phénotype atteint est similaire ou correspond à des anomalies déjà connues, il peut être judicieux dans un premier temps de regarder les gènes/mutations impliqué(e)s dans ce phénotype.

La première stratégie consiste à séquencer des gènes candidats pour détecter de nouveaux polymorphismes. Par exemple, en Tunisie, un troupeau de mouton Barbarine a été sélectionné sur la prolificité. Suite à l'observation de brebis hyperprolifiques et de stérilité dans le troupeau, la ségrégation d'une nouvelle mutation dans les gènes de prolificité *BMP15* et *GDF9* était alors envisagée. En effet, une nouvelle mutation dans le gène *BMP15* a été identifiée. Elle correspond à une mutation complexe combinant un SNP, une insertion et délétion appelée *FecX^{Bar}* : OARX (Oar_rambouillet_v1.0) : g.[56,600,937insG ; 56,600,945_56,600,947del ; 56,600,948C>A] (p.Ala101Cysfs*113) (OMIA 002306-9940) (Lassoued et al., 2017).

La deuxième stratégie consiste à séquencer quelques animaux (atteints et sains) tout génome et de filtrer les variants présents dans des gènes d'intérêts. Elle a été adoptée pour l'anomalie épidermolyse bulleuse identifiée en mouton vendéen. Les individus atteints présentaient les signes cliniques d'une épidermolyse bulleuse. Les mutations déjà connues en ovins, identifiées dans d'autres races dans les gènes *LAMC2* (Mömke et al., 2011) et *ITBG4* (Suárez-Vega et al., 2015), ont été testées mais aucune ne ségrégeaient dans la population. Une approche "tout génome" par re-séquençage haut-débit a donc été réalisée à partir de trois individus (2 individus atteints et un plein frère sain d'un individu atteint). Avec une hypothèse d'anomalie récessive, l'ensemble des variants a été filtré de telle sorte que les individus atteints étaient porteurs homozygotes variants et le frère non atteint était hétérozygote ou homozygote référence. La recherche de variants s'est focalisée dans une dizaine de gènes candidats responsables d'épidermolyse bulleuse et a abouti à la découverte d'une mutation non-sens OAR11 (Oar_rambouillet_v1.0) : g.7,412,626C>T (p.Arg885*) dans le gène *ITGB4* (Fabre et al., 2020) (voir [paragraphe 2.2.2](#)).

4.2.2 Détection de QTL : Analyses de liaison/association tout génome

Principe général. Un des principaux objectifs en génétique est de comprendre pour un caractère, quelle part de la variation phénotypique observée est expliquée par des différences génétiques (Ytournal et al., 2008). Cela repose sur l'identification de région QTL (*Quantitative Trait Loci*), définie comme une région du génome qui est associée à un caractère quantitatif par

l'utilisation de marqueurs génétiques. D'un point de vue statistique, la détection de QTL repose sur un modèle de régression simple où l'on suppose une relation linéaire entre le génotype au QTL et le phénotype d'intérêt :

$$y = \mu + \mathbf{X}\beta + \mathbf{Z}a + \varepsilon$$

où y est le vecteur des phénotypes du caractère, μ la moyenne du caractère dans la population, X le vecteur des génotypes au QTL, β les effets fixes additifs au QTL, Z la matrice d'incidence de y associée à a , a le vecteur des effets polygéniques aléatoires qui suit une loi Gaussienne $a \sim \mathcal{N}(0, A\sigma_A^2)$ avec A la matrice de parenté, σ_A^2 la variance génétique additive et ε la résiduelle (autres effets non captés par le modèle) qui suit une loi Gaussienne $\varepsilon \sim \mathcal{N}(0, I\sigma_\varepsilon^2)$ avec I la matrice identité et σ_ε^2 la variance résiduelle.

La cartographie de régions génomiques d'intérêt repose principalement sur le déséquilibre de liaison (noté DL) entre marqueurs. Le DL est défini comme l'association non-aléatoire d'allèles à des locus différents. Plus les marqueurs sont proches, plus ils ont des chances d'être transmis ensemble à la descendance. Les modèles statistiques à utiliser dépendent du dispositif expérimental et du choix de modélisation. La cartographie d'un QTL peut se faire par analyse de liaison intra-famille ou d'une association populationnelle.

Analyse de liaison. Cette méthode repose sur la construction d'une structure de population apparentée. En petit ruminants, de nombreuses familles QTL père-filles ont été réalisées. L'objectif est de suivre dans chaque famille la ségrégation des segments chromosomiques chez les descendants pour mettre en lien des différences de performances. Dans le cas de ce dispositif, ce sont de grands fragments de chromosome qui sont transmis à la descendance rendant la localisation des QTLs peu précise.

Analyse d'association. Cette méthode (*Genome Wide Association Study*, GWAS) repose sur l'analyse directe de la relation entre chaque génotype observé à un polymorphisme et le phénotype dans une population (Uffelmann et al., 2021). L'hypothèse sous-jacente impose que chacun des marqueurs testé est causal ou en fort déséquilibre de liaison avec la mutation causale. Afin d'être efficace, il est nécessaire d'avoir (*i*) une densité importante de marqueurs, possible par par l'essor des nouveaux outils (puce moyenne/haute densité de SNP, séquençage de génomes entiers) et un grand nombre d'individus non-apparentés. Cette méthode a été développée chez l'Homme et transposée chez les animaux de rente. Cependant, contrairement à l'espèce humaine, les populations d'élevage sont plus apparentées. Pour limiter cette structure de pedigree, une matrice de parenté peut être incluse dans le modèle afin d'éviter la détection de faux-positifs. Des modèles plus complexes avec des régressions bayésiennes simples ou multi-marqueurs peuvent également être utilisés (X. Zhou et al., 2013).

Analyse Cas-Contrôle. Dans le cadre d'une anomalie génétique, l'analyse d'association repose très souvent sur un phénotype binaire (individus atteints vs individus sains). Ainsi, les modèles statistiques GWAS peuvent être simplifiés en utilisant un modèle de régression logistique (Uffelmann et al., 2021). Sachant que les individus atteints d'une anomalie génétique sont issus d'un même ancêtre commun, la structure du pedigree est source de biais. Pour gagner en puissance, il est important de choisir un dispositif expérimental pertinent i.e. pour un individu atteint (supposé

homozygote variant), il est important de prélever un plein frère sain (supposé hétérozygote ou homozygote référence). Dans ce cas de figure, une analyse cas-contrôle peut être réalisée à partir de 50 individus (25 cas/25 contrôles). Un exemple de GWAS cas-contrôle a été utilisé pour mettre en évidence une région du génome impliquée dans la microphthalmie en race ovine Texel (OMIA 000649-9940) (Becker et al., 2010). Le GWAS est illustré dans la Figure 4.1. Le séquençage d'animaux atteints a ensuite permis de mettre en évidence une mutation faux-sens OAR22 (Oar_rambouillet_v1.0) : g.25497953C>G (p.Arg113Pro) dans le gène *PITX3* (*Paired Like Homeodomain 3*).

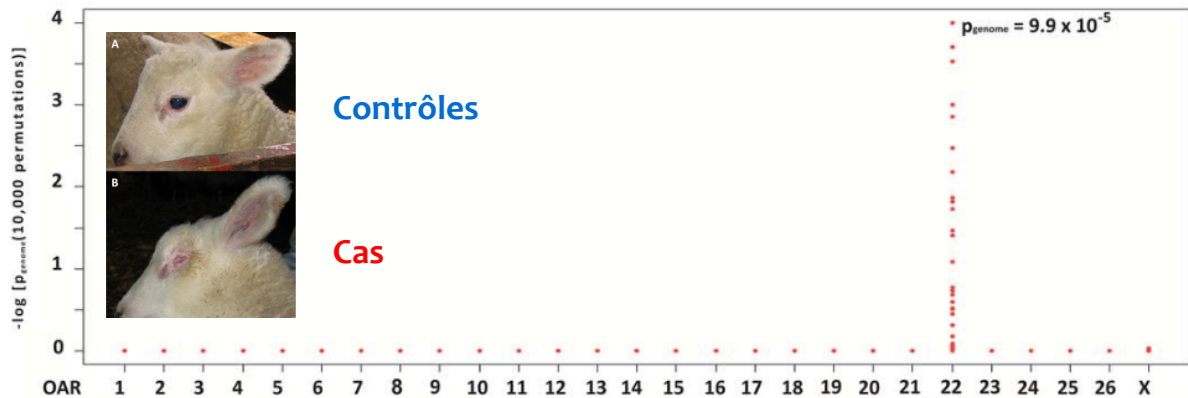


FIGURE 4.1 – **Analyse d'association de type cas-contrôles.** 23 individus sains et 23 atteints de microphthalmie ont été génotypés sur puce 50K pour réaliser l'analyse. Un intervalle de 3,6 Mb a été mis en évidence, susceptible d'abriter la mutation causale. *Extrait de* Becker et al. (2010).

4.2.3 Cartographie par homozygotie

La méthode de cartographie par homozygotie a été développée par Lander et Botstein (1987) et a été utilisée en productions animales pour la première fois par Charlier et al. (2008). La cartographie par homozygotie permet d'identifier une portion du génome homozygote identique entre individus atteints. Son principe est décrit dans la Figure 4.2. Cette méthode repose sur deux hypothèses : (i) le déterminisme génétique de l'anomalie est récessif et (ii) les individus atteints sont porteurs de la même mutation, héritée d'un même ancêtre commun. La méthode de cartographie par homozygotie peut être menée de manière indépendante avec une analyse d'association classique mais est souvent complémentaire. A la suite de l'analyse GWAS, Becker et al. (2010) ont permis de restreindre la recherche à un plus petit fragment chromosomique de 2,4 Mb qui est homozygote et identique chez tous les individus atteints (*Identical by descent*, IBD). Cette méthode peut être utilisée lorsque l'on dispose seulement de quelques cas. Cependant, plusieurs régions chromosomiques peuvent être identiques entre individus atteints. Pour gagner en puissance, il est judicieux de prendre des individus atteints les moins apparentés possible, d'augmenter le nombre de cas afin de réduire l'intervalle d'étude à seulement une seule région génomique et que le nombre de générations séparant les individus atteints et l'ancêtre commun soit élevé. L'ensemble de ces paramètres permettra d'avoir une région beaucoup plus courte liée aux recombinaisons (Duchesne et al., 2016).

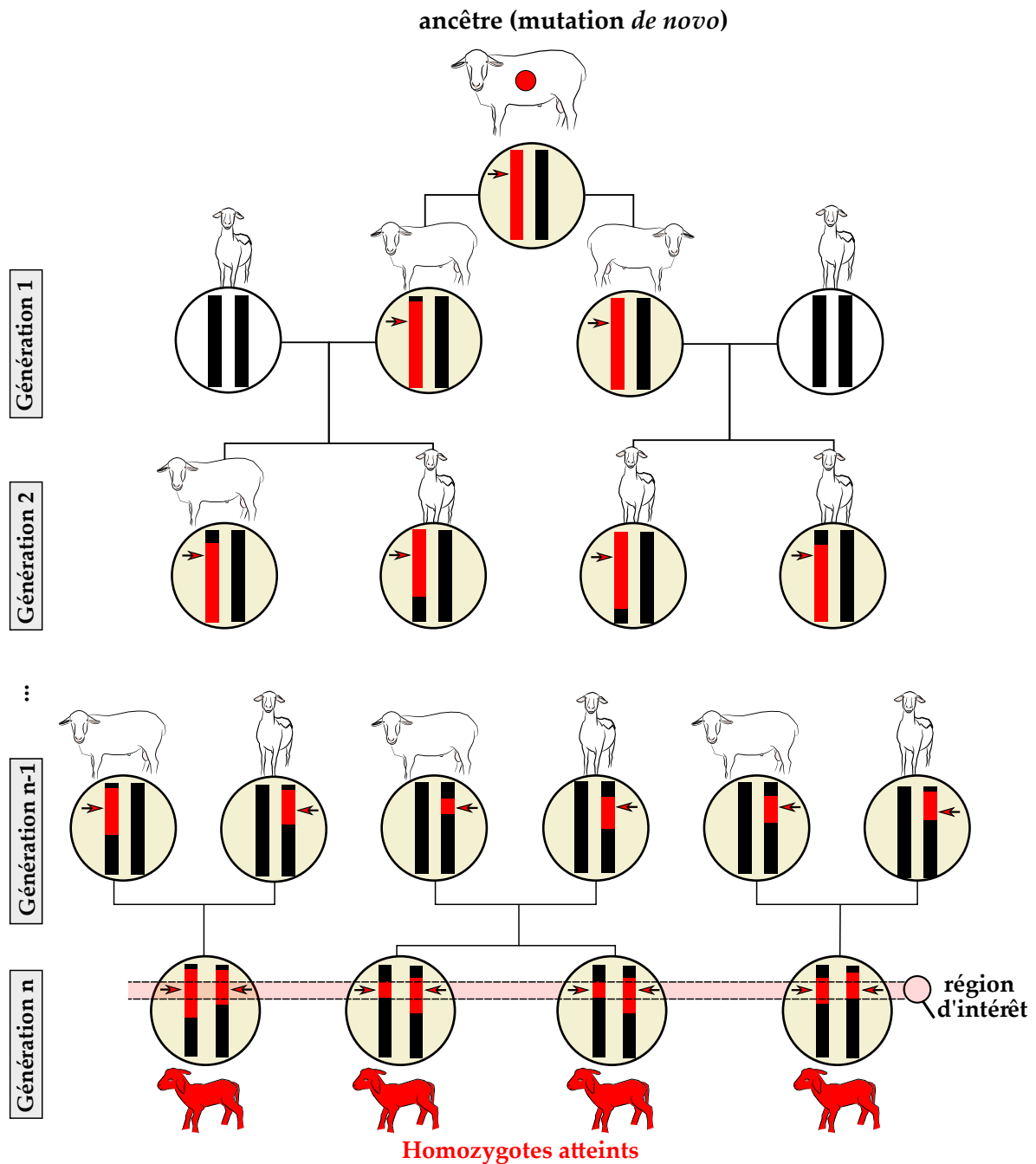


FIGURE 4.2 – Principe de la cartographie par homozygotie pour identifier des anomalies génétiques récessives. Une mutation *de novo* apparaît chez un individu (la localisation de la mutation est représentée par la flèche). Il va ensuite transmettre la mutation à l'état hétérozygote et donc du segment chromosomique entourant la mutation (phénotypiquement normal). Au cours des générations, le fragment entourant la mutation devient plus court sous l'effet des recombinaisons. A partir d'un certain nombre de générations, des animaux apparentés, porteurs hétérozygotes de la même mutation, vont s'accoupler et donner naissance à des individus atteints porteurs homozygotes de la mutation. La mutation est alors cartographiée dans un petit segment homozygote et identique chez tous les individus atteints. *Adapté de* Georges et al. (2019) ; Duchesne et al. (2016) et Ytournel et al. (2008).

4.3 Du génotype au phénotype : la génétique inverse

Pour réaliser les approches décrites précédemment, il est nécessaire d'avoir accès au phénotype, mais ceci n'est pas toujours évident. Dès lors que l'anomalie conduit à une mortalité embryonnaire, par exemple, l'acquisition des phénotypes et des échantillons biologiques est impossible. D'un point de vue zootechnique, une anomalie génétique conduisant à une mortalité embryonnaire est confondue avec une baisse de fertilité et passe généralement inaperçue (VanRaden et al., 2011 ; Fritz et al., 2013). Profitant de l'essor des données haut-débit (génotypage et re-séquençage d'un grand nombre d'animaux), des méthodes dites de **génétique inverse** (*Reverse genetic screen*), basées uniquement sur l'analyse des données sans connaissance au préalable de phénotypes, ont été développées pour mettre en évidence des régions du génome abritant des mutations causales délétères.

4.3.1 Exploitation des données de séquence pour identifier des variants délétères rares

Une stratégie de génétique inverse a vu le jour avec l'accumulation des données de re-séquençage tout génome notamment disponibles dans les projets 1 000 génomes dans les différentes espèces (Daetwyler et al., 2014 ; Denoyelle et al., 2021). La stratégie consiste à identifier les variants prédits délétères susceptibles d'avoir un fort impact biologique (Charlier et al., 2016 ; MacArthur et al., 2012). Son principe est détaillé dans la [Figure 4.3](#).

Par comparaison avec la séquence de référence, plusieurs millions de variants sont identifiés. La localisation des variants permise par les annotations positionnelle et fonctionnelle des génomes et la prédiction *in silico* de l'effet des variants par les outils bioinformatiques (voir [paragraphe 2.1.2.1](#)) permettent d'identifier des variants candidats supposés hautement délétères (Fritz et al., 2016). L'information de phénotypes associée aux gènes, disponible dans les bases de données humaines (OMIM), souris (MGI) ou animales (OMIA) permet de prédire les phénotypes attendus. Une fois que la liste des mutations candidates a été établie, il est nécessaire de génotyper les mutations à très grande échelle, rendu possible par l'ajout de ces dernières sur les puces SNP. Le test d'équilibre de Hardy-Weinberg permet d'identifier les mutations responsables de mortalité embryonnaire ou foetale. La causalité peut être validée par (*i*) le suivi d'accouplements à risque entre porteurs hétérozygotes ou (*ii*) la réalisation au laboratoire de tests fonctionnels *in vitro*. Par cette approche, Charlier et al. (2016) ont identifié 9 nouvelles mutations responsables de mortalité embryonnaire en bovins.

4.3.2 Identification de régions du génome présentant des déficits d'animaux homozygotes à partir de données de génotypage

L'accumulation des données de génotypage disponibles dans le cadre de la sélection génomique a permis à VanRaden et al. (2011) de proposer une méthode pour rechercher des haplotypes présentant un déficit en animaux homozygotes. Ces régions sont supposées porter des mutations récessives létales. En effet, le déficit d'homozygotes s'expliquerait par la mort précoce des animaux (stades embryonnaire, foetal, néonatal ou juvénile) avant leur génotypage. En France, une approche similaire a été faite par Fritz et al. (2013) dans les trois principales races bovines laitières : Holstein,

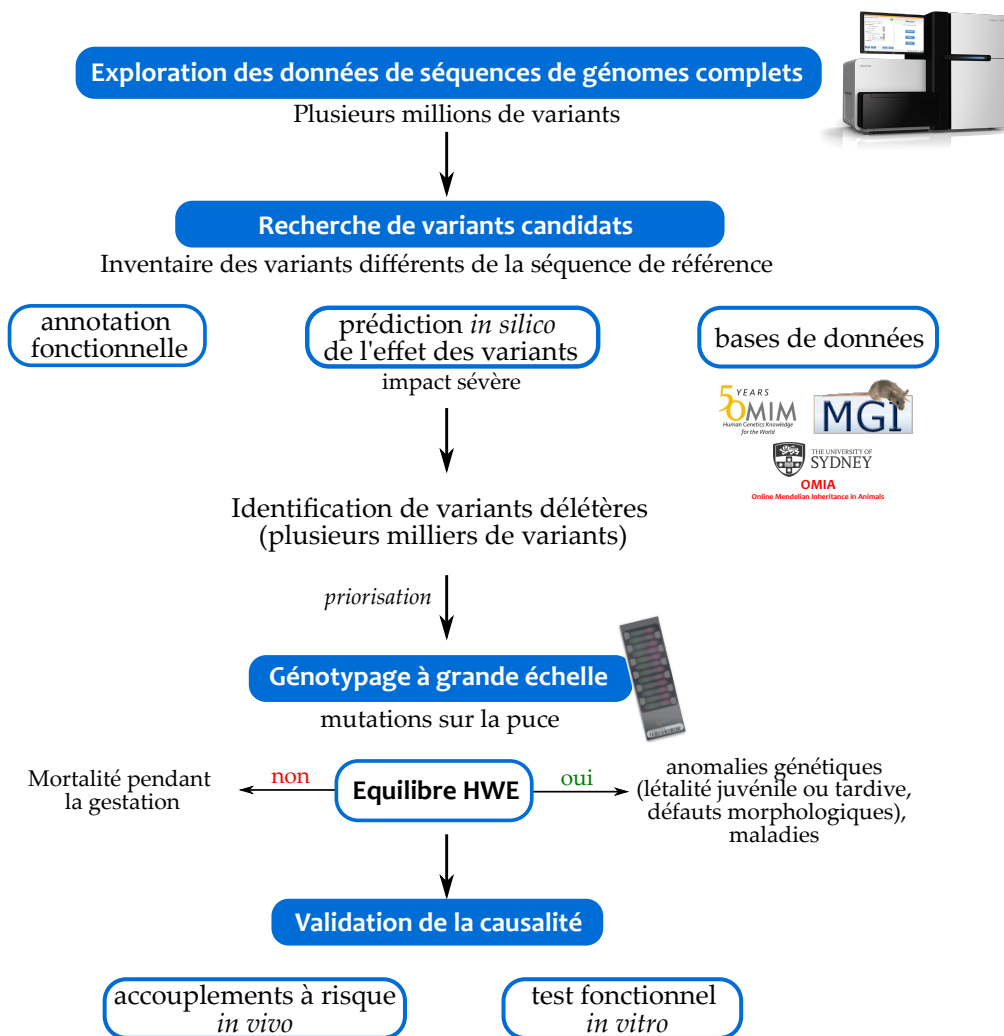


FIGURE 4.3 – Principe de la recherche de variants délétères rares à partir de l’exploitation des données de séquences. HWE : Hardy-Weinberg. *Adapté de Michot (2017)*

Montbéliarde et Normande. Cette méthode est basée sur le principe d’écart à la loi de Hardy-Weinberg (Voir paragraphe 3.1.4.1 - Encadré 3. Principe de l’équilibre de Hardy-Weinberg). Pour rappel, chacun des SNP présents sur la puce est choisi avec une fréquence de l’allèle minimal $>5\%$ et respectant l’équilibre de HWE. Or, une mutation hautement délétère, par définition, a une fréquence faible et peut déroger à l’équilibre de HWE. Ainsi, un marqueur sur la puce ne peut être causal. Pour l’approche de génétique inverse, il est nécessaire de raisonner en terme d’haplotypes qui sont composés d’un ensemble de marqueurs adjacents phasés (généralement composés de 20 marqueurs SNP) qui ont le même comportement que la mutation causale par le phénomène de déséquilibre de liaison.

Le principe de cette méthode repose sur un balayage complet du génome des individus génotypés dans une population sans *a priori* sur un phénotype (Figure 4.4). On fait l’hypothèse qu’un haplotype ségrège dans une population, seulement présent à l’état hétérozygote et n’étant jamais observé à l’état homozygote. Une fois de tels haplotypes identifiés, le statut (porteur, non-porteur) est déterminé pour chacun des individus génotypés. L’analyse des données de reproduction (réussite IA, mortinatalité ...) va permettre de confirmer la létalité par comparaison des accouplements

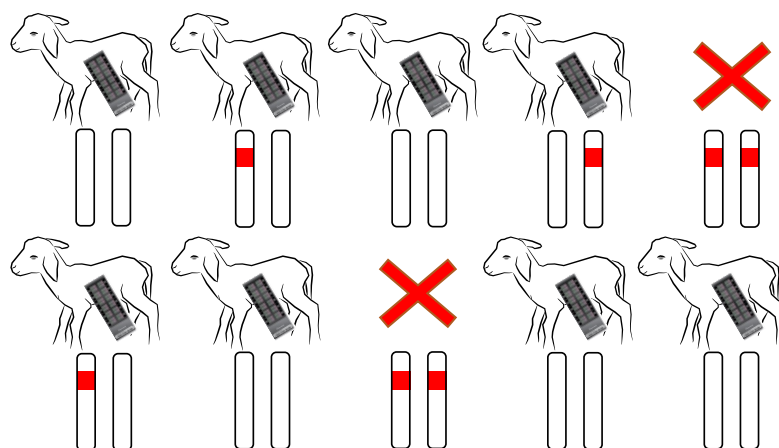


FIGURE 4.4 – Principe de la recherche d’haplotype en déficit d’animaux homozygotes. Un haplotype (représenté en rouge) ségrège dans la population génotypée mais n’est jamais retrouvé à l’état homozygote. Les animaux homozygotes à cet haplotype ne sont pas candidats à la sélection et supposés porteurs homozygotes d’une mutation délétère récessive.

à risque (mâles porteurs x femelles porteuses) et non à risque (autres combinaisons). L’étape suivante consiste à identifier la mutation causale localisée dans la région en déficit (séquençage d’animaux hétérozygotes à l’haplotype) en recherchant la mutation chez les animaux hétérozygotes porteurs de l’haplotype à l’état hétérozygote mais homozygote référence chez les individus non porteurs de l’haplotype.

A ce jour, 31 mutations causales (Table 4.1) localisées dans des haplotypes en déficit d’animaux homozygotes ont été publiées en bovins laitiers pour lesquelles l’accumulation des données de la sélection génomique a commencé en 2009 (VanRaden, 2020). A l’état homozygote, ces mutations sont responsables de mortalité embryonnaire précoce, d’avortements tout au long de la gestation ou de mortalité périnatale à juvénile. Par ce type de raisonnement, il est également possible d’identifier des mutations délétères responsables de défauts morphologiques qui écartent de fait les animaux de la sélection. C’est le cas pour un retard de croissance des veaux associé à une mutation faux-sens BTA26 (ARS-UCD1.2) : g.32940521C>G, (p.Asn176Lys) dans le gène *ACSL5* (*Acyl-CoA Synthetase Long Chain Family Member 5*) en Brown swiss (Häfliger, Seefried, Spengeler, & Drögemüller, 2021). Par cette approche, des régions en déficit peuvent également signer des mutations déjà connues comme par exemple les haplotypes HBY et HHC associés aux anomalies *Brachyspina* (Charlier et al., 2012) et *Complex Vertebral Malformation* (Thomsen et al., 2006), respectivement (Table 4.1).

Bien qu’initialement utilisé dans les populations bovines laitières où le nombre de génotypes est le plus important, ce raisonnement de génétique inverse s’est aujourd’hui généralisé dans plusieurs autres dispositifs de sélection : bovins allaitants (Jenko, 2019), porc (Derks et al., 2017 ; Derks, Lopes, et al., 2018), volaille (Derks, Megens, et al., 2018), dinde (Abdalla et al., 2020) et cheval (Todd et al., 2020). Cependant, aucun travail de recherche en petits ruminants n’a été publié avant mes travaux de thèse.

TABLE 4.1 – Liste des mutations causales identifiées par recherche de régions en déficit d’animaux homozygotes en bovins laitiers. Att= nombre d’animaux homozygotes attendus, DEL= délétion, Freq= fréquence de porteurs hétérozygotes de l’haplotype, FS= décalage du cadre de lecture (*frameshift*), MS= missense (faux-sens), NS= non-sens (codon-stop), Obs= nombre d’animaux homozygotes observés, SL= absence du codon start (*start loss*), SS= site d’épissage (*splice site*).

Races	Haplotype	Déficit (Att./Obs.) (%)	Freq (%)	OMIA ID	Gène	Type de variant	Phénotype	Références
Ayshire	AH1	14/0	26,1	001934-9913	<i>UBE3B</i>	SS (saut de l'exon 23)	Mortalité juvénile - Syndrome de ptosis, déficience intellectuelle, retard de croissance et mortalité (PIRM)	(T. A. Cooper et al., 2014)
	AH2	23/0	18,9	002134-9913	<i>RPAP2</i>	SS (splice acceptor)	Létalité embryonnaire	(Null et al., 2017)
Brown Swiss	BH2	41/1	13,0	001939-9913	<i>TUBD1</i>	MS (His>Arg)	Mortalité périnatale/juvénile - Syndrome respiratoire	(VanRaden et al., 2011 ; Schwarzenbacher et al., 2016)
	BH6	56/0	6,8	002517-9913	<i>MARS2</i>	MS (Arg>Gln)	Létalité embryonnaire ?	(Häfliger, Seefried, Spengeler, & Drögemüller, 2021)
	BH14	45/0	6,4	002518-9913	<i>MRRPL55</i>	NS	Létalité embryonnaire ?	(Häfliger, Seefried, Spengeler, & Drögemüller, 2021)

BH24	26/4	5,2	002519-9913	<i>CPT1C</i>	MS (Gly>Asp)	Mortalité juvénile	(Häffiger, Seefried, Spengeler, & Drögemüller, 2021)
BH34	27/7	4,8	002226-9913	<i>ACSL5</i>	MS (Asn>Lys)	Défauts morphologiques - Retard de croissance	(Häffiger, Seefried, Spengeler, & Drögemüller, 2021)
Fleckvieh							
FH2	37/2	8,2	001958-9913/ 000366-9913	<i>SLC2A2</i>	FS ->active SS cryptiques	Mortalité juvénile - Fanconi-Bickel syndrome	(Pausch et al., 2015)
FH4	33/0	6,6	001960-9913	<i>SUGT1</i>	MS (Trp>Arg)	Létalité embryonnaire	(Pausch et al., 2015)
Holstein							
<i>HH0/HBY*</i>	22/0	7	000151-9913	<i>FANCI</i>	DEL 3,3 kpb	Avortement tardif, mortalité périnatale - Brachyspina	(VanRaden et al., 2011; Fritz et al., 2013; Sahana et al., 2013)
<i>HHC*</i>	23/0	1,6	001340-9913	<i>SLC35A3</i>	MS (Val>Phe)	Avortement tardif, mortalité périnatale - CVM (Complex vertebral malformation)	(VanRaden et al., 2011; Fritz et al., 2013)
HH1	30/0	4,5	000001-9913	<i>APAF1</i>	NS	Létalité embryonnaire	(VanRaden et al., 2011; Fritz et al., 2013)
HH2	10/0	4,6	001823-9913	<i>IFT80</i>	FS	Létalité embryonnaire	(VanRaden et al., 2011; McClure et al., 2014)

HH3	33/0	4,7	001824-9913	<i>SMC2</i>	MS (Phe>Ser)	Létalité embryonnaire	(VanRaden et al., 2011; Fritz et al., 2013; McClure et al., 2014)
HH4	49/0	7,2	001826-9913	<i>GART</i>	MS (N>T)	Létalité embryonnaire	(Fritz et al., 2013)
HH5	42/0	2,80	001941-9913	<i>TFB1M</i>	DEL 138kbp	Létalité embryonnaire	(T. Cooper et al., 2013)
HH6	31/0	2,2	002149-9913	<i>SDE2</i>	SL	Létalité embryonnaire	(Fritz et al., 2018)
HH7	18/0	2,2	001830-9913	<i>CENPU</i>	DEL 4bp	Létalité embryonnaire	(Hozé et al., 2020)
HH13	22/2	4,1	001836-9913	<i>KIR2DS1</i>	NS	Mortalité juvénile	(Häffiger et al., 2022)
HH21	157/0	11	001908-9913	<i>NOTCH3</i>	DEL 3pb (inframe deletion)	Mortalité juvénile	(Häffiger et al., 2022)
HH25	20/5	4	002547-9913	<i>RIOX1</i>	DEL 30pb (inframe deletion)	?	(Häffiger et al., 2022)
HH35	31/7	4,8	002548-9913	<i>PCDH15</i>	MS (Leu>Val)	Mortalité périnatale	(Häffiger et al., 2022)
Jersey	JH1	23,4	001697-9913	<i>CWC15</i>	NS	Létalité embryonnaire	(VanRaden et al., 2011)
Montbéliarde	MH1	18	001827-9913	<i>PFAS</i>	MS (Arg>Cys)	Létalité embryonnaire	(Fritz et al., 2013)
	MH2	14	001828-9913	<i>SLC37A2</i>	NS	Létalité embryonnaire	(Fritz et al., 2013)

Normande	NH7	51/1	7,4	002201-9913	CAD	MS (Tyr>Cys)	Avortement tardif	(Mesbah-Uddin et al., 2019)
Original Braunvieh	OH2	20/0	9	002515-9913	TUBGCP5	MS (Thr>Lys)	Mortalité juvénile	(Häfliger, Seefried, Spengeler, & Drögemüller, 2021)
	OH4	15/0	7,8	002516-9913	LIG3	FS	Létalité embryonnaire	(Häfliger, Seefried, Spengeler, & Drögemüller, 2021)
Simmental	SH5	38/0	12,4	002505-9913	DIS3	FS	?	(Häfliger, Seefried, & Drögemüller, 2021)
	SH8	11/0	6,8	002508-9913	CYP2B6	MS (Ile>Asn)	?	(Häfliger, Seefried, & Drögemüller, 2021)
	SH9	19/2	8,6	002509-9913	NUBPL	MS (Ser>Tyr)	?	(Häfliger, Seefried, & Drögemüller, 2021)

A retenir Avant toute analyse, il est important de savoir si l'anomalie remontée est d'ordre génétique. Avec l'avènement des outils haut-débit (puces SNP, séquences tout génome), des approches de génétique classique (analyses d'association tout génome, cartographie par homozygotie) sont utilisées pour caractériser rapidement une anomalie génétique à partir de quelques cas. Ces outils offrent également la possibilité de rechercher des anomalies génétiques avant même leurs observations par des raisonnements de génétique inverse (déficit en homozygotes, recherche de variants rares). ■

5. Situation du sujet et objectifs de la thèse

L'avènement de la sélection génomique dans les différentes races de petits ruminants en sélection ouvre la possibilité de réaliser un raisonnement de génétique inverse pour identifier des mutations délétères par recherche de régions du génome en déficit d'animaux homozygotes. La synthèse bibliographique laisse penser que des mutations létales ségrégent dans les populations de petits ruminants ([Partie I](#)).

Avant mon travail de thèse, une étude préliminaire avait été faite en 2015 par Petit et al. (2015) pour rechercher de telles régions dans les populations d'ovins laitiers Lacaune (6 352 génotypages) et Manech Tête Rousse (1 589 génotypages). La méthode décrite par Fritz et al. (2013) avait été simplifiée pour estimer le nombre d'animaux homozygotes attendus avec l'utilisation de duos à la place de trios, car le nombre de génotypages disponibles était assez limité. Par ce travail, 3 haplotypes en Lacaune et 5 en Manech Tête Rousse avait été mis en évidence montrant bien la possibilité d'une telle approche dans les populations ovines. Un fait marquant de ce travail exploratoire a été la mise en évidence d'un haplotype en Lacaune en fort lien avec la mutation en sélection dans le gène *SOCS2* responsable d'une susceptibilité aux mammites (Rupp et al., 2015).

La mise en place de la sélection génomique en ovins (dès 2015 en Lacaune, 2017 en ROLP) et en caprins laitiers (2018) permet aujourd'hui d'avoir accès à un plus grand nombre de génotypages pour réaliser ce raisonnement de génétique inverse. Ainsi, dans ce manuscrit, je vais décrire les résultats obtenus durant mon stage de fin d'études d'ingénieur agronome (Mars à Août 2019) et de ma thèse (Octobre 2019 à Décembre 2022).

Ce travail de thèse s'articule autour de trois objectifs principaux. Le premier consiste à utiliser l'approche de génétique inverse dans les populations laitières de petits ruminants engagées dans un programme de sélection génomique pour détecter des régions en déficit d'animaux homozygotes, supposées abriter des mutations récessives létales. Le deuxième objectif est ensuite d'étudier l'effet de ces haplotypes sur les caractères de reproduction (afin de confirmer une possible létalité) et de production (identification d'un avantage sélectif potentiel des hétérozygotes). Le troisième et dernier objectif consiste à identifier les mutations causales sous-jacentes de ces régions en déficit.

Les résultats seront présentés en quatre parties : recherche de régions en déficit d'animaux homozygotes à partir des données de génotypages disponibles en petits ruminants ([Partie II](#)), effet des haplotypes sur les caractères ([Partie III](#)), recherche des mutations causales localisées dans les régions en déficit ([Partie IV](#)) et validation de la causalité des mutations identifiées ([Partie V](#)). L'ensemble des résultats sera discuté dans la discussion générale ([Partie VI](#)).

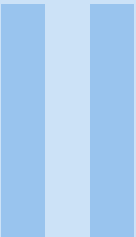
L'ensemble des résultats présentés repose sur deux articles publiés, deux articles rédigés et des résultats non publiés :

Article 1 (publié) - Ben Braiek et al., 2021 - *Genetics Selection Evolution* - **Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep.** doi : 10.1186/s12711-021-00634-1.

Article 2 (publié) - Ben Braiek et al., 2022 - *Genes* - **A Nonsense Variant in *CCDC65* Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep.** doi : 10.3390/genes13010045.

Article 3 - Ben Braiek et al., 2023a - bioRxiv, rédigé pour *Plos Genetics* - **Homozygous haplotype deficiency in Manech Tête Rouse dairy sheep revealed a nonsense variant in *MMUT* gene affecting newborn lamb viability.** doi : 10.1101/2023.03.10.531894.

Article 4 - Ben Braiek et al., 2023b - bioRxiv, rédigé pour *Animal genetics* - **A single base pair duplication in *SLC33A1* gene causes fetal losses and neonatal lethality in Manech Tête Rouse dairy sheep.** doi : 10.1101/2023.03.13.532360.



Identification d'haplotypes en déficit d'animaux homozygotes à partir des données de génotypages

6	Méthodes	111
6.1	Description des animaux génotypés	
6.2	Contrôle qualité, imputation et phasage des génotypages	
6.3	Identification des haplotypes en déficit d'animaux homozygotes	
7	Résultats : Identification de DHH en petits ruminants	117

L'ensemble des résultats décrits dans cette **Partie II**, repose sur les articles suivants (disponibles dans la [Partie VII - Productions scientifiques](#)) :

Article 1 (publié) - Ben Braiek et al., 2021 - *Genetics Selection Evolution* - **Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep.**

Article 3 - Ben Braiek et al., 2023a - bioRxiv, rédigé pour *Plos Genetics* - **Homozygous haplotype deficiency in Manech Tête Rousse dairy sheep revealed a nonsense variant in *MMUT* gene affecting newborn lamb viability.**

Des résultats complémentaires ont également été ajoutés.



6. Méthodes

6.1 Description des animaux génotypés

Les données de génotypages utilisées pour rechercher des régions en déficit d'animaux homozygotes sont extraites de la base de données nationale BDIR (Base de Données pour l'Indexation et la Recherche). Plusieurs extractions ont été réalisées au cours de ma thèse : Mars 2019 (Lacaune : 19 102 génotypages disponibles) ; Août 2021 (MTR : 6 845) et Septembre 2022 (Lacaune : 35 221, MTR : 7 405, BB : 2 464, MTN : 1 384 , Corse : 2 255, Alpine : 4 003, Saanen : 3 025).

Ces données correspondent à des animaux de millésimes 1991 à 2022 (Figure 6.1). En Lacaune lait, les animaux génotypés sont issus des schémas de deux entreprises de sélection : Ovitest (Saint-Léon, France) et Confédération Générale de Roquefort (Millau, France), qui représentent 54% et 46% de l'effectif de la race, respectivement. Depuis la création des deux schémas dans les années 1970, très peu d'échanges de reproducteurs sont effectués. Dans chacune des autres races, une seule entreprise de sélection est en charge de la sélection génétique (CDEO, CORSIA et Capgènes). Les données de génotypages ont été obtenues dans le cadre de différents programmes de recherche avant la mise en place la sélection génomique dans les différentes races, puis du programme en cours de sélection génomique. Le pedigree des animaux génotypés a été extrait du système officiel français de données sur l'élevage (Systèmes Nationaux d'Information Génétique, France Génétique Elevage, Paris, France).

Les données sont issues principalement de mâles (>80% pour les races ovines) tandis qu'en caprins on observe un équilibre entre mâles/femelles (Figure 6.1 A). De plus, les ovins ont été génotypés sur une puce SNP de moyenne densité (MD) (Illumina Ovine SNP50 BeadChip, 54 241 SNP, génotypés entre 1 et 12 mois d'âge) et/ou une puce SNP basse densité (LD) (SheepLD v.1-2-3, 15 000 SNP, génotypés entre 1 à 5 mois). Depuis la mise en place de la puce LD en 2017, les agneaux candidats à la sélection sont génotypés sur la puce LD puis suite à l'indexation génomique, les meilleurs entrent en CIA et sont génotypés sur puce MD (participent à l'entretien de la population de référence, voir paragraphe 1.2.3.2). En caprins, les animaux ne sont génotypés que sur une puce MD (GoatSNP50 Illumina v1, Goat_IGGC_65K_v2, génotypage vers 1 mois) (Figure 6.1 B). Les génotypages sont réalisés dans les laboratoires Labogena (<http://www.labogena.fr/>) ou Aveyron Labo (<http://www.aveyron-labo.com/>).

6.2 Contrôle qualité, imputation et phasage des génotypages

Le contrôle qualité de chaque individu génotypé repose sur le taux de SNP correctement génotypés (ou call rate), c'est-à-dire le pourcentage de SNP génotypés / nombre total de SNP, que ce soit au niveau du génome global et par chromosome. Si le typage global a un score inférieur

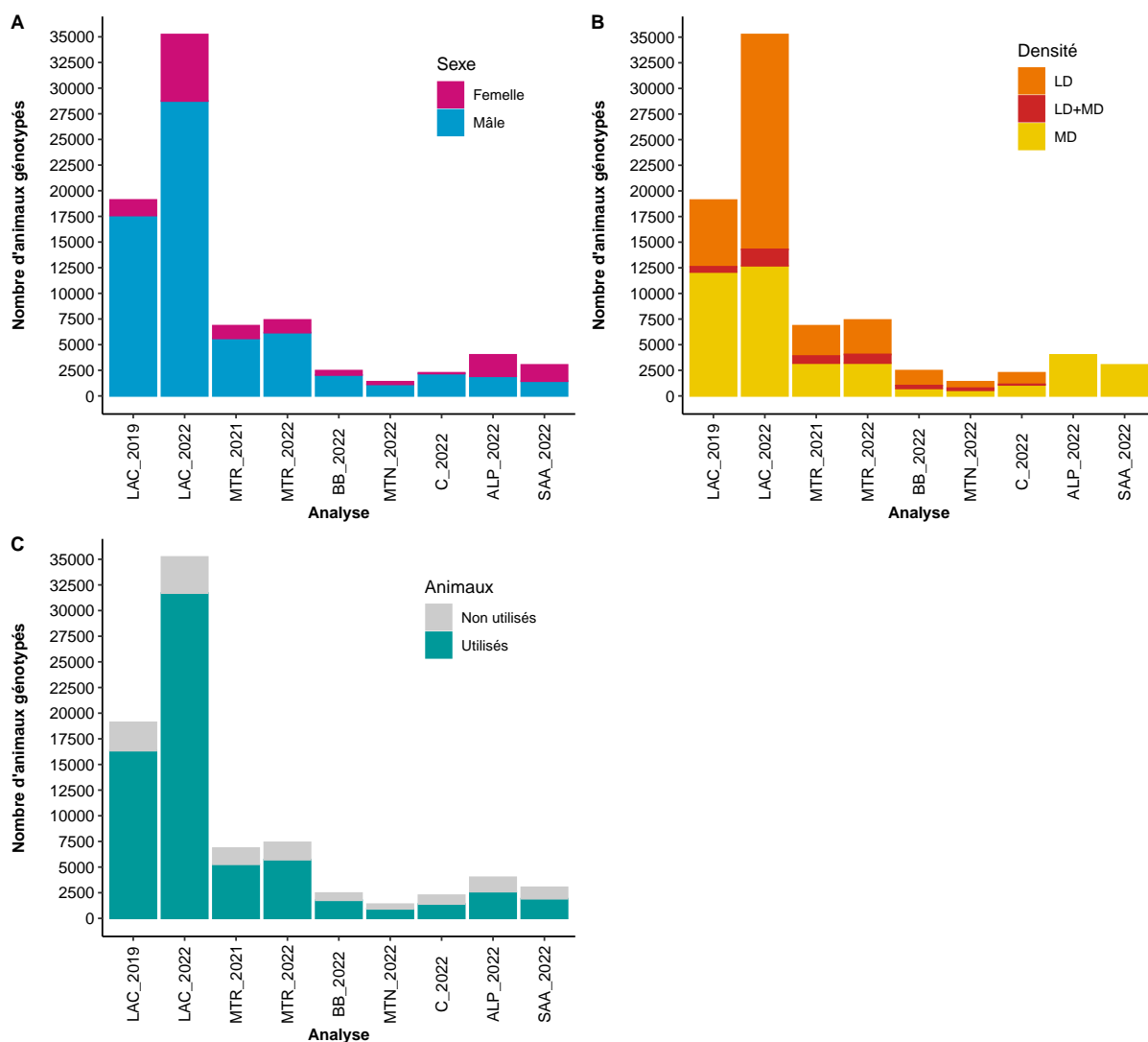


FIGURE 6.1 – Description des données de génotypage utilisées pour rechercher des régions du génome en déficit d'animaux homozygotes (A) Répartition du nombre d'animaux génotypés par sexe. (B) Répartition du nombre d'animaux génotypés par type de puce de génotypage. Les animaux génotypés LD+MD correspondent aux agneaux qui ont été génotypés LD (candidats à la sélection) puis génotypés MD en entrant en CIA. (C) Répartition du nombre d'animaux génotypés informatifs pour l'analyse de déficit. Un animal est utilisé lorsqu'il appartient à un trio d'individus génotypés (descendant, père et mère ou descendant, père et grand-père maternel). L'axe des abscisses correspond à la race et la date de l'analyse. LAC= Lacaune, MTR= Manech Tête Rousse, BB= Basco-Béarnaise, MTN= Manech Tête Noire, C= Corse, ALP= Alpine et SAA= Saanen, LD= puce basse densité 15k (*low density*), MD= puce moyenne densité 50K (*medium density*). Remarque : 2019 : extraction BDIR Mars 2019 ; 2021 : extraction BDIR Août 2021 et 2022 : extraction BDIR Septembre 2022.

à 95%, il est rejeté. Le contrôle qualité pour chacun des SNP est basé sur trois critères : (i) un call rate par SNP supérieur à 97% (% d'animaux génotypés pour chaque SNP), (ii) une fréquence de l'allèle mineur supérieure (MAF) à 1%, et (iii) respectant l'équilibre de Hardy-Weinberg ($P > 10^{-5}$) (Astruc et al., 2016a).

L'imputation des génotypes LD à MD et le phasage de tous les génotypes ont été réalisés avec le logiciel FImpute v2.2 (Sargolzaei et al., 2014) ; les puces LD et MD ovines ont 11 342 SNP communs. La précision de l'imputation LD à MD des génotypes en ovins laitiers a été évaluée par Larroque et Astruc (2017). En fonction des races, le call rate par animal varie de 96,56 (Corse) à 99,05% (Lacaune) et le call rate par SNP varie de 96,81 (corse) à 99,12% (Lacaune). Le coefficient de corrélation de Pearson au carré r^2 entre les génotypes SNP vrais et imputés est supérieur à 0,90 en Lacaune et ROLP tandis qu'en Corse ce taux est moins élevé (0,82). En caprins, il n'y a pas à proprement parler d'étape d'imputation puisqu'il n'existe qu'une puce MD.

Après contrôles qualités, il reste 38 696 SNP répartis sur les 26 chromosomes autosomiques et cartographiés sur l'assemblage du génome *Ovis aries* Oar_v2.0 ((Jiang et al., 2014), GigaDB, les coordonnées Oar_v2.0 sont disponibles sur <http://dx.doi.org/10.5524/100023>). A partir de Janvier 2021, les marqueurs (nombre réduit à 38 523 SNP) ont été mis à jour pour être cartographiés sur un assemblage plus récent Oar_v3.1 (les coordonnées génomiques sont disponibles sur https://figshare.com/articles/dataset/Mapping_of_ISGC_SNP_chip_probes/8424935/2). En caprins, 46 707 SNP disponibles après contrôles qualités, sont cartographiés sur la version du génome CHIR1 (Tosser-Klopp et al., 2014).

6.3 Identification des haplotypes en déficit d'animaux homozygotes

Sur la base des génotypes SNP phasés et imputés, les génomes ont été balayés en utilisant une approche de fenêtre glissante (Figure 6.2) de 20 SNP consécutifs pour identifier des haplotypes en déficit d'animaux homozygotes (appelés HHD : *Homozygous Haplotype Deficiency*) sur les 26 autosomes ovins et 29 autosomes caprins en comparant le nombre d'animaux homozygotes observés et attendus en utilisant la méthode développée par Fritz et al. (2013) chez les bovins laitiers français, et adaptée à mes données en petits ruminants.

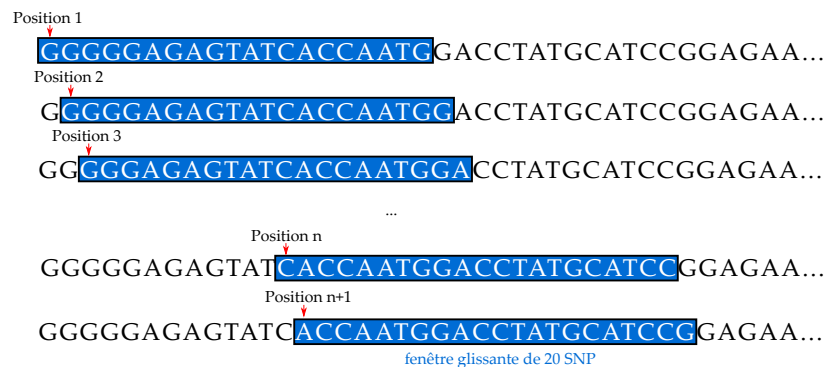


FIGURE 6.2 – Principe de la fenêtre glissante.

Pour chaque fenêtre de 20 SNP consécutifs (représentant des régions d'environ 1,0 à 1,5 Mb), les fréquences de tous les haplotypes observés sont calculées à partir de la phase maternelle, qui

est associée à une plus grande diversité d'haplotypes. Seuls les haplotypes avec une fréquence supérieure à 1% sont sélectionnés.

Pour chaque haplotype k sélectionné, le nombre d'animaux homozygotes observés $N_{Obs}(k)$ est comparé au nombre d'animaux homozygotes attendus $N_{Att}(k)$. Ce nombre $N_{Att}(k)$ est estimé en utilisant la probabilité de transmission intra-trio avec la formule décrite dans Fritz et al. (2013) (voir [Encadré 4. Calcul du nombre d'animaux homozygotes attendus \$E_{Att}\(k\)\$](#)). Deux types de trios sont considérés ([Figure 6.3](#)) : les trios descendance-père-mère (la probabilité de transmission de l'haplotype k est estimée à partir du génotype du père et de la mère) et les trios descendance-père-grand-père maternel (la probabilité de transmission de l'haplotype k est estimée à partir des génotypes du père et du grand-père maternel). Le nombre de génotypages pris en compte dans l'analyse de déficit est décrit dans la [Figure 6.1 C](#) et [Table 6.1](#).

Encadré 4. Calcul du nombre d'animaux homozygotes attendus $E_{Att}(k)$

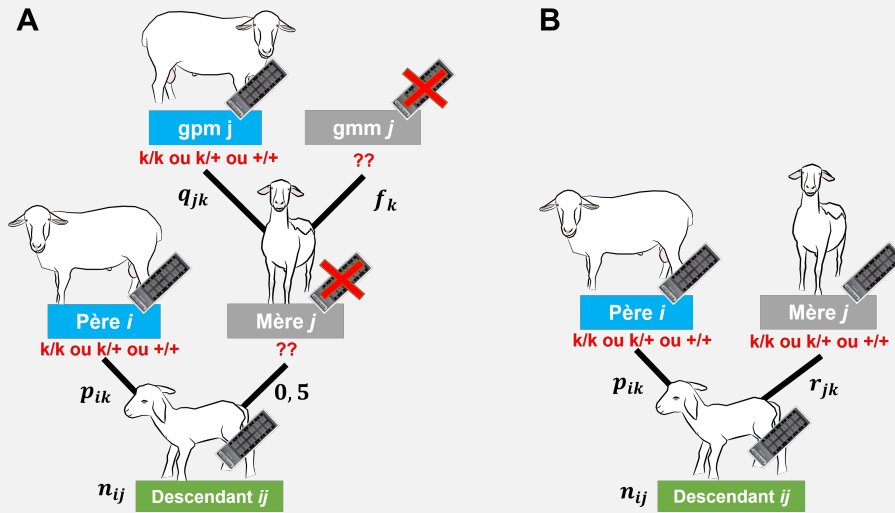


FIGURE 6.3 – Calcul du nombre d'homozygotes attendus à partir de trios génotypés

Dans le cas de trios génotypés descendant, père et grand-père maternel ([Figure 6.3 A](#)), la formule est :

$$E_{Att}(k) = \sum_{i=1}^{n_p} p_{ik} \sum_{j=1}^{n_{gpm}} 0.5 \left[\underbrace{(q_{jk}(1-f_k) + f_k(1-q_{jk}))}_{\text{gpm porteur}} + \underbrace{2q_{jk}f_k}_{\text{gpm et gmm porteurs}} \right] n_{ij} \quad (1)$$

(1) se simplifie en :

$$E_{Att}(k) = \sum_{i=1}^{n_p} p_{ik} \sum_{j=1}^{n_{gpm}} 0.5(q_{jk} + f_k) n_{ij} \quad (1')$$

Lorsque la mère est génotypée (trios descendant, père et mère) ([Figure 6.3 B](#)), la formule se simplifie et devient :

$$E_{Att}(k) = \sum_{i=1}^{n_p} p_{ik} \sum_{j=1}^{n_m} r_{jk} n_{ij} \quad (2)$$

avec,

- gmm = grand-mères maternelles,
- n_p = nombre de pères,
- n_m = nombre de mères,
- n_{gpm} = nombre de grand-pères maternels,
- p_{ik} = probabilité de transmission de l'haplotype k par le père i à ses descendants (1 si l'animal est porteur homozygote de l'haplotype k , 0.5 si l'animal est porteur hétérozygote de l'haplotype k et 0 si non porteur),
- q_{jk} = probabilité de transmission de l'haplotype k par le grand-père maternel j à ses filles (1 si l'animal est porteur homozygote de l'haplotype k , 0.5 si l'animal est porteur hétérozygote de l'haplotype k et 0 si non porteur),
- r_{ik} = probabilité de transmission de l'haplotype k par la mère i à ses descendants (1 si l'animal est porteur homozygote de l'haplotype k , 0.5 si l'animal est porteur hétérozygote de l'haplotype k et 0 si non porteur),
- f_k = fréquence de l'haplotype k dans la population,
- n_{ij} = nombre de descendants issus du père i et du grand-père maternel j pour (1,1') ou la mère j pour (2)

Remarque : Les formules (1') et (2) peuvent être combinées dans l'analyse de déficit.

La probabilité d'observer des homozygotes "q" avec une espérance "lambda" est estimée à l'aide de la distribution de Poisson et calculée avec la fonction `ppois` du logiciel RStudio (R version 4.1.3), $ppois(q = N_{Obs}(k), lambda = N_{Att}(k))$, comme décrit précédemment dans Mesbah-Uddin et al. (2019). Chaque haplotype k est supposé être significativement en déficit d'animaux homozygotes lorsque la P-valeur est inférieure à $1,9 \times 10^{-4}$ en ovins et $1,7 \times 10^{-4}$ en caprins. Ces seuils ont été obtenus par correction de Bonferroni pour les tests multiples à un seuil de significativité de 5‰ en supposant que le nombre de tests indépendants était égal au nombre de chromosomes ($n = 26$ en ovins et $n=29$ en caprins). Parmi les haplotypes significatifs, seuls ceux présentant un déficit important allant de 75 à 100%, sont retenus comme HHD, défini par :

$$\frac{N_{Att}(k) - N_{Obs}(k)}{N_{Att}(k)} \geq 0,75$$

Lorsque les HHD significatifs de 20 SNP sont consécutifs (c'est-à-dire décalés du précédent par un seul SNP) et ont le même nombre minimum d'animaux homozygotes observés, ils sont regroupés pour définir un haplotype plus grand (tous ces HHD primaires sont en déséquilibre de liaison les uns avec les autres), que nous appelons "deficient homozygous haplotype" (DHH) précédé du préfixe de la race étudiée (L : Lacaune, MTR : Manech Tête Rousse ...). Le statut homozygote, hétérozygote et non porteur de chaque haplotype constituant une région DHH est ensuite déterminé pour chaque animal de la population étudiée.

TABLE 6.1 – Description du nombre de génotypes utilisés dans les analyses de déficit. p= père, gpm=grand-père maternel.

Espèce	Race	Date extraction BDIR	Nombre de génotypes disponibles	Nombre de descendants p+m	Nombre de descendants p+gpm	Nombre de descendants dans l'étude	Nombre d'animaux utilisés dans l'étude
Ovins	LAC	Mars 2019 (Sept 2022)	19 102 (35 221)	38 (1 035)	15 530 (29 871)	15 568 (30 906)	16 346 (31 703)
	MTR	Août 2021 (Sept 2022)	6 845 (7 405)	77 (80)	4 799 (5 216)	4 876 (5 296)	5 271 (5 721)
	BB	Sept 2022	2 464	22	1 592	1 614	1 765
	MTN	Sept 2022	1 384	7	790	797	928
	Corse	Sept 2022	2 255	52	1 192	1 244	1 400
Caprins	Alpine	Sept 2022	4 003	826	1 240	2 066	2 606
	Saanen	Sept 2022	3 025	430	1 169	1 599	1 916

Le déséquilibre de liaison est estimé entre deux régions DHH présentes sur un même chromosome par la mesure du coefficient r^2 introduit par Hill et Robertson (1968). Chaque région DHH est définie comme un locus bi-allélique avec un allèle 1 (haplotype en déficit, DHH) et un allèle 2 (tous les autres haplotypes identifiés pour la même région, +). En ovins, les coordonnées des SNP inclus dans chaque haplotype, obtenues à partir des positions sur les assemblages des génomes ovins Oar_v2.0 (extraction BDIR avant 2021) ou Oar_3.1 (extraction BDIR après 2021), sont repositionnées sur l'assemblage plus récent du génome Oar_rambouillet_v1.0 pour la suite des analyses génétiques.

7. Résultats : Identification de DHH en petits ruminants

En utilisant cette approche de génétique inverse, 266 haplotypes en déficit d'animaux homozygotes (HHD) de 20 marqueurs en Lacaune (analyse de déficit en Mars 2019), 150 HHD en Manech Tête Rousse (analyse de déficit en Août 2021) et 5 HHD en Basco-Béarnaise (analyse de déficit en Septembre 2022) ont été mis en évidence (fréquence $\geq 1\%$, déficit $\geq 75\%$ et P-Poisson $\leq 1,9 \times 10^{-4}$) (Table 7.1). Cependant, aucun HHD n'est significatif en races laitières ovines Manech Tête Noire et Corse et en races caprines Alpine et Saanen. Une représentation sous forme de Manhattan plot est proposée dans la Figure 7.1. L'ensemble des HHD peut être regroupé en 11 régions en Lacaune (LDHH, "Lacaune déficient homozygous haplotype"), 5 régions en Manech Tête Rousse (MTRDHH, "Manech Tête Rousse déficient homozygous haplotype") et 1 région en Basco-Béarnaise (BBDHH, "Basco-Béarnaise déficient homozygous haplotype") (Table 7.3). Les DHH sont construits à partir de HHD chevauchants qui présentent le même nombre d'animaux homozygotes observés (Figure 7.2).

TABLE 7.1 – **Nombre de HHD identifiés en petits ruminants.** Lac= Lacaune, MTR= Manech Tête Rousse, BB= Basco-Béarnaise, MTN= Manech Tête Noire, ALP= Alpine, SAA= Saanen. Les HHD sont des haplotypes de 20 marqueurs SNP.

Date extraction BDIR	Lac	MTR	BB	MTN	Corse	ALP	SAA
Mars 2019	266	-	-	-	-	-	-
Août 2021	-	150	-	-	-	-	-
Septembre 2022	311	175	5	0	0	0	0

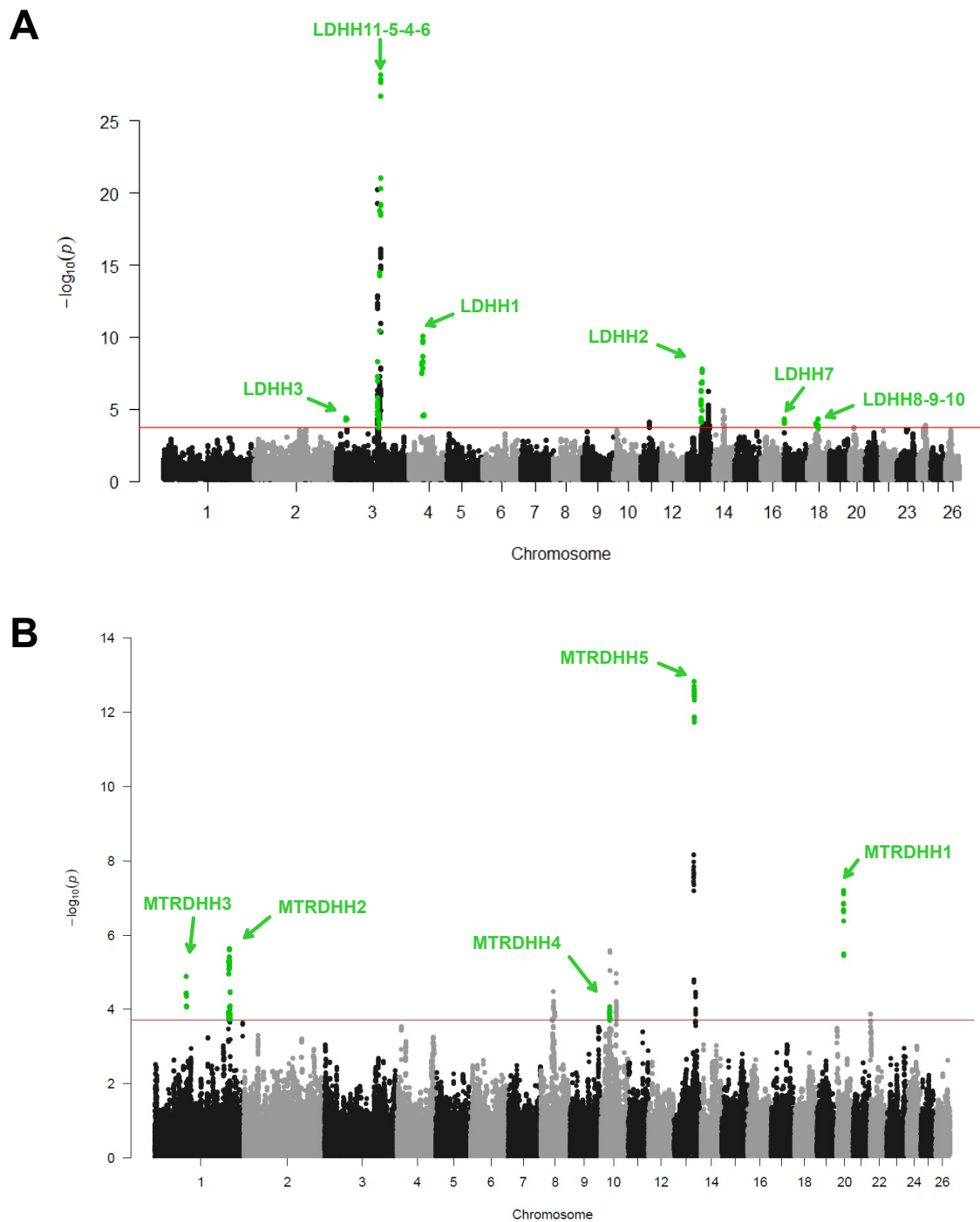


FIGURE 7.1 – Représentation sous forme de Manhattan plot des HHD identifiés en races ovines laitières Lacaune (A) et Manech Tête Rouge (B). Chaque point représente un haplotype de 20 marqueurs avec une fréquence $\geq 1\%$. La ligne rouge représente le seuil de P-valeur ($1,9 \times 10^{-4}$) pour considérer un haplotype significativement en déficit d'animaux homozygotes. Seuls les HHD avec un déficit en homozygotes $\geq 75\%$ (représentés par des points verts) ont été sélectionnés et ont abouti à l'identification de 266 et 150 HHD significatifs regroupés dans 11 (LDHH1 à 11) et 5 régions (MTRDHH1 à 5) en Lacaune et Manech Tête Rouge, respectivement. Les coordonnées génomiques sont basées sur les génomes de référence ovins Oar_v2.0 en Lacaune et Oar_v3.1 en Manech Tête Rouge.

TABLE 7.3 – Liste des haplotypes en déficit d’animaux homozygotes identifiés en ovins laitiers.. En Lacaune 11 LDHH (1 à 11) ont été identifiés par l’analyse de déficit en Mars 2019 (19 102 génotypes, version génome Oar_v2.0), 5 MTRDHH ont été identifiés par l’analyse de déficit en Août 2021 (6 845 génotypes, version génome Oar_v3.1) et 1 BBDHH (BBDHH1) a été identifié par l’analyse de déficit en septembre 2022 (2 464 génotypes, version génome Oar_v3.1). Les valeurs entre parenthèses correspondent à la mise à jour des caractéristiques des haplotypes identifiés par la nouvelle analyse de déficit de Septembre 2022 dans les races ovines Lacaune et Manech Tête Rousse.

¹ Nombre de marqueurs composant les haplotypes.

² Les positions sont données sur la version du génome ovins Oar_rambouillet_v1.0.

² Fréquence de porteurs hétérozygotes dans l’ensemble de la population génotypée.

⁴ Nombre d’animaux homozygotes attendus.

⁵ Nombre d’animaux homozygotes observés.

(-) Haplotype non retrouvé par la nouvelle recherche de déficit de Septembre 2022 dans les races Lacaune et Manech Tête Rousse.

Haplotype	OAR	Nmarq ¹	Pos. début ²	Pos. fin ²	Taille (Mb)	Fréq. He ³ (%)	NAAtt ⁴	Nobs ⁵	Déficit	P-Poisson
LDHH1	4	46	49 015 021	52 306 310	3,3	6,7 (6,8)	21 (44)	0	100%	$7,6 \times 10^{-10}$ ($7,8 \times 10^{-20}$)
LDHH2	13	28	47 202 438	49 206 217	2,0	6,2 (6,0)	17 (33)	0	100%	$4,1 \times 10^{-8}$ ($4,7 \times 10^{-15}$)
LDHH3	3	48	33 708 561	36 596 529	2,9	4,3 (3,5)	10 (12)	0 (1)	100% (92%)	$4,5 \times 10^{-5}$ ($8,0 \times 10^{-5}$)
LDHH4	3	24	142 047 396 (141 279 184)	143 599 873 (143 118 180)	1,6 (1,8)	3,8 (3,0)	10 (11)	0	100%	$4,5 \times 10^{-5}$ ($1,7 \times 10^{-5}$)
LDHH5	3	29	140 770 458 (-)	142 399 985 (-)	1,6 (-)	3,7 (-)	9 (-)	0 (-)	100% (-)	$1,2 \times 10^{-4}$ (-)
LDHH6	3	27	146 243 481	147 946 399	1,7	12,1 (11,9)	72 (140)	3 (5)	96%	$3,5 \times 10^{-27}$ ($7,3 \times 10^{-53}$)
LDHH7	17	29	621 608 (-)	2 173 903 (-)	1,6 (-)	4,7 (-)	12 (-)	1 (-)	92% (-)	$8,0 \times 10^{-5}$ (-)

LDHH8	18	23	24 624 301 (-)	26 429 906 (-)	1,8 (-)	5,5 (-)	14 (-)	2 (-)	86% (-)	$9,4 \times 10^{-5}$ (-)
LDHH9	18	37	30 211 568 (30 982 269)	32 489 999 (33 208 337)	2,3 (2,2)	4,4 (4,3)	13 (23)	2	85% (91%)	$2,2 \times 10^{-4}$ ($3,0 \times 10^{-8}$)
LDHH10	18	26	31 985 974 (-)	33 989 971 (-)	2,0 (-)	5,7 (-)	17 (-)	3 (-)	82% (-)	$4,1 \times 10^{-5}$ (-)
LDHH11	3	28	138 443 834 (138 726 532)	140 585 367 (141 216 374)	2,1 (2,5)	7,2 (5,2)	19 (23)	4 (5)	79% (78%)	$3,8 \times 10^{-5}$ ($6,9 \times 10^{-6}$)
MTRDHH1	20	32	22 996 023	24 950 669	2,0	9,7 (9,4)	13	0	100%	$2,3 \times 10^{-6}$
MTRDHH2	1	66	251 858 291	256 412 373	4,6	8,7 (8,8)	10 (11)	0	100%	$4,5 \times 10^{-5}$ ($1,7 \times 10^{-5}$)
MTRDHH3	1	39	103 828 253	106 639 185	2,8	7,8 (7,9)	9 (10)	0	100%	$1,2 \times 10^{-4}$ ($4,5 \times 10^{-5}$)
MTRDHH4	10	26 (30)	30 476 913 (30 561 021)	31 527 996 (32 325 300)	1,1 (1,8)	8,7 (8,4)	11	1	91%	$2,0 \times 10^{-4}$
MTRDHH5	13	29	64 305 288	67 234 998	2,9	16,6 (16,1)	49 (51)	8	84%	$5,1 \times 10^{-13}$ ($9,5 \times 10^{-14}$)
BBDHH1	10	33	88 743 412	90 499 204	1,8	12,2	9	0	100%	$1,2 \times 10^{-4}$

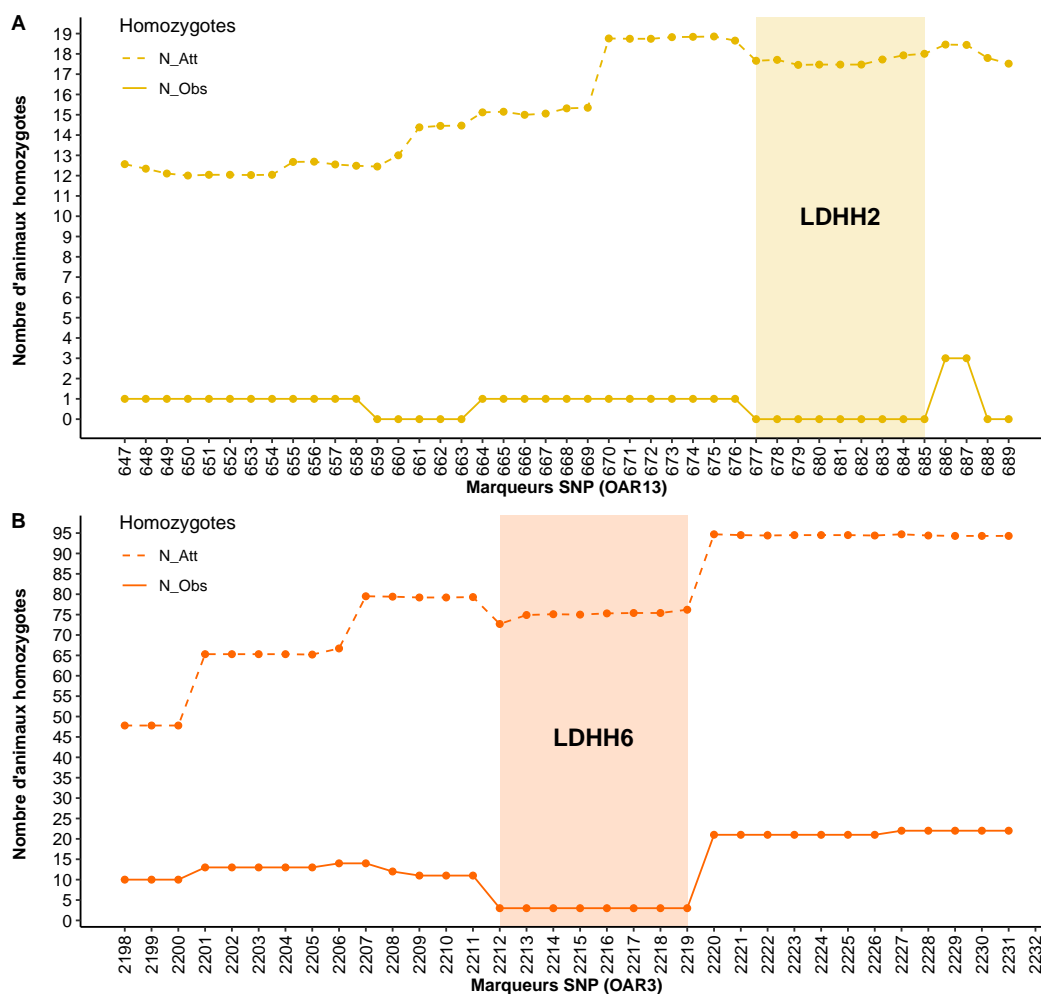


FIGURE 7.2 – Exemples de construction des DHH à partir des HHD chevauchants. Le numéro des marqueurs en abscisse représente le numéro du premier marqueur du HHD (haplotype de 20 SNP). La construction de ces DHH repose sur l'analyse de Mars 2019. N_Att= Nombre d'animaux homozygotes attendus, N_Obs= Nombre d'animaux homozygotes observés. (A) Construction de l'haplotype LDHH2 qui présente un déficit total. Sur le chromosome OAR13, 43 HHD présentent chacun un déficit d'animaux homozygotes. Parmi ces haplotypes, 20 HHD sont chevauchants et présentent un déficit total avec zéro animaux homozygotes observés, ils sont alors regroupés en un haplotype de plus grande taille (LDHH2, composé de 28 SNP). (B) Construction de l'haplotype LDHH6 qui présente un déficit partiel. Les mêmes règles de regroupement s'appliquent. Parmi les 38 HHD de la région identifiée sur OAR3, 8 HHD sont chevauchants et présentent un déficit partiel avec le même minimum d'animaux homozygotes observés (N_Obs=3), ils sont alors regroupés en un haplotype de plus grande taille (LDHH6, composé de 27 SNP).

En Lacaune, l'analyse de Mars 2019 (19 102 génotypages disponibles dont 16 346 animaux appartiennent à des trios) a permis de mettre en évidence 11 LDHH (Table 7.3). Parmi ces haplotypes, cinq LDHH présentent un déficit complet d'animaux homozygotes observés (LDHH1 à 5) et six LDHH présentent un déficit partiel allant de 79 à 96% du nombre d'animaux homozygotes attendus (LDHH6 à 11). La longueur des haplotypes identifiés varie de 1,6 à 3,3 Mb sur le génome Oar_rambouillet_v1.0.

Les haplotypes LDHH3, 4, 5, 6 et 11 sont tous localisés sur OAR3. Sur ce chromosome, seuls les haplotypes LDHH4 (24 SNP) et LDHH5 (29 SNP) sont en déséquilibre de liaison important (71%). Bien qu'ils partagent neuf SNP à leurs extrémités, LDHH4 et LDHH5 n'ont pas été regroupés à

l'origine puisque certains HDD consécutifs de 20 SNP entre les deux LDHH ne suivaient pas notre règle de regroupement (non significatifs, un nombre minimum d'animaux homozygotes différent). Sur OAR18, LDHH8 est situé à 3,8 Mb de LDHH9 et à 5,6 Mb de LDHH10, et présente un déséquilibre de liaison modéré de 55% et 40% avec ces deux haplotypes, respectivement. Par contre, LDHH9 (37 SNP) et LDHH10 (26 SNP) sont eux en déséquilibre de liaison élevé (72%). Ils partagent 11 SNP, mais n'ont pas été regroupés initialement pour la même raison que LDHH4 et 5. D'autres LDHH sont situés sur OAR4 (LDHH1), OAR13 (LDHH2) et OAR17 (LDHH7).

De nouveaux génotypages alimentent constamment la base de données. Une nouvelle analyse de déficit a été conduite en Septembre 2022 (35 221 génotypages disponibles dont 31 703 animaux appartiennent à des trios). Cela a permis de consolider les haplotypes LDHH1, LDHH2, LDHH3, LDHH6 (même marqueurs qui composent les haplotypes) et d'affiner la localisation des haplotypes LDHH4, LDHH9 et LDHH11 (Table 7.3). Comme le montre la Figure 7.3, les haplotypes LDHH4 et 5 qui étaient en déséquilibre de liaison, ne signent plus qu'une seule région génomique (LDHH4, OAR3 : 141,3-143,1 Mb, composée de 29 marqueurs) et LDHH8, 9 et 10 signent également une unique région (LDHH9, OAR18 : 31,0-33,2 Mb, composée de 36 marqueurs). Les coordonnées de l'haplotype LDHH11 sont légèrement modifiées (OAR3 : 138,7-141,2 Mb, composée de 33 marqueurs). Cependant, cette nouvelle analyse ne permet pas de retrouver l'haplotype LDHH7 (OAR17) préalablement identifié. Il s'agit probablement d'un faux-positif détecté dans l'analyse de Mars 2019, il ne sera donc pas considéré pour les analyses génétiques suivantes.

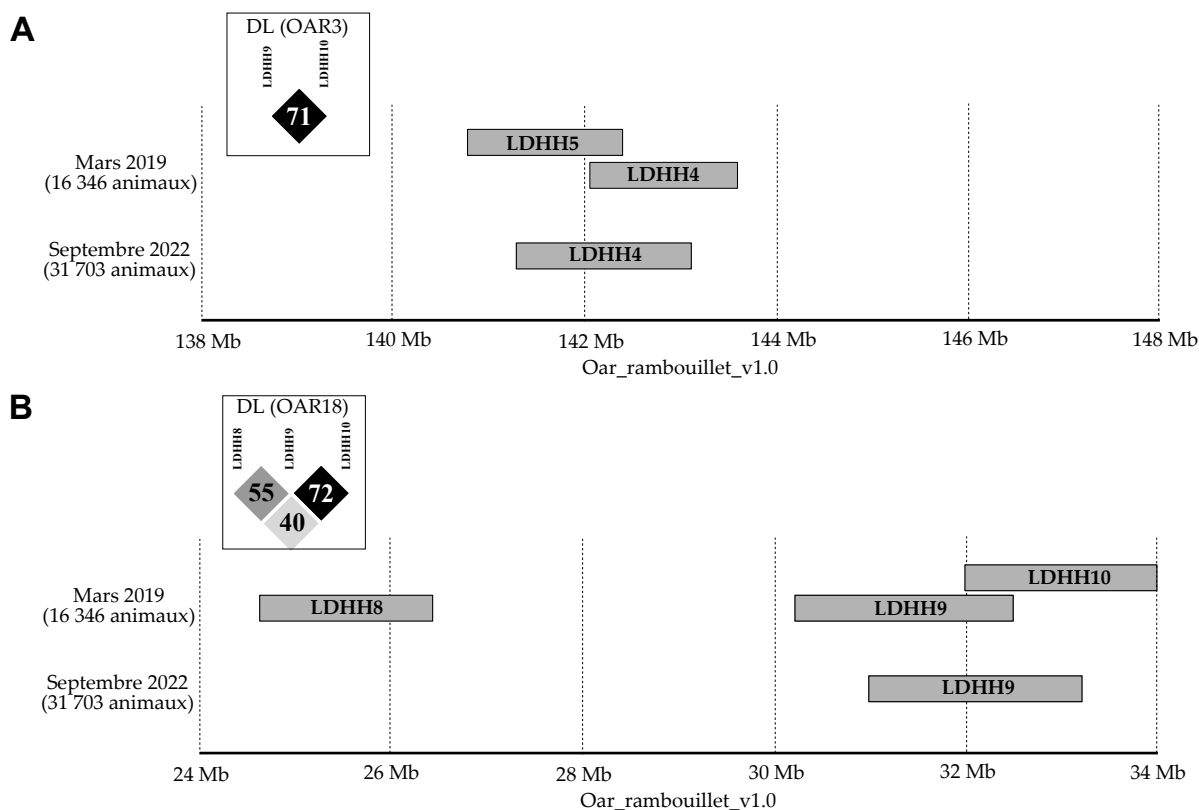


FIGURE 7.3 – Affinage de la cartographie des haplotypes LDHH4-5 (OAR3) et LDHH8-9-10 (OAR18)

Pour la suite en Lacaune, sept régions génomiques indépendantes supposées contenir des mutations délétères sont considérées, localisées sur OAR3 (LDHH3 : 33,7-36,6 Mb ; LDHH11 :

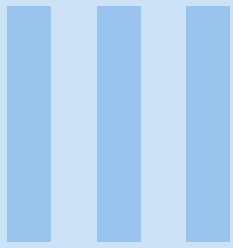
138,7-141,2 Mb ; LDHH4 : 141,3-143,1 Mb et LDHH6 : 146,2-147,9 Mb), OAR4 (LDHH1 : 49,0-52,3 Mb), OAR13 (LDHH2 : 47,2-49,2 Mb) et OAR18 (LDHH9 : 31,0-33,2 Mb) de la version du génome Oar_rambouillet_v1.0. A partir de l'ensemble de la population génotypée (n=35 221), la fréquence des porteurs hétérozygotes varie de 3,0 à 6,8% pour les LDHH avec un déficit total d'homozygotes, et de 4,3 à 11,9% pour les LDHH avec un déficit partiel.

En Manech Tête Rousse, l'analyse d'Août 2021 (6 845 génotypages disponibles dont 5 271 animaux appartiennent à des trios) a permis de mettre en évidence 5 MTRDHH. Trois haplotypes sont en déficit total d'animaux homozygotes (MTRDHH1, 2 et 3), alors que deux haplotypes, MTRDHH4 et 5, présentent un déficit partiel (84% et 91%) avec 1 et 8 animaux homozygotes génotypés alors que 11 et 49 sont attendus, respectivement. Les différents MTRDHH sont localisés sur OAR20 (MTRDHH1 : 23,0-25,0 Mb), OAR1 (MTRDHH2 : 251,9-256,4 Mb et MTRDHH3 : 103,8-106,6 Mb), OAR10 (MTRDHH4 : 30,5-31,6 Mb) et OAR13 (MTRDHH5 : 64,3-67,2 Mb). Les longueurs des haplotypes varient de 1,1 à 4,6 Mb sur Oar_rambouillet_v1.0. Les fréquences observées de porteurs hétérozygotes sont comprises entre 7,8% et 16,6%. MTRDHH2 et MTRDHH3 sont tous les deux situés sur OAR1, mais ne sont pas en déséquilibre de liaison. Par conséquent, les cinq MTRDHH identifiés sont susceptibles d'héberger cinq mutations causales indépendantes.

Une nouvelle analyse de déficit a été conduite en Septembre 2022 (7 405 génotypages disponibles dont 5 721 animaux appartiennent à des trios). Cela a également permis de consolider les haplotypes MTRDHH1, 2, 3 et 5. Cependant, les coordonnées de l'haplotype MTRDHH4 sont légèrement modifiées (OAR10 : 30,8-32,2 Mb, composé de 30 marqueurs).

En race Basco-Béarnaise, l'analyse de déficit de Septembre 2022 (2 464 génotypages disponibles dont 1 765 animaux appartiennent à des trios) a permis de mettre en évidence un unique haplotype BBDHH1 (OAR10 : 88,7-90,5 Mb) en déficit total (9 animaux homozygotes attendus vs 0 observé). La fréquence de porteurs hétérozygotes de l'haplotype est estimée à 12,2%.

A retenir Les analyses de déficit en homozygotes ont permis de mettre en évidence 7 LDHH en Lacaune, 5 MTRDHH en Manech Tête Rousse et 1 BBDHH en Basco-Béarnaise. On fait l'hypothèse que chacune de ces régions est en déséquilibre de liaison avec une mutation délétère. L'accumulation de données de génotypages exploitables (trios) permet d'assurer la découverte d'haplotypes en déficit avec une meilleure spécificité. ■



Effets des haplotypes sur les caractères

8	Fertilité et mortalité	129
8.1	Méthodes	
8.2	Résultats : confirmation de la létalité	
9	Caractères en sélection	135
9.1	Méthodes	
9.2	Résultats : potentiel avantage sélectif des animaux hétérozygotes aux DHH	
10	Évolution des fréquences des haplotypes	141
10.1	Évolution de la fréquence des LDHH en Lacaune	
10.2	Évolution de la fréquence des MTRDHH en Manech Tête Rousse	

L'ensemble des résultats décrits dans cette **Partie III**, repose sur les articles suivants (disponibles dans la [Partie VII - Productions scientifiques](#)) :

Article 1 (publié) - Ben Braiek et al., 2021 - *Genetics Selection Evolution* - **Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep.**

Article 3 - Ben Braiek et al., 2023a - bioRxiv, rédigé pour *Plos Genetics* - **Homozygous haplotype deficiency in Manech Tête Rousse dairy sheep revealed a nonsense variant in *MMUT* gene affecting newborn lamb viability.**

Des résultats complémentaires ont également été ajoutés.



8. Fertilité et mortinatalité

L'analyse des caractères de fertilité a été faite avec les 11 LDHH identifiés par l'analyse de déficit de Mars 2019 en Lacaune et les 5 MTRDHH identifiés par l'analyse d'août 2021 en MTR. Ces haplotypes sont supposés porter des mutations récessives létales. En effet, les déficits d'homozygotes s'expliqueraient par la mort précoce des animaux (stade embryonnaire, foetal, néonatal ou juvénile) avant l'âge du génotypage (Figure 8.1). Cependant, ces déficits pourraient également être liés à des défauts morphologiques ou une non-conformité aux standards de la race qui amèneraient à ne pas génotyper de tels animaux. Afin d'approcher ces hypothèses de létalité ou de choix par l'homme (contre-sélection), j'ai analysé les caractères de fertilité et de mortinatalité, qui devraient permettre de mettre en évidence des effets négatifs dans les accouplements entre individus porteurs de l'haplotype où des descendants homozygotes sont générés. J'appellerai ces accouplements "à risque". La base de données des inséminations artificielles issues des brebis en CLO a été mobilisée dans cet objectif.

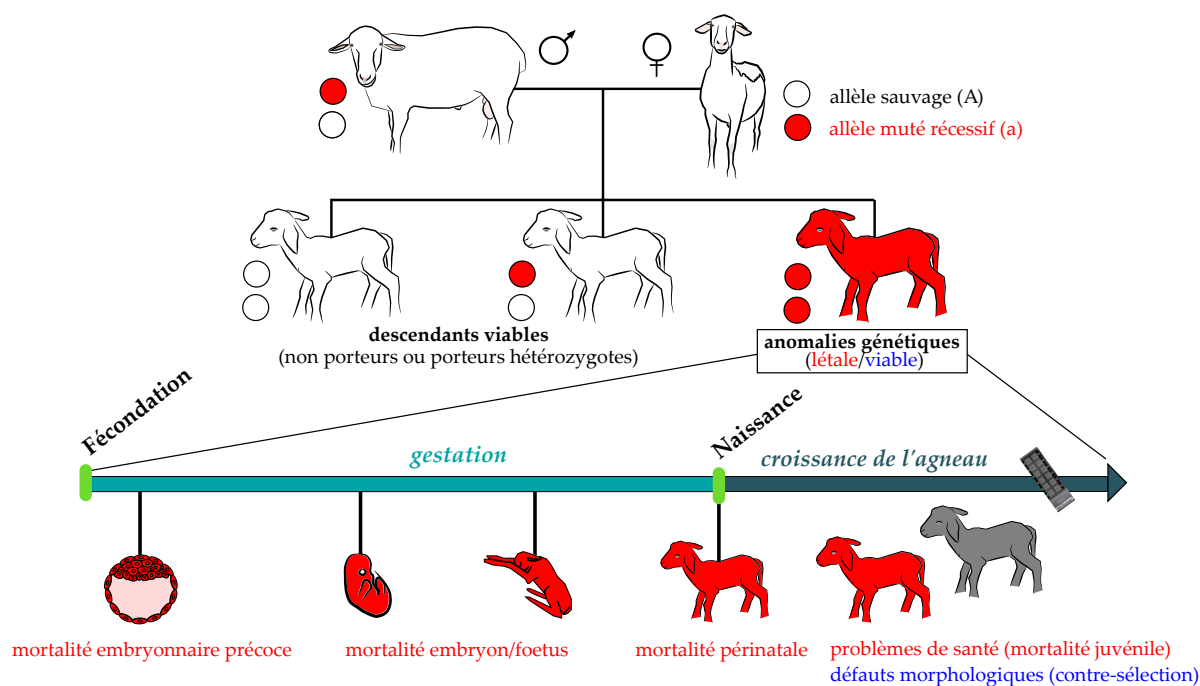


FIGURE 8.1 – Causes possibles de déficit en homozygotes.

8.1 Méthodes

Les données de deux caractères de fertilité sur une vingtaine d'années ont été extraites de la base nationale : le succès à l'insémination artificielle (AIS, *artificial insemination success*) et le

TABLE 8.1 – Paramètres du modèle pour l'analyse des caractères de fertilité. AIS= *artificial insemination success*, SBR= *stillbirth rate* et L=Lactation.

Race	Caractère (nombre d'accouplements)	Effet fixe testé	Effets fixes	Effets aléatoires
Lacaune	AIS (n= 1 155 835)	Type accouplements (Risque, Non à risque)	Mois IA (Mars à Septembre); Numéro de lactation (L1 à L4+)	Troupeau X Année (470 troupeaux entre 2006 et 2018)
	SBR (n= 804 577)	Type accouplements (Risque, Non à risque)	Mois IA (Mars à Septembre); Numéro de lactation (L1 à L4+); Prolificité (1,2, 3+)	Troupeau X Année (470 troupeaux entre 2006 et 2018)
MTR	AIS (n= 330 844)	Type accouplements (Risque, Non à risque)	Mois IA (Printemps, Eté); Numéro de lactation (L1 à L4+)	Troupeau X Année (313 troupeaux entre 2006 et 2019)
	SBR (n= 201 637)	Type accouplements (Risque, Non à risque)	Mois IA (Printemps, Eté); Numéro de lactation (L1 à L4+); Prolificité (1,2, 3+)	Troupeau X Année (313 troupeaux entre 2006 et 2019)

taux de mortalité (SBR, *stillbirth rate*). Parmi tous les enregistrements, je me suis concentré uniquement sur les accouplements entre brebis issues de père génotypé et inséminées avec un bélier génotypé, le père et le bélier d'IA ayant tous les deux un statut connu à chaque DHH (Table 8.1). L'AIS est codé "1" pour la réussite à l'IA et "0" pour l'échec en comparant la date d'agnelage et la durée théorique de la gestation à partir du jour de l'IA (147 ± 5 jours en Lacaune et 151 ± 7 jours en MTR). Le SBR est déterminé uniquement dans le groupe des brebis gestantes (Réussite IA = 1) et codé "1" s'il y a au moins un mort-né dans la portée ou "0" si tous les agneaux sont nés vivants. Quatre types d'accouplements différents sont possibles pour chaque DHH (Figure 8.2) : (i) bélier porteur \times brebis issue d'un père porteur, (ii) bélier non porteur \times brebis issue d'un père porteur, (iii) bélier porteur \times brebis issue d'un père non porteur et (iv) bélier non porteur \times brebis issue d'un père non porteur. Le type d'accouplement (i) est considéré comme un accouplement à risque, et les types d'accouplements (ii), (iii) et (iv) sont considérés comme non à risque.

Un modèle binaire de seuil logistique avec une fonction de lien logit a été utilisé pour comparer l'AIS et le SBR entre les accouplements à risque et non à risque en utilisant la fonction GLMIX du logiciel SAS (version 9.4; SAS Institute Inc., Cary, NC). Le modèle utilisé est $y = X\beta + Z\gamma + e$, où y est un vecteur de "0" ou "1" codant pour AIS ou SBR en considérant les observations correspondantes Z ($Z = \{0, 1\}$ pour AIS, $Z = \{0, 1, 2+\}$ pour SBR) et un seuil pour chaque variable, de sorte que $y = 1$ si $Z \geq 1$ ou $y = 0$ "sinon"; X est la matrice d'incidence des effets

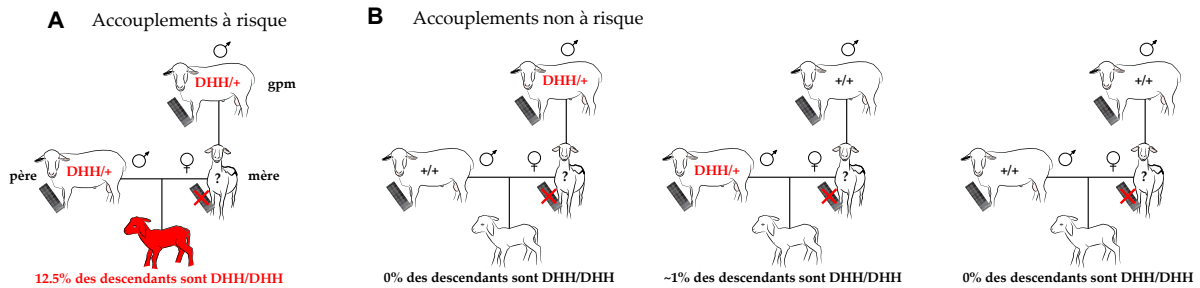


FIGURE 8.2 – **Accouplements à risque (A) vs non à risque (B)**. Les animaux DHH/+ sont porteurs hétérozygotes de l'haplotype en déficit. Les brebis ne sont pas génotypées. gpm= grand-père maternel.

fixes; β est un vecteur d'effets fixes; Z est une matrice d'incidence d'effets aléatoires; γ est un vecteur d'effets aléatoires et e est un vecteur d'effets résiduels. Les effets fixes et aléatoires du modèle sont décrits dans la [Table 8.1](#).

Les caractères sont considérés significativement différents lorsque l'effet fixe "type d'accouplement" a une valeur $P \leq 6,3 \times 10^{-4}$ en Lacaune et $P \leq 1,0 \times 10^{-2}$ en MTR, après correction de Bonferroni pour les tests multiples avec un niveau de significativité α à 1% en Lacaune et 5% en MTR. Ces seuils ont été obtenus en divisant le niveau de signification α par le nombre de tests correspondant au nombre d'haplotypes indépendants \times nombre de caractères testés ($n=8$ en Lacaune, $n=2$ caractères) et uniquement par le nombre d'haplotypes indépendants en MTR ($n=5$).

8.2 Résultats : confirmation de la létalité

Les deux caractères étudiés nous permettent d'émettre des hypothèses sur le moment de la létalité des différentes DHH. En effet, l'AIS peut être un bon indicateur des pertes embryonnaires au cours des premières semaines suivant l'IA tandis que le SBR peut être un bon indicateur pour identifier des DHH qui causent de la létalité périnatale.

En Lacaune, le taux de réussite à l'IA moyen est de 69,6%. L'analyse des accouplements à risque met en évidence pour quatre LDHH (1, 2, 8 et 9) un effet négatif significatif sur l'AIS, entraînant une diminution du taux de réussite allant de 2,2% pour LDHH1 ($P = 3,2 \times 10^{-6}$) à 3,2% pour LDHH2 ($P = 2,6 \times 10^{-9}$). Dans la population gestante, 5,1% des portées comportent au moins un mort-né. L'analyse du SBR met en évidence cinq LDHH (3, 6, 8, 9 et 10) significativement associés à une augmentation du taux de mortinatalité allant de 1,1% pour LDHH6 ($P = 5,9 \times 10^{-7}$) à 3,4% pour LDHH9 ($P = 8,5 \times 10^{-12}$) en comparant les accouplements à risque et non à risque.

En Manech Tête Rousse, le taux de réussite à l'IA moyen est de 60,9%. En comparant les accouplements à risque et non à risque, seul MTRDHH2 montre une diminution significative de -3,3% du taux de réussite à l'IA ($P = 3,5 \times 10^{-4}$) dans les accouplements à risque. Pour le second caractère étudié, le taux de mortinatalité moyen dans la population est de 7,5%. Les MTRDHH1 et 2 montrent une augmentation importante avec +7,5% ($P = 4,0 \times 10^{-24}$) et +4,3% ($P = 1,3 \times 10^{-6}$) de mort-nés dans les accouplements à risque par rapport aux accouplements non à risque, respectivement. Les trois autres haplotypes n'ont aucun impact significatif sur les

caractères de fertilité étudiés.

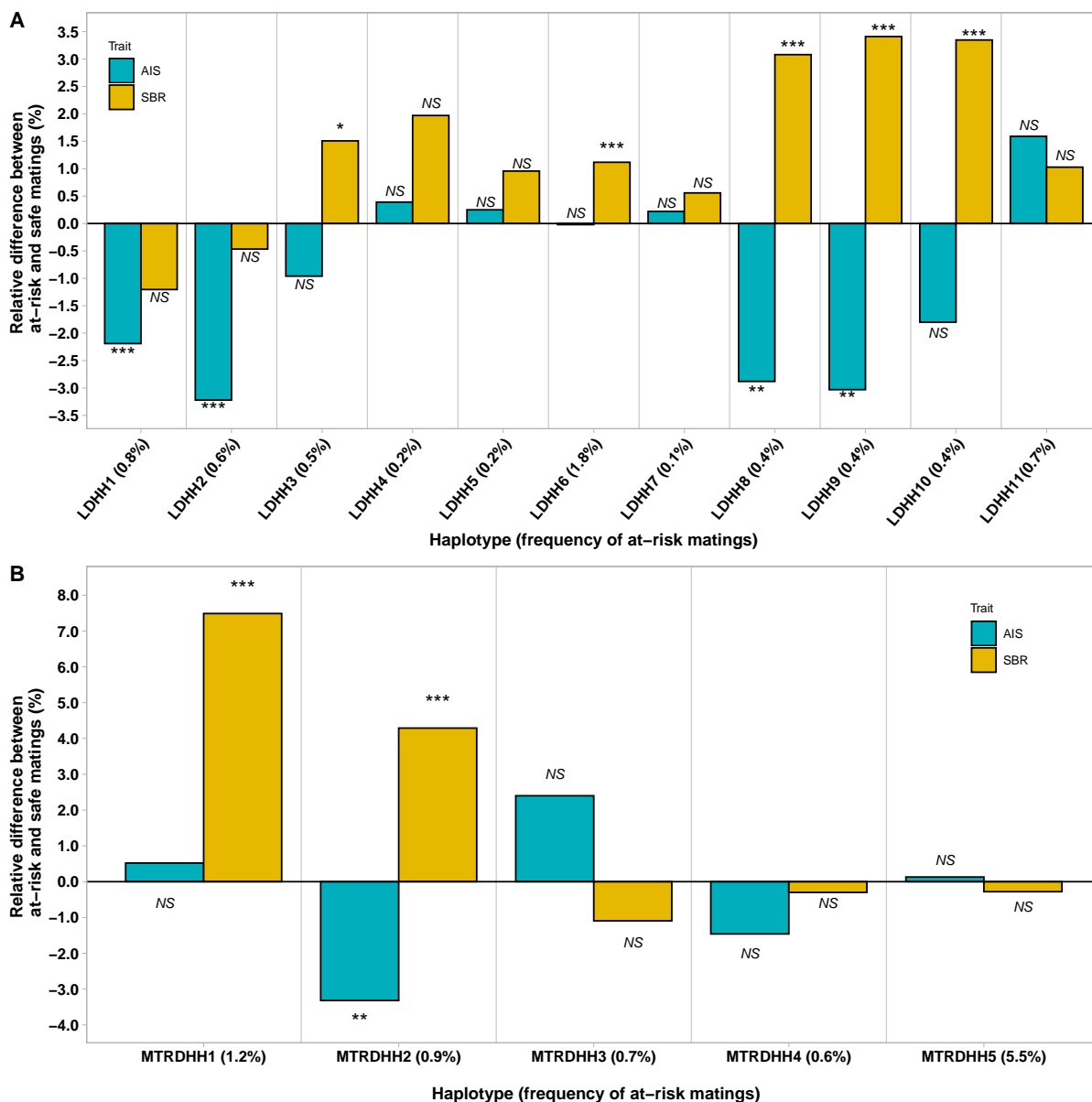


FIGURE 8.3 – Effets des LDHH (A) et MTRDHH (B) sur le taux de réussite à l'insémination artificielle et sur le taux de mortalité dans les accouplements à risque par rapport aux accouplements non à risque. Réussite à l'insémination artificielle (AIS), taux de mortalité (SBR). Pour chaque DHH, la fréquence des accouplements à risque est indiquée entre parenthèses. Les effets significatifs sont indiqués par la P-valeur corrigée pour plusieurs tests avec un seuil fixé à $\alpha = 1\%$ en Lacaune et $\alpha = 5\%$ en MTR.

A retenir L'analyse des caractères de fertilité permet de confirmer que certains des haplotypes détectés semblent bien être associés à des mutations délétères, possiblement létales. Il est désormais possible de classer les haplotypes en quatre catégories en fonction des effets négatifs sur les caractères de fertilité et mortalité. Le premier groupe concerne les haplotypes qui affectent seulement le taux de réussite à l'IA (LDHH1, LDHH2) pour lesquels on fait l'hypothèse que ces régions abritent des mutations embryonnaires létales. Le second groupe est composé des haplotypes LDHH8-9-10 (qui signe la même mutation) et MTRDHH2 qui

ont des effets significatifs sur les deux caractères étudiés. On fait l'hypothèse que ces deux régions abritent des mutations qui conduisent à de la létalité tout au long de la gestation (stade embryonnaire à foetal et périnatal). Le troisième groupe correspond aux haplotypes LDHH3, LDHH6 et MTRDHH1 qui ont un impact uniquement sur le taux de mort-nés. On fait l'hypothèse que ces haplotypes contiennent des mutations qui conduisent à une mortalité pendant la période périnatale. Le dernier groupe contient le reste des haplotypes pour lesquels aucun effet n'a été mis en évidence à partir des deux caractères étudiés. Ces haplotypes peuvent être liés à des mutations qui conduisent à de la létalité plus tardive ou associés à des défauts morphologiques/standard de races contre-sélectionnés. ■

9. Caractères en sélection

A travers ce chapitre, j'ai cherché à identifier un potentiel avantage sélectif des haplotypes mis en évidence.

9.1 Méthodes

9.1.1 Analyse des Daughter Yield Deviation

Les caractères qui entrent dans l'objectif de sélection sont décrits sous forme de DYD (*Daughter Yield Deviation*), correspondant à la performance moyenne des filles de chaque bélier, corrigée pour les effets du milieu et la valeur génétique moyenne des mères (Astruc et al., 2016a). Les six caractères étudiés à partir des évaluations génétiques (GenEval, Jouy-en-Josas, France) sont la production laitière (MY, *milk yield*), les taux de matières grasses (FC, *fat content*) et protéiques (PC, *protein content*), les quantités de matières grasses (FY = MY×FC, *fat yield*) et protéiques (PY = MY×PC, *protein yield*) et le score de cellules somatiques dans le lait (LSCS, *lactation somatic cell score*). Le LSCS est un caractère de santé, qui correspond au SCS moyen par lactation, c'est-à-dire la transformation logarithmique du nombre de cellules somatiques (SCC) le jour du test, défini par $SCS = (\log_2 \frac{SCC}{1000}) + 3$ (Barillet et al., 2016 ; IDELE, 2016 ; Rupp et al., 2003). Pour comparer l'ensemble des caractères sur une même échelle, chaque DYD a été divisé par son écart-type génétique et appelé DYD standardisé (sDYD). Seuls les béliers génotypés avec des enregistrements sur au moins 20 filles ont été inclus dans l'analyse afin d'obtenir des valeurs de DYD suffisamment précises (n = 5 400 béliers en Lacaune et n = 2 570 béliers en MTR). Pour chaque caractère, on compare les béliers porteurs et non porteurs des DHH par analyse de variance en utilisant la procédure GLM dans le logiciel SAS (version 9.4 ; SAS Institute Inc., Cary, NC). Le modèle à effets fixes est $y = X\beta + e$, où y est un vecteur de données sDYD pour chaque caractère ; X est une matrice d'incidence d'effets fixes ; β est un vecteur des effets fixes et e est un vecteur des erreurs résiduelles. Les effets fixes sont le statut génétique (porteur, non-porteur) et l'année de naissance (2000 à 2016) pour corriger du gain génétique annuel. Les caractères diffèrent significativement entre les béliers porteurs et non porteurs lorsque l'effet du statut du DHH est significatif, c'est-à-dire avec une P-valeur inférieure à $3,1 \times 10^{-4}$ en Lacaune et $1,0 \times 10^{-2}$ en MTR. Ces seuils ont été obtenus par correction de Bonferroni pour les tests multiples à un niveau de significativité $\alpha = 1\%$ en Lacaune et $\alpha = 5\%$ en MTR. En Lacaune, ce seuil est corrigé par le nombre de tests indépendants (supposé être le nombre de régions LDHH significatives (n=8) × par les quatre caractères indépendants étudiés (MY, FC, PC et LSCS)). En MTR, seul le nombre de MTRDHH (n=5) est pris en compte.

9.1.2 Analyse de l'Index de Synthèse Ovins Laitiers en Manech Tête Rousse

L'Index de Synthèse Ovins Laitiers génomique (appelé "ISOLg") a été extrait de l'évaluation génomique officielle (GenEval, Jouy-en-Josas, France) pour l'ensemble des 714 agneaux MTR candidats à la sélection pour le millésime 2021. L'ISOLg est déterminé par une combinaison de quatre caractères sélectionnés : MY, FC, PC et LSCS. Pour chaque MTRDHH, les ISOLg d'agneaux porteurs et non porteurs hétérozygotes sont comparés par un test de comparaison de moyenne non-paramétrique de Wilcoxon sous l'hypothèse nulle avec un risque $\alpha=5\%$ en utilisant la fonction "wilcox.test" du logiciel R (version 4.1.3, R Core Team, 2022).

9.2 Résultats : potentiel avantage sélectif des animaux hétérozygotes aux DHH

Pour les 11 régions LDHH détectées dans la race Lacaune, nous avons testé s'il existait un potentiel avantage sélectif à l'état hétérozygote pour les six caractères laitiers inclus dans les évaluations génomiques. La [Figure 9.1](#) montre les différences relatives de sDYD calculées entre les béliers hétérozygotes et non porteurs des DHH. Pour tous les caractères de production laitière, des valeurs positives indiquent une amélioration des caractères sélectionnés, alors que pour le score des cellules somatiques, une valeur positive indique une détérioration de la santé de la mamelle des brebis. Parmi les 11 haplotypes, sept ont un effet significatif sur au moins un caractère. Les filles de béliers porteurs de LDHH1 ont un taux protéique dans le lait plus élevé (sDYD +0,11), tandis que les filles de porteurs de LDHH3 ont une quantité de matière protéique (sDYD -0,12) et un taux protéique (sDYD -0,13) significativement plus faibles. De plus, LDHH3 était significativement associé à un faible score de cellules somatiques dans le lait (sDYD -0,10). Les porteurs de LDHH4 et LDHH5 montrent une augmentation significative de la production laitière (sDYD +0,22 et +0,21, respectivement) et de la quantité de matière protéique (sDYD +0,14 et +0,13, respectivement) mais ont des effets négatifs sur les taux butyreux (sDYD -0,19 et -0,21, respectivement) et protéiques (sDYD -0,20 pour les deux). Les porteurs de LDHH7 et LDHH9 ont une meilleure production laitière avec des sDYD de +0,17 et +0,11, respectivement. LDHH11 est associé à une augmentation de +0,12 de sDYD pour la production de laitière, et à un spectre d'effets plus large avec un effet positif sur les quantités de matières grasses et protéiques (sDYD avec +0,09 et +0,14, respectivement), mais un effet négatif sur le score cellulaire (+0,12) et sur le taux de matière grasse (sDYD -0,12).

Parmi les cinq haplotypes identifiés en race Manech Tête Rousse, trois sont associés à un effet significatif sur au moins un caractère ([Figure 9.2](#)). Les filles des béliers porteurs de MTRDHH2 ont une meilleure production laitière (sDYD +0,06) mais une diminution de la quantité de matière grasse (sDYD -0,11). Pour MTRDHH4, on observe une augmentation significative du score des cellules somatiques (sDYD -0,13). Les filles des porteurs de MTRDHH5 ont une quantité de matière grasse plus élevée (sDYD -0,06), tandis que MTRDHH1 et 3 n'ont aucun effet sur les caractères étudiés. Afin d'estimer si ces haplotypes peuvent avoir un impact sur le gain génétique des caractères sélectionnés, l'index génomique synthétique a été extrait pour chaque agneau mâle de la cohorte de sélection génomique 2021. Cependant, aucune différence significative n'a été observée sur cet index entre les agneaux hétérozygotes porteurs et non porteurs pour chaque MTRDHH ([Figure 9.3](#)).

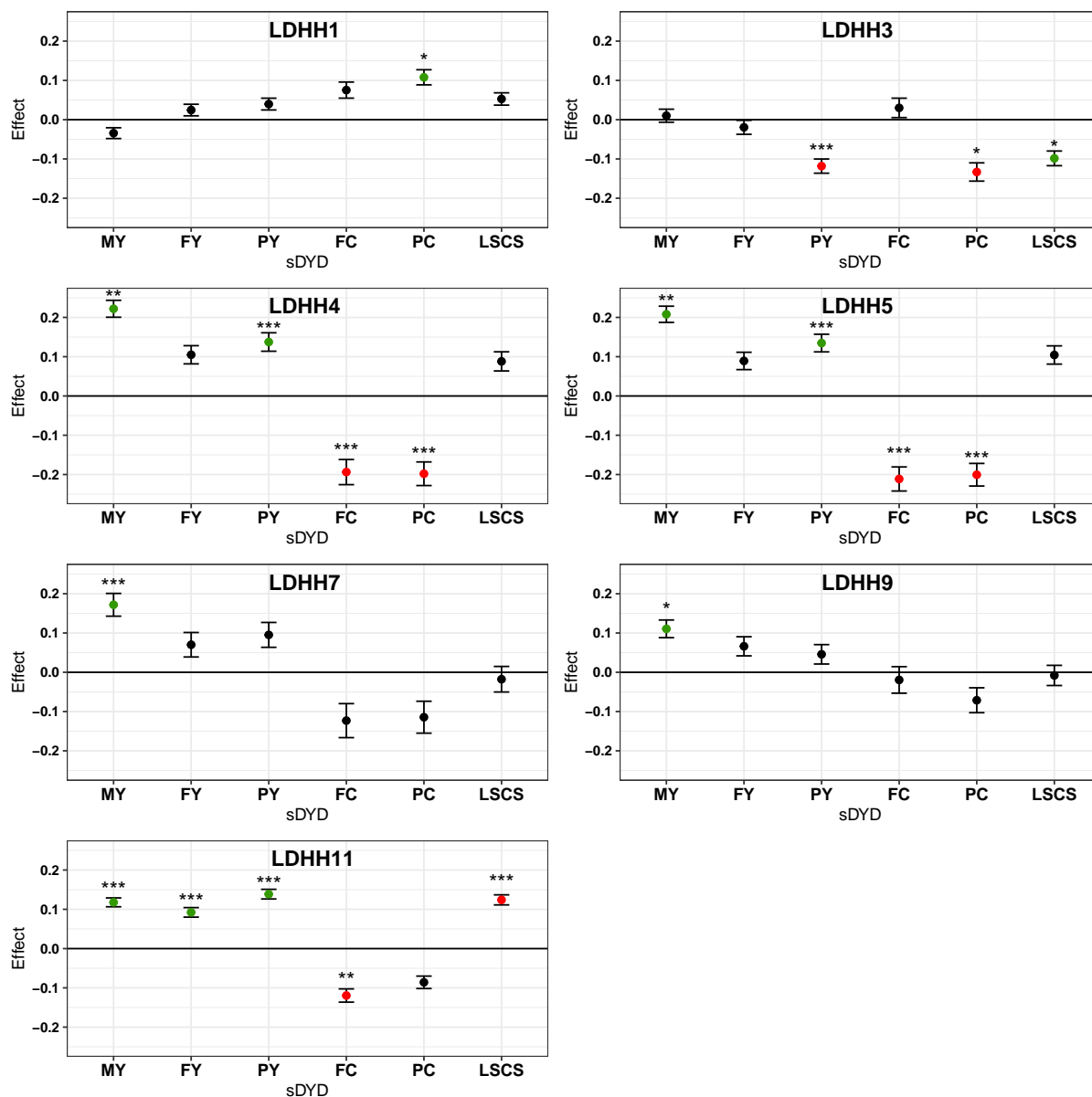


FIGURE 9.1 – Différence relative des sDYG entre les béliers hétérozygotes et non porteurs pour 6 caractères sélectionnés en race Lacaune. Production laitière (MY), quantité de matière grasse (FY), quantité de matière protéique (PY), taux butyreux (FC), taux protéique (PC), score de cellules somatiques dans le lait (LSCS), sDYG= écart standardisé du DYG (DYG divisé par l'écart type génétique). Les effets significatifs sont indiqués par la P-valeur corrigée pour des tests multiples avec un seuil fixé à $\alpha=1\%$ (*), 0,1% (**), 0,01% (***). Les barres d'erreurs indiquent les erreurs standards. Les effets favorables significatifs des hétérozygotes sont en vert et les effets défavorables significatifs sont en rouge. A titre informatif en Lacaune, une différence de sDYG de 0,1 correspond à 3,7 litres de lait (MY), une quantité de matière grasse de 2,3 hg (FY), une quantité de matière protéique de 1,6 hg (PY), un taux butyreux de 0,4 g/L (FC) et un taux protéique de 0,2 g/L (PC).

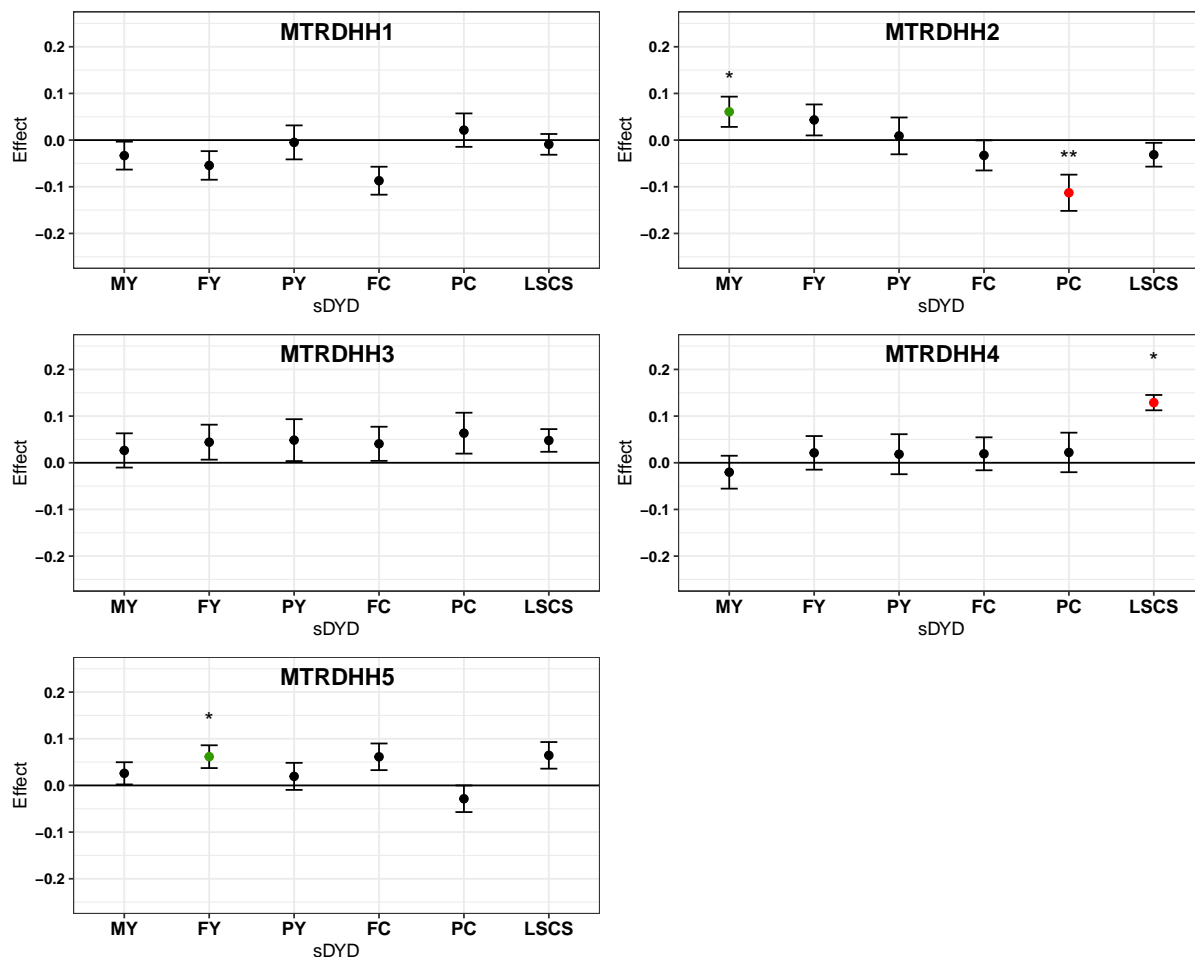


FIGURE 9.2 – Différence relative des sDYD entre les béliers hétérozygotes et non porteurs pour 6 caractères sélectionnés en race Manech Tête Rousse. Production laitière (MY), quantité de matière grasse (FY), quantité de matière protéique (PY), taux butyreux (FC), taux protéiques (PC), score de cellules somatiques dans le lait (LSCS), sDYD= écart standardisé du DYD (DYD divisé par l'écart type génétique). Les effets significatifs sont indiqués par la P-valeur corrigée pour des tests multiples avec un seuil fixé à $\alpha = 5\%$ (*), 1% (**), $0,1\%$ (***) . Les barres d'erreur indiquent les erreurs standard. Les effets favorables significatifs des hétérozygotes sont en vert tandis que les effets défavorables significatifs sont en rouge. A titre informatif en MTR, une différence de sDYD de 0,1 correspond à 3,0 litres de lait (MY), une quantité de matière grasse de 1,5 hg (FY), une quantité de matière protéique de 1,1 hg (PY), un taux butyreux de 0,4 g/L (FC) et un taux protéique de 0,3 g/L (PC).

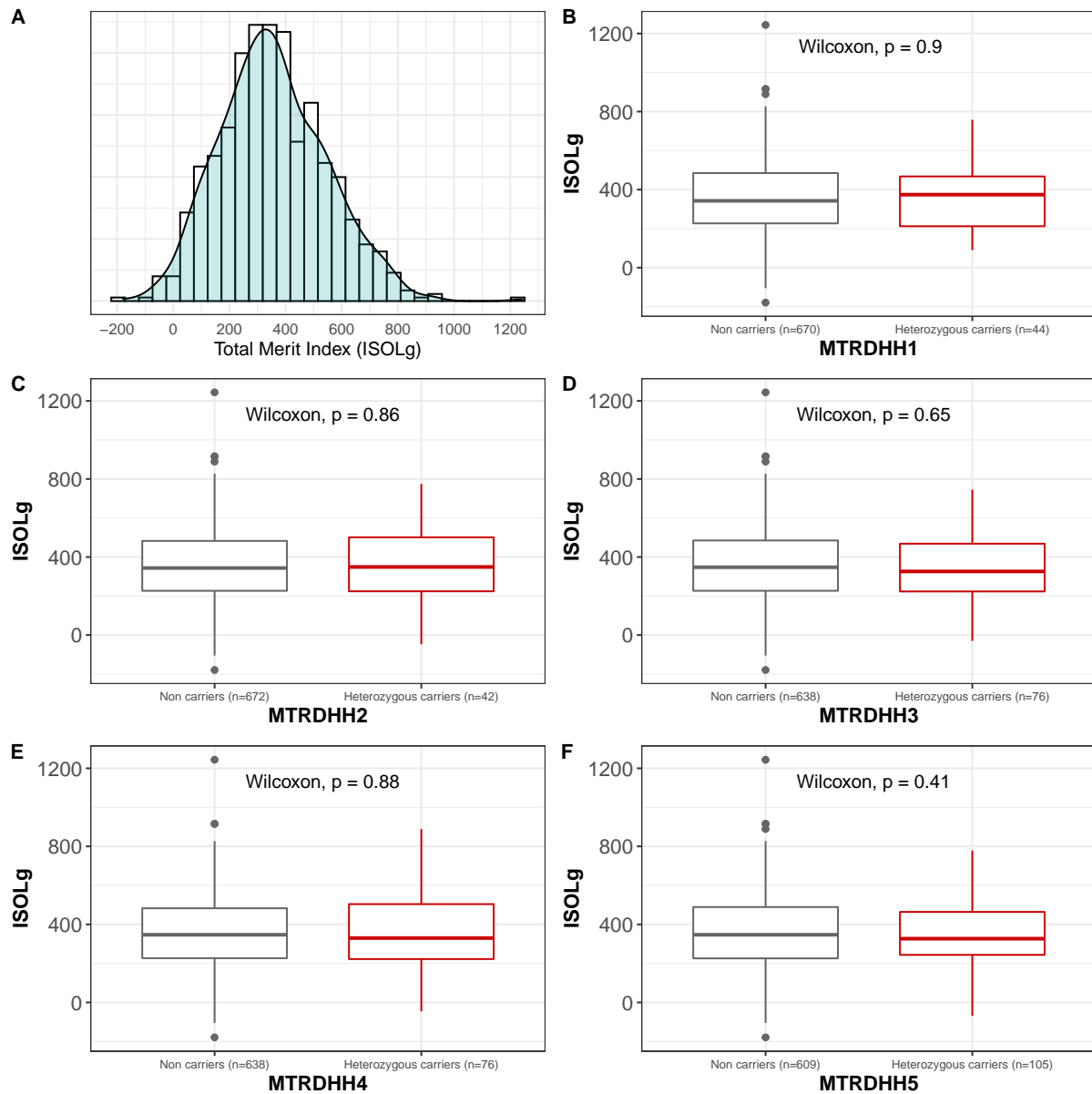


FIGURE 9.3 – Comparaison des index de synthèse en fonction de leur statut aux MTRDHH des agneaux Manech Tête Rousse génomiques de la cohorte 2021 ($n = 714$). (A) Distribution de l'index de synthèse ovins laitier génomique (ISOLg) chez les agneaux Manech Tête Rousse. L'indice ISOLg est déterminé par une combinaison de quatre caractères sélectionnés : MY, FC, PC et LSCS. (B, C, D, E et F) Comparaison de l'index ISOLg en fonction de leur statut à chaque MTRDHH.

A retenir Cinq haplotypes en Lacaune (LDHH1, 3, 4, 9 et 11) et 2 haplotypes en Manech Tête Rousse présentent un avantage sélectif pour des caractères laitiers et de santé. Ces avantages sélectifs ont pu contribuer dans le passé à faire émerger les mutations dans les schémas. ■

10. Évolution des fréquences des haplotypes

L'objectif de ce chapitre est de suivre l'évolution de la fréquence des haplotypes dans les schémas de sélection. Afin d'avoir une représentation de cette évolution, seuls les millésimes à partir de 2017 ont été considérés. En effet, ce n'est qu'à partir de 2017 que les agneaux candidats à la sélection sont systématiquement génotypés sur la puce basse densité dès le plus jeune âge et ce qui permet de donner une bonne représentativité de la population génotypée.

10.1 Évolution de la fréquence des LDHH en Lacaune

L'évolution des fréquences de porteurs hétérozygotes des sept LDHH indépendant (analyse Septembre 2022) en Lacaune est relativement stable sur ces six dernières années. La fréquence moyenne de chacun des haplotypes est de 6,8% (LDHH1), 6,0% (LDHH2), 3,5% (LDHH3), 3,0% (LDHH4), 11,9% (LDHH6), 4,3% (LDHH9) et 5,2% (LDHH11) (Figure 10.1). Cependant dans cette race, les animaux génotypés sont issus de deux schémas distincts : Ovitest (breeding company 1, BC1) et Confédération générale de Roquefort (breeding company 2, BC2). Ces deux schémas sont issus d'une même population qui s'est divisée dans les années 1970. De ce fait, il est plus pertinent d'étudier l'évolution des haplotypes intra-entreprise de sélection. Ainsi, on observe notablement que les haplotypes LDHH1 et LDHH2 (hypothèse de mutations embryonnaires létales) ségrégent principalement dans le schéma de Confédération tandis que les haplotypes LDHH3, LDHH6 et LDHH9 (avortement tardif, mortalité périnatale) ségrégent dans le schéma Ovitest. Dans le schéma Ovitest, on retrouve tous les haplotypes identifiés (parfois à des fréquences faibles) tandis que dans le schéma Confédération, un nombre négligeable d'animaux (fréquence <0,2%) est porteur des haplotypes LDHH3, LDHH4 et LDHH9 qui semblent être plus représentés dans le schéma Ovitest.

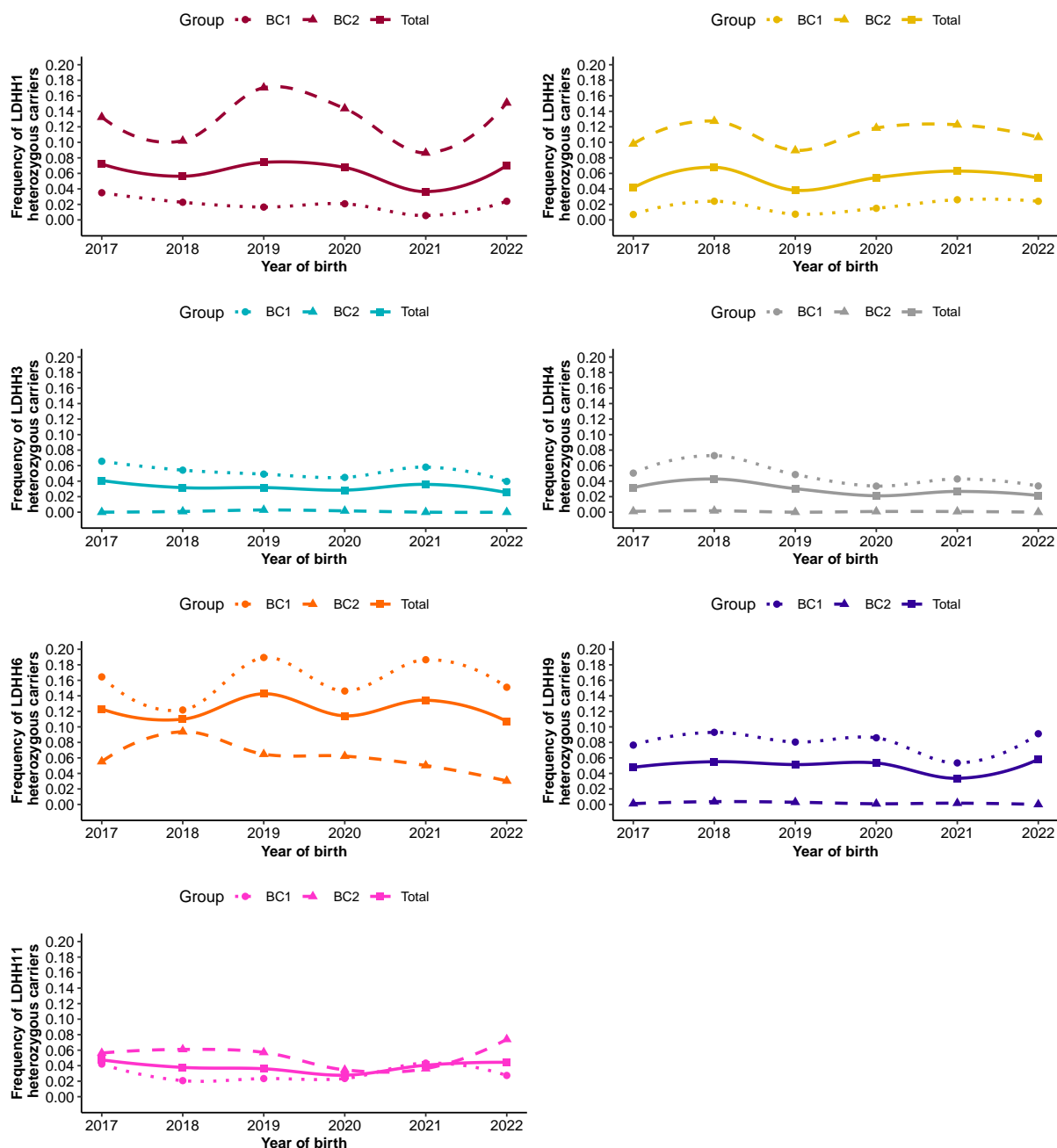


FIGURE 10.1 – Évolution des fréquences des porteurs hétérozygotes des LDHH entre 2017 et 2022 chez les agneaux mâles Lacaune. L'évolution des fréquences a été calculée dans les populations de mâles génotypés dans les deux entreprises de sélection : breeding company 1 (BC1, 60% de l'effectif) et breeding company 2 (BC2, 40% de l'effectif). Pour chaque millésime (2017 à 2022), les effectifs sont de 2 528, 2 594, 2 803, 3 030, 3 177 et 3 722 mâles génotypés, respectivement.

10.2 Évolution de la fréquence des MTRDHH en Manech Tête Rousse

Les évolutions des fréquences de porteurs hétérozygotes des MTRDHH sont stables au cours de ces quatre dernières années autour de 6,8% (MTRDHH1), 7,4% (MTRDHH2), 9,3% (MTRDHH3), 10,2% (MTRDHH4) et 16,1% (MTRDHH5) (Figure 10.2). Néanmoins, on peut remarquer une augmentation spectaculaire de la fréquence de MTRDHH3 entre 2017 et 2018 passant de 2,8% à 10,9%.

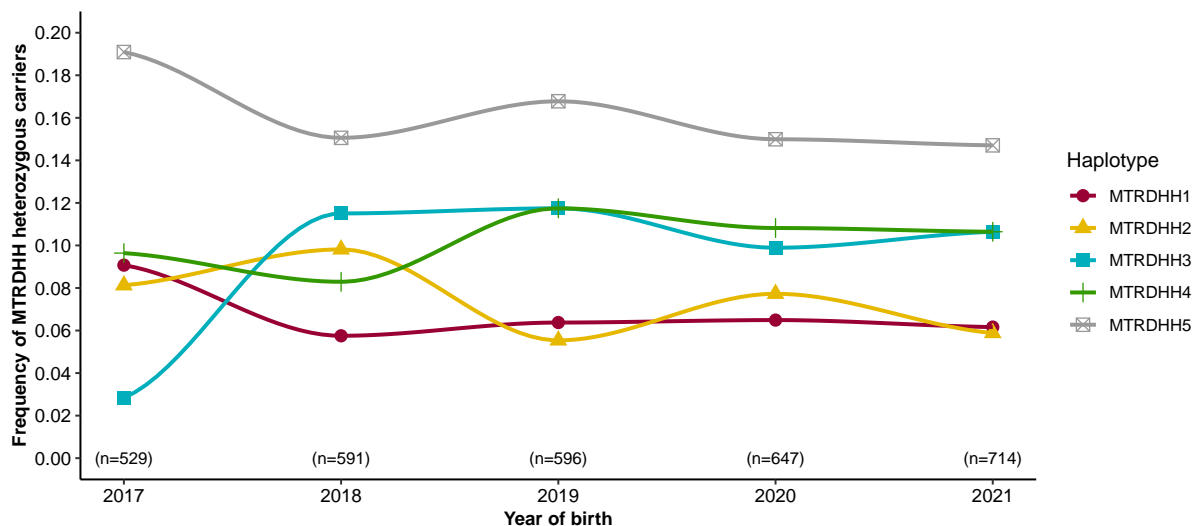


FIGURE 10.2 – Évolution des fréquences des porteurs hétérozygotes des cinq MTRDHH entre 2017 et 2021 chez les agneaux mâles Manech Tête Rousse.

A retenir La fréquence des porteurs hétérozygotes des différents haplotypes est stable (Lacaune : 3,0% à 11,9% - MTR : 6,8% à 16,1%) au cours de ces dernières années. Cela suggère que le maintien de ces différentes mutations s'explique par le hasard (dérive génétique). Néanmoins, aucun lien évident n'est identifié entre les fréquences des haplotypes et les effets significatifs sur les caractères en sélection. ■

IV

Recherche des gènes candidats et des mutations pour chaque haplotype

11	Recherche de gènes candidats	149
11.1	Méthodes	
11.2	Résultats : identification de gènes candidats fonctionnels	
12	Recherche des mutations candidates	155
12.1	Méthodes	
12.2	Résultats : identification des mutations candidates localisées dans les DHH	
12.3	Développement de tests de génotypages spécifiques aux mutations candidates	

L'objectif de cette partie est d'identifier les gènes candidats et les mutations causales localisées dans les sept haplotypes indépendants en Lacaune (LDHH1, 2, 3, 4, 6, 9 et 11) et les cinq haplotypes en Manech Tête Rousse (MTRDHH1, 2, 3, 4 et 5).

L'ensemble des résultats décrits dans cette **Partie IV**, repose sur les articles suivants (disponibles dans la [Partie VII - Productions scientifiques](#)) :

Article 2 (publié) - Ben Braiek et al., 2022 - *Genes* - **A Nonsense Variant in *CCDC65* Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep.**

Article 3 - Ben Braiek et al., 2023a - bioRxiv, rédigé pour *Plos Genetics* - **Homozygous haplotype deficiency in Manech Tête Rousse dairy sheep revealed a nonsense variant in *MMUT* gene affecting newborn lamb viability.**

Article 4 - Ben Braiek et al., 2023b - bioRxiv, rédigé pour *Animal genetics* - **A single base pair duplication in *SLC33A1* gene causes fetal losses and neonatal lethality in Manech Tête Rousse dairy sheep.**

Des résultats complémentaires ont également été ajoutés.



11. Recherche de gènes candidats

A partir des gènes localisés dans les différentes régions identifiées en déficit, l'objectif est de mettre en évidence des candidats susceptibles de porter les mutations causales et dont l'altération du fonctionnement serait impliquée dans des phénotypes de mortalité connus chez les mammifères en tirant parti des bases de données disponibles chez la souris, l'homme et les espèces d'élevage.

11.1 Méthodes

Tous les gènes situés et annotés dans les régions LDHH et MTRDHH étendues de 1 Mb de chaque côté ont été extraits du génome ovin Oar_rambouillet_v1.0 (OAR1 : NC_040252.1, OAR3 : NC_040254.1, OAR4 : NC_040255.1, OAR10 : NC_040261.1, OAR13 : NC_040264.1, OAR18 : NC_040269.1, et OAR20 : NC_040271.1) via le logiciel CLC (QIAGEN CLC Main Workbench 7.9). Les gènes avec un phénotype "knock-out" connu chez la souris pour de la mortalité et/ou du vieillissement précoce (embryonnaire, prénatal, périnatal, néonatal, postnatal, mortalité avant le sevrage, mortalité prématurée et diminution du taux de survie) ou associé à des anomalies génétiques connues chez les mammifères disponibles dans les bases de données Mouse Genome Informatics (MGI, <http://www.informatics.jax.org>), International Mouse Phenotyping Consortium (IMPC, <https://www.mousephenotype.org>), Online Mendelian Inheritance in Man (OMIM, <https://omim.org>) et Online Mendelian Inheritance in Animal (OMIA, <https://omia.org>) (Consulté le 27 mai 2022), ont été extraits et filtrés à l'aide du package R "biomaRt" (version 2.52.0, <https://doi.org/doi:10.18129/B9.bioc.biomaRt>). Les gènes candidats pertinents ont été présentés sous forme de heatmap à l'aide du package R "pheatmap" (version 1.0.12).

11.2 Résultats : identification de gènes candidats fonctionnels

Au sein des sept régions indépendantes en Lacaune (LDHH1, 2, 3, 4, 6, 9 et 11) et des cinq régions en Manech Tête Rousse (MTRDHH1, 2, 3, 4 et 5) étendues de 1 Mb de chaque côté, 945 gènes (537 pour les régions Lacaune et 408 pour les régions MTR) codant pour des protéines sont annotés. Pour chacun de ces gènes, les informations de mortalité chez la souris et leur implication dans des anomalies génétiques connues chez les mammifères ont été extraites lorsqu'elles sont disponibles dans les bases de données MGI, IMPC, OMIM et OMIA. Parmi les 945 gènes, 177 sont impliqués dans des phénotypes de mortalité chez la souris (113 en Lacaune et 64 en MTR) et 131 gènes (86 en Lacaune et 45 en MTR) sont associés à des troubles génétiques connus chez l'homme. A partir de ces informations, j'ai considéré comme candidats de premier plan, les gènes connus pour être associés à de la mortalité chez la souris et associés à des syndromes connus chez l'homme. L'intersection de ces informations conduit à mettre en évidence 76 gènes candidats (53

en Lacaune et 23 en MTR). De plus, un certain nombre d'entre eux est également présent dans la base de données OMIA (mutations causales identifiées dans ces gènes pour des phénotypes connus chez les espèces d'élevage), ce qui représente une liste réduite de 15 gènes en Lacaune (dont 5 gènes uniquement présent dans OMIA) et 7 gènes en MTR (dont 5 gènes uniquement présent dans OMIA). Ainsi, l'ensemble des 85 gènes (57 en Lacaune et 28 en MTR) candidats retenus, est décrit dans les [Figures 11.1 et 11.2](#) pour les sept régions Lacaune et dans la [Figure 11.3](#) pour les cinq régions Manech Tête Rousse.

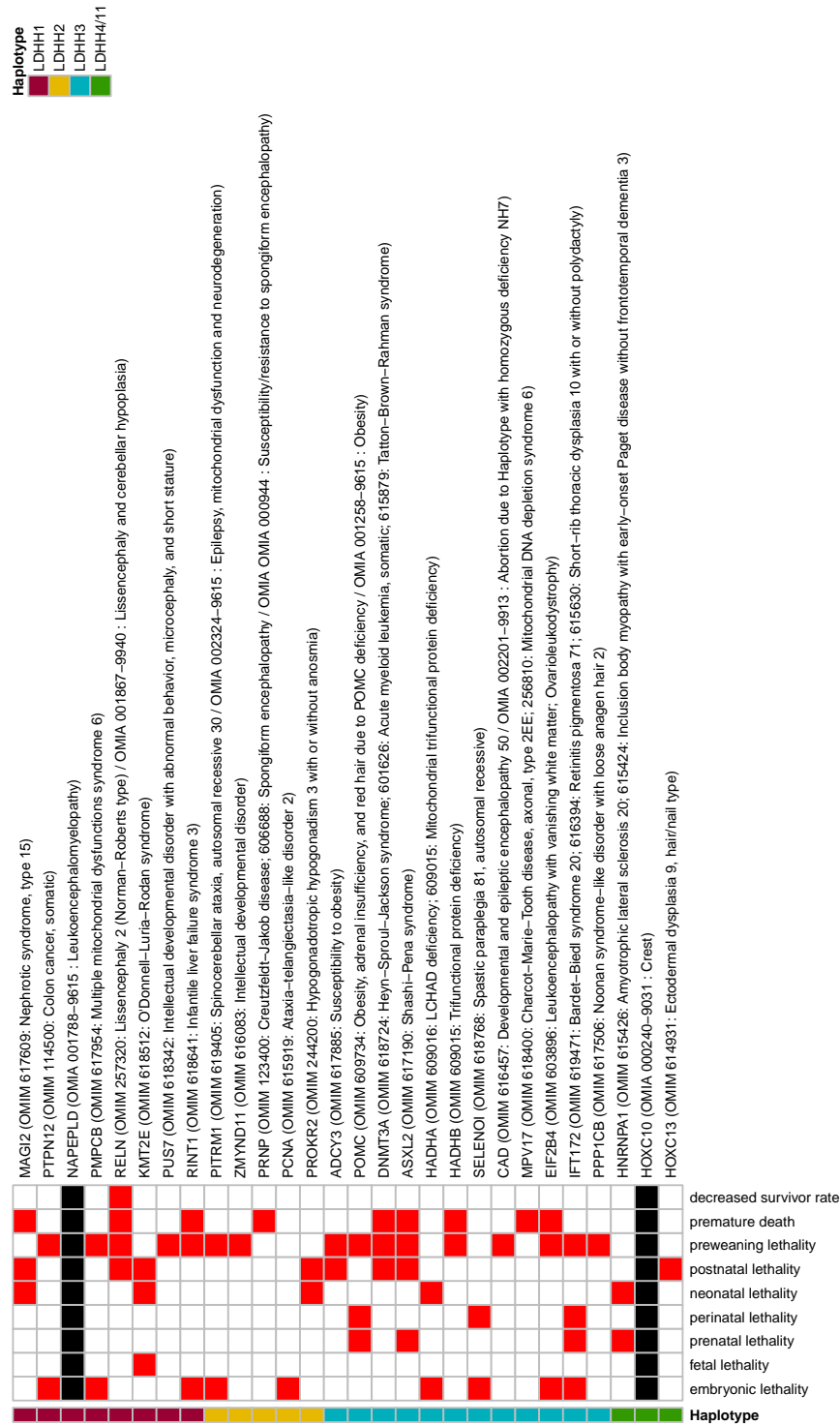


FIGURE 11.1 – Gènes candidats positionnels localisés dans les LDHH impliqués dans des phénotypes de mortalité chez la souris et/ou associés à des anomalies connues chez les mammifères. Les cases rouges indiquent les stades de létalité connus chez la souris lorsque le gène est muté, données issues des bases de données MGI/IMPC. Lorsque le gène est associé à des anomalies connues chez les mammifères, les phénotypes de OMIM et/ou OMIA sont décrits entre parenthèses. Les cases noires indiquent que le gène est impliqué dans un phénotype rapporté dans OMIA ou OMIM mais aucun phénotype létal chez la souris n'est connu. Remarque : en élargissant la recherche de gènes à 1 Mb de part et d'autre de chaque région, les haplotypes LDHH4 et LDHH11 partagent une région commune de 2 Mb.

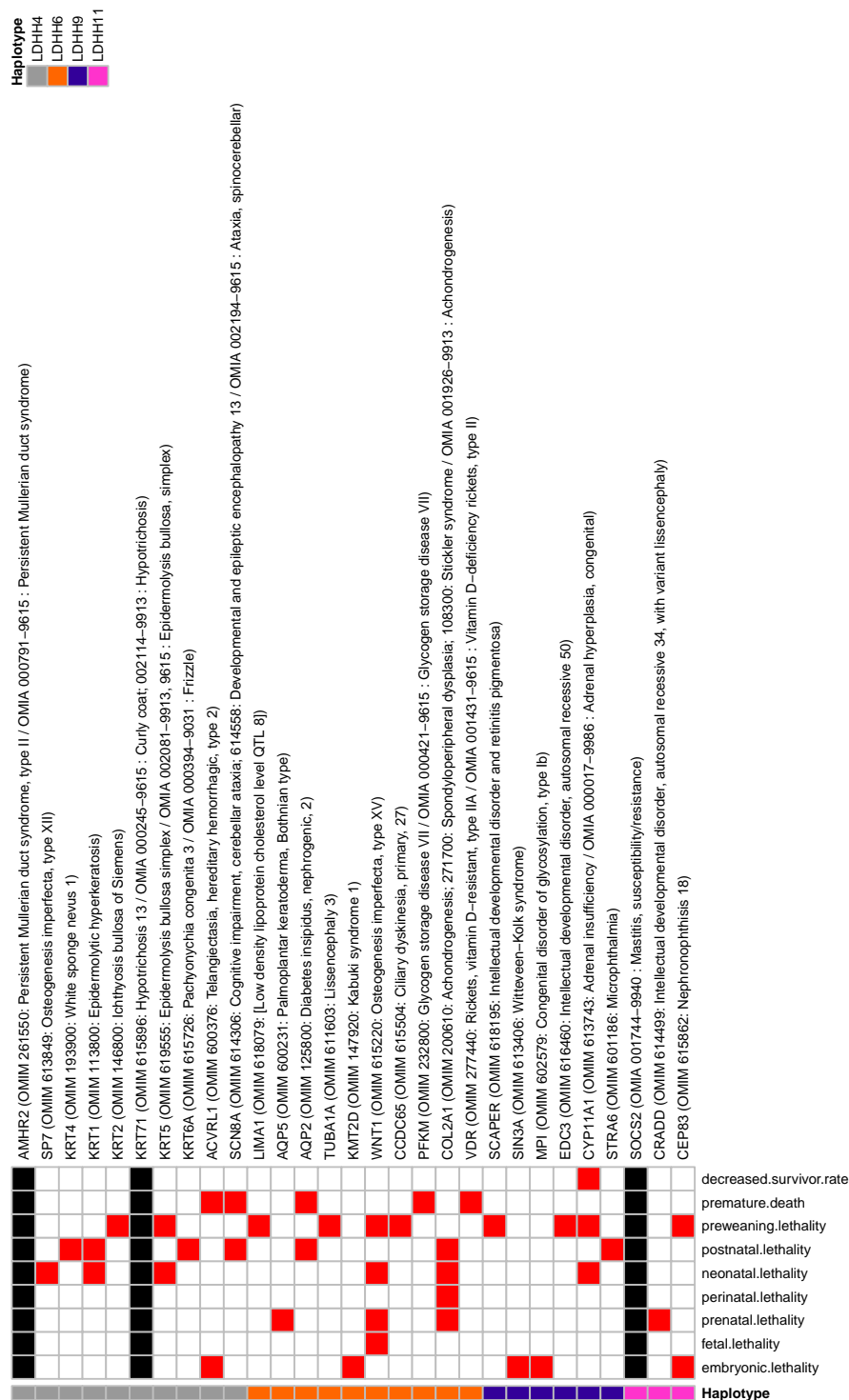


FIGURE 11.2 – Gènes candidats positionnels localisés dans les LDHH impliqués dans des phénotypes de mortalité chez la souris et/ou associés à des anomalies connues chez les mammifères. Les cases rouges indiquent les stades de létalité connus chez la souris lorsque le gène est muté, données issues des bases de données MGI/IMPC. Lorsque le gène est associé à des anomalies connues chez les mammifères, les phénotypes de OMIM et/ou OMIA sont décrits entre parenthèses. Les cases noires indiquent que le gène est impliqué dans un phénotype rapporté dans OMIA ou OMIM mais aucun phénotype léthal chez la souris n'est connu.

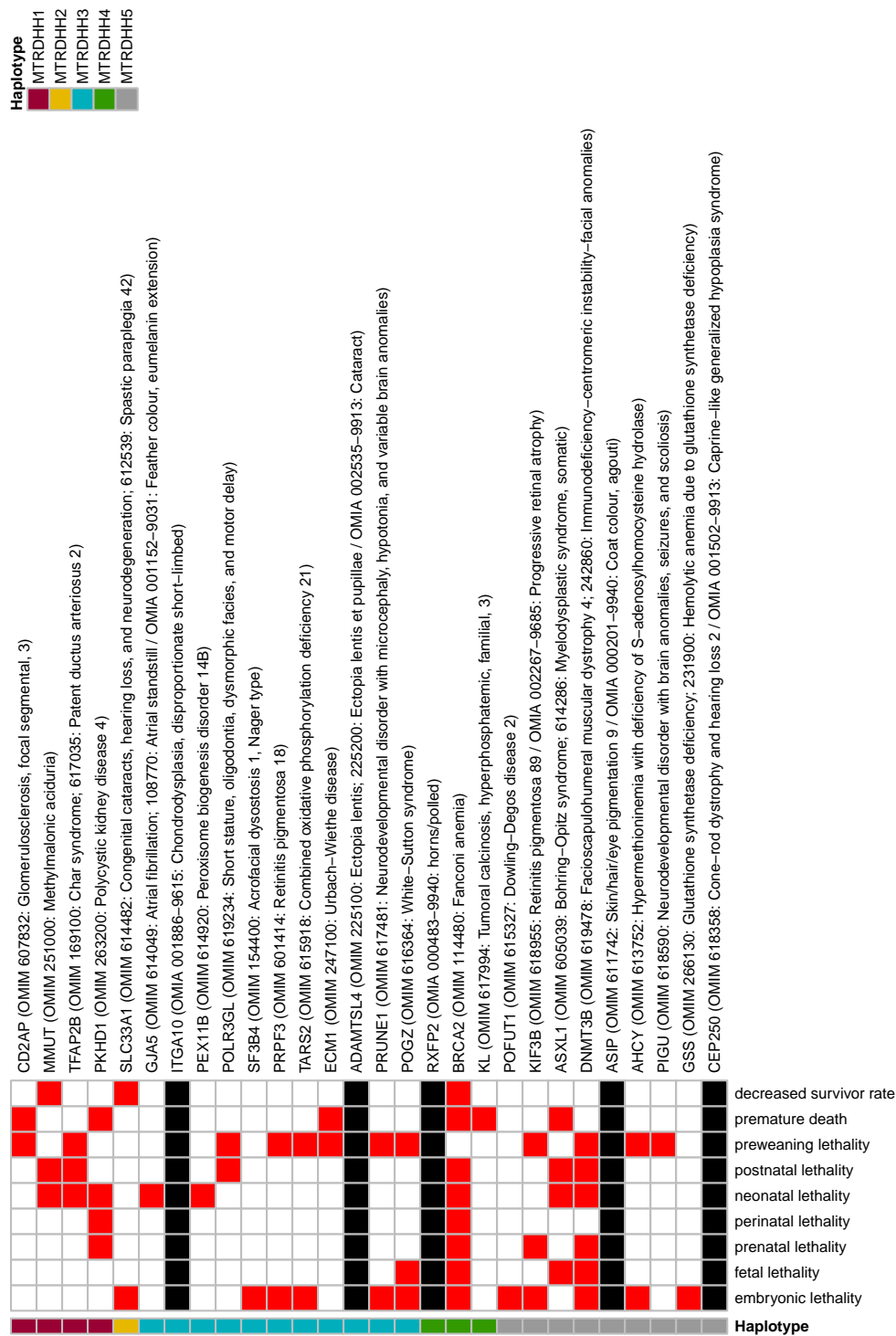


FIGURE 11.3 – Gènes candidats positionnels localisés dans les MTRDHH impliqués dans des phénotypes de mortalité chez la souris et/ou associés à des anomalies connues chez les mammifères. Les cases rouges indiquent les stades de létalité connus chez la souris lorsque le gène est muté, données issues des bases de données MGI/IMPC. Lorsque le gène est associé à des anomalies connues chez les mammifères, les phénotypes de OMIM et/ou OMIA sont décrits entre parenthèses. Les cases noires indiquent que le gène est impliqué dans un phénotype rapporté dans OMIA ou OMIM mais aucun phénotype létal chez la souris n'est connu.

A retenir A partir de l'exploitation des bases de données disponibles, il est possible de mettre en évidence des gènes candidats impliqués dans de la mortalité chez la souris et connus pour être associés à des anomalies chez les mammifères (homme, espèces d'élevage). A ce stade de la démarche, le détail des différents gènes candidats ne présente pas d'intérêt. Ils seront des régions prioritaires dans l'analyse des données de séquence tout génome pour la détection de polymorphismes pouvant représenter les mutations causales influençant les caractères de fertilité et mortinatalité. ■

12. Recherche des mutations candidates

L'objectif de ce chapitre est d'identifier les mutations candidates et possiblement causales associées aux régions en déficit. En l'absence de matériel biologique issu d'animaux homozygotes, la recherche des mutations ne peut se faire que chez les animaux hétérozygotes porteurs des haplotypes.

12.1 Méthodes

12.1.1 Description des données de séquençage

Au laboratoire, 100 séquences de génomes ovins complets (WGS) Illumina HiSeq/NovaSeq à lectures courtes sont disponibles et représentent 15 races (Figure 12.1). Les données de 95 génomes ont été générées préalablement à ma thèse dans différents projets de recherche INRAE et du Teagasc en Irlande. Cinq séquences supplémentaires ont été générées pour répondre aux objectifs de ma thèse (3 séquences Lacaune, 2 séquences MTR). Parmi l'ensemble de ces séquences, 31 sont issues d'animaux Lacaune et 22 d'animaux Manech Tête Rousse. Une description des animaux séquencés (effectif par race, numéro ENA, Numéro du projet EBI) est disponible en [Annexe : Figure 21.1](#).

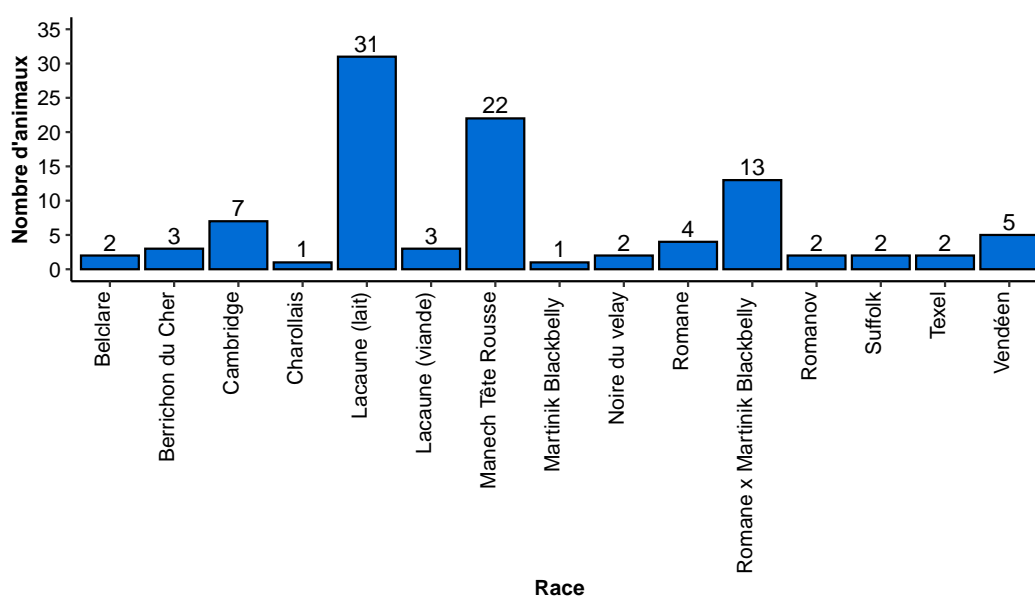


FIGURE 12.1 – Distribution par race des 100 séquences ovines WGS disponibles au laboratoire.

12.1.2 Détection de variants et annotation

L'alignement des lectures et la détection des variants ont été effectués à l'aide de Nextflow v20.10.0 et du pipeline Sarek v2.6.1 (Garcia et al., 2020) avec le support de l'équipe bio-informatique SIGENAE (P. Bardou, INRAE). Brièvement, les différentes étapes réalisées sont les suivantes : l'alignement des lectures avec BWA v0.7.17-r1188 (Li & Durbin, 2010) contre l'assemblage du génome ovin Oar_rambouillet_v1.0 (GCF_002742125.1), le marquage des lectures en double (MarkDuplicates), le recalibrage de la qualité de base (BQSR) et la détection de petits variants germinaux (HaplotypeCaller en mode GVCF) avec GATK v4.1.7.0 (McKenna et al., 2010), l'annotation de petits variants avec SnpEff v4.3t (Cingolani et al., 2012) et le contrôle qualité avec MultiQC v1.8 (Ewels et al., 2016).

12.1.3 Identification des mutations candidates

Tous les SNP et petites insertions/délétions (InDels) situés dans les régions des différents DHH étendues de 1 Mb de chaque côté ont été extraits (Oar_rambouillet_v1.0) à l'aide de SnpSift (Cingolani et al., 2012). On fait les hypothèses fortes que les mutations candidates sont en parfait déséquilibre de liaison avec l'haplotype et sont spécifiques de race. Le filtre des variants repose sur trois critères : (i) la mutation est un variant par rapport au génome de référence de race Rambouillet (variant non porté par l'animal Rambouillet); (ii) présente à l'état hétérozygote chez tous les moutons de la race qui sont porteurs hétérozygotes du DHH (0/1); (iii) mais absente, c'est-à-dire à l'état homozygote référence (0/0), chez tous les moutons de la race qui sont non porteurs du DHH et chez les autres races plus lointaines d'un point de vue phylogénétique (0/0).

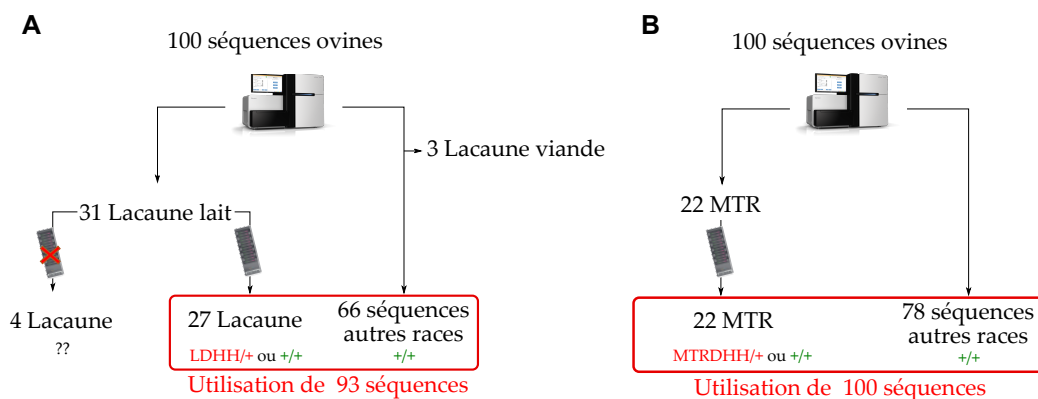


FIGURE 12.2 – Stratégie d'utilisation des données de séquence pour identifier les mutations candidates dans les régions en déficit d'animaux homozygotes en Lacaune (A) et Manech Tête Rousse (B). Parmi les animaux séquencés, 27 Lacaune lait et 22 Manech Tête Rousse sont également génotypés sur puce SNP dans le cadre de la sélection génomique avec un statut connu aux différents DHH.

L'exploitation des 100 séquences est différente selon la race (Figure 12.2). En Lacaune, 31 séquences sont disponibles mais seulement 27 animaux sont également génotypés sur puce SNP (statut connu aux différents LDHH). Parmi les autres races, on dispose de 3 séquences de Lacaune viande. J'ai considéré que ces séquences peuvent être source de biais (race génétiquement proche de la Lacaune lait). Ainsi, les 4 animaux Lacaune lait non génotypés et les 3 Lacaune viande ont été écartés de l'analyse. Au final, 93 séquences sont retenues pour rechercher les mutations

candidates en Lacaune (**Figure 12.2 A**). En Manech Tête Rousse, 22 séquences d'animaux également génotypés sont disponibles avec un statut connu aux différents MTRDHH. Pour cette race, les 100 séquences sont utilisées dans l'analyse (**Figure 12.2 B**). Pour déterminer les mutations candidates, une analyse de corrélation a été réalisée pour chacun des haplotypes entre le statut au DHH (0, 1 et 2) et le dosage des allèles pour les variants bi-alléliques (également codés 0, 1 ou 2) pour les animaux séquencés à l'aide de la commande `geno-r2` de *VCFtools* (Danecek et al., 2011). Les variants candidats (corrélation=1) ont été vérifiés manuellement à l'aide du logiciel de visualisation des fichiers BAM *Integrative Genomics Viewer (IGV)* (Thorvaldsdóttir et al., 2013). L'intervalle de recherche et le nombre d'animaux séquencés hétérozygotes pour chacun des DHH sont décrits dans la **Table 12.1**. Lorsque la mutation est de type faux-sens, l'impact de la mutation est évalué par les outils Polyphen-2 (<http://genetics.bwh.harvard.edu/pph2/>) et SIFT (https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html) et la conservation de la séquence protéique est réalisée par un alignement MUSCLE (CLC Main Workbench 7.9.1) pour différentes espèces.

TABLE 12.1 – Description du nombre de variants identifiés dans les DHH.

Haplotype	OAR	Intervalle de recherche (pb)	Nombre d'animaux hétérozygotes	Nombre de variants	Nombre de variants (r=1)	Nombre de variants (r=1) avec impact HIGH ou MODERATE
LDHH1	4	48 015 021 - 53 306 310	2	57 668	15	0
LDHH2	13	46 202 438 - 50 206 217	2	53 264	27	1
LDHH3	3	32 708 561 - 37 596 529	2	58 356	121	2
LDHH4	3	140 279 184 - 144 118 180	2	47 567	36	0
LDHH6	3	145 243 481 - 148 946 399	2	43 610	11	1
LDHH9	18	29 982 269 - 34 208 337	2	60 417	79	1
LDHH11	3	137 726 532 - 142 216 374	2	61 538	6	0
MTRDHH1	20	21 996 023 - 25 950 669	2	78 019	4	1
MTRDHH2	1	250 858 291 - 257 412 373	5	111 984	189	1
MTRDHH3	1	102 828 253 - 107 639 185	3	82 990	100	2
MTRDHH4	10	29 561 021 - 33 325 300	2	54 312	41	1
	9	48 632 692 - 48 976 498		3 951	0	0
MTRDHH5	13	63 305 288 - 68 234 998	5	52 297	0	0

12.2 Résultats : identification des mutations candidates localisées dans les DHH

Le nombre de variants de type SNP et petites InDels localisés dans les 12 régions (7 régions Lacaune et 5 régions MTR) étendues de 1 Mb de chaque côté varie de 23 610 (LDHH6) à 111 984 (MTRDHH2). Pour l'ensemble des régions, on dispose d'un nombre d'animaux séquencés hétérozygotes variable entre 2 et 5. Le calcul de corrélation pour chacun des haplotypes entre le statut au DHH (0, 1 et 2) et le dosage des allèles pour les variants bi-alléliques (codés 0, 1 ou 2) permet de mettre en évidence des variants qui ne sont présents qu'à l'état hétérozygote uniquement chez les porteurs hétérozygotes des haplotypes. Le nombre de variants ayant une corrélation parfaite entre le statut au DHH et le dosage des allèles varie entre 4 (MTRDHH1) et 189 (MTRDHH2). Cependant, aucun variant identifié n'a une corrélation parfaite ($r=1$) pour l'haplotype MTRDHH5. Par la suite, seuls les variants ayant une corrélation parfaite ($r=1$) et prédits avec un impact fort (HIGH) ou modéré (MODERATE) par SnpEff (Cingolani et al., 2012), sont considérés comme mutations causales candidates. Ce filtre permet de réduire le nombre de mutations candidates à seulement 1 à 2 par région. Néanmoins, aucune mutation candidate n'est identifiée pour les régions LDHH1, LDHH4, LDHH11 en Lacaune et MTRDHH5 en Manech Tête Rousse. L'ensemble de résultats est décrit dans la [Table 12.1](#). Pour les régions où aucune mutation candidate n'a été identifiée par cette stratégie, j'ai tenté d'identifier visuellement des mutations candidates dans les CDS des gènes à partir des fichiers BAM des animaux hétérozygotes des différents DHH sous IGV. Par ce travail, j'ai pu mettre en évidence une duplication d'une paire de bases OAR4 :g.51,214,530dupT dans le gène *ORC5* (*Origin Recognition Complex Subunit 5*) dans LDHH1 qui n'était pas listée dans les variants de cette région par la chaîne d'analyse bioinformatique ([Figure 12.3](#)). Au final, j'ai pu identifier 11 mutations candidates : 6 mutations faux-sens, 2 mutations non-sens, 2 mutations entraînant un décalage du cadre de lecture (*frameshift*, liées à des duplications d'une seule paire de bases) et une causant la perte d'un codon (*inframe deletion*). L'ensemble de ces mutations est décrit dans la [Table 12.2](#), et comme nous le verrons, elles ciblent de nouveaux gènes candidats non révélés par la stratégie des phénotypes décrits dans les bases de données.

TABLE 12.2 – Liste des mutations candidates identifiées dans les DHH (corrélation=1). REF/ALT= Référence/Alternative ; QC=Quality Control ; ORC5=Origin Recognition Complex Subunit 5 ; IDH1=Isopentenyl-Diphosphate Delta Isomerase 1 ; PREB=Prolactin Regulatory Element Binding ; GPN1=GPN-Loop GTPase 1 ; CCDC65=Coiled-Coil Domain Containing 65 ; EDC3=Enhancer Of MRNA Decapping 3 ; MMUT=Methylmalonyl-CoA Mutase ; SLC33A1=Solute Carrier Family 33 Member 1 ; FCGR1A=Fc Gamma Receptor 1a ; RPRD2=Regulation Of Nuclear Pre-MRNA Domain Containing 2 ; RXFP2=Relaxin Family Peptide Receptor 2. La prédiction des mutations faux-sens est réalisée avec deux outils Polyphen-2 et SIFT. Le score PolyPhen-2 est compris entre 0 (mutation prédictée neutre) et 1 (mutation prédictée délétère). Le score SIFT est compris entre 0 (mutation prédictée délétère) et 1 (mutation prédictée neutre).

Haplotype	OAR : position	REF/ALT	QC	Localisation	Conséquences fonctionnelles	PolyPhen-2	SIFT Score
LDHH1	OAR4 :g.51,214,530	dupT	-	Exon, <i>ORC5</i> (c.839dupA)	HIGH, duplication 1pb (p.Lys280Lysfs*6)	-	-
LDHH2	OAR13 :g.47,939,558_47,939,560	delAAG	431,52	Exon, <i>LOC101108851/IDH1</i> (c.287_289delAAG)	MODERATE, délétion 3pb (p.Lys96del)	-	-
LDHH3	OAR3 :g.35,745,589	G>A	380,96	Exon, <i>PREB</i> (c.997C>T)	MODERATE, faux-sens (p.Leu333Phe)	1	0,04
LDHH3	OAR3 :g.36,185,884	G>A	215,8	Exon, <i>GPN1</i> (c.212G>A)	MODERATE, faux-sens (p.Arg71His)	1	0,01
LDHH6	OAR3 :g.147,207,999	C>A	506,34	Exon, <i>CCDC65</i> (c.331G>T)	HIGH, non-sens (p.Glu111*)	-	-
LDHH9	OAR18 :g.31,922,931	T>G	363,84	Exon, <i>EDC3</i> (c.29T>G)	MODERATE, faux-sens (p.Val10Gly)	1	0

MTRDHH1	OAR20 :g.23,776,347	G>A	145,95	Exon, <i>MMUT</i> (c.1225C>T)	HIGH, non-sens (p.Gln409*)	-	-
MTRDHH2	OAR1 :g.252,649,023	dupG	492,62	Exon, <i>SLC33A1</i> (c.735dupG)	HIGH, duplication 1pb (p.Arg246Alafs*3)	-	-
MTRDHH3	OAR1 :g.105,662,733	T>A	657,95	Exon, <i>FCGR1A</i> (c.973T>A)	MODERATE, faux-sens (p.Leu325Ile)	0,999	0
MTRDHH3	OAR1 :g.106,260,059	G>A	461,62	Exon, <i>RPRD2</i> (c.3899G>A)	MODERATE, faux-sens (p.Arg1300His)	0	0,64
MTRDHH4	OAR10 :g.31,038,357	G>A	432,84	Exon, <i>RXFP2</i> (c.77C>T)	MODERATE, faux-sens (p.Ser26Leu)	0,663	0

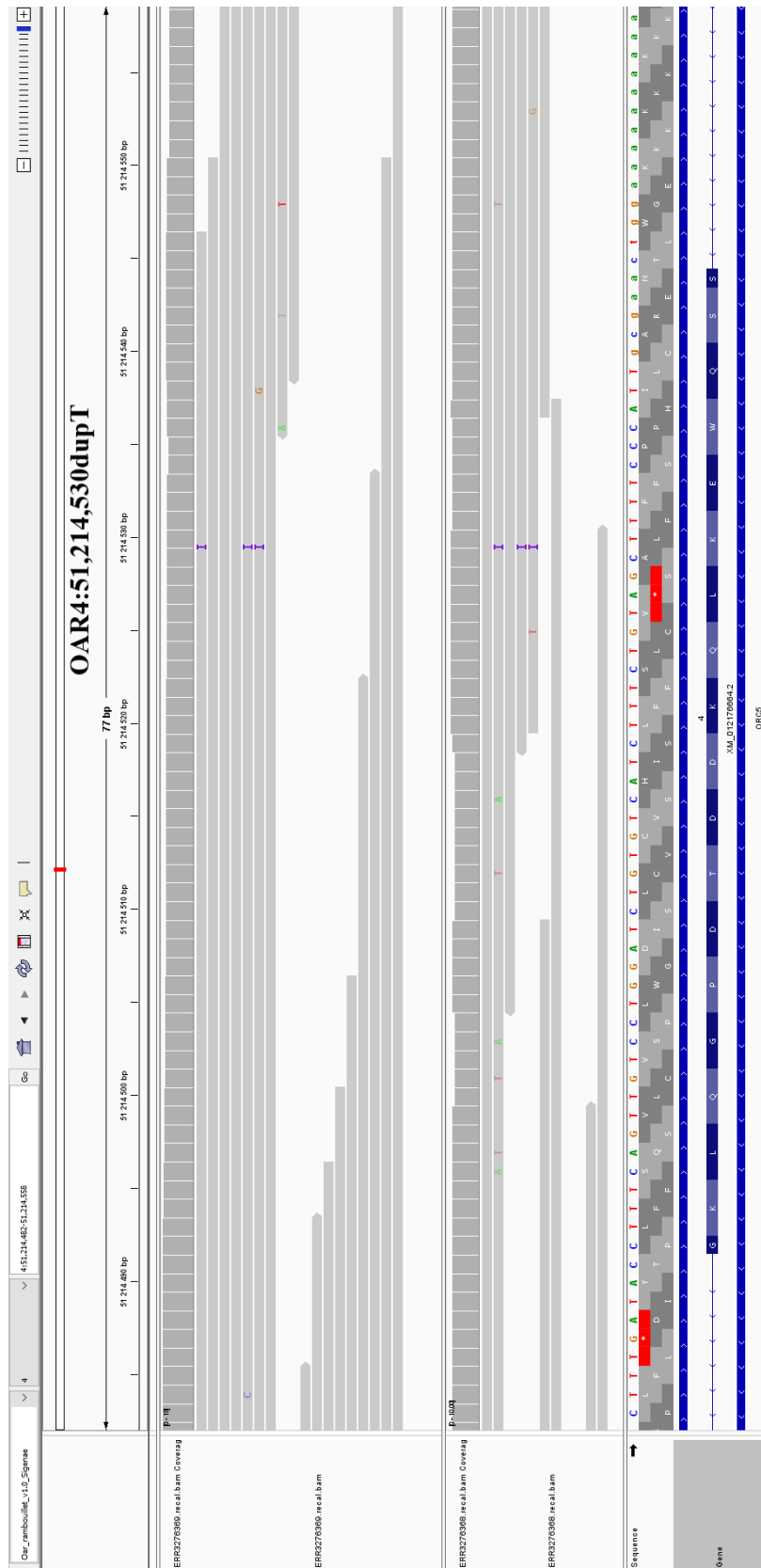


FIGURE 12.3 – Capture d'écran IGV des BAM des deux animaux hétérozygotes à l'haplotype LDHH1. La détection de variant n'a pas permis d'identifier la mutation OAR4 :g51,214,530dupT. La recherche de la duplication a été visualisée sur l'ensemble des animaux séquencés mais n'est présente que chez les deux animaux hétérozygotes LDHH1/+. Des erreurs de séquences sont identifiées dans la plupart des lectures contenant la duplication.

12.2.1 Description des mutations identifiées dans les régions associées à de la mortalité embryonnaire

L'analyse des caractères de fertilité et mortinatalité nous a permis de faire l'hypothèse que les haplotypes identifiés en Lacaune LDHH1 et LDHH2 (premier groupe) contiennent tous les deux des mutations supposées létales embryonnaires à l'état homozygote. En effet, la recherche de variants chez les animaux porteurs hétérozygotes permet de mettre en évidence une duplication d'une paire de bases OAR4 :g.51,214,530 dans le gène *ORC5* (*Origin Recognition Complex Subunit 5*) dans la région LDHH1 (Figure 12.4) et une délétion de 3pb avec conservation du cadre de lecture (*inframe deletion*) OAR13 :g.47,939,558_47,939,560 dans le gène *LOC101108851/IDI1* (*Isopentenyl-Diphosphate Delta Isomerase 1*) dans la région LDHH2 (Figure 12.5).

Dans la région LDHH1, la mutation c.839dupA dans *ORC5* crée un décalage du cadre de lecture et conduit à l'apparition prématurée d'un codon stop en position 285 (p.Lys280Lysfs*6), avec seulement 65% de la protéine traduite (Figure 12.4). Le gène *ORC5* code pour une sous-unité, très conservée dans toutes les cellules eucaryotes, du complexe d'initiation de la réplication de l'ADN et l'expression des gènes (transcription silencieuse) (Quintana et al., 1998 ; Ishiai et al., 1997). Ce complexe se lie aux origines de réplication de l'ADN de manière dépendante de l'ATP et sert d'échafaudage pour l'assemblage du complexe de préréplication. Il est également impliqué dans diverses fonctions cellulaires telles que la formation d'hétérochromatine, la condensation et la ségrégation des chromosomes, la duplication du centrosome, la cytokinèse et l'exportation nucléaire des ARNm (Popova et al., 2018). Il a également été montré que le knock-out du gène *Orc5* murin entraîne de graves altérations de la croissance et de la ramification des dendrites dans les neurones de l'hippocampe en culture (Huang et al., 2005). La suppression de l'une des 6 sous-unités conduit à de la létalité (Ishiai et al., 1997). Chez le mouton, ce gène a une expression ubiquiste (SheepAtlas ; <http://biogps.org>) (Clark et al., 2017).

Dans la région LDHH2, la délétion de 3 paires de bases dans le gène *IDI1* crée la perte de la Lysine 96 dans un site actif très conservé du domaine hydrolase de la protéine (Figure 12.5). Le gène code pour une enzyme localisée dans les peroxysomes qui catalyse l'interconversion de l'isopentényl diphosphate (IPP) en son isomère hautement électrophile, le diméthylallyl diphosphate (DMAPP) (Breitling et al., 2003). Il s'agit d'un intermédiaire conduisant à la biosynthèse des stéroïdes et des acides biliaires à partir du cholestérol. *IDI1* est très exprimé dans le cerveau, cœur, muscles squelettiques, reins et foie (Clizbe et al., 2007). Chez le mouton, ce gène a une expression élevée au stade embryonnaire blastocyste, et dans le cortex et la médulla de la glande surrénale chez l'adulte (SheepAtlas ; <http://biogps.org>) (Clark et al., 2017). Chez la souris, le knock-out de ce gène conduit à de la mortalité avant le sevrage (MGI :2442264). Chez l'homme, des mutations dans ce gène conduisent au syndrome de Zellweger (dysmorphie faciale, crises d'épilepsie, dysfonctionnements hépatiques et rénaux) et à de l'adronoleucodystrophie néonatale, tous deux liés à une diminution de l'activité de *IDI1* dans le foie (Krisans et al., 1994).

Par leur implication dans des processus biologiques de réplication de l'ADN (*ORC5*) et du métabolisme du cholestérol (*IDI1*), les deux mutations identifiées sont donc de très bonnes candidates des régions LDHH1 et LDHH2 pour être causales de mortalité embryonnaire chez le mouton, mais ces deux gènes ne faisaient pas partie de la liste des candidats potentiels (Figure 11.1).

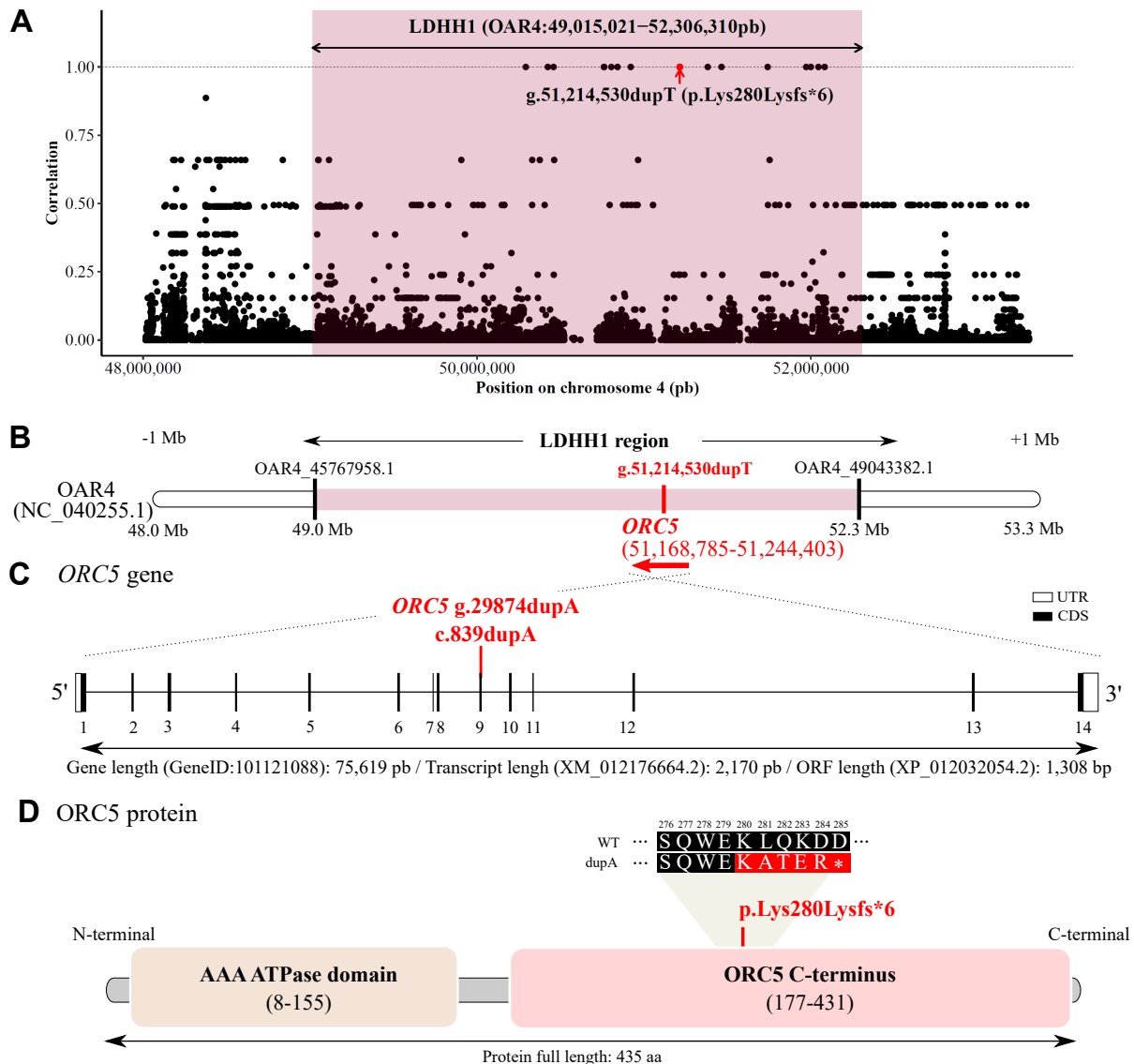


FIGURE 12.4 – Identification d’une duplication d’une paire de bases (*frameshift*) OAR4 :g.51,214,530dupT dans le gène *ORC5* associée à l’haplotype LDHH1. (A) Nuage de points montrant la corrélation entre le statut LDHH1 et les génotypes des variants pour les 93 animaux WGS. Chaque point représente un variant. (B) Haplotype LDHH1 (NC_040255.1, OAR4 : 49 015 021 - 52 306 310pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et le dernier marqueur de la puce Illumina Ovine SNP50 définissant l’haplotype LDHH1. (C) Structure du gène *ORC5* (GeneID : 101121088) et localisation de la mutation *ORC5* g.29874dupA (c.839dupA) identifiée dans le 9^{ème} exon (XM_012176664.2). (UTR : région non traduite; CDS : séquence codante). (D) Protéine *ORC5* (XP_012032054.2) composée des domaines AAA ATPase (pfam13191) et *ORC5* C-terminus (pfam14630). La mutation crée un décalage du cadre de lecture et conduit à l’apparition prématurée d’un codon stop en position 285 (p.Lys280Lysfs*6).

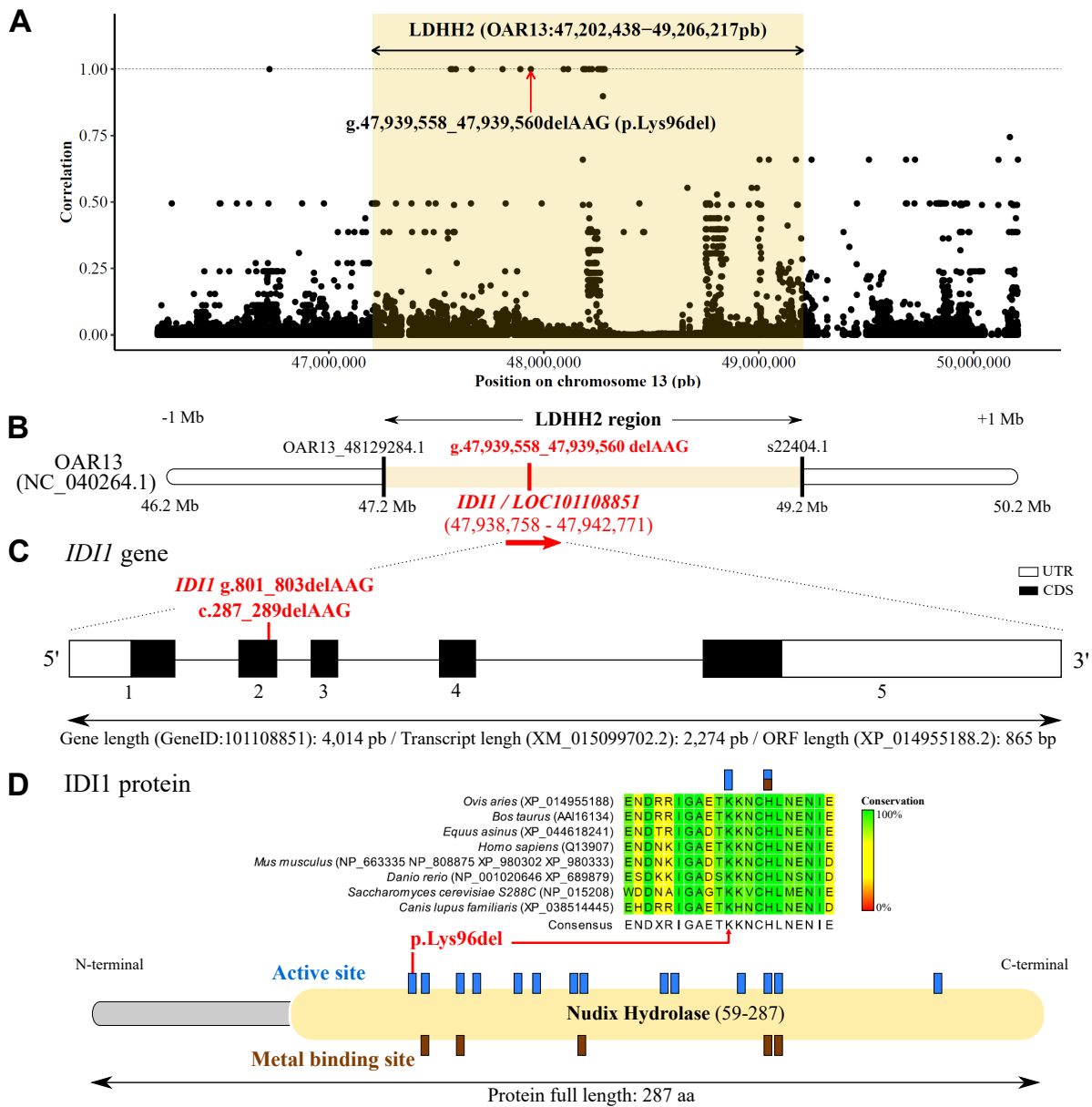


FIGURE 12.5 – Identification d'une délétion de 3 paires de bases (*inframe deletion*) OAR13 :g.47,939,558_47,939,560delAAG dans le gène *ID11* associée à l'haplotype LDHH2 (A) Nuage de points montrant la corrélation entre le statut LDHH2 et les génotypes des variants pour les 93 animaux WGS. Chaque point représente un variant. (B) Haplotype LDHH2 (NC_040264.1, OAR13 : 47 202 438 - 49 206 217pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et le dernier marqueur de la puce Illumina Ovine SNP50 définissant l'haplotype LDHH2. (C) Structure du gène *ID11* (GeneID : 101108851) et localisation de la mutation *ID11* g.801_803delAAG (c.287_289delAAG) identifiée dans le 2^{ème} exon (XM_015099702.2). (UTR : région non traduite; CDS : séquence codante). (D) Protéine *ID11* (XP_014955188.2) composée d'un domaine hydrolase Nudix (c100447), de 14 sites actifs (db_xref : CDD :239218) et de 5 sites de liaison aux métaux (db_xref : CDD :239218). La mutation supprime une Lysine (position 96).

12.2.2 Description des mutations identifiées dans les régions associées à de la létalité pendant toute la gestation

L'analyse des caractères de fertilité et mortinatalité a permis de mettre en évidence un groupe d'haplotypes, composé des haplotypes LDHH9 et MTRDHH2, qui ont des effets significatifs sur les deux caractères étudiés. On fait l'hypothèse que ces deux régions abritent des mutations qui conduisent à de la létalité tout au long de la gestation (stade embryonnaire à foetal et périnatal). La recherche de variants candidats chez les animaux porteurs hétérozygotes a permis de mettre en évidence une mutation faux-sens OAR18 :g.31,922,931T>G dans le gène *EDC3* (*Enhancer Of MRNA Decapping 3*) localisée dans la région LDHH9 (Figure 12.6) et une duplication d'une paire de bases (*frameshift*) OAR1 :g.252,649,023dupG dans le gène *SLC33A1* (*Solute Carrier Family 33 Member 1*) associée à l'haplotype MTRDHH2 (Figure 12.7).

Dans la région LDHH9, la mutation c.29T>G dans le gène *EDC3* entraîne une substitution d'une Valine en Glycine (p.Val10Gly) (Figure 12.6). Le gène *EDC3* code pour une protéine qui intervient dans un complexe de décapsulation des ARNm requis pour l'élimination de la coiffe en 5' avant leur dégradation (Zhang et al., 2021). La mutation impacte le premier acide aminé du motif Sm1, localisé dans le domaine LDm16_N. Ce domaine LSm assure la liaison avec DCP1 (une autre enzyme de décapsulation de l'ARNm) et la localisation des P-bodies (microgranules ribonucléoprotéiques intervenant dans la régulation des ARN) (Tritschler et al., 2007). L'alignement entre espèces montre une très bonne conservation de la séquence protéique. La mutation identifiée est prédite délétère par les algorithmes PolyPhen-2 et SIFT. Elle semble donc perturber le bon fonctionnement du complexe de dégradation des ARNm. La recherche de gènes candidats avait également permis de révéler *EDC3* comme gène candidat fonctionnel (Figure 11.2). En effet, *EDC3* est impliqué dans un trouble du développement intellectuel (OMIM 616460) (Weil et al., 2020; Ahmed et al., 2015) et associé chez la souris à de la létalité avant sevrage lorsque le gène est inactivé (MGI :2142951). Chez le mouton, ce gène a une expression très importante au stade embryonnaire blastocyste (SheepAtlas; <http://biogps.org>) (Clark et al., 2017).

Dans la région MTRDHH2, j'ai identifié une duplication d'une paire de bases (c.735dupG) dans le gène *SLC33A1* (Figure 12.7). La mutation crée un décalage du cadre de lecture et conduit à l'apparition prématurée d'un codon stop en position 248 (p.Arg246Alafs*3), avec seulement 45% de la protéine traduite entraînant une perte de 6 des 9 domaines transmembranaires de *SLC33A1*. *SLC33A1* est le seul gène de la région MTRDHH2 qui avait été retenu comme gène candidat fonctionnel (Figure 11.3). *SLC33A1* code pour un transporteur d'un métabolite essentiel au métabolisme énergétique : l'acétyl-Coenzyme A (acétyl-CoA). La protéine est localisée dans la membrane du réticulum endoplasmique. Chez la souris, l'inactivation de ce gène conduit à de la létalité embryonnaire ou à une mort précoce de l'animal (MGI :1332247). Chez l'homme, l'altération de ce gène par des mutations récessives conduit à de la mortalité infantile (entre 2 et 6 ans) associée à une cataracte congénitale, une perte auditive, une faible teneur en cuivre et céruloplasmine sérique, un retard de développement sévère (OMIM 614482) (Huppke et al., 2012) et à de la paraplégie spastique (OMIM 612539) (P. Lin et al., 2008). De plus, avec le modèle zebrafish, P. Lin et al. (2008) ont mis en évidence une atteinte du développement qui se

traduit par une queue courbée et une excroissance défectueuse des axones de la moelle épinière. L'acétyl-CoA pourrait jouer un rôle majeur dans la régulation de la croissance cellulaire, de la prolifération et de l'apoptose, suggérant que les défaillances métaboliques de l'acétyl-CoA sont responsables de mortalité embryonnaire (Tsuchiya et al., 2014). Chez le mouton, ce gène est exprimé de manière ubiquiste (SheepAtlas; <http://biogps.org>) (Clark et al., 2017).

Les deux mutations identifiées pour les régions LDHH9 et MTRDHH2 sont de très bonnes candidates pour être associées à de la mortalité embryonnaire ou néonatale par leur implication dans des processus biologiques de dégradation de l'ARNm (*EDC3*) et de transport de l'acétyl-CoA (*SLC33A1*). Ces deux gènes faisaient partie de la liste des candidats potentiels (Figures 11.2 et 11.3).

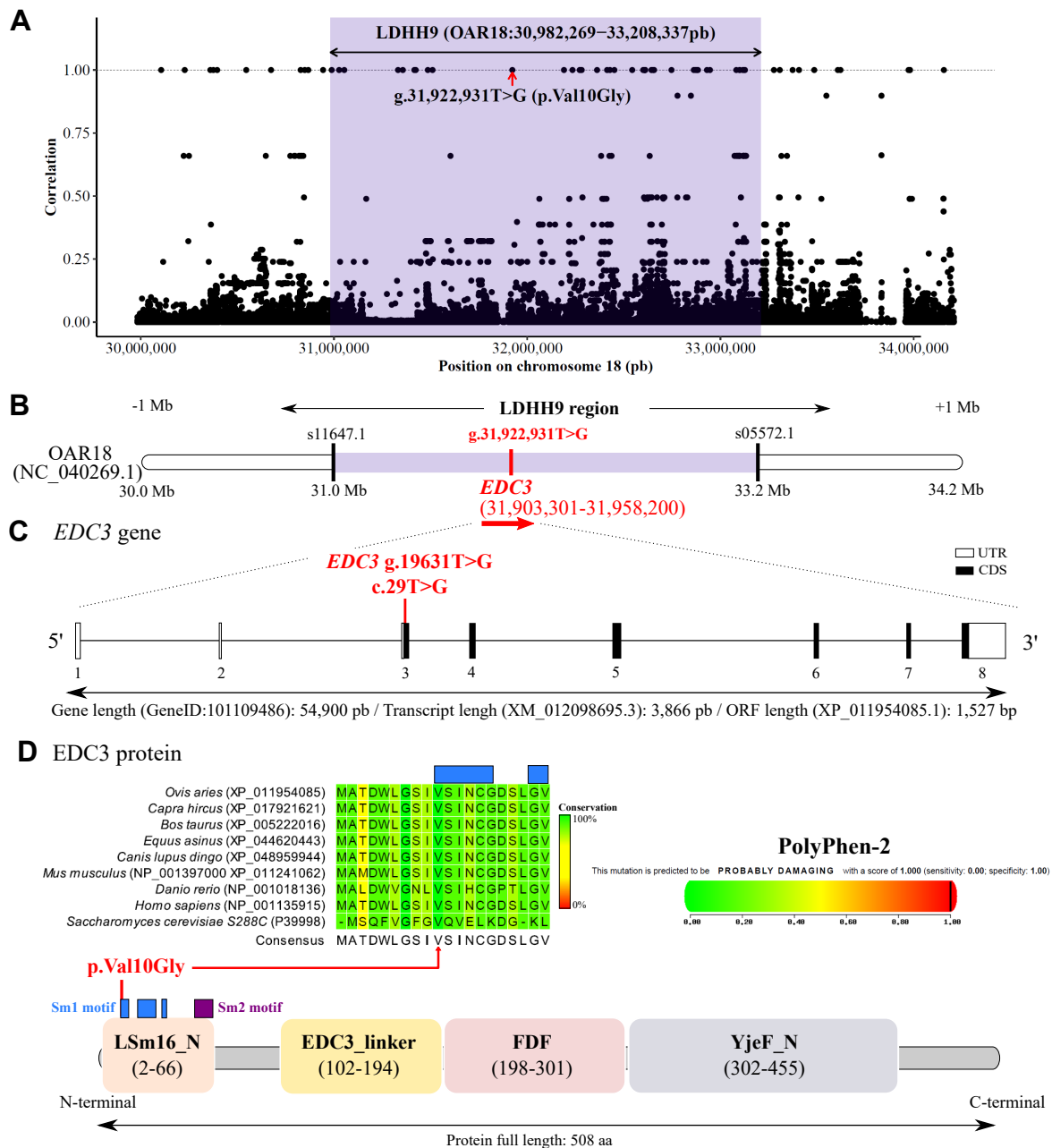


FIGURE 12.6 – Identification d'une mutation faux-sens OAR18 :g.31,922,931T>G dans le gène *EDC3* associée à l'haplotype LDHH9 (A) Nuage de points montrant la corrélation entre le statut LDHH9 et les génotypes des variants pour les 93 animaux WGS. Chaque point représente un variant. **(B)** Haplotype LDHH9 (NC_040269.1, OAR18 : 30 982 269 - 33 208 337pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et le dernier marqueur de la puce Illumina Ovine SNP50 définissant l'haplotype LDHH9. **(C)** Structure du gène *EDC3* (GeneID : 101109486) et localisation de la mutation *EDC3* g.19631T>G (c.29T>G) identifiée dans le 3^{ème} exon (XM_012098695.3). (UTR : région non traduite; CDS : séquence codante). **(D)** Protéine *EDC3* (XP_011954085.1) composée d'un domaine "Like-Sm protein 16, N-terminal domain" (LSm16_N; cd01737), un domaine "Linker region of enhancer of mRNA-decapping protein 3" (*EDC3*_linker; pfam16598), "FDF (Phe-Asp-Phe) domain" (FDF; pfam09532) et un domaine "YjeF-related protein N-terminus" (YjeF_N; pfam03853). La protéine possède également des motifs Sm1 et Sm2. La mutation entraîne une substitution d'une Valine en Glycine (p.Val10Gly) et correspond au premier acide aminé du motif Sm1.

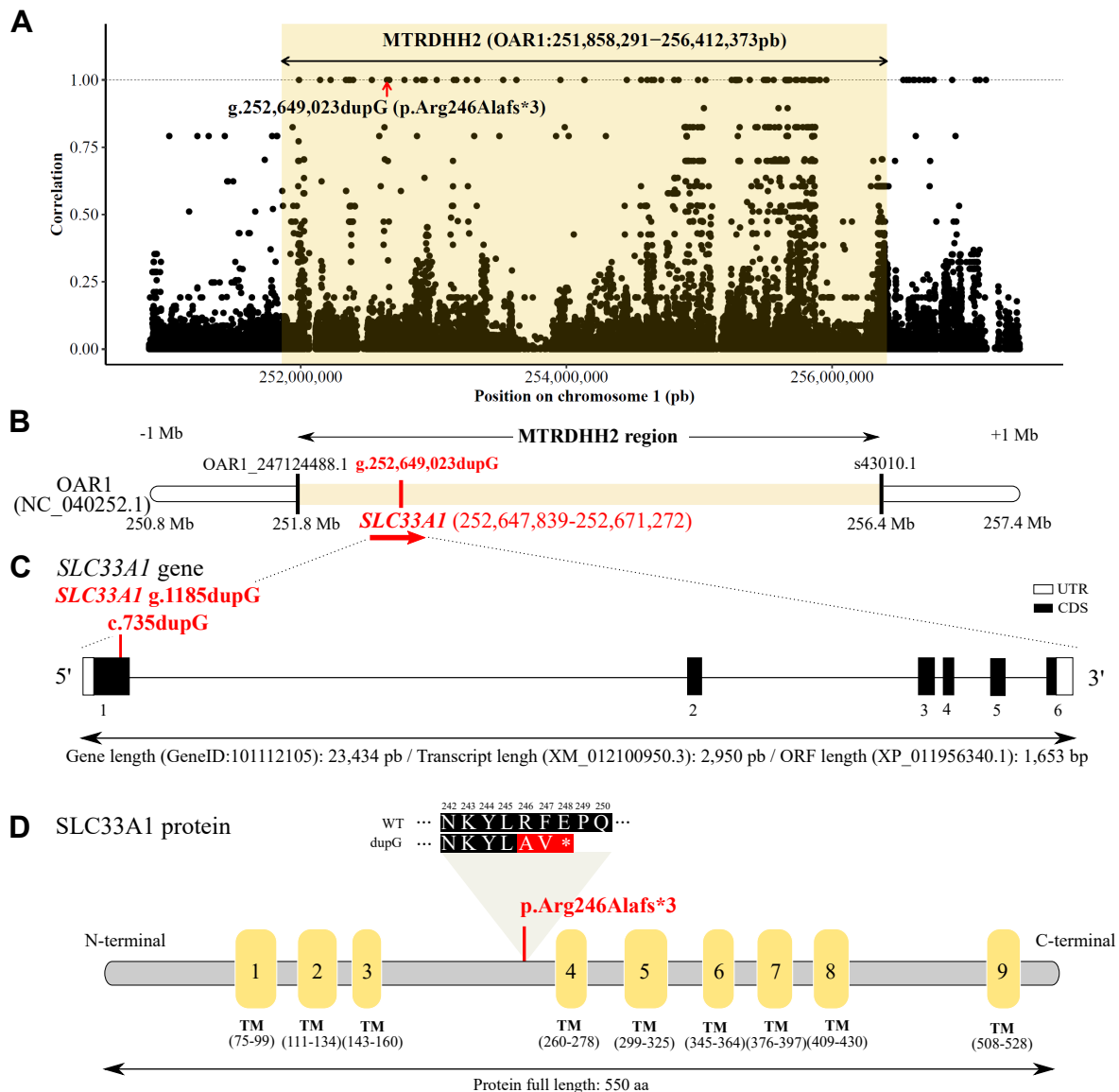


FIGURE 12.7 – Identification d'une duplication d'une paire de bases (*frameshift*) OAR1 :g.252,649,023dupG dans le gène *SLC33A1* associée à l'haplotype MTRDHH2 (A) Nuage de points montrant la corrélation entre le statut MTRDHH2 et les génotypes des variants pour les 100 animaux WGS. Chaque point représente un variant. (B) Haplotype MTRDHH2 (NC_040252.1, OAR1 : 251 858 291 - 256 412 373pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et dernier marqueur de la puce Illumina Ovine SNP50 définissant l'haplotype MTRDHH2. (C) Structure du gène *SLC33A1* (GeneID : 101112105) et localisation de la mutation *SLC33A1* g.1185dupG (c.735dupG) identifiée dans le 1^{er} exon (XM_012100950.3). (UTR : région non traduite; CDS : séquence codante). (D) Protéine *SLC33A1* (XP_011956340.1) avec annotation des domaines par UniProt (A0A6P3TI15_SHEEP) composée de 9 domaines transmembranaires (TM). La mutation crée un décalage du cadre de lecture et conduit à l'apparition prématurée d'un codon stop en position 248 (p.Arg246Alafs*3).

12.2.3 Description des mutations identifiées dans les régions associées à de la mortalité périnatale

L'analyse des caractères de fertilité et mortinatalité a permis de mettre en évidence un troisième groupe d'haplotypes, composé des haplotypes LDHH3, LDHH6 et MTRDHH1, qui ont tous un effet significatif uniquement sur l'augmentation du taux de mort-nés. On fait l'hypothèse que ces haplotypes contiennent des mutations qui conduisent à une mortalité pendant la période périnatale. La recherche de variants candidats chez les animaux porteurs hétérozygotes a permis de mettre en évidence deux mutations faux-sens candidates dans la région LDHH3 (OAR3 :g.35,745,589G>A) dans le gène *PREB* (*Prolactin Regulatory Element Binding*) et OAR3 :g.36,185,884G>A dans le gène *GPN1* (*GPN-Loop GTPase 1*). Pour l'haplotype LDHH6, c'est une mutation non-sens OAR3 :g.147,207,999C>A dans le gène *CCDC65* (*Coiled-Coil Domain Containing 65*) qui a été identifiée en Lacaune. Enfin, pour MTRDHH1 en Manech Tête rousse, une mutation non-sens OAR20 :g.23,776,347G>A est localisée dans le gène *MMUT* (*Methylmalonyl-CoA Mutase*).

Dans la région LDHH3, j'ai identifié deux mutations faux-sens, c.997C>T dans le gène *PREB* et c.212G>A dans le gène *GPN1* (Figures 12.8 et 12.9), toutes deux candidates pour être causales. Le gène *PREB* code pour une protéine régulatrice à motif de liaison WD capable de se fixer sur le promoteur du gène de la prolactine et intervient dans la régulation de sa transcription (Fliss et al., 1999). Cette mutation localisée dans l'exon 7 du gène *PREB* entraîne une substitution d'une Leucine en Phénylalanine (p.Leu333Phe, acides aminés apolaires) affectant un domaine protéique de type WD40 (Figure 12.8). Ce type de domaine intervient dans un grand nombre de processus biologiques tels que la signalisation cellulaire et la régulation de la transcription. La mutation identifiée est prédite délétère par les algorithmes PolyPhen-2 et SIFT et est également localisée dans une zone très conservée entre espèces. Chez la souris, le knock-out de ce gène conduit à de la mortalité embryonnaire et de la mortalité avant le sevrage (MGI :1355326). Chez le mouton, l'expression de ce gène est ubiquiste (SheepAtlas ; <http://biogps.org>) (Clark et al., 2017). La présence de la mutation pourrait perturber l'action du facteur de transcription PREB agissant sur la régulation de la prolactine. Cette hormone est un régulateur de croissance pour de nombreux tissus, y compris les cellules du système immunitaire. Elle peut également jouer un rôle dans la survie des cellules en limitant l'apoptose, et est essentielle pour la lactation (Taylor Clelland et al., 2000). La seconde mutation candidate dans le gène *GPN1* est localisée dans l'exon 3 (Figure 12.9). La mutation entraîne une substitution d'une Arginine en Histidine (p.Arg71His, acides aminés polaires chargés positivement) et elle est prédite délétère par les algorithmes PolyPhen-2 et SIFT. Elle est également localisée dans une zone très conservée entre espèces. Le gène *GPN1* code pour une enzyme à activité guanosine triphosphatase et intervient dans la réparation de l'ADN et dans l'activation de la transcription (Reyes-Pardo et al., 2012). Aucun phénotype knock-out n'est disponible chez la souris pour ce gène, et aucune altération du gène humain n'est rapportée dans OMIM. Chez le mouton, l'expression de ce gène est ubiquiste (SheepAtlas ; <http://biogps.org>) (Clark et al., 2017). De par le rôle de ces gènes, les deux mutations semblent être candidates pour la région LDHH3.

Dans la région LDHH6, j'ai identifié une unique mutation non-sens c.331G>T dans le gène *CCDC65* (Figure 12.10). Le variant identifié devrait aboutir à un codon stop prématuré et à

une protéine tronquée en position 111 (p.Glu111*). Alors que la totalité de la protéine CCDC65 traduite est composée de 498 acides aminés, la forme mutante ne représenterait seulement que 22% de la protéine. La mutation est localisée dans le premier domaine Coiled-Coil (CC) dans une région hautement conservée entre les espèces de mammifères et même avec les algues (*Chlamydomonas reinhardtii*) où le rôle de CCDC65 a été initialement étudié (Heuser et al., 2009; Gui et al., 2019; J. Lin et al., 2011; Bower et al., 2013). Ce gène joue un rôle essentiel dans l'assemblage du complexe nexine-dynéine (N-DRC) intervenant dans régulation de la motilité des cils et des flagelles (Horani et al., 2013). La protéine tronquée résultante a conservé la majeure partie du domaine NYD-SP28 partagé avec CCDC164 faisant également partie du N-DRC (Austin-Tse et al., 2013) mais a perdu deux domaines CC connus pour être impliqués dans les interactions protéine-protéine (J. Lin et al., 2011; Bower et al., 2013). Ce gène avait été précédemment retenu comme gène candidat pour LDHH6 (Figure 11.2). En effet, *CCDC65* est associé au syndrome de dyskinésie ciliaire chez l'homme (OMIM 615504) et conduit à une mortalité avant l'âge du sevrage chez les souris knock-out (MGI :2146001). Les symptômes se traduisent par des situs inversus, des infections chroniques des voies respiratoires et une infertilité chez le mâle. Chez le mouton, l'expression de ce gène est plus importante dans les tissus/organes avec cellules ciliées/flagellées tels que les poumons et les testicules (SheepAtlas; <http://biogps.org>) (Clark et al., 2017).

Dans la région MTRDHH1, j'ai identifié une unique mutation non-sens c.1225C>T dans le gène *MMUT* générant une protéine tronquée (p.Gln409*) avec seulement 55% de la protéine traduite (Figure 12.11). *MMUT* intervient dans la conversion du méthylmalonyl-CoA en Succinyl-CoA qui alimente le cycle de Krebs permettant la récupération d'énergie à partir des sucres, des lipides et des protéines. Cette réaction nécessite l'action d'un cofacteur AdoCbl (forme coenzyme de la vitamine B12) (Baumgartner et al., 2014). Le variant conduit à réduire d'un tiers le domaine d'activité mutase et entraîne la perte du domaine de liaison de la vitamine B12. Ce gène fait également partie de la liste de l'analyse des gènes candidats dans la région MTRDHH1 (Figure 11.3). En effet chez l'homme, l'altération de *MMUT* par des mutations hétérozygotes et homozygotes est associée au syndrome d'acidémie méthylmalonique (OMIM 251000). Chez la souris, le knock-out de *Mmut* conduit à la mortalité avant sevrage (MGI :97239). Si la conversion du méthylmalonyl-CoA en Succinyl-CoA n'est pas possible, cela entraîne une accumulation d'acide méthylmalonique dans les fluides (sang, urine) qui devient létale pour l'animal (Deodato et al., 2006; Lucienne et al., 2020). Chez le mouton, l'expression de ce gène est importante dans le foie et les reins (SheepAtlas; <http://biogps.org>) (Clark et al., 2017).

Une fois de plus, les quatre mutations identifiées pour les régions LDHH3, LDHH6 et MTRDHH1 apparaissent être de bonnes candidates pour expliquer l'augmentation de la mortalité à l'état homozygote. Les gènes affectés, même s'ils ne sont pas tous dans la liste des candidats précédemment établie, interviennent dans des processus biologiques importants de la régulation de la transcription (*PREB*), de réparation de l'ADN (*GPN1*), de fonctionnement de cellules ciliées (*CCDC65*) ou encore de métabolisme énergétique (*MMUT*). Pour la région LDHH3, à ce stade, il est difficile de privilégier l'une ou l'autre des mutations dans *PREB* ou *GPN1*.

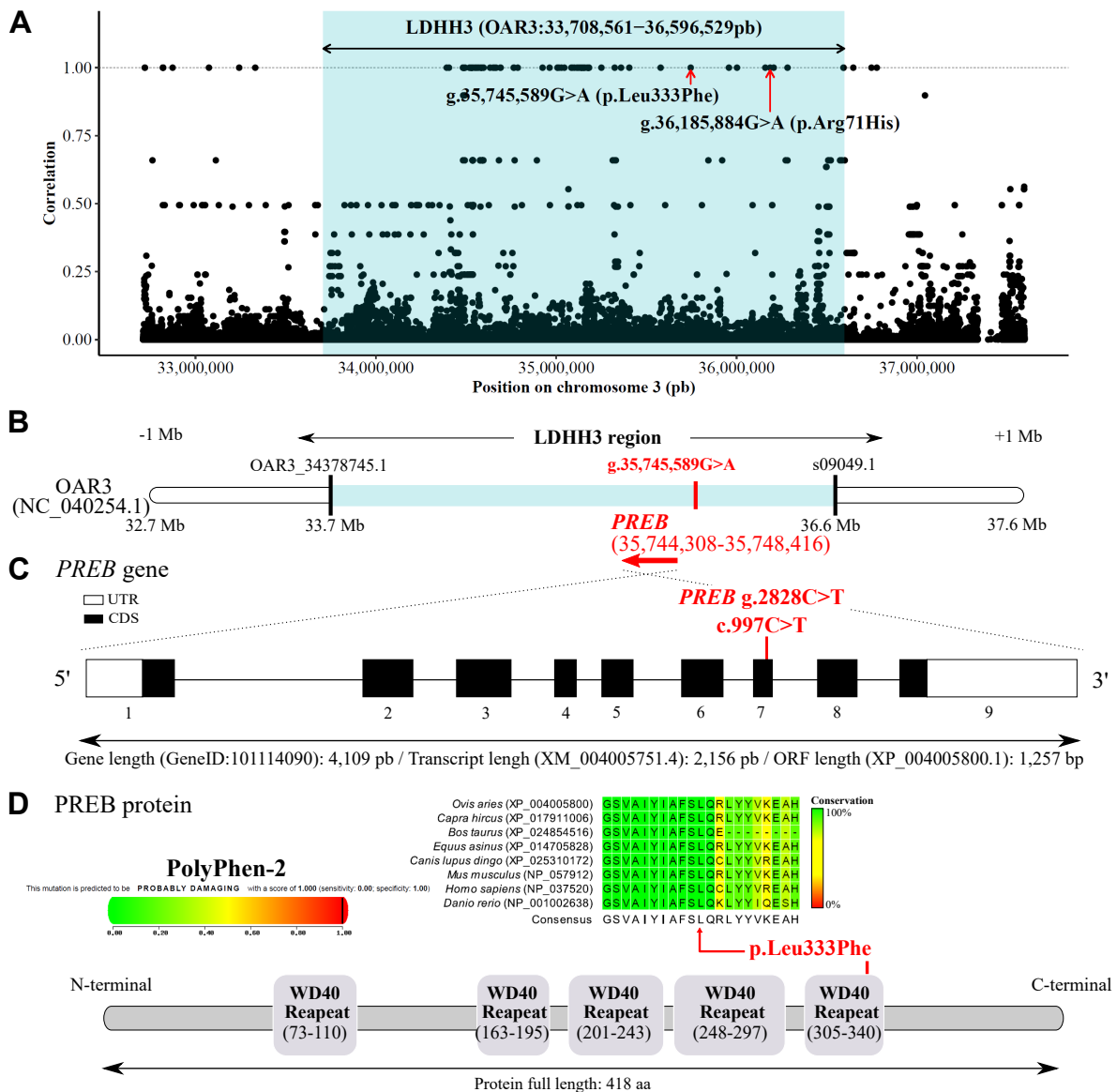


FIGURE 12.8 – Identification de deux mutations faux-sens OAR3 :g.35,745,589G>A dans le gène *PREB* et OAR3 :g.36,185,884G>A dans le gène *GPN1* associées à l'haplotype LDHH3 (A) Nuage de points montrant la corrélation entre le statut LDHH3 et les génotypes des variants pour les 93 animaux WGS. Chaque point représente un variant. (B) Haplotype LDHH3 (NC_040254.1, OAR3 : 33 708 561 - 36 596 529pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et dernier marqueur de la puce Illumina Ovine SNP50 définissant l'haplotype LDHH3. (C) Structure du gène *PREB* (GeneID : 101114090) et localisation de la mutation *PREB* g.2828C>T (c.997C>T) identifiée dans le 7^{ème} exon (XM_004005751.4). (UTR : région non traduite ; CDS : séquence codante). (D) Protéine *PREB* (XP_004005800.1) composée de quatre domaines "WD40 repeat" (db_xref : CDD :293791). La mutation entraîne une substitution d'une Leucine en Phénylalanine (p.Leu333Phe).

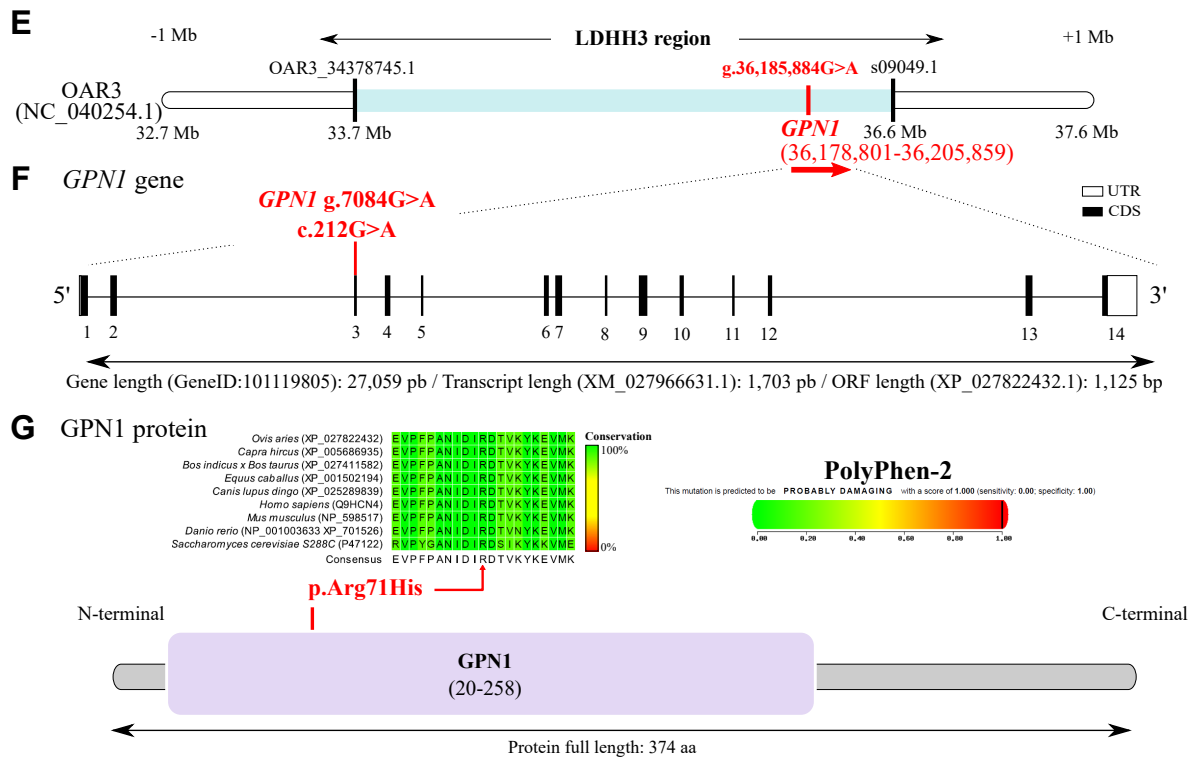


FIGURE 12.9 – Identification de deux mutations faux-sens OAR3 :g.35,745,589G>A dans le gène *PREB* et OAR3 :g.36,185,884G>A dans le gène *GPN1* associées à l'haplotype LDHH3 (E) Haplotype LDHH3 (NC_040254.1, OAR3 : 33 708 561 - 36 596 529pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et dernier marqueur de la puce Illumina Ovine SNP50 définissant l'haplotype LDHH3. (F) Structure du gène *GPN1* (GeneID : 101119805) et localisation de la mutation *GPN1* g.7084G>A (c.212G>A) identifiée dans le 3^{ème} exon (XM_027966631.1). (UTR : région non traduite; CDS : séquence codante). (G) Protéine *GPN1* (XP_027822432.1) composée d'un seul domaine *GPN1* (GPN-loop GTPase 1; cd17870). La mutation entraîne une substitution d'une Arginine en Histidine (p.Arg71His).

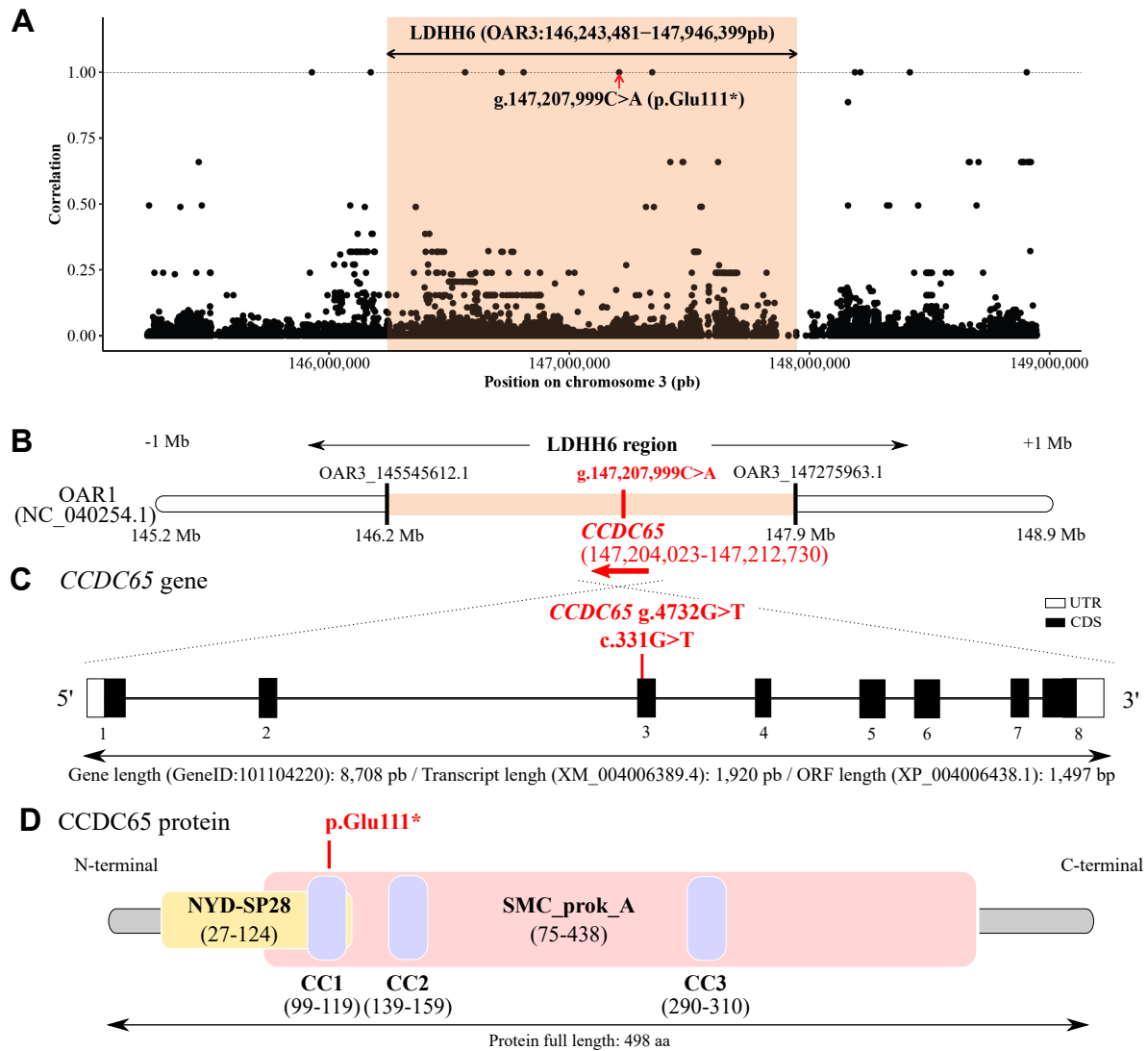


FIGURE 12.10 – Identification d'une mutation non-sens OAR3 :g.147,207,999C>A dans le gène *CCDC65* associée à l'haplotype LDHH6 (A) Nuage de points montrant la corrélation entre le statut LDHH6 et les génotypes des variants pour les 93 animaux WGS. Chaque point représente un variant. (B) Haplotype LDHH6 (NC_040254.1, OAR3 : 146 243 481 - 147 946 399pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et dernier marqueur de la puce Illumina Ovine SNP50 définissant l'haplotype LDHH6. (C) Structure du gène *CCDC65* (GeneID : 101104220) et localisation de la mutation *CCDC65* g.4732G>T (c.331G>T) identifiée dans le 3^{ème} exon (XM_004006389.4). (UTR : région non traduite; CDS : séquence codante). (D) Protéine *CCDC65* (XP_004006438.1) composée d'un domaine NYD-SP28 (Sperm tail; pfam14772), un domaine SMC_prok_A (chromosome segregation protein SMC, primarily archaeal type; TIGR02169) et de trois domaines "Coiled-Coil" (UniProtKB W5QCQ9). La mutation entraîne l'apparition prématurée d'un codon stop en position 111 (p.Glu111*).

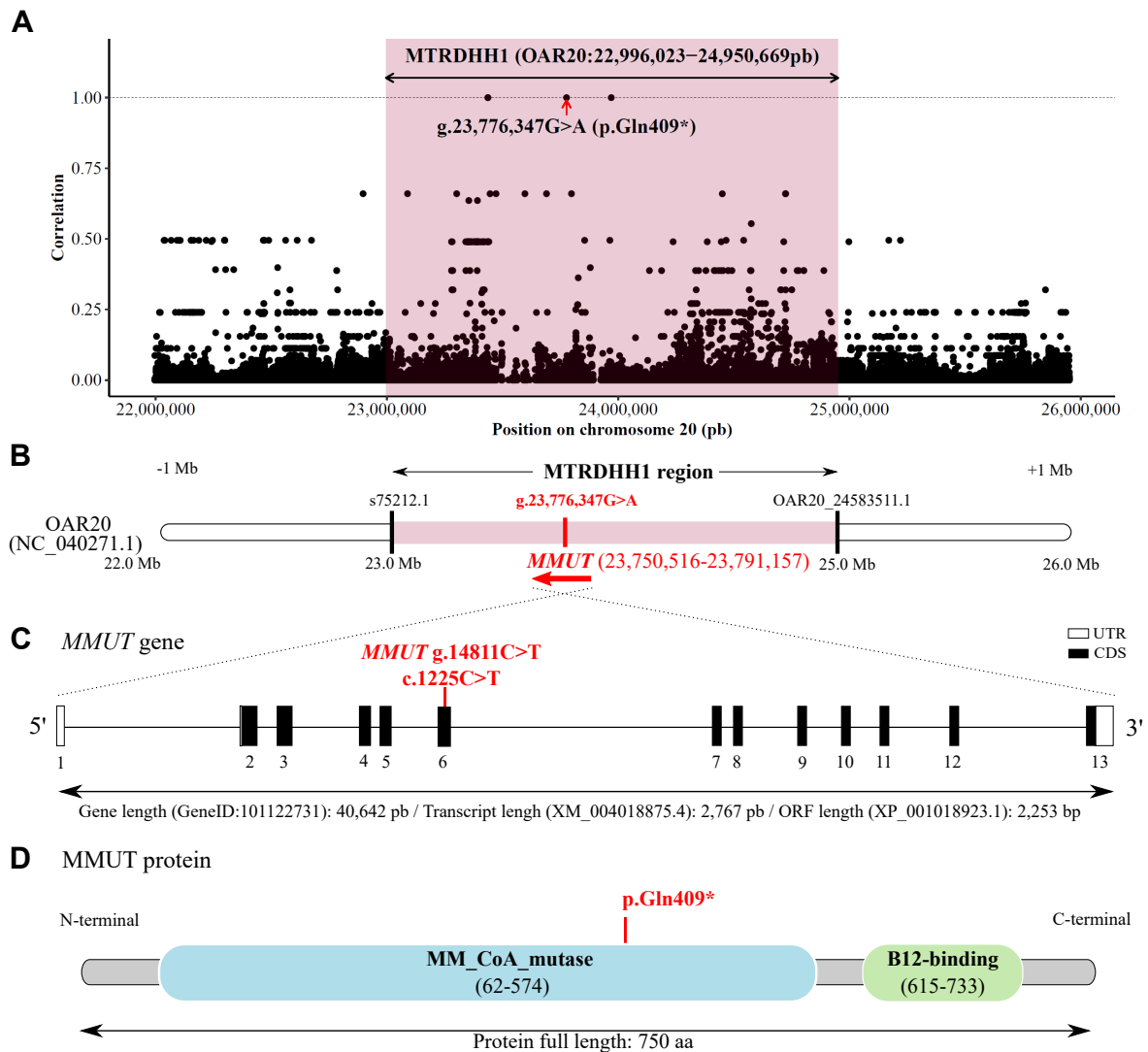


FIGURE 12.11 – Identification d'une mutation non-sens OAR20 :g.23,776,347G>A dans le gène *MMUT* associée à l'haplotype MTRDHH1 (A) Nuage de points montrant la corrélation entre le statut MTRDHH1 et les génotypes des variants pour les 100 animaux WGS. Chaque point représente un variant. (B) Haplotype MTRDHH1 (NC_040271.1, OAR20 : 22 996 023 - 24 950 669pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et dernier marqueur de la puce Illumina Ovine SNP50 définissant l'haplotype MTRDHH1. (C) Structure du gène *MMUT* (GeneID : 101122731) et localisation de la mutation *MMUT* g.14811C>T (c.1225C>T) identifiée dans le 6^{ème} exon (XM_004018875.4). (UTR : région non traduite ; CDS : séquence codante). (D) Protéine *MMUT* (XP_004018923.1, Pfam : A0A6P3T7X3_SHEEP) composée d'un domaine "MM_CoA_mutase" (PF01642) et d'un domaine "B12-binding" (PF02310). La mutation entraîne l'apparition prématurée d'un codon stop en position 409 (p.Gln409*).

12.2.4 Description des mutations identifiées dans les régions associées à de la mortalité juvénile ou à des défauts morphologiques

Le dernier groupe d'haplotypes contient MTRDHH3, MTRDHH4, MTRDHH5, LDHH4 et LDHH11 pour lesquels aucun effet n'a été mis en évidence à partir des caractères de fertilité et mortinatalité. Ces haplotypes pourraient être liés à des mutations qui conduisent à de la mortalité plus tardive (entre quelques jours après la naissance et avant l'âge du génotypage) ou associés à des défauts contre-sélectionnés qu'ils soient morphologiques ou de standard de races. Parmi ces régions, je n'ai pu mettre en évidence des mutations candidates que pour deux haplotypes seulement, MTRDHH3 et MTRDHH4 en Manech Tête Rousse.

Pour MTRDHH3, j'ai identifié deux mutations faux-sens c.973T>A dans le gène *FCGR1A* (*Fc Gamma Receptor Ia*) et c.3899G>A dans le gène *RPRD2* (*Regulation Of Nuclear Pre-mRNA Domain Containing 2*) (Figures 12.12 et 12.13), toutes deux candidates pour être causales. *FCGR1A* intervient dans la régulation de la réponse immunitaire. Il code pour la protéine Fc γ RIA (CD64), un récepteur de haute affinité pour les fragments Fc (*Fragment cristallisable*) des immunoglobulines gamma. Ce récepteur fait partie de la famille des gènes *FCGR1* hautement homologues entre eux (*FCGR1A*, *FCGR1B* et *FCGR1C*) et exprimés à la surface des cellules immunitaires (Nimmerjahn & Ravetch, 2008). Le récepteur Fc γ RIA est fortement exprimé à la surface des monocytes, des macrophages et des leucocytes polymorphonucléaires (Pincetic et al., 2014). Chez la souris, des mutations homozygotes récessives entraînent des anomalies de la réponse immunitaire qui se traduisent par une réponse inflammatoire réduite (MGI :95498). Récemment, des mutations de *FCGR1A* ont été associées à une susceptibilité accrue à la sarcoïdose (Wu et al., 2022) et conduisent à une destruction osseuse précoce dans l'arthrite inflammatoire (Zuo & Deng, 2021 ; Boross et al., 2008). En ovins, la mutation identifiée, entraînant une substitution d'une Leucine en Isoleucine (p.Leu325Ile, acides aminés polaires chargés positivement), est prédite délétère par les algorithmes PolyPhen-2 et SIFT. Elle est également localisée dans une zone très conservée entre espèces. Chez le mouton, ce gène est fortement exprimé dans les macrophages dérivés de la moelle osseuse (SheepAtlas ; <http://biogps.org>). La seconde mutation candidate pour MTRDHH3 se situe dans le dernier exon du gène *RPRD2* (Figure 12.13). *RPRD2* aiderait à la liaison du complexe ARN polymérase II pour la transcription (Ni et al., 2011). Chez le mouton, l'expression de ce gène est ubiquiste (SheepAtlas ; <http://biogps.org>). Chez la souris, le knock-out de ce gène conduit à de la létalité embryonnaire et à de la mortalité avant le sevrage (MGI :1922387). Cependant, la mutation substitue une Argine en Histidine, deux acides aminés polaires chargés positivement (pArg1300His). Cette substitution conservative est prédite sans effet par les algorithmes PolyPhen-2 et SIFT et l'Arginine 1300 impactée n'est pas conservée dans les autres espèces (Figure 12.13 G). Ainsi, la mutation dans le gène *FCGR1A* représenterait la meilleure candidate pour MTRDHH3.

Dans la région MTRDHH4, j'ai identifié une mutation faux-sens c.77C>T dans le gène *RXFP2* (*Relaxin Family Peptide Receptor 2*). Ce gène code pour un récepteur couplé aux protéines G à sept domaines membranaires (GPCR). Il intervient dans la signalisation des peptides de la famille de la relaxine et de l'insulin-like 3 (*INSL3*). La signalisation INSL3/RXFP2 est essentielle à la descente des testicules (Esteban-Lopez & Agoulnik, 2020). Chez la souris, les homozygotes

mâles pour une mutation perte de fonction ciblée de *Rxfp2* présentent une cryptorchidie intra-abdominale bilatérale et une stérilité associée à un défaut de différenciation des gubernacules (ligaments qui contrôlent le mouvement testiculaire au cours du développement) (MGI :2153463, (Gorlov et al., 2002)). En ovins, la mutation identifiée entraîne la substitution non conservative d'une Sérine en Leucine (p.Ser26Leu, passage d'un acide aminé polaire à apolaire) et est localisée dans le premier exon du gène. Son impact est prédit modéré par les algorithmes PolyPhen-2 et SIFT. De plus, l'acide aminé impacté est localisé dans une zone non conservée, même chez les espèces appartenant à la famille des *Pecora* (Mammifères ruminants). Avec ces éléments, on peut penser que la mutation candidate pour MTRDHH4 n'est pas causale. Néanmoins, ce gène *RXFP2* avait été retenu comme gène candidat par son implication dans le phénotype de cornage (présence/absence) en ovins (rétroinsertion de 1,8kb dans le 3'UTR du gène *RXFP2*) (Wiedemar & Drögemüller, 2015). On fait l'hypothèse que le déficit d'animaux homozygotes pour MTRDHH4 dans la race Manech Tête Rousse pourrait être associé à un choix de standard de race par rapport au cornage (déclassement des femelles cornues comme mères à béliers) ou à un défaut morphologique associé à la cryptorchidie (agneaux cryptorchides écartés de la sélection des reproducteurs).

Avec les filtres stricts utilisés pour identifier les mutations candidates (corrélation parfaite entre génotype à la mutation et génotype du DHH, annotation dans le codant, effet prédit fort sur la fonction de la protéine), aucune mutation répondant à ses critères n'a été mise en évidence dans les régions MTRDHH5, LDHH4 et LDHH11 (Figure 12.16). Pour MTRDHH5, je n'ai pas de trouvé de variants dont les allèles sont en association totale avec le génotype de cet haplotype. Cependant, cette région sur OAR13 contient le gène *ASIP* (*Agouti Signaling Protein*) dans lequel j'ai retrouvé deux mutations, g.66,475,132_66,475,136del (délétion de 5pb) et g.66,474,980T>A (SNP, p.Cys126Ser) qui conduisent à une toison de couleur noire (OMIA 000201-9940) (Norris & Whan, 2008 ; Smit et al., 2002). Ces deux mutations ségrégent dans le groupe des animaux MTR séquencés, mais en association imparfaite avec MTRDHH5. Il existe par contre des variants en association parfaite avec LDHH4 (36 variants) et LDHH11 (6 variants), mais l'annotation fonctionnelle de ces variants ne permet pas de cibler des candidats évidents. De façon intéressante, dans la région couverte par l'haplotype LDHH11 sur OAR3, on retrouve le gène *SOCS2* (*Suppressor Of Cytokine Signaling 2*) pour lequel une mutation faux-sens (g.139,302,270C>T, p.Arg96Cys) a été mise en évidence en association avec la sensibilité aux mammites chez les brebis Lacaune (OMIA 001744-9940) (Rupp et al., 2015). La mutation est défavorable pour la santé de la mamelle mais a un effet pléiotrope conduisant également à augmenter le gabarit, la croissance et la production laitière des animaux homozygotes mutés par rapport aux animaux non porteurs. Dans la région LDHH4 (située à proximité de LDHH11), on ne retrouve pas de mutations sous sélection déjà connues en ovins (base de données OMIA). Néanmoins, on note la présence d'un cluster de gènes homéotiques (*HOXC*) et de kératine (*KRT*) mais aucune mutation n'a été découverte dans ces régions (Figures 11.1 et 11.2). Le lien entre ces haplotypes et les mutations connues liées à la morphologie ou à des gènes sous sélection sera approfondi dans le Chapitre 15.

Finalement, la recherche de mutations candidates pour les haplotypes non associées à des effets

sur la fertilité et mortinatalité s'avère plus difficile. Un premier groupe d'haplotypes permettrait de faire l'hypothèse d'un lien avec des mutations connues qui seraient sous sélection (ou contre-sélectionnées) en Manech Tête Rousse ou en Lacaune. Il s'agit de MTRDHH4 en lien possible avec une mutation dans *RXFP2* et affectant le cornage ou la cryptorchidie, MTRDHH5 en lien possible avec les mutations dans *ASIP* et la toison de couleur noire, et enfin LDHH11 en lien possible avec la mutation dans *SOCS2* et son action sur la sensibilité aux mammites et sur le gabarit. Pour les haplotypes MTRDHH3 (mutation dans *FCGR1A*, défaut immunitaire) et LDHH4 (gènes candidats dans les clusters de gènes *HOXC* et *KRT*), ils sont peut-être associés à des problèmes sanitaires et de la mortalité plus tardive, qui écarteraient de fait ces animaux du génotypage pour la sélection génomique.

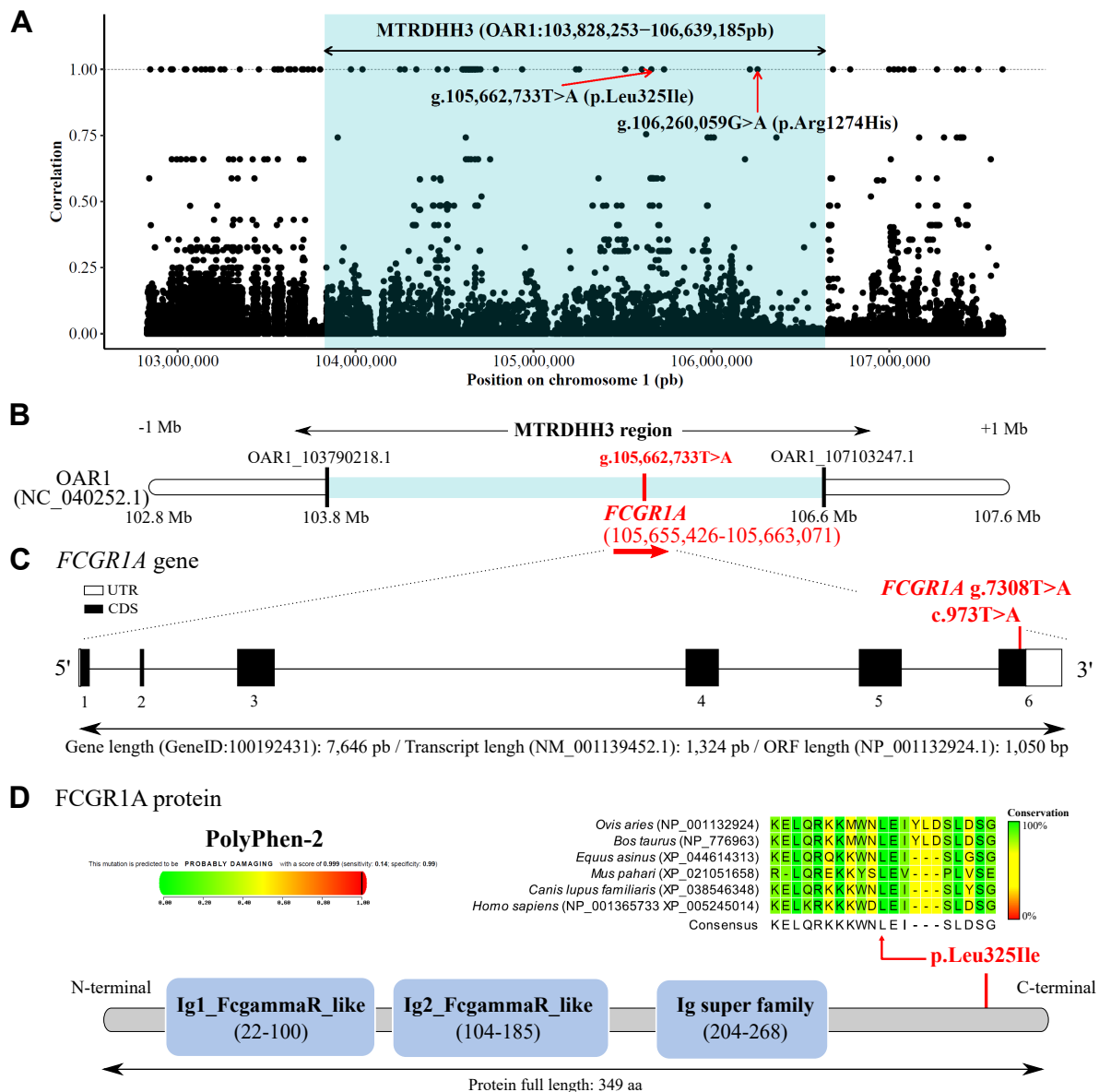


FIGURE 12.12 – Identification de deux mutations faux-sens OAR1 :g.105,662,733T>A dans le gène *FCGR1A* et OAR1 :g.106,260,059G>A dans le gène *RPRD2* associées à l'haplotype MTRDHH3 (A) Nuage de points montrant la corrélation entre le statut MTRDHH3 et les génotypes des variants pour les 100 animaux WGS. Chaque point représente un variant. (B) Haplotype MTRDHH3 (NC_040252.1, OAR1 : 103 828 253 - 106 639 185pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et dernier marqueur de la puce Illumina Ovine SNP50 définissant l'haplotype MTRDHH3. (C) Structure du gène *FCGR1A* (GeneID : 100192431) et localisation de la mutation *FCGR1A* g.7308T>A (c.973T>A) identifiée dans le dernier exon (NM_001139452.1). (UTR : région non traduite; CDS : séquence codante). (D) Protéine FCGR1A (NP_001132924.1) composée de trois domaines : "First immunoglobulin (Ig)-like domain of Fcgamma-receptors (FcgammaRs), and similar domains" (Ig1_FcgammaR_like ; cd05752), "Second immunoglobulin (Ig)-like domain of Fcgamma-receptors (FcgammaRs), and similar domains" (Ig2_FcgammaR_like ; cd05753), "Immunoglobulin domain" (Ig super family ; cl11960). La mutation entraîne une substitution d'une Leucine en Isoleucine (p.Leu325Ile).

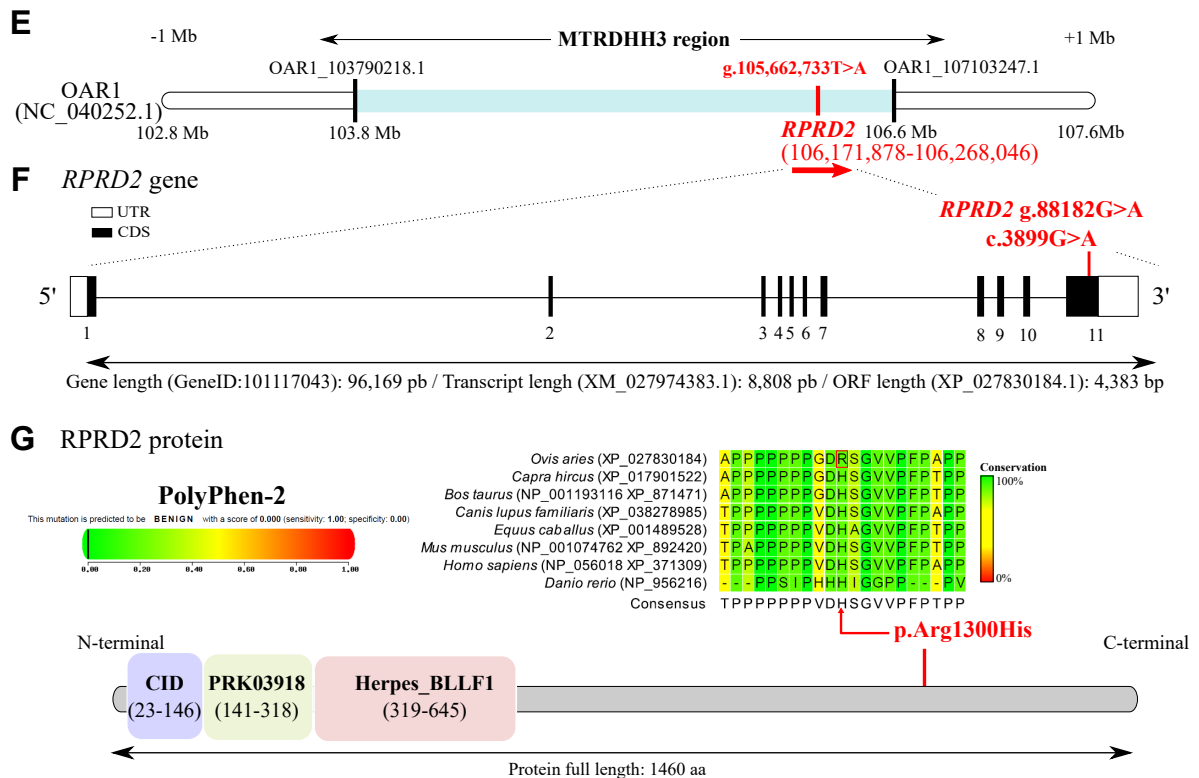


FIGURE 12.13 – (Identification de deux mutations faux-sens OAR1 :g.105,662,733T>A dans le gène *FCGR1A* et OAR1 :g.106,260,059G>A dans le gène *RPRD2* associées à l'haplotype MTRDHH3 (E) Haplotype LDHH3 (NC_040254.1, OAR3 : 33 708 561 - 36 596 529pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et dernier marqueur de la puce Illumina Ovine SNP50 définissant l'haplotype LDHH3. (F) Structure du gène *RPRD2* (GeneID : 101117043) et localisation de la mutation *RPRD2* g.88182G>A (c.3899G>A) identifiée dans le dernier exon (XM_027974383.1). (UTR : région non traduite ; CDS : séquence codante). (G) Protéine RPRD2 (XP_027830184.1) composée de trois domaines : "CID (CTD-Interacting Domain) of Regulation of nuclear pre-mRNA domain-containing protein 2" (CID_RPRD2; cd17001), "DNA double-strand break repair ATPase Rad50" (PRK03918) et "Herpes virus major outer envelope glycoprotein (BLLF1)" (Herpes_BLLF1); pfam05109). La mutation entraîne une substitution d'une Arginine en Histidine (p.Arg1300His).

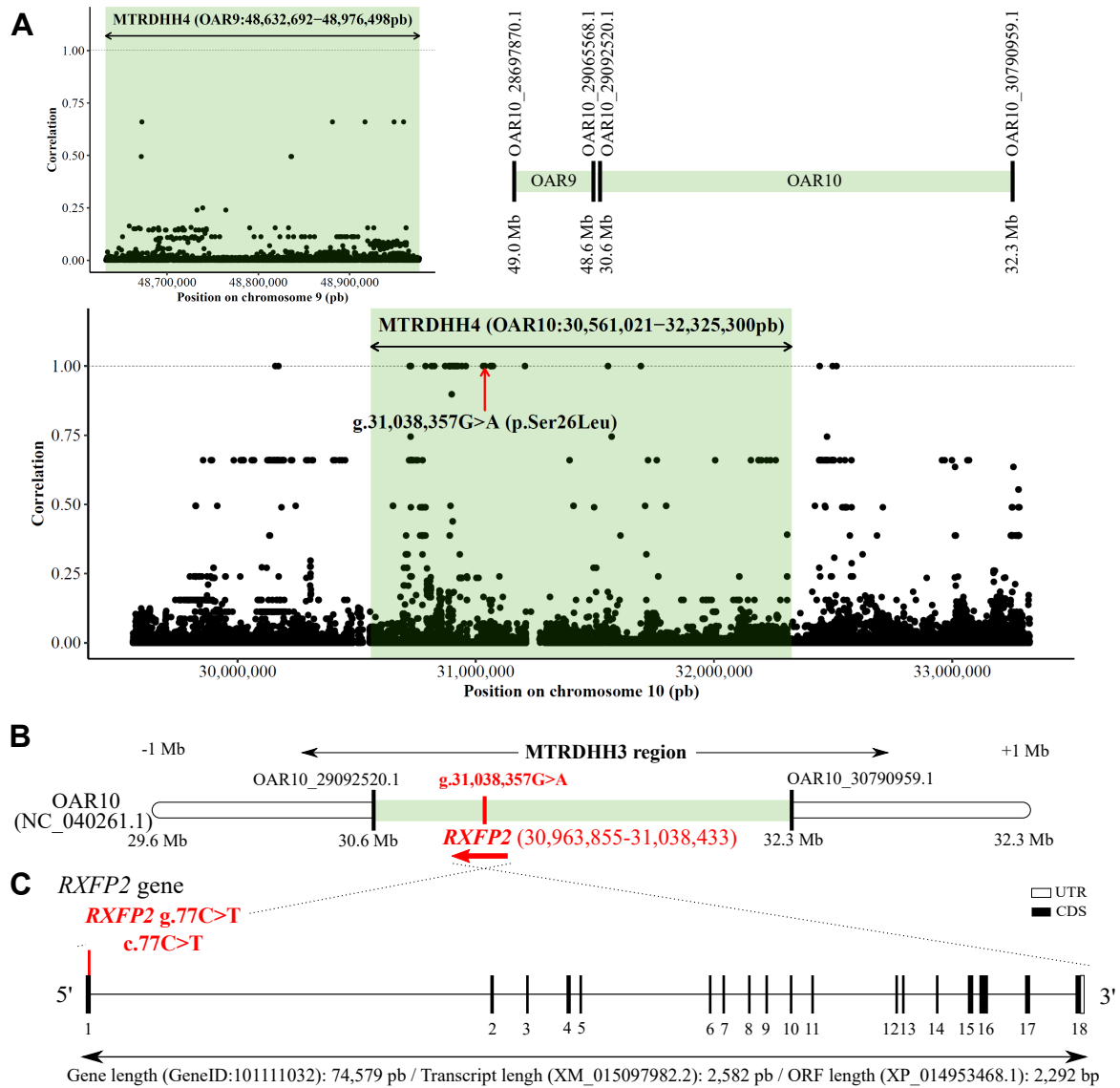


FIGURE 12.14 – Identification d’une mutation faux-sens OAR10 :g.31,038,357G>A dans le gène *RAFP2* associée à l’haplotype MTRDHH4 (A) Nuage de points montrant la corrélation entre le statut MTRDHH4 et les génotypes des variants pour les 100 animaux WGS. Chaque point représente un variant. (B) Haplotype MTRDHH4 (NC_040261.1, OAR10 : 30 561 021 - 32 325 300pb et OAR9 : 48 632 692 - 48 976 498pb) étendu de chaque côté de 1 Mb. Les barres noires indiquent le premier et dernier marqueur de la puce Illumina Ovine SNP50 définissant l’haplotype MTRDHH4. (C) Structure du gène *RAFP2* (GeneID : 101111032) et localisation de la mutation *RAFP2* g.77C>T (c.77C>T) identifiée dans le premier exon (XM_015097982.2). (UTR : région non traduite ; CDS : séquence codante).

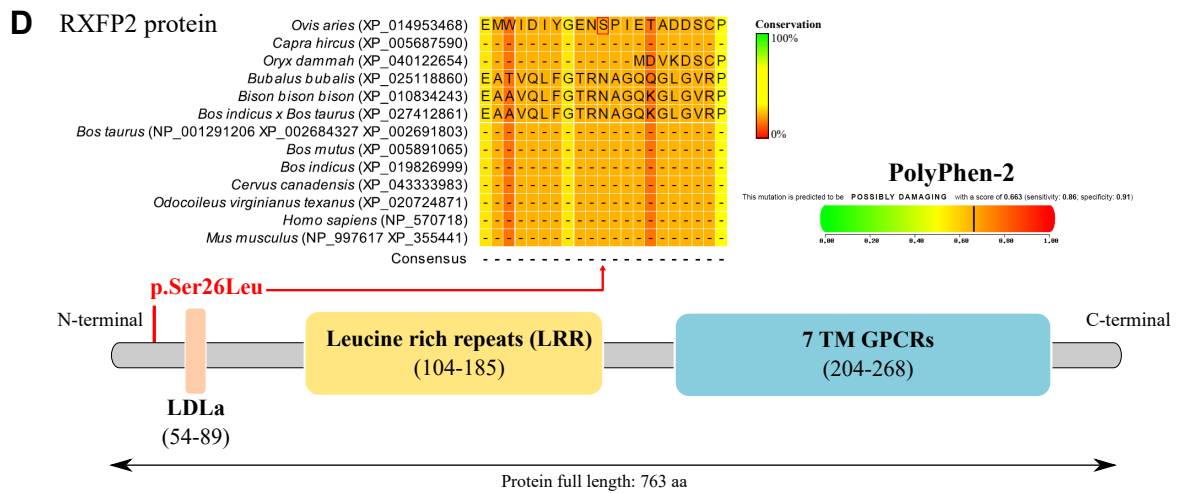


FIGURE 12.15 – Identification d’une mutation faux-sens OAR10 :g.31,038,357G>A dans le gène *RXFP2* associée à l’haplotype MTRDHH4 (D) Protéine RXFP2 (XP_014953468.1) composée d’un domaine "Low Density Lipoprotein Receptor Class A domain" (LDLa; cd00112), d’un domaine Leucine-rich repeat (LRR) protein (LRR; COG4886) et de 7 domaines trans-membranaires "seven-transmembrane G protein-coupled receptor superfamily" (7tm_GPCRs; cl28897). La mutation entraîne une substitution d’une Serine en Leucine (p.Ser26Leu).

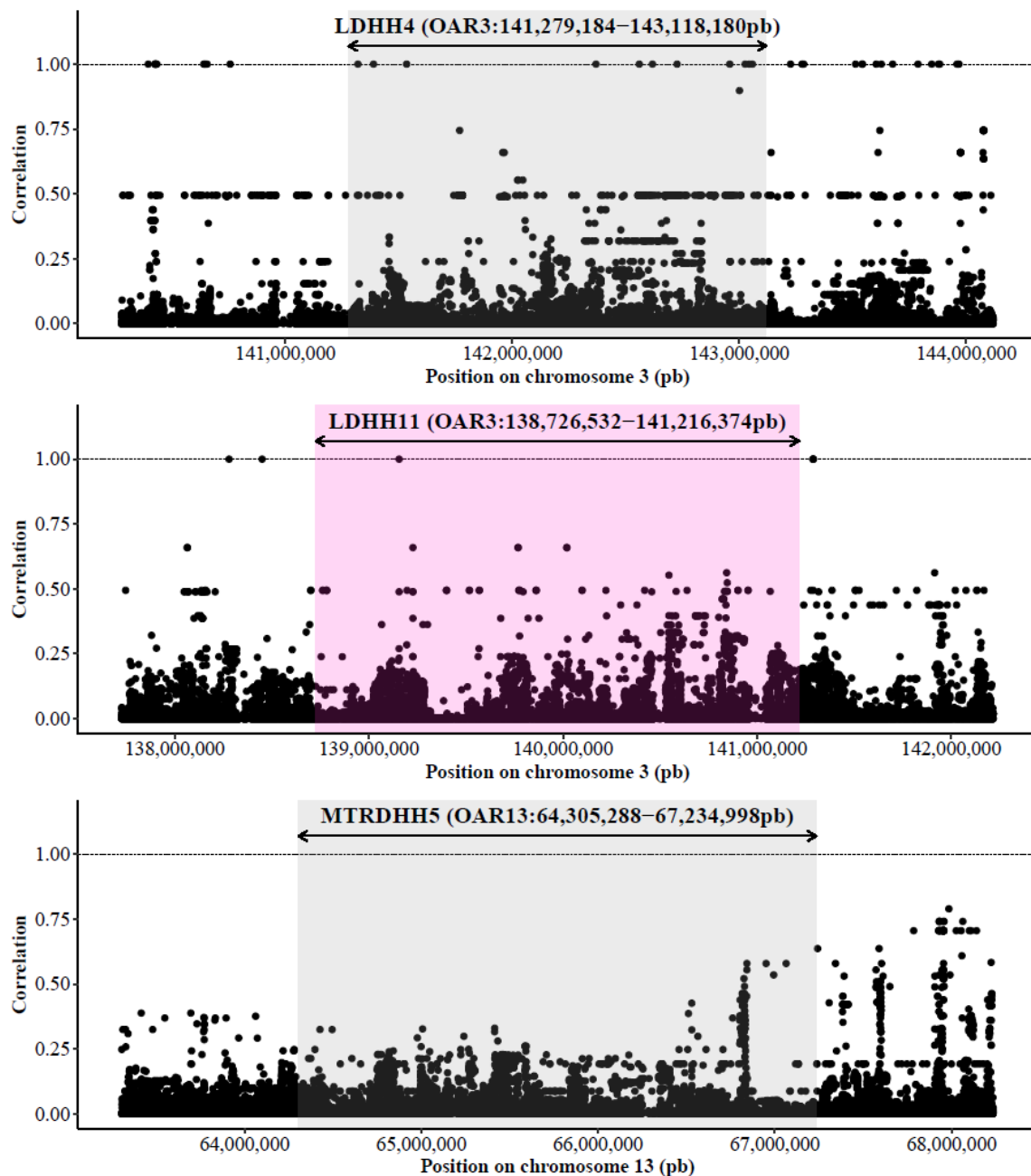


FIGURE 12.16 – **Haplotypes ne présentant pas de variants candidats.** Nuage de points montrant la corrélation entre le statut et les génotypes des variants pour les animaux WGS pour les haplotypes LDHH4, LDHH11 et MTRDHH5. Chaque point représente un variant. A noter que 36 variants (LDHH4) et 6 variants (LDHH11) sont en parfait déséquilibre de liaison avec les haplotypes correspondant mais aucun de ces variants n'est prédit avec un impact fort ou modéré. Pour MTRDHH5, aucun variant ne présente une corrélation de $r=1$.

12.3 Développement de tests de génotypages spécifiques aux mutations candidates

12.3.1 Développement d'un test spécifique de génotypage

Afin de pouvoir étudier les mutations d'intérêts, des tests de génotypages spécifiques ont été développés. Pour cela, deux méthodes sont utilisées pour génotyper une mutation : amplification par PCR puis digestion par une enzyme de restriction (RFLP, *Restriction Fragment Length Polymorphism*) et le génotypage par PCR allèle spécifique (PACE, *PCR allele competitive extension*). Pour amplifier chacun des fragments d'intérêts, un couple d'amorces a été défini à l'aide de l'outil Primer3plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) et l'enzyme de restriction a été choisie à l'aide de l'outil WatCut (<http://watcut.uwaterloo.ca/template.php>). Pour le génotypage allèle spécifique, le choix des amorces est fourni par 3CR Bioscience (<https://3crbio.com>). L'ensemble des méthodes de génotypages est décrit dans l'Annexe : Table 21.1.

12.3.2 Validation à grande échelle des mutations candidates

La validation fonctionnelle des mutations candidates est une étape longue qui nécessite d'avoir accès à des échantillons biologiques d'animaux homozygotes et qui peut s'avérer difficile dans le cas de mutations létales. Toutefois, une validation statistique sur un grand nombre d'individus génotypés permet de confirmer à grande échelle le lien entre le statut à l'haplotype et le statut à la mutation candidate. Nous avons acquis les ADN de la cohorte 2021 des ovins laitiers génotypés ($n=4\ 470$ agneaux dont 3 004 Lacaune et 743 MTR), et ceux des béliers actifs des haras des CIA Lacaune et Manech Tête Rousse permettant de réaliser cette association.

12.3.2.1 Haplotypes en lien avec des mutations supposées létales

Grâce à cette banque d'ADN, j'ai pu mesurer le degré d'association entre le statut aux haplotypes en déficit d'animaux homozygotes et les génotypes aux mutations candidates pour les régions supposées abriter des mutations létales dans les gènes *ORC5* (LDHH1), *IDI1* (LDHH2), *CCDC65* (LDHH6) en Lacaune et *MMUT* (MTRDHH1), *SLC33A1* (MTRDHH2) et *FCGR1A* (MTRDHH3) en Manech Tête Rousse. Pour l'ensemble de ces régions, on constate une très bonne association entre le statut à l'haplotype et les génotypes à la mutation pour chacune des six mutations candidates (Figure 12.17). Pour chaque haplotype, tous les porteurs hétérozygotes de l'haplotype (DHH/+) sont hétérozygotes à la mutation et les animaux non porteurs sont homozygotes non porteurs de la mutation. À noter que pour LDHH6, un animal est homozygote à l'haplotype et à la mutation (Figure 12.17 C). Je reviendrai sur cet agneau dans la Partie V. Pour la plupart des régions, on retrouve quelques animaux non porteurs des haplotypes qui sont hétérozygotes aux mutations. Une attention particulière à ces animaux permet de mettre en évidence qu'ils sont porteurs de versions recombinantes plus courtes des haplotypes en déficit.

12.3.2.2 Haplotypes en lien avec des mutations connues en ovins associées à des caractères de morphologie

Pour quatre haplotypes en déficit (LDHH4, LDHH11 en Lacaune et MTRDHH4, MTRDHH5 en MTR), on fait l'hypothèse que les déficits sont liés à des mutations connues, en sélection dans les populations.

Dans le cas de LDHH4 et LDHH11 proches sur OAR3, j'ai supposé un lien avec la mutation

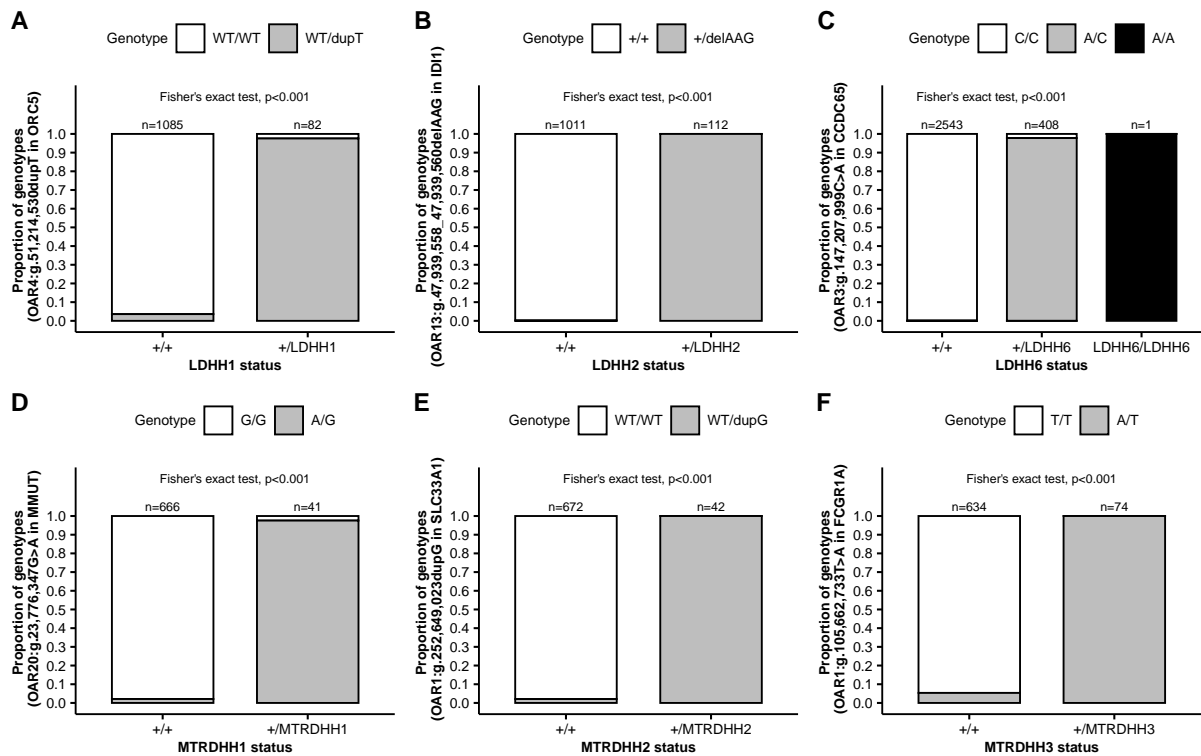


FIGURE 12.17 – Lien entre le statut aux haplotypes et les mutations candidates identifiées pour les régions associées à des mutations létales.

faux-sens g.139,302,270C>T (Oar_rambouillet_v1.0) identifiée dans le gène *SOCS2* en race Lacaune. Cette mutation est présente sur la puce de génotypage basse densité. En faisant l'association entre le statut aux haplotypes LDHH4 et LDHH11 avec la mutation, on constate que tous les animaux hétérozygotes porteurs des haplotypes sont hétérozygotes (C/T) à la mutation (Figure 12.18 A et B). Ainsi, je fais l'hypothèse que ces deux haplotypes sont liés partiellement à la mutation dans *SOCS2*. Dans le groupe des non porteurs, on retrouve 20% des animaux C/T (LDHH4 et LDHH1), suggérant que plusieurs haplotypes signent la mutation.

Dans le cas de MTRDHH4, on suppose un lien avec l'insertion de 1,8kb dans le 3'UTR du gène *RXFP2* responsable de l'absence de cornes. Cette mutation est également présente sur la puce basse densité. L'analyse d'association entre le statut à MTRDHH4 et la mutation dans *RXFP2* montre qu'aucun animal n'est sauvage (délétion à l'état homozygote, +/+). Le déficit lié à MTRDHH4 semble être en relation avec l'allèle inséré (As) mais l'association n'est pas totale. Il est donc difficile de conclure sur le lien direct avec cette mutation connue.

Pour MTRDHH5, j'avais émis l'hypothèse que ce déficit pourrait être en lien avec la couleur noire de toison puisque les animaux noirs sont disqualifiés. C'est en effet un critère de non-conformité avec le standard de la race MTR. Ces animaux ne sont donc jamais génotypés pour être sélectionnés comme reproducteurs. A ce jour, deux mutations connues dans le gène *ASIP* : OAR13 :g.66,475,132_66,475,136del (Délétion 5pb) et g.66,474,980T>A (SNP, p.Cys126Ser) ségrégent dans la race Manech Tête Rousse. La délétion est un variant présent sur la puce basse densité et un test de génotypage spécifique a été développé pour le SNP. L'analyse d'association suggère un lien entre MTRDHH5 et la délétion dans *ASIP* puisque 88% des MTRDHH5/+ sont hétérozygotes à la mutation et aucun animal ins/ins n'est hétérozygote à l'haplotype. A l'inverse,

MTRDHH5 ne semble pas être lié au SNP dans *ASIP* puisque l'on retrouve une distribution des génotypes similaire entre les porteurs hétérozygotes et non-porteurs de l'haplotype.

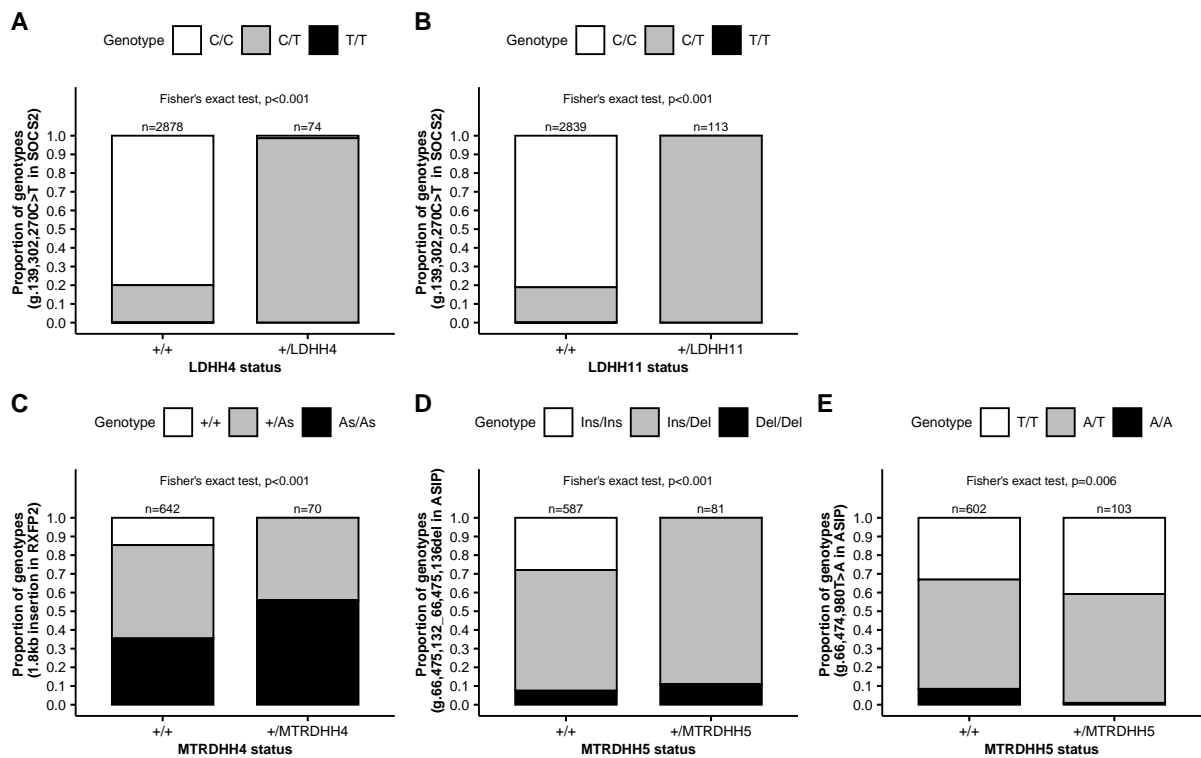


FIGURE 12.18 – Lien entre le statut aux haplotypes et les mutations connues en ovins pour les régions associées à des caractères de morphologie.

A retenir Dans les douze haplotypes en déficit détectés (LDHH1, LDHH2, LDHH3, LDHH4, LDHH6, LDHH9 et LDHH11 en Lacaune et MTRDHH1, MTRDHH2, MTRDHH3, MTRDHH4 et MTRDHH5 en MTR), j'ai pu mettre en évidence des mutations candidates dans les régions supposées abriter des mutations létales embryonnaires (LDHH1/*ORC5*, LDHH2/*IDI1*), ou tout au long de la gestation (LDHH9/*EDC3*, MTRDHH2/*SLC33A1*), ou au stade périnatal (LDHH3/*PREB/GPN1*, LDHH6/*CCDC65* et MTRDHH1/*MMUT*), ou au stade juvénile (MTRDHH3/*FCGR1A*). Cependant, aucune mutation candidate pertinente n'a été mise en évidence dans les régions LDHH4, LDHH11, MTRDHH4 et MTRDHH5. Toutefois, des mutations déjà connues affectant des gènes associés à des caractères de morphologie semblent pouvoir conduire aux déficits observés dans ces régions : sensibilité/résistance aux mammites et croissance (lien LDHH4/LDHH11 avec *SOCS2*) et la couleur de toison noire (entre MTRDHH5 et *ASIP* (del5pb)). Il est difficile de conclure au lien entre MTRDHH4 et la mutation cornes/sans cornes dans *RXFP2*. Il est à noter que tous les haplotypes qui seraient liés à des caractères de morphologie sont en déséquilibre partiel avec les mutations en sélection. ■

V Validation de la causalité des mutations candidates

13 Anomalies génétiques induisant de la mortalité pendant la gestation 193

- 13.1 Croisements *ORC5* et *IDI1* en Lacaune
- 13.2 Croisements *SLC33A1* en Manech Tête Rousse

14 Anomalies génétiques induisant de la mortalité périnatale . 207

- 14.1 Croisements *MMUT* en Manech Tête Rousse
- 14.2 Croisements *CCDC65* en Lacaune

15 Anomalies génétiques en lien avec des défauts morphologiques 217

- 15.1 *SOCS2* : stature et sensibilité aux mammites
- 15.2 *ASIP* : coloration de la toison
- 15.3 *RXFP2* : cornage et cryptorchidie

L'ensemble des résultats décrits dans cette **Partie V**, repose sur les articles suivants (disponibles dans la [Partie VII - Productions scientifiques](#)) :

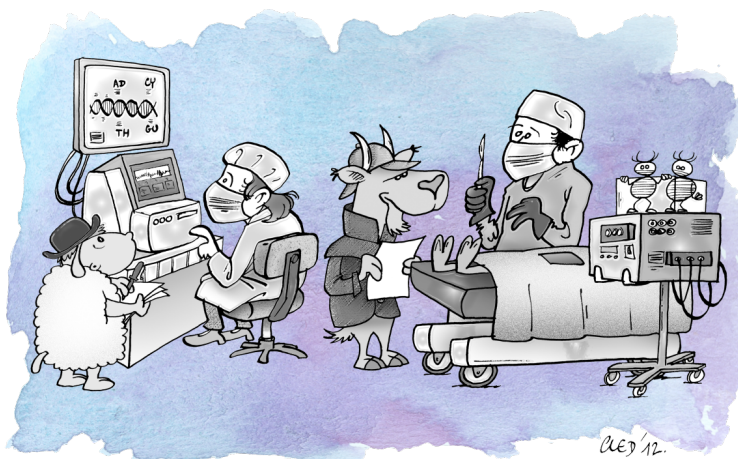
Article 1 (publié) - Ben Braiek et al., 2021 - *Genetics Selection Evolution* - **Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep.**

Article 2 (publié) - Ben Braiek et al., 2022 - *Genes* - **A Nonsense Variant in *CCDC65* Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep.**

Article 3 - Ben Braiek et al., 2023a - bioRxiv, rédigé pour *Plos Genetics* - **Homozygous haplotype deficiency in Manech Tête Rousse dairy sheep revealed a nonsense variant in *MMUT* gene affecting newborn lamb viability.**

Article 4 - Ben Braiek et al., 2023b - bioRxiv, rédigé pour *Animal genetics* - **A single base pair duplication in *SLC33A1* gene causes fetal losses and neonatal lethality in Manech Tête Rousse dairy sheep.**

Des résultats complémentaires ont également été ajoutés.



13. Anomalies génétiques induisant de la mortalité pendant la gestation

Dans ce chapitre, l'objectif est d'apporter des éléments permettant de valider la causalité des mutations identifiées dans les gènes *ORC5* (LDHH1), *IDI1* (LDHH2) identifiés en Lacaune et *SLC33A1* (MTRDHH2) identifié en Manech Tête Rousse. D'après les analyses précédentes, elles sont supposées altérer le développement embryonnaire précoce ou plus tardivement pendant la gestation. Pour répondre à cet objectif, des accouplements à risque entre animaux hétérozygotes porteurs des mutations *ORC5* ou *IDI1* ou *SLC33A1* ont été réalisés afin de confirmer la baisse de réussite à l'IA. Les diagnostics de gestation ont été réalisés par échographie entre 45 et 60 jours de gestation, et par un test moléculaire précoce à 15 jours de gestation. L'ensemble de ces procédures expérimentales sur animaux a fait l'objet d'une autorisation de projet (APAFIS30615-2021032318054889 v4).

13.1 Croisements *ORC5* et *IDI1* en Lacaune

13.1.1 Accouplements à risque et diagnostic de gestation par échographie

TABLE 13.1 – Réussite à l'insémination et nombre d'agneaux obtenus à la mise-bas dans les accouplements à risque *ORC5*, *IDI1* et témoin.

Mutation étudiée	Nombre de brebis inséminées	Taux réussite IA (%)	Nombre d'agneaux	+/+	m/+	m/m	Prolificité
c.839dupA <i>ORC5</i>	9	77,8%	11	8	3	0	1,6
c.287_289delAAG <i>IDI1</i>	4	50,0%	6	1	5	0	3,0
Témoin	13	46,2%	8	-	-	-	1,3

Un troupeau de Lacaune lait de 80 animaux (brebis de réforme achetées pour un autre protocole expérimental) était présent sur l'installation expérimentale de Langlade (INRAE GenPhySE, Pompertuzat, 31450). Ces animaux ont été génotypés aux mutations d'intérêts. Ainsi, 9 brebis étaient hétérozygotes à la mutation dans le gène *ORC5* et 4 brebis étaient hétérozygotes à la mutation dans le gène *IDI1* (Table 13.1). Ces brebis ont été inséminées avec de la semence de béliers hétérozygotes pour chacune des mutations. Un groupe témoin de 13 brebis (brebis non porteuses des deux mutations) a été inséminé avec de la semence de béliers non porteurs des

deux mutations. Afin de déterminer l'état de gestation des 26 brebis, la mise au point d'un diagnostic moléculaire de gestation a été envisagé à partir de prises de sang total réalisées tous les 7 jours à partir de l'IA jusqu'à J49, en complément d'un diagnostic par échographie à J45 après IA. Les résultats échographiques indiquent que le taux de réussite à l'IA est de 77,8% dans les accouplements *ORC5*, 50% dans les accouplements *IDI1* et de 46,2% dans le lot témoin. Sachant que le taux de réussite à l'IA dans la race Lacaune est de 70%, la réussite à l'IA des accouplements *ORC5* (n=9) est augmentée de 8% mais est réduite de 20% dans les accouplements *IDI1* (n=4). De manière surprenante, le taux de réussite à l'IA du lot témoin est seulement de 46,2%. Ces résultats restent exploratoires puisque les effectifs des croisements à risque sont faibles. Toutefois, à la mise-bas, aucun animal homozygote aux mutations dans *ORC5* et *IDI1* n'est observé (Table 13.1).

13.1.2 Choix d'un test moléculaire de gestation précoce

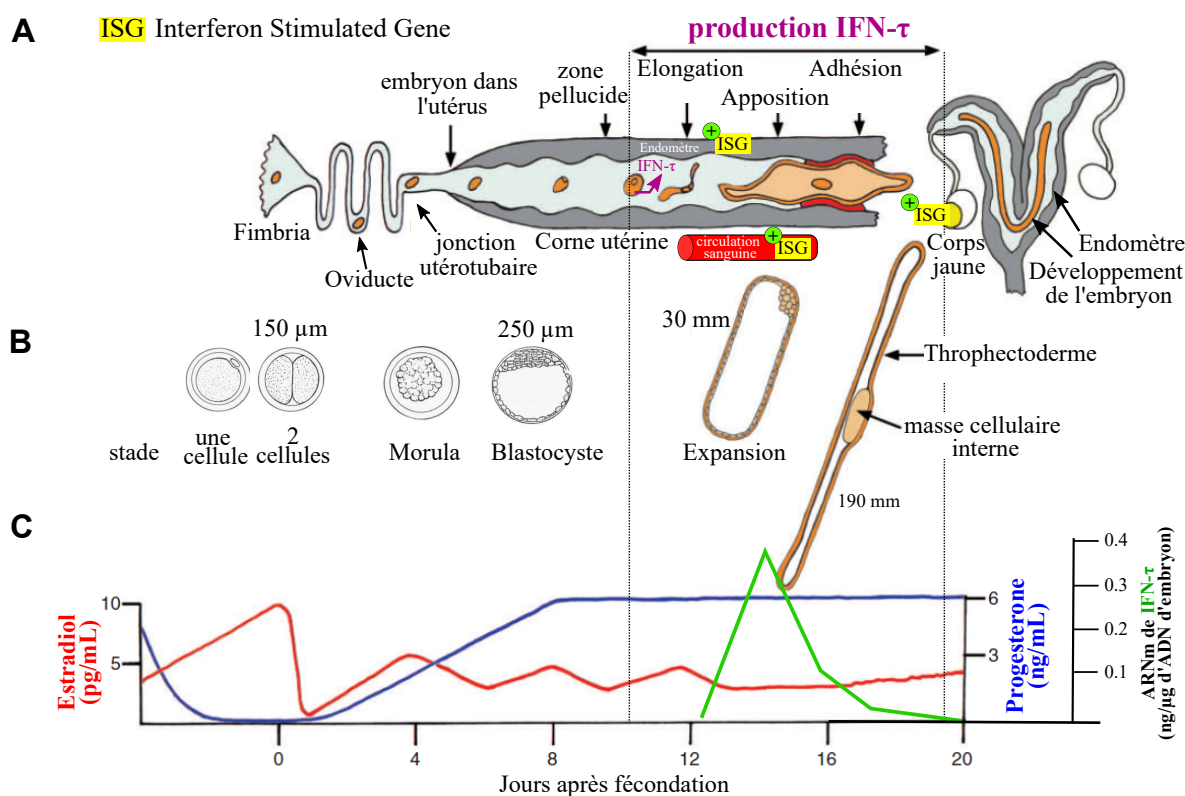


FIGURE 13.1 – Développement du conceptus au cours des 20 premiers jours après fécondation dans le tractus génital maternel ovin. (A) Évolution du développement du conceptus au cours de la gestation (B) Différents stades de développement de l'embryon. Chez les ruminants, le blastocyste s'allonge lors des dernières étapes après l'implantation. (C) Evolution de la concentration en estradiol, progestérone et d'IFN- τ au cours de la gestation. *Adapté de* (Spencer et al., 2006).

Chez la brebis, un diagnostic de gestation négatif réalisé classiquement par échographie entre 45 et 60 jours post-fécondation (visualisation nette des fœtus) peut se traduire par un échec de fécondation ou par des pertes embryonnaires. Dans la littérature, la plupart des pertes embryonnaires est attendue pendant la période allant de la fécondation (jour 0) à l'implantation du conceptus (jours 12-16) (Bindon, 1971 ; Spencer et al., 2008 ; Wilmut et al., 1986). Lors de

l'implantation, l'embryon se lie à l'endomètre et entraîne la sécrétion d'interféron τ (IFN- τ) comme signal de reconnaissance de la gestation (Bazer, 2013) (Figure 13.1). L'IFN- τ est une cytokine produite lors de la réponse immunitaire de la brebis qui considère l'embryon comme non soi. En réponse, l'expression des gènes stimulés par l'interféron (*ISG*, *Interferon Stimulated Gene*) est fortement augmentée dans l'endomètre mais également dans les cellules immunitaires circulantes dans le sang (Ott & Gifford, 2010). Parmi ces *ISG*, les gènes *MX1* (*Myxovirus-influenza virus resistance 1*), *STAT1* (*Signal transducer and activator of transcription 1*) et *CXCL10* (*Chemokine C-X-C motif ligand 10*), fortement exprimés dans le sang, ont été utilisés pour prédire l'état de gestation au jour 14-15 chez la brebis (Mauffré et al., 2016). De plus, l'implantation est également contrôlée par une hormone stéroïdienne, la progestérone, qui joue un rôle majeur dans l'utérus pour stimuler la survie, la croissance et le développement des blastocystes (Spencer et al., 2008) (Figure 13.1 C).

Afin de répondre à l'objectif de mise au point d'un diagnostic précoce, les prises de sang sériées tous les 7 jours entre le jour de l'IA (J0) et J49 (quelques jours après l'échographie) ont été réalisées sur quelques animaux pour doser la progestérone dans le plasma et suivre par qPCR l'expression des gènes *MX1* et *STAT1*, stimulés par l'IFN- τ .

13.1.2.1 Dosage de la progestérone

La progestérone a été dosée (ELISA, DRG Instruments GmbH, EIA-1561) dans le plasma sanguin sur cinq brebis gestantes et cinq brebis vides à l'échographie (J45) entre la période J0 et J49 tous les 7 jours (Figure 13.2). En dosant la progestérone, on est capable de séparer les brebis gestantes et vides seulement au 21^{ème} et 42^{ème} jour de gestation (test non-paramétrique de Wilcoxon, P-valeur < 0,05).

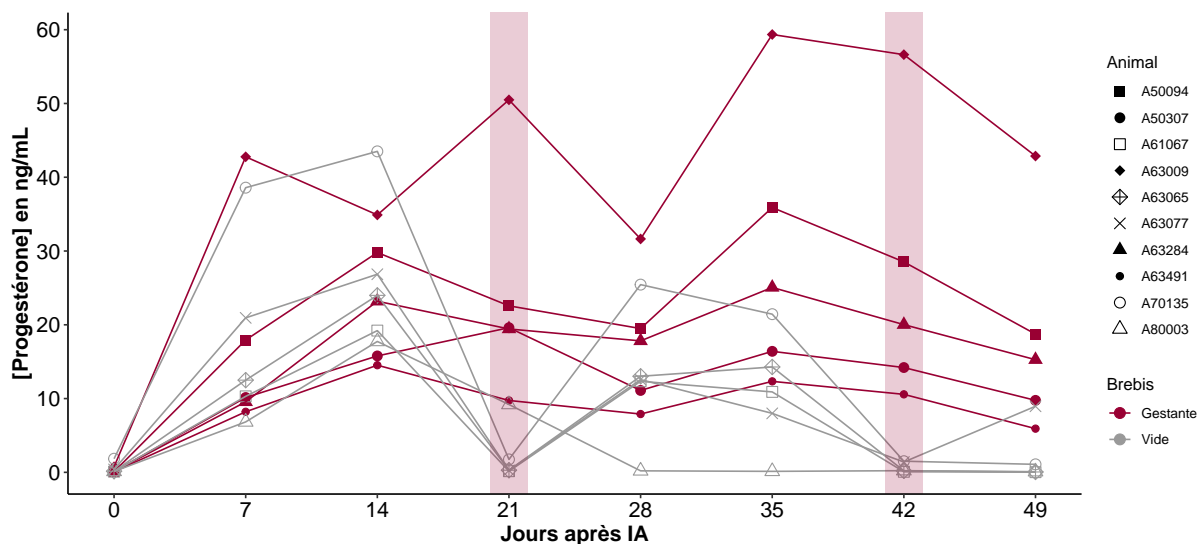


FIGURE 13.2 – Évolution de la progestérone entre 0 à 49 jours après IA

13.1.2.2 Expression par qPCR des gènes *MX1* et *STAT1*

Les ARN totaux ont été extraits du sang des brebis avec le kit Nucleospin® RNA Blood (Macherey-Nagel, Ref 740200.50) selon le protocole du fabricant en partant de 800 μ L de sang total avec un traitement de digestion DNaseI pour éliminer l'ADN génomique contaminant.

Les ARN ont été quantifiés par spectrophotométrie (spectrophotomètre NanoDrop® ND-8000, ThermoFischer) et conservés à -80°C . Après contrôle qualitatif et quantitatif des échantillons d'ARN, la transcription inverse a été réalisée à partir de 500ng d'ARN total en ajoutant une amorce oligo(dT) T22V ancree (1 μL à 100 μM), une amorce oligo-dN9 aléatoire (1 μL à 100 μM) et des dNTP (2 μL à 10mM) dans un volume réactionnel de 54 μL . Ce mélange a été incubé à 65°C pendant 5 min dans un thermocycleur ABI2700 (Applied Biosystems) puis descendu à 4°C . Un second mélange réactionnel (18,5 μL /réaction) contenant le tampon de réaction (14 μL de First strand Buffer 5X, Invitrogen, France), du DTT (Dithiothréitol, 3 μL à 0,1M), de la Rnasine (0,5 μL , 40 unités/ μL , Promega, France) et de la transcriptase inverse Superscript II (1 μL , 200 unités/ μL , Invitrogen, France) a été ajouté à la solution d'ARN dénaturés (volume final de réaction de 72,5 μL) puis incubés 50 minutes à 42°C et placés 15 minutes à 70°C . La solution d'ADN complémentaire (ADNc) obtenue a été directement diluée à un rapport de 1:2 et stockée à -20°C . La PCR quantitative (qPCR) a été réalisée en utilisant 3 μL d'ADNc, 5 μL de SYBR Green PCR en temps réel Master Mix 2X (Applied Biosystems) et 2 μL d'amorces à 3 μM dans un volume de réaction total de 10 μL . Chaque échantillon a été testé en quatre exemplaires. La qPCR a été réalisée sur un système de PCR en temps réel QuantStudio 6 Flex (ThermoFisher). Pour chaque paire d'amorces, l'efficacité d'amplification a été évaluée par $E = e^{-1/\alpha}$ où α est la pente d'une courbe linéaire obtenue à partir d'une dilution en série d'ADNc et des valeurs Ct (seuil de cycle) correspondantes. L'abondance des transcrits d'ARN a été quantifiée à l'aide de la méthode delta Ct (ΔCt) corrigée par deux gènes de référence (*GAPDH*, *YWHAZ*). Les amorces ont été conçues à l'aide de Beacon Designer 8 (Premier Biosoft). La liste des séquences d'amorces qPCR, la longueur d'amplification et l'efficacité sont disponibles dans l'Annexe : Table 21.2.

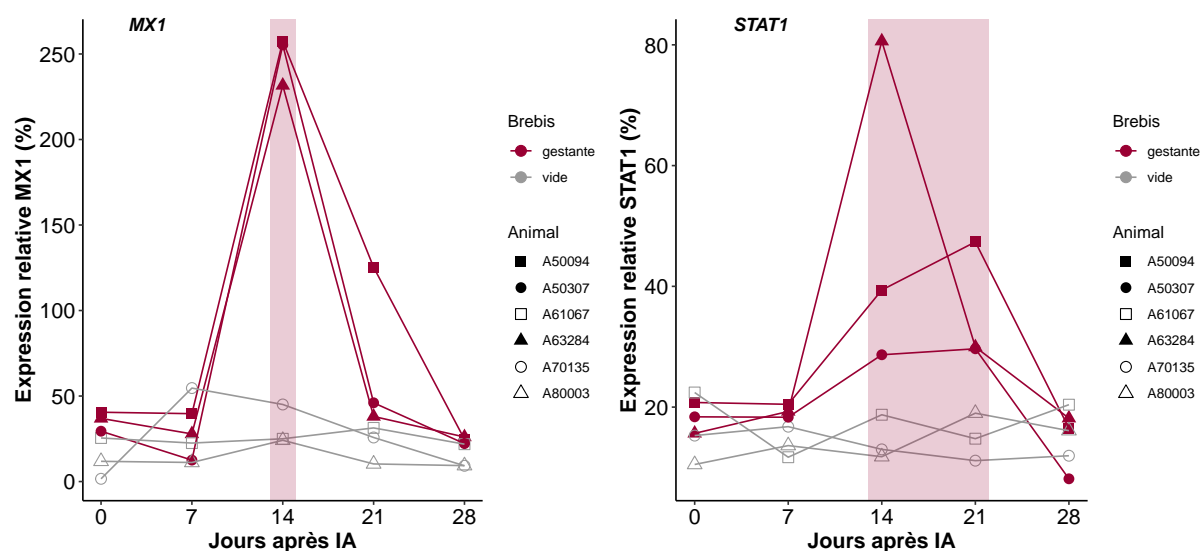


FIGURE 13.3 – Expression relative des gènes *MX1* et *STAT1* entre le 0 et 28^{ème} jour après IA chez des brebis gestantes et vides.

Dans un premier temps, les expressions relatives des gènes *MX1* et *STAT1* entre J0 et J28 après IA de 6 brebis vides et 6 brebis gestantes ont été analysées (Figure 13.3). L'analyse d'expression montre une forte surexpression du gène *MX1* à 14 jours qui permet de distinguer significativement les brebis gestantes et vides (test non-paramétrique de Wilcoxon, P-valeur <

0,05). L'expression de *STAT1* est significativement augmentée entre le 14^{ème} et le 21^{ème} jour après l'IA (test non-paramétrique de Wilcoxon, P-valeur < 0,05).

Avec nos effectifs trop faibles, les croisements à risque effectués ne permettent pas de prouver l'impact sur la réussite à l'IA des mutations identifiées dans les gènes *ORC5* et *ID11*. Cependant, les échantillons biologiques collectés sur les 26 brebis me permettent d'étudier la faisabilité du test moléculaire de gestation basé sur l'expression des gènes *MX1* et *STAT1* permettant un diagnostic plus précoce que la progestérone (14 jours après l'IA contre 21 jours).

13.1.3 Analyse du pouvoir de diagnostic du test moléculaire par qPCR sur l'état de gestation

13.1.3.1 Analyses statistiques

Le statut de gestation de la brebis (gestante/non gestante) a d'abord été déterminé par le diagnostic échographique classiquement utilisé à J45 post-IA. L'évaluation du test moléculaire de diagnostic de gestation (GDMT) est basée sur l'expression des gènes *MX1* et *STAT1* à J14. Les valeurs d'expression de *MX1* et *STAT1* entre les brebis gestantes et vides ont été comparées à l'aide d'un test non-paramétrique de Wilcoxon, et la prédiction du diagnostic a été réalisée par analyse de courbe ROC (*Receiver Operating Characteristic*) à l'aide de l'outil easyROC (Goksuluk et al., 2016). L'aire sous la courbe (AUC) a été évaluée et comparée à la valeur attendue de 0,5 sous l'hypothèse nulle. La détermination de la valeur seuil (discriminer les brebis gestantes et vides) est déterminée par l'index de Youden (qui maximise sensibilité et spécificité). Cette valeur seuil a été utilisée pour classer les animaux en quatre catégories : vrai positif (TP c'est-à-dire GDMT+ et gestantes), faux positif (FP c'est-à-dire GDMT+ et vides), vrai négatif (TN c'est-à-dire GDMT- et gestantes) et faux négatif (FN c'est-à-dire GDMT- et vides).

13.1.3.2 Résultats

Les expressions relatives des gènes *MX1* et *STAT1* sont significativement plus fortes chez les 16 brebis gestantes comparées aux 10 brebis vides (Figure 13.4 A et B). Les analyses ROC donnent des aires sous la courbe de 0,944 (P-valeur= $1,3 \times 10^{-23}$) pour *MX1* et de 0,922 (P-valeur= $6,6 \times 10^{-12}$) pour *STAT1* (Table 13.3). Ces aires sont significativement différentes de 0,5, indiquant que le test moléculaire de diagnostic de gestation basé sur l'expression de *MX1* ou *STAT1* est performant. Le test avec *MX1* possède une sensibilité de 88% et une spécificité de 100% alors que le test avec *STAT1* possède une sensibilité de 73% et une spécificité de 100%. Avec ces valeurs, il est possible de déterminer un seuil d'expression de 156% pour *MX1* et de 9% pour *STAT1* (Figure 13.4 C, D, F et G), permettant de prédire les brebis gestantes avec ce test moléculaire (Figure 13.4 E et H).

TABLE 13.3 – Paramètres d’analyses de la courbe ROC pour le test de diagnostic moléculaire de gestation basé sur l’expression du niveau d’ARNm de *MX1* et *STAT1*. La valeur seuil d’expression (cut-off point) est déterminée par l’index de Youden (Yd). AUC= aire sous la courbe ; se=erreur standard ; Se=Sensibilité ; Sp= Spécificité.

Gène	AUC ± se	P-value	Se	Sp	Yd	Cut-off point
<i>MX1</i>	0,944 ± 0,044	$1,3 \times 10^{-23}$	88	100	0,875	156
<i>STAT1</i>	0,922 ± 0,061	$6,6 \times 10^{-12}$	73	100	0,733	9

L’utilisation de ce test ouvre alors des perspectives pour identifier des pertes embryonnaires précoces quand le test moléculaire de diagnostic de gestation est positif à J14, mais négatif lors du diagnostic échographique réalisé classiquement entre J45 et J60 en élevage.

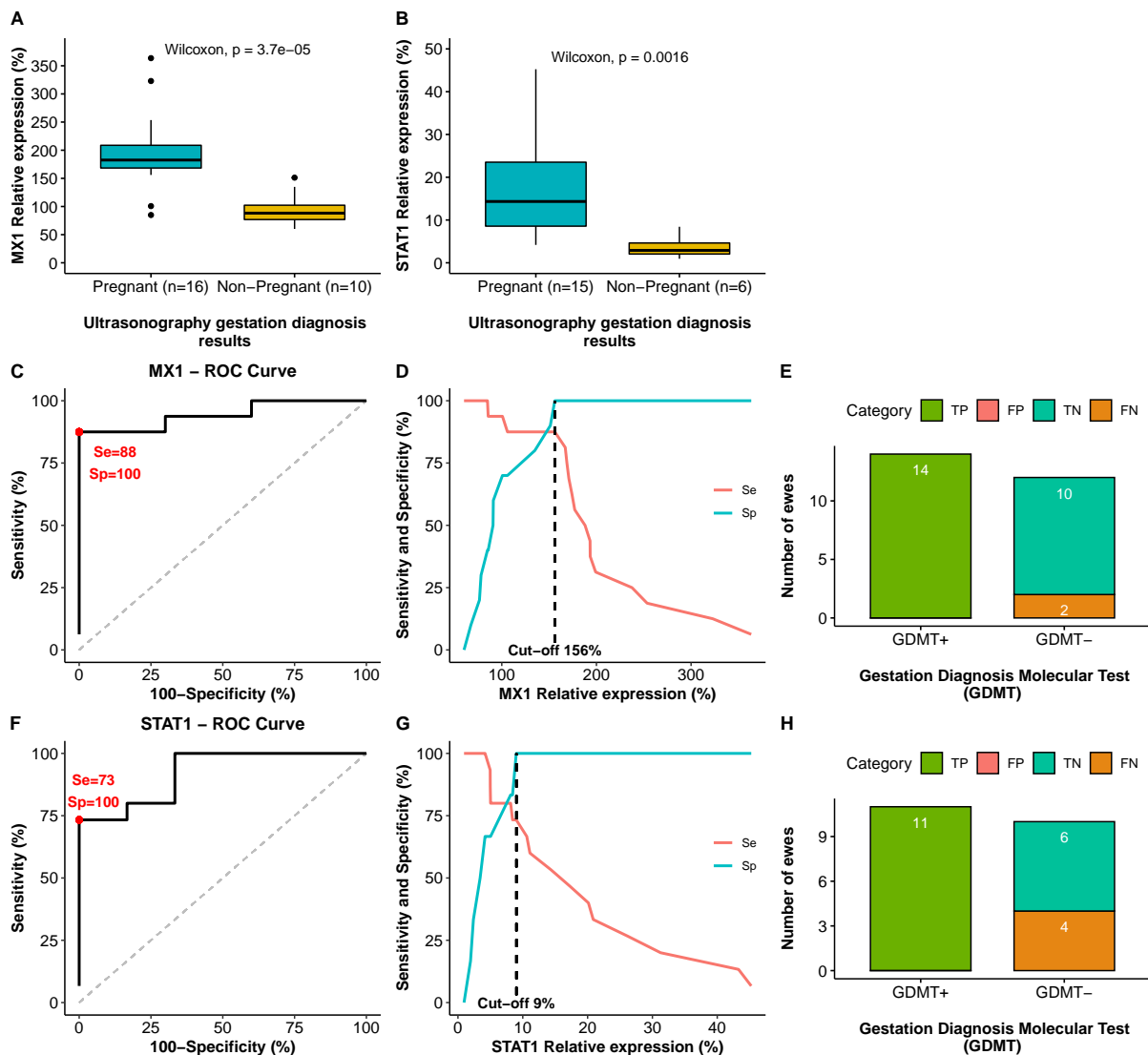


FIGURE 13.4 – Diagnostic moléculaire de gestation précoce basé sur l'expression des gènes *MX1* et *STAT1* en Lacaune. Expression relative des gènes *MX1* (A) et *STAT1* (B) 15 jours après l'IA entre les brebis gestantes ou vides, évaluées 45 jours après l'IA par échographie. (C, F) Courbes ROC représentant la sensibilité (Se) et la spécificité (Sp) du test moléculaire de gestation. (D, G) Détermination de la valeur seuil d'expression relative de *MX1* (156%) associée à une Se=88% et Sp=100% en utilisant la méthode de Youden qui maximise sensibilité et spécificité. (E) Répartition des brebis gestantes et vides selon le diagnostic moléculaire de gestation précoce basé sur l'expression relative de *MX1* (G) Détermination de la valeur seuil d'expression relative de *STAT1* (9%) associée à un Se=73% et Sp=100% en utilisant la méthode de Youden. (H) Répartition des brebis gestantes et vides selon le diagnostic moléculaire de gestation précoce basé sur l'expression relative de *STAT1*.

TP = vrai positif; FP = faux positif, TN = vrai négatif; FN = faux négatif; Yd= Youden index.

13.2 Croisements *SLC33A1* en Manech Tête Rousse

Afin d'étudier la mutation dans *SLC33A1*, un gène qui code pour un transporteur d'un métabolite essentiel au métabolisme énergétique l'acétyl-Coenzyme A, des croisements à risque entre porteurs hétérozygotes de la mutation c.735dupG ont été réalisés dans 6 élevages privés. Pour rappel, cette mutation, appelée par la suite *SLC33A1_dupG*, est supposée associée à de la mortalité tout au long de la gestation (perte fœtale et mortalité périnatale). Cette mutation fait l'objet de l'article 4 - Ben Braiek et al., 2023b - bioRxiv, rédigé pour *Animal genetics* - **A single base pair duplication in *SLC33A1* gene causes fetal losses and neonatal lethality in Manech Tête Rousse dairy sheep**. Les principaux résultats sont repris ci-dessous.

13.2.1 Dispositif expérimental d'accouplements des porteurs *SLC33A1_dupG*

Un ensemble de 419 brebis (dont des filles de béliers porteurs de MTRDHH2) réparties dans 6 élevages privés a été génotypé pour la mutation *SLC33A1_dupG*. Après génotypage, 137 brebis ont été sélectionnées pour constituer deux groupes d'accouplements, un groupe témoin d'accouplements non à risque avec 73 brebis non porteuses accouplées par IA avec des béliers non porteurs, et un groupe d'accouplements à risque avec 64 brebis hétérozygotes pour *SLC33A1_dupG* accouplées par IA avec des béliers hétérozygotes afin de générer des agneaux homozygotes *SLC33A1_dupG*. Pour définir le statut de gestation des brebis, une prise de sang a été réalisée à J15 après l'IA pour le diagnostic moléculaire de gestation et une échographie a été réalisée entre le J45 et J60, par les techniciens d'IA. Le dispositif expérimental est décrit dans la Figure 13.5.

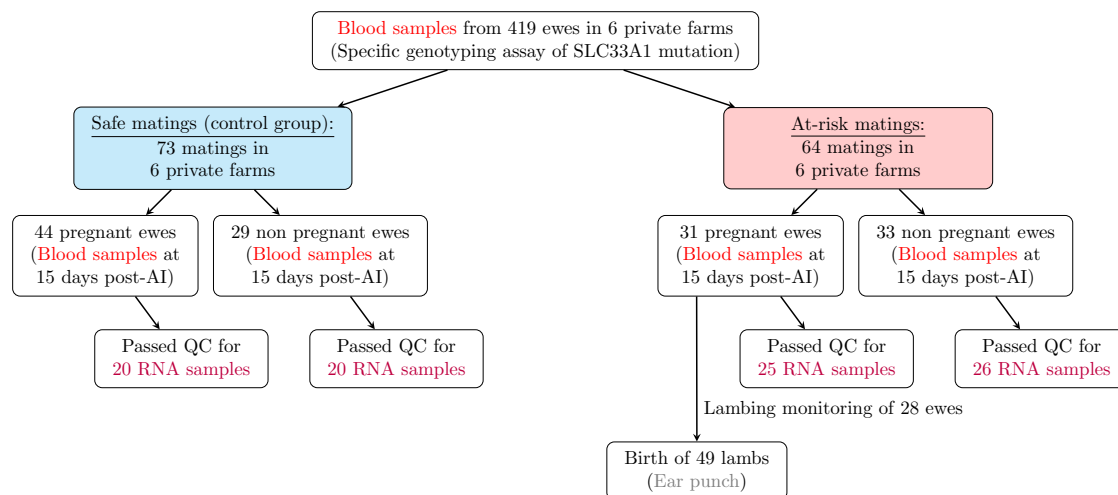


FIGURE 13.5 – Dispositif expérimental. QC = contrôle de la qualité et de la quantité.

13.2.2 Réussite à l'IA et mortalité périnatale

La réussite à l'IA mesurée par échographie est de 60,3% dans le groupe d'accouplements non à risque, ce qui correspond à la réussite globale de l'IA dans la population de Manech Tête Rousse. Cependant, la réussite à l'IA est réduite à 48,4% dans le groupe d'accouplements à risque entre porteurs hétérozygotes de *SLC33A1_dupG* (Figure 13.6). Ce résultat confirme les observations faites au niveau populationnel avec MTRDHH2 et laisse supposer des pertes embryonnaires avant

J45 de gestation, étudiées dans le [paragraphe 13.2.3](#).

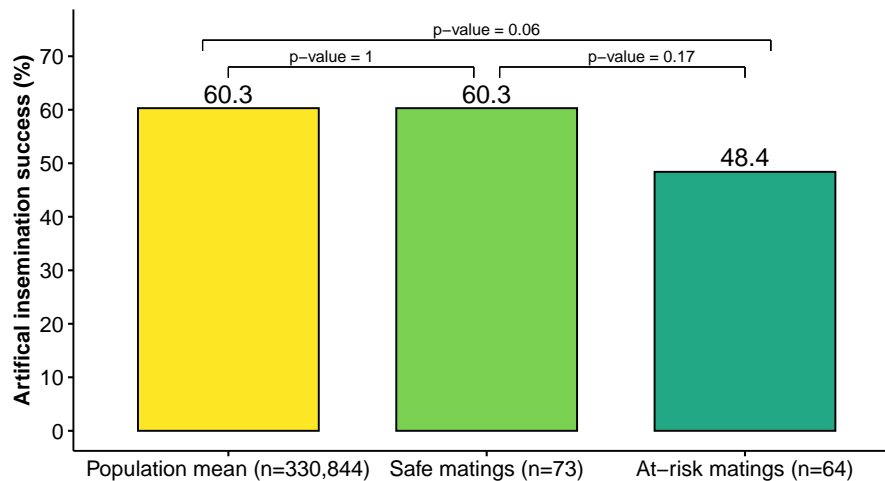


FIGURE 13.6 – **Taux de réussite à l’insémination artificielle des Manech Tête Rousse.** La moyenne de la population (*population mean*) a été établie dans l’article 3 - Ben Braiek et al., 2023a (*Plos Genetics*) à l’aide des enregistrements d’accouplements de la population (n = 330 844). Le succès de l’insémination artificielle est basé sur la date d’agnelage en fonction de la durée de gestation à partir du jour de l’IA (151 ± 7 jours). Dans les accouplements témoins (*safe mating*) et à risque (*at-risk mating*) réalisés dans les 6 élevages privés (n=137 brebis), le diagnostic de gestation a été réalisé par échographie entre les jours 45 et 60. Les p-valeurs sont obtenues par un test exact de Fisher.

Parmi les 31 brebis gestantes du groupe d’accouplements à risque, 28 brebis ont eu un suivi complet (observations attentives des éleveurs) tout au long de la gestation jusqu’au sevrage à l’âge de 1 mois. Aucun avortement n’a été observé entre le moment de l’échographie (jour 45-60 après IA) et la mise-bas. La durée de la gestation varie entre 147 et 154 jours conformément à la moyenne de la population de Manech Tête Rousse (151 ± 7 jours) et 49 agneaux sont nés ([Figure 13.7a](#) et [Table 13.5](#)). Au cours de la période de croissance de 0 à 30 jours, 18 agneaux sont morts, correspondant à un taux de mortalité de 34,7% (en excluant un animal au génotype indéterminé, aucun tissu n’a été prélevé) ([Figure 13.7b](#)). Le tableau de contingence ([Table 13.5](#)) entre les génotypes des agneaux et la viabilité (vivant ou mort) indique un taux de viabilité significativement plus faible pour les agneaux homozygotes dupG/dupG (test exact de Fisher, P-valeur<0,001). Les agneaux homozygotes *SLC33A1_dupG* contribuent à 47% de la mortalité dans les accouplements à risque et la mortalité est survenue principalement dans les cinq premiers jours après l’agnelage. Un agneau mort-né ([Figure 13.7c](#)) a été expulsé lors de la mise bas (dans le placenta) au terme de la gestation, et il a montré un arrêt de développement caractéristique d’un stade de mi-gestation. La plupart des animaux homozygotes est morte entre 1 et 5 jours ([Figure 13.7d](#)). Un seul agneau homozygote est mort entre 5 et 30 jours. Cet agneau présentait une faiblesse et une raideur des muscles des jambes et des déformations de la colonne vertébrale entraînant de graves problèmes locomoteurs ([Figure 13.7e](#)).

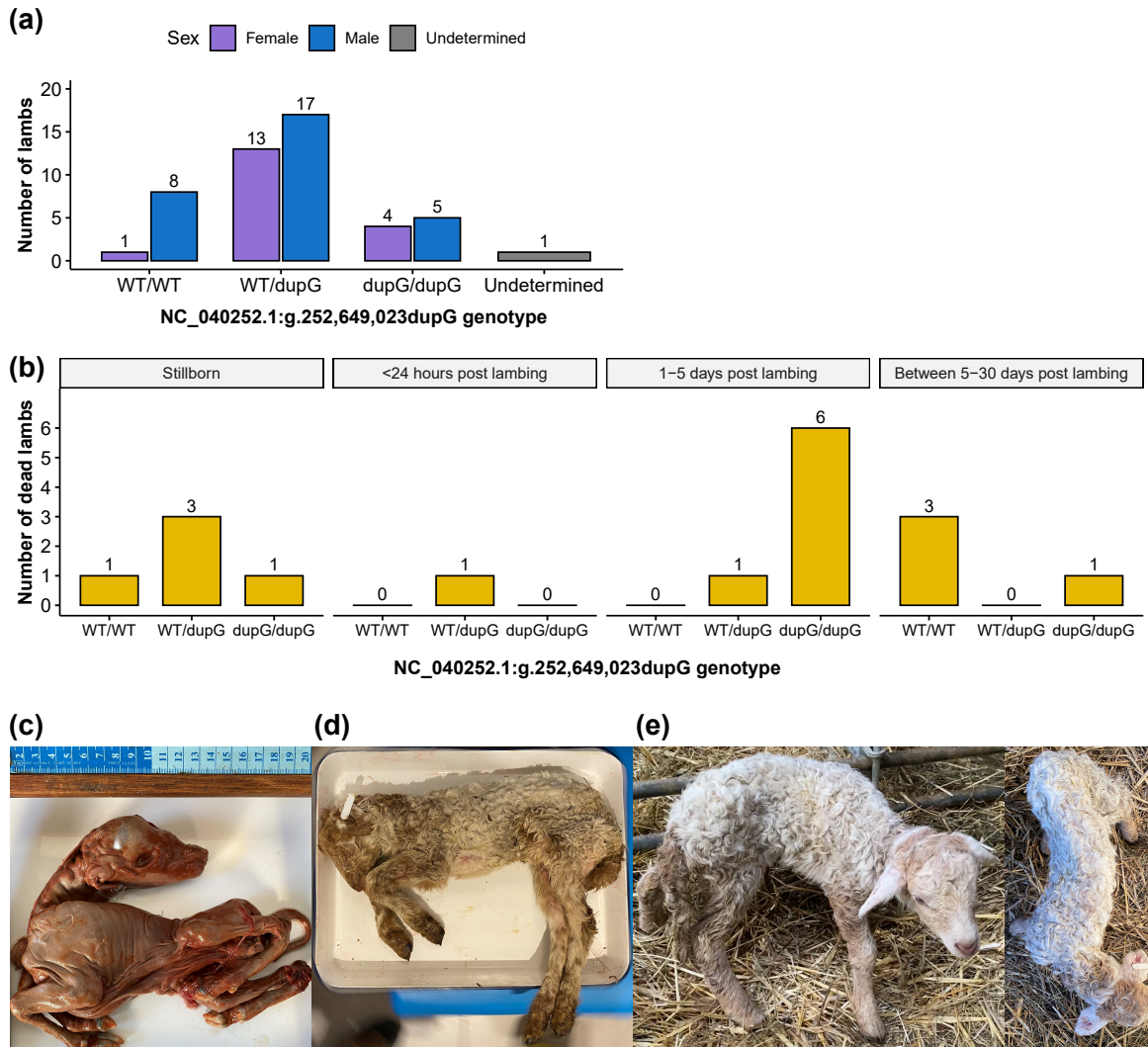


FIGURE 13.7 – **Bilan des mises-bas des accouplements à risque.** (a) Répartition des 49 agneaux issus de 28 brebis gestantes suivant les génotypes à la mutation *SLC33A1_dupG*. Le génotype d'un agneau mort-né reste indéterminé. (b) Répartition des 17 agneaux morts entre la naissance et l'âge de sevrage (30^{ème} jour après IA) en fonction du stade de létalité et des génotypes. (c) Agneau mort-né. (d) Agneau mort entre 1 et 5 jours après l'agnelage (1,750 kg). (e) Agneau mort entre 5 et 30 jours après l'agnelage, souffrant de faiblesse et de raideur dans les muscles des jambes et présentant des déformations de la colonne vertébrale.

TABLE 13.5 – **Tableau de contingence entre les génotypes à la mutation *SLC33A1_dupG* et la viabilité des agneaux.** -/- : génotype indéterminé. Cet animal n'est pas pris en compte pour le calcul du taux de mortalité. WT/WT= génotype homozygote sauvage, WT/dupG=animal hétérozygote et dupG/dupG= animal homozygote muté.

Viabilité	WT/WT	WT/dupG	dupG/dupG	-/-	Total
Vivant	5	25	1	0	31
Mort	4	5	8	1	18
Total	9	30	9	1	49

13.2.3 Pertes fœtales

La diminution de la réussite à l'IA dans l'accouplement à risque entre les porteurs hétérozygotes *SLC33A1_dupG* suggère des pertes d'embryons homozygotes au cours de la période de gestation entre l'IA et le diagnostic échographique au jour 45-60. Le test moléculaire de diagnostic de gestation (GMDT) développé dans le [paragraphe 13.1.3](#), a donc été utilisé pour estimer une éventuelle perte d'embryons en développement entre J15 et J45-60. Afin de recalibrer le test sur ce croisement (changement de race), son efficacité a été évaluée sur les brebis témoins (croisements non à risque) puis les paramètres du test ont été généralisés afin de prédire les résultats que l'on devrait obtenir dans le groupe des croisements à risque (sans impact de la mutation), et que l'on va comparer avec l'observé (avec impact de la mutation).

Ainsi, les échantillons d'ARN sanguin de 91 brebis (n= 40 dans le groupe d'accouplements non à risque, n= 51 dans le groupe d'accouplements à risque) prélevés 15 jours (avec un jour de flottement) après l'IA ont été analysés par RT-qPCR ([Figure 13.8a](#)). L'état de gestation de référence (gestante/vide) est établi entre 45 et 60 jours après l'IA par échographie. Dans le groupe témoin (non à risque), l'expression relative de *MX1* chez les brebis gestantes (84,3%) est significativement plus élevée que chez les brebis vides (67,7%, P-valeur = 0,043, test de Wilcoxon) ([Figure 13.8a](#)). La même tendance est observée pour l'expression de *STAT1*, même si la différence n'est pas significative (P-valeur=0,076, test de Wilcoxon, [Figure 13.8b](#)). Ainsi pour la suite, seule l'expression relative de *MX1* est conservée pour le GDMT et la fiabilité du diagnostic est testée en utilisant la courbe ROC sur les données du groupe d'accouplements non à risque comme données d'apprentissage. Le graphique ROC montre une AUC de 0,687, différant significativement de 0,5 (P-valeur = 0,033). A partir de la méthode ROC01 qui favorise la sensibilité (Perkins & Schisterman, 2006), le seuil de décision est fixé à 63% et il est caractérisé par une sensibilité de 75% et une spécificité de 60% ([Table 13.7](#) et [Figure 13.8c et d](#)).

La prévalence (Pr) de la réussite à l'IA a été fixée à 60,3% (moyenne de la population de Manech Tête Rousse) et m'a permis de calculer la valeur prédictive positive (VPP = 74%) et la valeur prédictive négative (VPN = 61%). La valeur seuil a été utilisée pour classer les brebis en quatre classes (TP, FP, TN et FN) selon le GDMT (+ ou -) et le diagnostic échographique à J45-60

TABLE 13.7 – Paramètres d’analyses de la courbe ROC pour le diagnostic moléculaire de gestation basé sur l’expression de *MX1*.

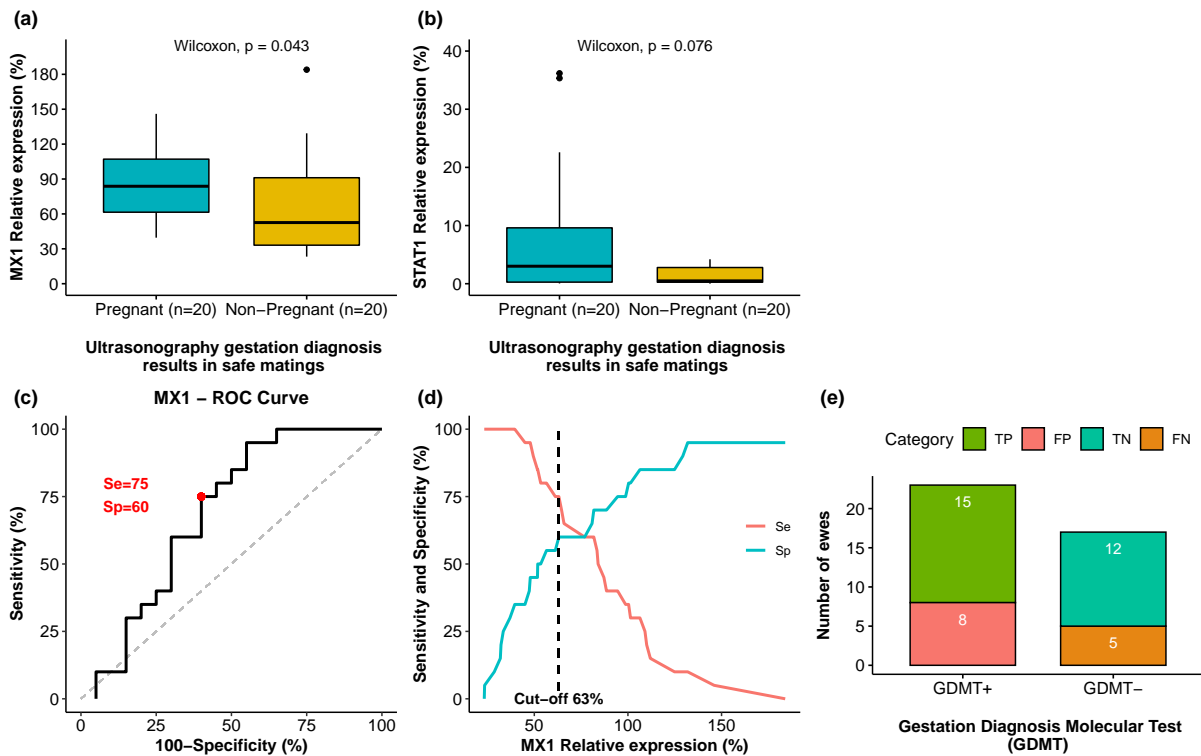
¹ La méthode ROC01 (minimise la distance entre la courbe ROC et le point (100-Sp=0, Se=100)) a été utilisée pour une coupure optimale. ² La VPP et la VPN ont été calculées comme décrit dans la section Matériel et méthodes de l’article 4 en prenant en compte les valeurs de Sensitivité, Spécificité et la Prévalence. AUC = Aire sous la courbe ; se = erreur standard ; Se = sensibilité ; Sp = Spécificité ; Pr = Prévalence, par exemple, succès de l’insémination artificielle chez la brebis laitière Manech Tête Rousse ; VPP = valeur prédictive positive ; VPN = Valeur prédictive négative.

Gene	AUC±se	P-value	Se (%)	Sp (%)	Cut-Off (%) ¹	Pr (%)	PPV (%) ²	NPV (%) ²
<i>MX1</i>	0,687±0,088	0,033	75	60	63	60,3	74	61

(Figure 13.8e). Le GDMT positif a permis de détecter 65% (TP/(TP+FP)) de brebis gestantes dans le groupe d’accouplements témoins. En revanche, aucune différence significative n’est observée pour l’expression de *MX1* entre les brebis gestantes et vides du groupe d’accouplements à risque (P-valeur = 0,27, test de Wilcoxon, Figure 13.8f).

Ainsi, les paramètres du diagnostic moléculaire (seuil = 63%, Pr = 60,3%, VPP = 74% et VPN = 61%) tirés des accouplements témoins ont été transposés aux données d’expression de *MX1* du groupe d’accouplements à risque. Dans ce groupe, la comparaison du nombre observé et attendu de brebis avec un GDMT positif (expression relative *MX1* > 63%) met en évidence une forte tendance statistique à réduire le nombre de brebis gestantes observées puisque 16 brebis gestantes étaient attendues avec un GDMT positif alors que seulement 12 brebis gestantes sont observées (Test d’ajustement à une distribution théorique par un test du Chi², X² = 3,6667, df= 1, P-valeur = 0,055). Aucune différence significative n’est mise en évidence entre le nombre observé et attendu de brebis gestantes avec un GDMT négatif (test du Chi², X² = 0,58586, df= 1, P-valeur = 0,44). Ainsi, grâce au diagnostic moléculaire, je mets vraisemblablement en évidence des pertes fœtales dans le groupe d’accouplements à risque entre 15 jours et 60 jours de gestation, avec quatre brebis supposées héberger des fœtus homozygotes pour le variant *SLC33A1_dupG*.

Safe Matings (Control group)



At-risk Matings

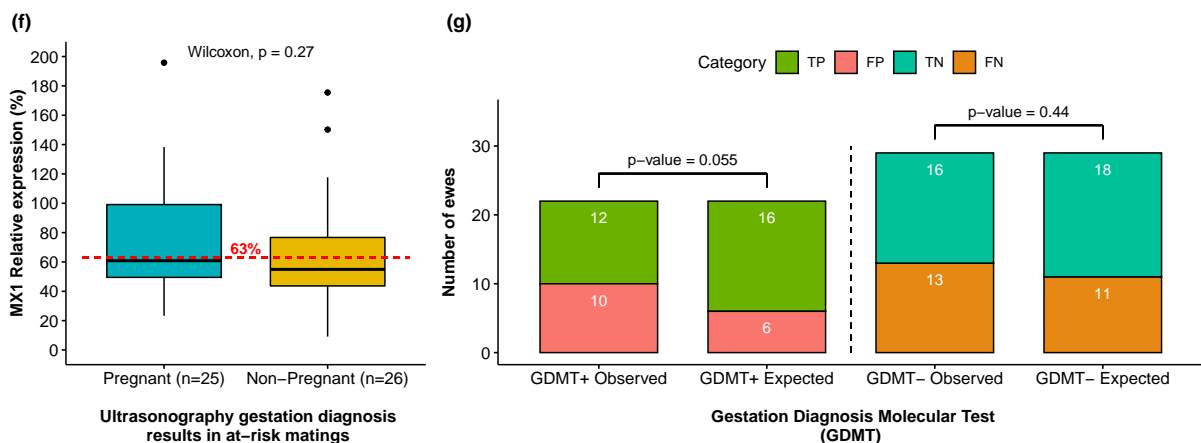


FIGURE 13.8 – Diagnostic moléculaire de gestation précoce (GDMT) basé sur l'expression des gènes *MX1* et *STAT1* en Manech Tête Rousse. Résultats de qPCR pour l'expression relative de l'ARNm de *MX1* (a) et *STAT1* (b) entre les brebis gestantes et vides dans les accouplements témoins (non à risque). (c) Courbe ROC représentant la sensibilité (Se) et la spécificité (Sp) du GDMT en utilisant l'expression relative de *MX1*. (d) Détermination de la valeur seuil d'expression de *MX1* (63%) associée à Se=75% et Sp = 60% en utilisant la méthode ROC01. (e) Répartition des brebis gestantes et vides selon le GDMT utilisant l'expression de *MX1*. (f) Résultats de qPCR pour l'expression relative de *MX1* et *STAT1* entre les brebis gestantes et vides dans les accouplements à risque entre porteurs de *SLC33A1_dupG*. (g) Répartition des brebis gestantes et vides observées et attendues selon le GDMT utilisant l'expression de *MX1*. Les nombres observés ont été calculés en appliquant la valeur seuil de 63% pour considérer un GDMT positif (expression *MX1* > 63%) ou négatif (expression *MX1* < 63%). Les nombres attendus ont été calculés sur la base de la prévalence (Pr = 60,3%), de la VPP = 74% et de la VPN = 61%. Les différences entre les nombres observés et attendus des résultats du GDMT (+ et -) ont été évaluées par un test du Chi². TP = vrai positif; FP = faux positif, TN = vrai négatif; FN = faux négatif, Pr = prévalence; VPP = valeur prédictive positive; VPN = valeur prédictive négative.

A retenir Les croisements à risque en Lacaune pour les mutations dans *ORC5* et *IDI1* réalisés en installation expérimentale n'ont pas permis de valider la causalité de ces mutations. Néanmoins, ces accouplements ont permis de développer un test moléculaire de diagnostic de gestation basé sur l'expression de gènes cibles de l'interféron τ , *MX1* et *STAT1*. Ce test, basé uniquement sur *MX1*, a pu être utilisé en élevages commerciaux de brebis Manech Tête Rousse pour caractériser des pertes fœtales liées à la mutation dans *SLC33A1*, également associée à une mortalité périnatale des agneaux homozygotes. ■

14. Anomalies génétiques induisant de la mortalité périnatale

Au travers de ce chapitre, l'objectif est de valider la causalité des mutations identifiées dans les gènes *MMUT* (MTRDHH1) en Manech Tête Rousse et *CCDC65* (LDHH6) en Lacaune. Les analyses précédentes sur l'effet des haplotypes laissaient supposer un déterminisme récessif léthal autour de la naissance. Pour répondre à cet objectif, des accouplements à risque entre animaux hétérozygotes porteurs des mutations *MMUT* ou *CCDC65* ont aussi été réalisés afin de confirmer la mortalité périnatale des agneaux homozygotes et tenter d'en définir les causes. L'ensemble de ces procédures expérimentales sur animaux a fait l'objet d'une autorisation de projet (APAFIS30615-2021032318054889 v4).

14.1 Croisements *MMUT* en Manech Tête Rousse

L'objectif est de déterminer la causalité de la mutation OAR20 :g.23,776,347G>A (appelée *single nucleotide variant*, SNV (G>A) *MMUT* par la suite) localisée dans l'haplotype MTRDHH1, supposée augmenter le taux de mort-nés dans les croisements à risque. Pour rappel, *MMUT* intervient dans la conversion du méthylmalonyl-CoA en Succinyl-CoA qui alimente le cycle de Krebs permettant la récupération d'énergie à partir des sucres, des lipides et des protéines. Chez l'homme, des mutations dans ce gène conduisent au syndrome d'acidémie méthylmalonique (OMIM 251000). L'ensemble des résultats présentés ci-dessous est détaillé dans l' [Article 3 - Ben Braiek et al., 2023a - bioRxiv, rédigé pour Plos Genetics - Homozygous haplotype deficiency in Manech Tête Rousse dairy sheep revealed a nonsense variant in MMUT gene affecting newborn lamb viability](#). Les principaux résultats sont repris ci-dessous.

14.1.1 Dispositif expérimental d'accouplements des porteurs du SNV *MMUT*

Pour valider l'impact de l'allèle muté du SNV *MMUT*, j'ai programmé des accouplements à risque entre porteurs hétérozygotes pour générer des agneaux homozygotes. Des échantillons de sang ont été prélevés sur 181 brebis MTR, filles de pères porteurs de MTRDHH1 réparties dans 6 élevages privés. Le test de génotypage spécifique du SNV *MMUT* a permis d'identifier 82 brebis hétérozygotes. Parmi ces brebis, 73 sont localisées dans les 6 fermes privées (Expérience 1) et 9 ont été transférées dans l'installation expérimentale INRAE de Langlade (Expérience 2). Toutes les brebis ont été inséminées avec de la semence fraîche de béliers hétérozygotes pour le SNV *MMUT*. Le dispositif expérimental est décrit dans [Figure 14.1](#).

14.1.2 Mortalité périnatale

Le diagnostic échographique permet d'identifier 49 brebis gestantes (44 dans l'expérience 1 et 5 dans l'expérience 2). Cela correspond à un taux de réussite à l'IA de 59,8%, conforme au taux

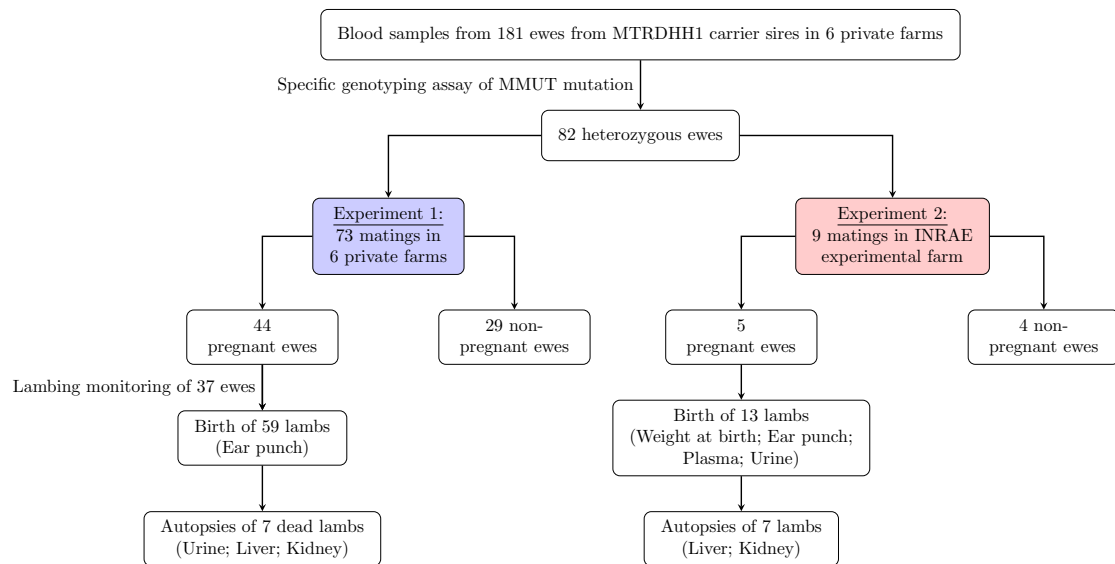


FIGURE 14.1 – Dispositif expérimental pour générer des agneaux homozygotes SNV *MMUT* (A/A) dans les 6 élevages privés et dans l’installation expérimentale INRAE.

moyen de 60,9% déterminé dans l’ensemble de la population lors de l’analyse des caractères de fertilité. Dans l’expérience 1, seules 37 des 44 brebis gestantes ont été suivies après le diagnostic de gestation et ont donné naissance à 59 agneaux (prolificité moyenne de 1,6, taille de portée allant de 1 à 3) avec une durée de gestation comprise entre 139 et 159 jours. Dans l’expérience 2, les 5 brebis gestantes ont donné naissance à 13 agneaux (prolificité moyenne de 2,6, taille de portée allant de 2 à 4) avec une durée de gestation comprise entre 151 et 157 jours. Aucun avortement au cours des cinq mois de gestation n’a été observé. Au total, 72 agneaux (52% de mâles et 48% de femelles) sont nés et des biopsies auriculaires ont été collectées pour le génotypage spécifique de la mutation (Table 14.1). La distribution des génotypes au SNV *MMUT* ne diffère pas entre les deux expériences (test exact de Fisher, P-valeur = 0,686) avec au total 21 agneaux homozygotes mutés A/A, 29 hétérozygotes G/A et 21 non-porteurs homozygotes G/G.

TABLE 14.1 – Résultats du géotypage au SNV *MMUT* des agneaux issus d'accouplements à risque dans les élevages privés et expérimentaux selon le sexe.

*Biopsie auriculaire non disponible pour cet individu.

^a Nombre d'agneaux vivants.

† Nombre d'agneaux morts (-/-= génotype indéterminé).

Category	G/G	G/A	A/A	-/-	Total
<i>Experiment 1 : Private farms (n=6)</i>					
Male	9	9	10	1	29
Female	8	16	5		29
Undetermined			1		1
Total	17	25	16	1	59
<i>Experiment 2 : Experimental farm (n=1)</i>					
Male	4	2	2		8
Female	0	2	3		5
Total	4	4	5		13
<i>All</i>					
Male	13	11	12	1	37
Female	8	18	8	0	34
Undetermined			1		1
Total	21 (19^a/2†)	29 (25^a/4†)	21 (2^a/19†)	1* (0^a/1†)	72

Tous les agneaux ont été suivis pendant la période de 0 à 30 jours jusqu'au sevrage. Vingt-cinq agneaux sont morts au cours de cette période, ce qui représente un taux de mortalité important de 34,7% (Figure 14.2). Le tableau de contingence entre le génotype des agneaux (A/A, G/A ou G/G) et la viabilité (vivant ou mort) indiquent un taux de mortalité plus élevé pour les agneaux homozygotes A/A (test exact de Fisher, P-valeur < 0,001, Table 14.1). En effet, les agneaux morts A/A représentent 78% de la mortalité totale des agneaux. La mort des agneaux homozygotes A/A est majoritairement survenue dans les premières 24 heures après l'agnelage.

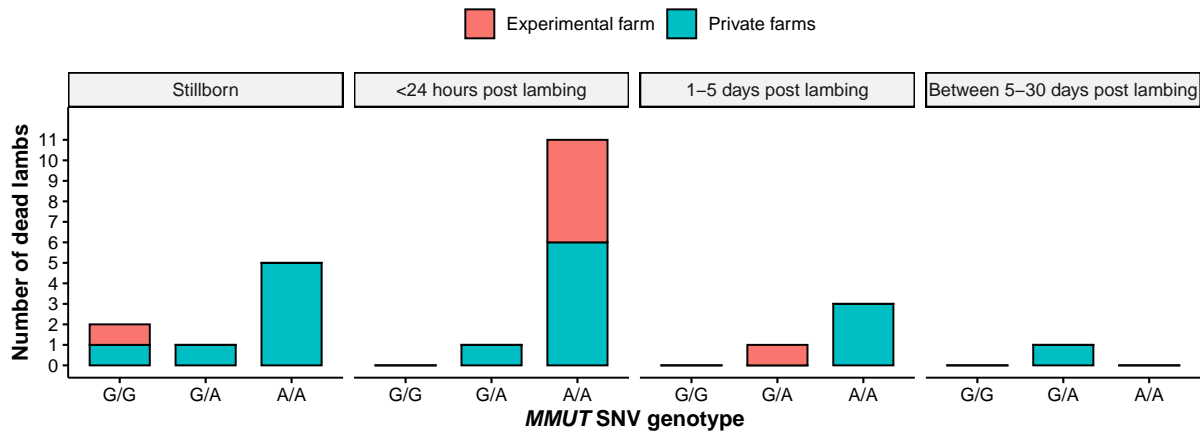


FIGURE 14.2 – Distribution des agneaux morts en fonction du temps et selon les génotypes au SNV *MMUT*.

L'examen clinique des agneaux morts n'a pas permis d'identifier des symptômes spécifiques. Deux agneaux homozygotes A/A ont passé l'âge du sevrage. Lors de l'expérience 2, les 13 agneaux ont été pesés à la naissance (mâles : $4,0 \pm 0,9$ kg, femelles : $2,9 \pm 1,4$ kg) et les agneaux A/A (quel que soit le sexe) avaient un poids à la naissance significativement inférieur à celui des autres génotypes (test non-paramétrique de Wilcoxon, P-value=0,019, Figure 14.3).

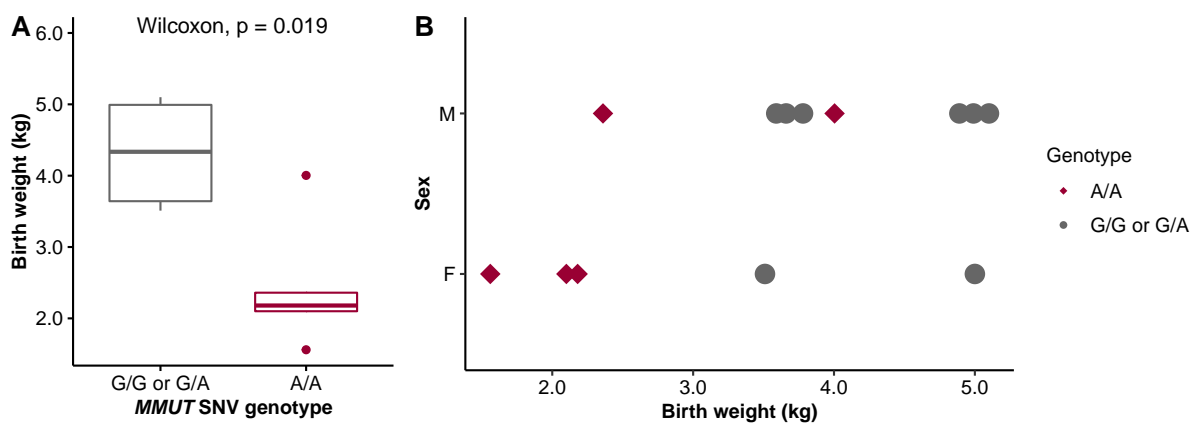


FIGURE 14.3 – Effet des génotypes du SNV *MMUT* sur le poids à la naissance des 13 agneaux nés à l'installation expérimentale INRAE. (A) Représentation Boxplot du poids à la naissance selon les génotypes (B) distribution du poids à la naissance par sexe (M : Mâle, F : Femelle) et génotypes. Les agneaux homozygotes atteints sont en rouge (génotype A/A).

14.1.3 Expression et activité de *MMUT*

D'après l'atlas d'expression des gènes ovins (<http://biogps.org/sheepatlas/>; consulté le 17 février 2022), *MMUT* apparaît fortement exprimé dans les reins et le foie (Clark et al., 2020). Les mutations faisant apparaître un codon stop prématuré, comme c'est le cas pour le SNV *MMUT*, en plus de produire une protéine tronquée, peuvent avoir un effet indirect sur la dégradation des ARNm du gène, phénomène connu sous le terme "*nonsense-mediated mRNA decay*" ou *NMD* (Embree et al., 2022).

Afin de mettre en évidence ces deux phénomènes sur la protéine et l'ARNm de *MMUT*, j'ai collecté des échantillons de foie et de rein sur les différents agneaux morts de différents génotypes au SNV *MMUT*. Par RT-qPCR (Figure 14.4 A), j'ai mis en évidence une réduction significative de l'expression de *MMUT* dans le foie chez les agneaux homozygotes mutés A/A (test non-paramétrique de Wilcoxon, P-valeur = 0,0015) mais pas dans les reins (test non-paramétrique de Wilcoxon, P-valeur = 0,66). J'ai également évalué l'expression de *MMUT* à partir d'extraits de protéines de foie et de rein chez deux agneaux homozygotes A/A et deux homozygotes G/G. L'analyse par Western-blot montre une expression de la protéine *MMUT* normale dans le rein et le foie alors que la protéine mutée n'a pas été détectée, au moins sous une forme complète dans les deux tissus (Figure 14.4 B). J'ai également essayé d'évaluer l'accumulation d'acide méthylmalonique (MLA) quantifié par ELISA dans l'urine et le sang d'agneaux mutés A/A prélevés post-mortem ou entre 1 et 5 jours après la naissance chez des animaux vivants (Figure 14.5). Cependant, aucune différence significative n'a été observée avec les taux de MLA des agneaux G/G ou G/A (test non-paramétrique de Wilcoxon, P-valeur = 0,54 dans l'urine, test non-paramétrique de Wilcoxon, P-valeur = 0,34 dans le plasma).

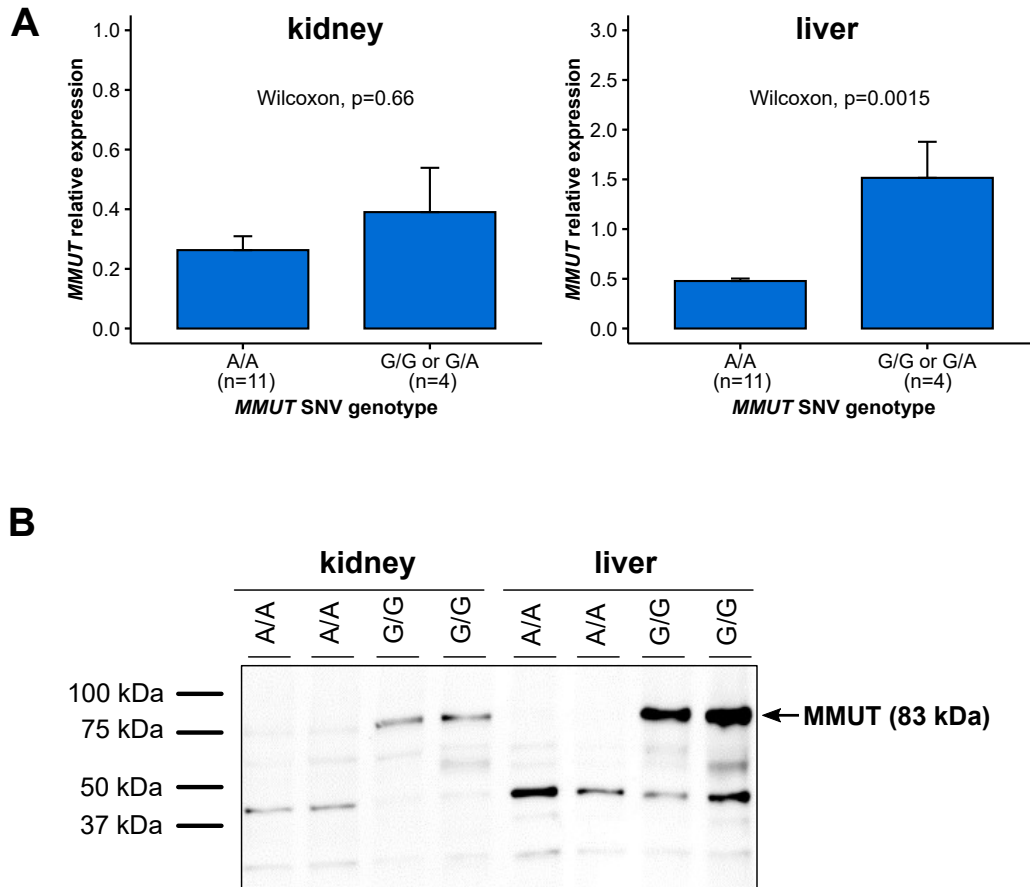


FIGURE 14.4 – Effet du SNV sur l'expression du gène *MMUT*. (A) Expression relative des ARNm *MMUT* dans les reins et le foie déterminée par RT-qPCR. (B) Expression de la protéine *MMUT* dans le rein et foie par Western-Blot avec un anticorps anti-*MMUT*. La comparaison de moyennes est effectuée par un test non-paramétrique de Wilcoxon. Les barres d'erreurs indiquent les erreurs standards.

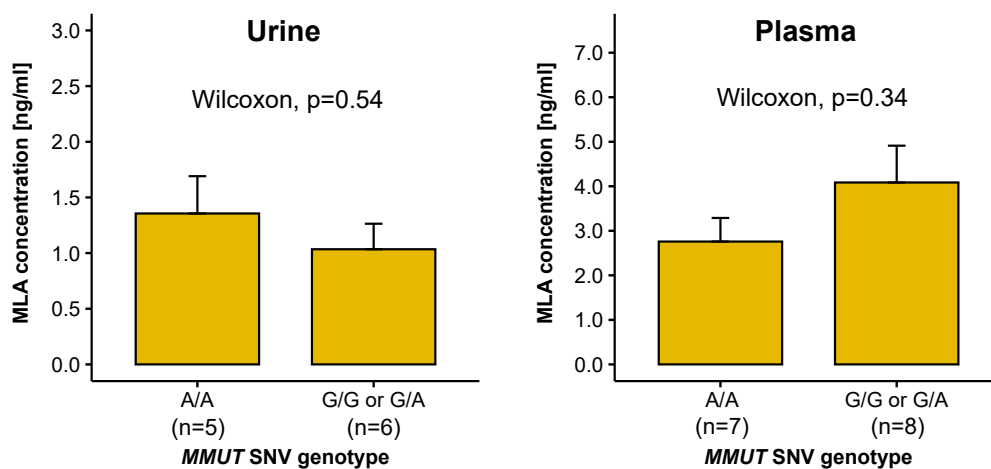


FIGURE 14.5 – Dosage ELISA de l'acide méthylmalonique (MLA). La comparaison de moyennes est effectuée par un test non-paramétrique de Wilcoxon. Les barres d'erreurs indiquent les erreurs standards.

14.2 Croisements *CCDC65* en Lacaune

Comme pour *MMUT*, afin d'étudier la causalité de la mutation non-sens OAR3 : g.147,207,999 C>A (c.331G>T) dans le gène *CCDC65* (LDHH6), j'ai également planifié des accouplements à risque entre animaux Lacaune porteurs hétérozygotes de la mutation SNV (C>A) *CCDC65*. Pour rappel, le SNV *CCDC65* est supposé augmenter le taux de mort-nés dans les croisements à risque. *CCDC65* joue un rôle essentiel dans l'assemblage du complexe nexine-dynéine (N-DRC) intervenant dans la régulation de la motilité des cils et des flagelles (Horani et al., 2013). Chez l'homme, des variants de ce gène conduisent au syndrome de Dyskinésie Ciliaire Primitive. Les résultats présentés ci-dessous sont décrits dans l' **Article 2** (publié) - Ben Braiek et al., 2022 - *Genes - A Nonsense Variant in *CCDC65* Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep.*

14.2.1 Dispositif expérimental d'accouplements des porteurs du SNV *CCDC65*

Des brebis laitières Lacaune (n=245) de deux troupeaux expérimentaux INRAE (Langlade et La Fage) ont été génotypées pour la mutation c.331G>T dans *CCDC65*. Après génotypage, 17 brebis hétérozygotes ont été inséminées artificiellement avec de la semence fraîche de béliers porteurs hétérozygotes (n=3) sélectionnés parmi les animaux génotypés pour la sélection génomique avec un statut connu à l'haplotype LDHH6 et confirmés hétérozygotes par le test de génotypage spécifique à la mutation. Un diagnostic échographique de gestation a été réalisé entre 45 et 60 jours après l'IA. Les agneaux ont été pesés à la naissance et 15 jours après la naissance. Des biopsies auriculaires (TSU Allflex) ont été prélevées sur les agneaux pour réaliser le génotypage au SNV *CCDC65*.

14.2.2 Mortalité périnatale

Sur les 17 brebis inséminées, 11 brebis ont été diagnostiquées gestantes par échographie. Cela correspond à une réussite à l'IA de 65%, en accord avec la réussite moyenne de 69% enregistrée pour l'ensemble de la population laitière Lacaune (Ben Braiek et al., 2021). La durée de gestation a été enregistrée comme normale (146 ± 2 jours) et aucun avortement ou mort-né n'a été observé. En fin de gestation, 16 agneaux sont nés (7 mâles et 9 femelles) et ont été génotypés. Cela représente 5 non porteurs G/G, 6 hétérozygotes G/T et 5 homozygotes mutés T/T. Tous les agneaux ont été pesés à la naissance (mâles : $4,9 \pm 1,0$ kg, femelles : $3,9 \pm 0,6$ kg), et aucune différence significative de poids à la naissance n'est observée entre les génotypes. Une deuxième pesée a été effectuée au jour 15 et le gain moyen quotidien (GMQ) a été calculé sur la période 0-15 jours. Sur la base du test non-paramétrique de Wilcoxon, les porteurs homozygotes (T/T) montrent un GMQ significativement plus faible par rapport aux autres agneaux (G/G ou G/T) (Figure 14.6).

L'examen clinique a révélé des problèmes respiratoires, tels que de la tachypnée et le nez qui coule pour 4 sur 5 agneaux T/T dès les premiers jours de vie. Malgré un traitement vétérinaire approprié pour la suspicion de maladies infectieuses pulmonaires (basé sur l'administration de glucocorticoïdes et d'antibiotiques), la détresse respiratoire récurrente a entraîné une baisse de l'état corporel général (attestée par les GMQ inférieurs) conduisant à la mort naturelle ou à

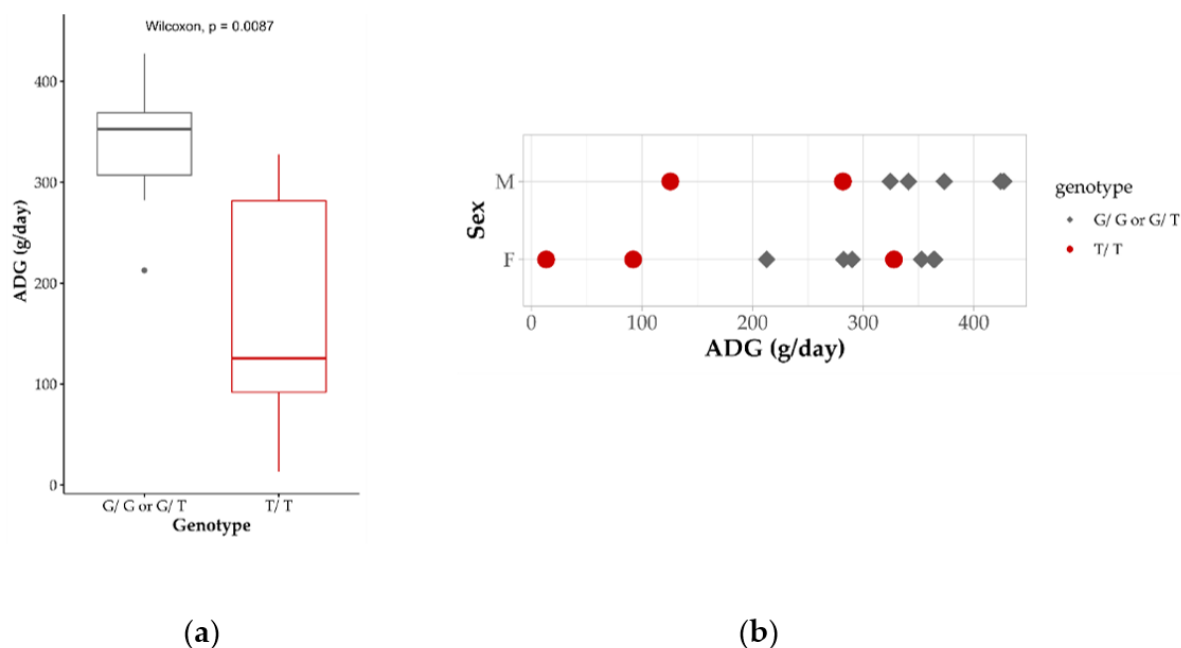
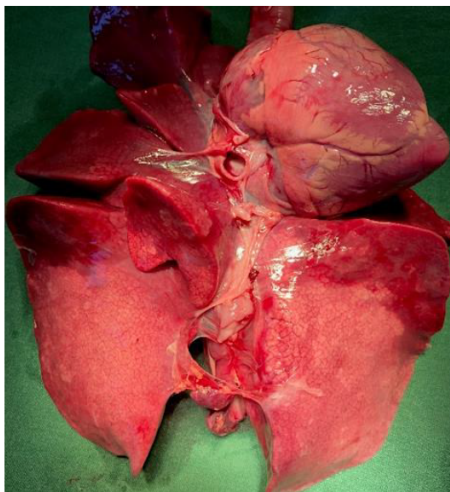


FIGURE 14.6 – Effet du génotype *CCDC65* sur le gain quotidien moyen (GMQ) pendant la période de 0 à 15 jours. GMQ moyen (g/jour) selon le génotype *CCDC65* c.521G>T (a) et la distribution des GMQ par sexe (M : Mâle ; F : Femelle) et génotype (b).

l'euthanasie des agneaux atteints entre 15 et 25 jours après la naissance. Une seule agnelle T/T avec le GMQ le plus élevé (327,8 g/jour pendant la période 0–15 jours) a dépassé l'âge de sevrage avec un syndrome respiratoire léger et a été vendue à un engraisseur. De plus, l'analyse précédente de la cohorte de sélection génomique 2021 avait permis d'identifier un animal homozygote pour le SNV *CCDC65* (Figure 14.7). Cet agneau mâle était présent dans un centre d'élevage, et nous avons obtenu l'information qu'il avait reçu plusieurs fois un traitement vétérinaire en raison de problèmes respiratoires. L'agneau est finalement mort à l'âge de cinq mois et demi. Une autopsie a été pratiquée sur les quatre agneaux T/T morts avant 25 jours. Une large hépatisation de certains lobes pulmonaires était observable (Figure 14.8) probablement liée à une pneumonie infectieuse.



FIGURE 14.7 – Béliet homozygote *CCDC65* âgé de 3 mois.



(a)



(b)

FIGURE 14.8 – Agneau atteint homozygote *CCDC65* avec lésions pulmonaires. (a) système cœur-poumons d'un agneau homozygote atteint (b) grossissement des lobes hépatisés.

A retenir La réalisation d'accouplements à risque en élevages ou en installations expérimentales a permis de valider la causalité et le caractère récessif létal des mutations découvertes dans *MMUT* et *CCDC65*. Les observations réalisées sur les agneaux homozygotes ont effectivement montré qu'elles conduisent à de la mortalité périnatale ou juvénile (*CCDC65*), mais dont il reste à connaître plus précisément les causes syndromiques, acidémie méthylmalonique pour *MMUT* et dyskinésie ciliaire pour *CCDC65* comme chez l'homme. ■

15. Anomalies génétiques en lien avec des défauts morphologiques

Au travers de ce chapitre, l'objectif est de mieux comprendre l'association entre les haplotypes en déficit (LDHH4, LDHH11, MTRDHH4 et MTRDHH5) liés à des mutations déjà connues dans les gènes *SOCS2*, *RXFP2* et *ASIP* décrits dans le [paragraphe 12.3.2.2](#), pour lesquels les génotypes sont disponibles et extraits des puces basse densité.

15.1 *SOCS2* : stature et sensibilité aux mammites

Nous avons vu précédemment que LDHH11 était associé partiellement à la mutation faux-sens g.139,302,270C>T (p.Arg96Cys) dans *SOCS2*. Pour rappel, cette mutation a été découverte pour son lien avec le comptage des cellules somatiques dans le lait et la sensibilité/résistance aux mammites chez la brebis Lacaune (Rupp et al., 2015). Elle exerce également un effet pléiotropique sur la croissance, le gabarit et la production laitière. Afin de mieux comprendre le lien entre LDHH11 et *SOCS2*, j'ai extrait tous les haplotypes (fréquence 1%, déficit > 0) présents dans cette région, nommés "LDHH11a" à "LDHH11f", triés par P-valeur croissante ([Figure 15.1](#)). Initialement lors de la première analyse en Mars 2019 (19 102 animaux Lacaune), j'ai identifié un seul haplotype qui correspondait aux critères (fréquence $\geq 1\%$, déficit $\geq 75\%$ et P-Poisson $\leq 1,9 \times 10^{-4}$) définissant LDHH11 (appelé par la suite LDHH11b). Cependant, lors de la nouvelle analyse de Septembre 2022 (35 221 animaux Lacaune), j'ai identifié un nouvel haplotype significativement en déficit dans la même région (nommé LDHH11a). L'analyse d'association entre le statut à LDHH11a et les génotypes à la mutation dans *SOCS2* montre également une forte association ([Figure 15.2 A](#)). Afin, d'identifier d'autres haplotypes associés à *SOCS2*, j'ai fait ce travail d'association pour les 7 haplotypes ayant une fréquence $\geq 1\%$. Ainsi, on observe trois haplotypes qui se comportent de la même manière : LDHH11a, b et c et qui signent la mutation *SOCS2*. En effet, tous les animaux hétérozygotes à ces haplotypes sont hétérozygotes à la mutation ([Figure 15.2 A, B et C](#)). De plus, si l'on regarde la composition en marqueurs de ces trois haplotypes ([Figure 15.1](#)), on constate qu'ils partagent un haplotype plus court de 7 marqueurs de la position 2104 à 2110 (appelé hap_*SOCS2* "AGGGGGG") juste avant la mutation dans *SOCS2*. L'association entre hap_*SOCS2* et les génotypes à la mutation montre clairement une association très forte ([Figure 15.2 H](#)). De plus, hap_*SOCS2* est en déséquilibre de liaison modéré avec les haplotypes LDHH11a, b et c (DL compris entre 12% et 24% ; [Figure 15.3](#)).

Pour rappel LDHH4, localisé à seulement 62kb de LDHH11, était également lié à *SOCS2* (voir [paragraphe 12.3.2.2](#)). Afin de confirmer ce lien, j'ai calculé le déséquilibre de liaison entre LDHH4 et les différents haplotypes signant *SOCS2* (LDHH11a, b et c, hap_*SOCS2*) ([Figure 15.3](#)). On constate que LDHH4 est en fort déséquilibre de liaison avec l'haplotype LDHH11c (81%)

Haplotype	NbAtt	NbObs	NbObs pop	Freq he	Déficit	P-Poisson	2102	2103	2104	2105	2106	2107	2108	2109	2110	<i>SOCS2</i>	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134						
LDHH11a	30	6	9	5,5%	80%	1.66E-07	G	G	G	G	G	G	G	G	G		A	A	A	A	A	A	A	A	G	G	G	C	A	C	G	A	G	A	A	G	A	A	A	G						
LDHH11b	23	5	6	5,2%	78%	8.59E-06	G	G	G	G	G	G	G	G	G		G	C	A	A	A	A	A	A	A	A	A	A	C	A	C	G	A	A	A	A	A	A	A	A	A	A				
LDHH11c	10	1	1	2,8%	90%	4.83E-04	G	G	A	G	G	G	G	G	G		A	A	A	A	A	A	A	A	A	A	A	A	C	A	C	A	A	A	A	A	A	A	A	A	A	A	A			
LDHH11d	46	32	34	7,1%	30%	2.09E-02	A	G	A	A	A	A	A	A	A		G	C	G	A	G	G	C	A	A	A	A	A	A	G	A	G	A	A	A	A	A	A	A	A	A	A	A	A		
LDHH11e	32	26	31	6,4%	19%	1.68E-01	G	G	A	A	A	A	A	A	A		A	C	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
LDHH11f	564	550	593	18,2%	2%	2.89E-01	G	C	G	A	A	A	A	A	A		A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
LDHH11g	2	1	3	2,1%	44%	4.64E-01	A	C	G	A	A	A	A	A	A		G	C	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

FIGURE 15.1 – Liste des 7 haplotypes localisés dans la région de LDHH11. NbAtt= nombre d’animaux homozygotes attendus; NbObs= nombre d’animaux homozygotes observés; NbObs pop= nombre d’animaux homozygotes observés dans l’ensemble de la population génotypée (n=35 221); Freq he= fréquence des porteurs hétérozygotes. La région LDHH11 est définie par 33 marqueurs SNP (marqueurs SNP à la position 2102 à 2134). La mutation OAR3 :g.139,302,270C>T dans *SOCS2* est située entre le marqueur 2110 (g.139,262,478A>G) et 2111 (g.139,590,385A>G).

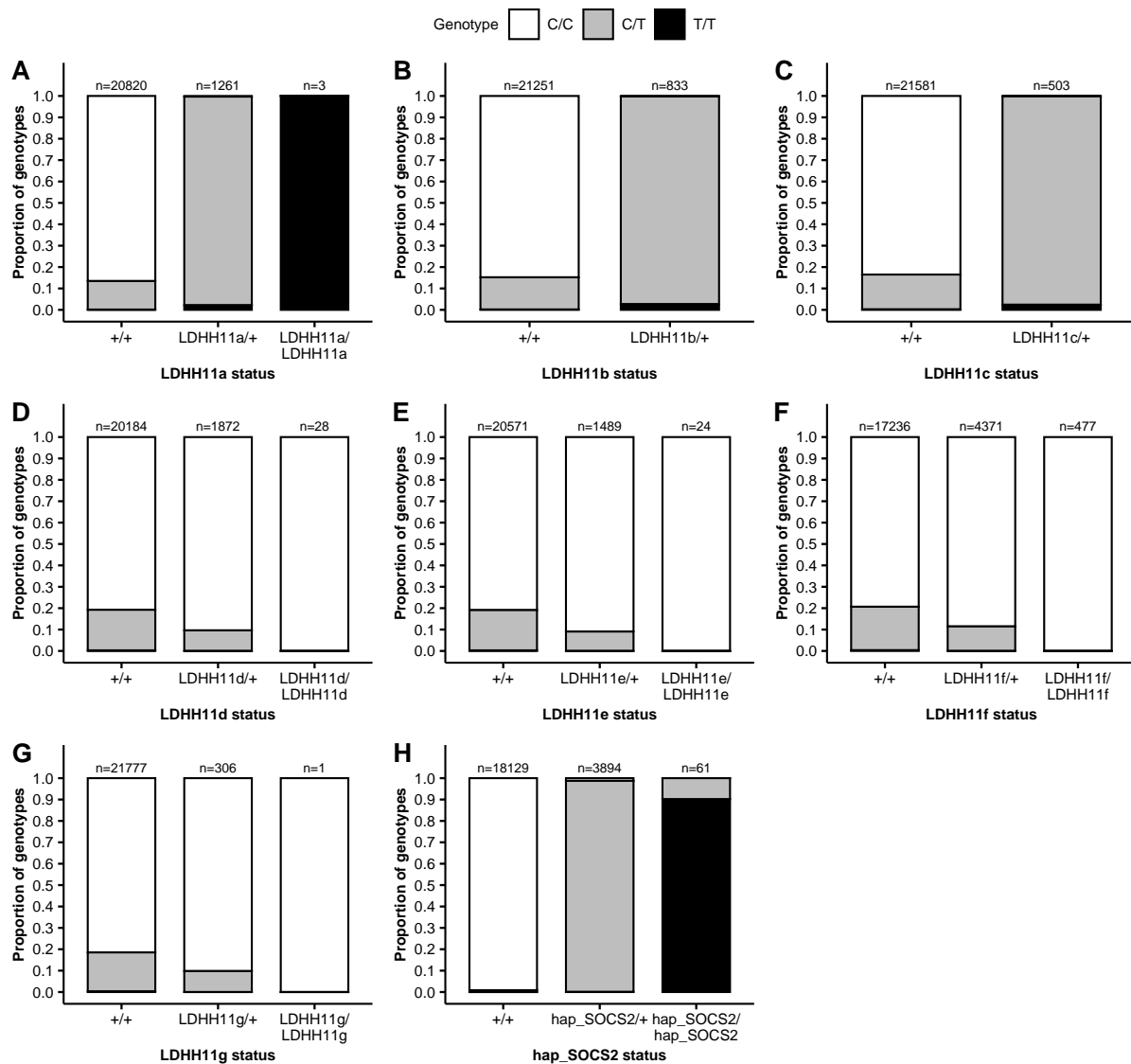


FIGURE 15.2 – Lien entre le statut aux haplotypes identifiés dans la région LDHH11 et la mutation OAR3 :g.139,302,270C>T dans *SOCS2*. Le génotypage spécifique de la mutation a été extrait de la puce basse densité à partir du millésime 2015 (n=22 084).

et modérément avec hap_SOCS2 (DL de 9% dans l'ensemble de la population). Cependant, LDHH4 n'existe que dans le schéma Ovitest et ce déséquilibre de liaison avec hap_SOCS2 passe à 17% si l'on ne considère que les animaux Ovitest (n=17 409). La mutation dans *SOCS2* n'est pas connue pour avoir des caractéristiques récessives létales et ne peut donc pas constituer une mutation causale. Je fais l'hypothèse que LDHH4 contient une mutation récessive létale (0 animaux homozygotes observés dans la population alors que 11 sont attendus) qui est apparue dans un haplotype contenant l'allèle muté g.139,302,270T et subit un phénomène d' "auto-stop" avec la mutation *SOCS2* en sélection.

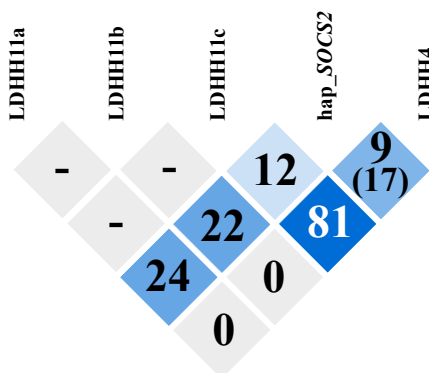


FIGURE 15.3 – Déséquilibre de liaison entre LDHH4 et les différents haplotypes signant *SOCS2* (LDHH1a, b et c ; hap_SOCS2). L'haplotype hap_SOCS2 (7 marqueurs) permet de prédire la phase de l'haplotype portant l'allèle muté de la mutation dans *SOCS2*.

15.2 ASIP : coloration de la toison

Précédemment, j'ai également fait l'hypothèse que MTRDHH5 était associé partiellement à une délétion connue de 5pb dans *ASIP* (*OAR13* :g.66,475,132_66,475,136del) liée à la couleur de toison noire à l'état homozygote. Afin de mieux comprendre pourquoi cette région contenant *ASIP* est en déficit d'homozygotes, j'ai extrait tous les haplotypes (fréquence $\geq 1\%$, déficit > 0) présents dans cette région, nommés MTRDHH5a à MTRDHH5k, triés par P-valeur croissante (Figure 15.4). L'haplotype MTRDHH5 initialement identifié correspond à l'haplotype MTRDHH5a, et il est le seul des 11 haplotypes à satisfaire les critères de sélection (fréquence $\geq 1\%$, déficit $\geq 75\%$ et P-Poisson $\leq 1,9 \times 10^{-4}$). L'analyse d'association entre le statut aux 11 haplotypes identifiés dans la région MTRDHH5 et la délétion de 5pb dans *ASIP*, montre un fort lien pour 8 d'entre-eux (MTRDHHa, c, d, f, i, j, k) (Figure 15.5). En effet, tous les animaux hétérozygotes porteurs des haplotypes sont hétérozygotes Ins/Del à la mutation. Trois haplotypes (MTRDHHb, e et g) ne sont pas liés à la mutation. Lorsque l'on regarde les différents marqueurs qui composent les haplotypes, il est difficile de distinguer un haplotype recombinant plus court signant la délétion comme identifié avec la mutation *SOCS2* (Figure 15.4). Seul le marqueur en position 921 (s13874.1, g.66795891A>C) permet de discriminer les deux groupes d'haplotypes (allèle A associé à la délétion et allèle C associé à l'insertion).

Ainsi, avec l'analyse de déficit, on est capable d'identifier l'haplotype le plus fréquent (16% d'animaux porteurs hétérozygotes pour MTRDHH5a), associé partiellement à la mutation. Dans l'ensemble de la population génotypée, 14 animaux homozygotes à MTRDHH5 sont observés

dont 12 sont des femelles. Pour les brebis MTR, nous avons accès aux informations de qualification/disqualification comme mères à béliers. Grâce à cette information, 10 brebis sur les 12 homozygotes sont disqualifiées pour être de couleur noire. En effet, le schéma de sélection génomique n'autorise pas le génotypage des agneaux noirs car non conformes aux standards de la race pour être futurs reproducteurs (Figure 15.6). Ainsi, la délétion dans *ASIP* partiellement associée à l'haplotype MTRDHH5 pourrait expliquer le déficit partiel observé d'animaux homozygotes par la contre-sélection des agneaux mâles de couleur noire.

Haplotype	NbAtt	NbObs	NbObs pop	Freq he	Déficit	P-Poisson	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	<i>ASIP</i>	918	919	920	921	922	923	924	925	926	927	
MTRDHH5a	51	8	14	16.1%	84%	7.2E-14	A	G	A	A	A	A	A	A	G	G	A	G	A	A	G	G	A	A	G	A	A	A	A	A	A	A	A	A	A	G	
MTRDHH5b	28	23	33	13.0%	17%	2.1E-01	G	G	C	A	A	A	A	A	A	A	A	A	A	A	G	G	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
MTRDHH5c	72	66	94	20.9%	9%	2.4E-01	A	G	A	A	A	A	A	A	A	G	A	G	A	A	G	G	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
MTRDHH5d	2	1	2	4.8%	57%	3.2E-01	G	G	C	G	C	A	A	A	A	A	G	G	G	C	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
MTRDHH5e	11	9	10	9.4%	19%	3.2E-01	A	G	A	G	C	A	A	A	A	G	A	G	A	A	G	G	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
MTRDHH5f	1	0	0	3.8%	100%	3.3E-01	G	G	C	G	C	A	A	A	A	G	G	G	G	C	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
MTRDHH5g	9	8	9	8.9%	8%	5.0E-01	A	A	A	A	A	A	A	A	A	G	A	G	A	C	G	G	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
MTRDHH5h	1	0	0	2.5%	100%	5.6E-01	G	G	A	A	C	A	A	A	A	A	A	G	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
MTRDHH5i	1	0	0	2.8%	100%	5.7E-01	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
MTRDHH5j	0	0	0	2.6%	100%	6.2E-01	A	G	C	G	C	A	A	A	A	G	G	G	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A
MTRDHH5k	0	0	0	1.8%	100%	6.6E-01	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A

FIGURE 15.4 – Liste des 11 haplotypes localisés dans la région de MTRDHH5. NbAtt= nombre d’animaux homozygotes attendus; NbObs= nombre d’animaux homozygotes observés; NbObs pop= nombre d’animaux homozygotes observés dans l’ensemble de la population génotypée (n=7 405); Freq he= fréquence des porteurs hétérozygotes. La région MTRDHH5 est définie par 29 marqueurs SNP (marqueurs SNP à la position 899 à 927). La mutation OAR13 :g.66,475,132_66,475,136del (Délétion 5pb) dans *ASIP* est située entre le marqueur 917 (g.66,160,896A>G) et 918 (g.66,693,760A>C).

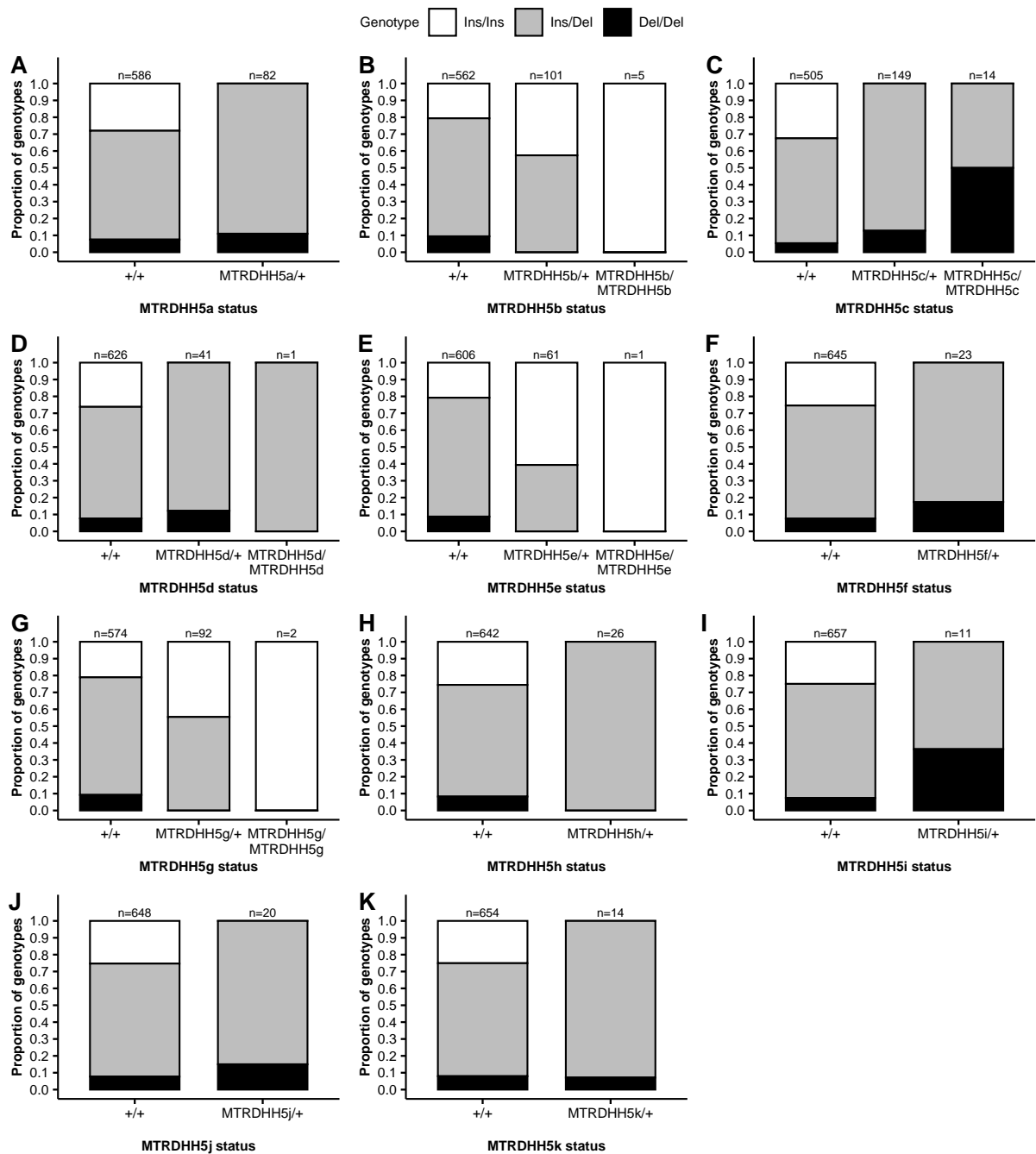


FIGURE 15.5 – Lien entre le statut aux haplotypes identifiés dans la région MTRDHH5 et la mutation OAR13 :g.66,475,132_66,475,136del (délétion 5pb) dans *ASIP*. Le génotypage spécifique de la mutation a été extrait de la puce basse densité pour les agneaux génomique de la cohorte 2021 (n=668).



FIGURE 15.6 – Agneau noir Manech Tête Rousse

15.3 *RXFP2* : cornage et cryptorchidie

Précédemment, nous avons pu voir un lien possible entre l'haplotype MTRDHH4 et deux polymorphismes du gène *RXFP2* : une nouvelle mutation faux-sens g.31,038,357G>A prédite avec un impact modéré (p.Ser26Leu, score Polyphen-2 0,66) et une insertion de 1,8kb dans le 3' UTR du gène connue pour être associée au phénotype sans cornes chez les ovins. En parallèle de mon travail sur les analyses de déficit en homozygotes, dans le cadre du projet PRESAGE de construction d'un observatoire des anomalies génétiques chez les petits ruminants, j'ai participé à l'Action 4 (Déterminisme génétique des anomalies) pour étudier un cas pilote portant sur l'étude du déterminisme génétique du cornage et de la cryptorchidie en Manech Tête Rousse. Si l'on retrace l'histoire de la race, les Manech Tête Rousse ont été importées en France dans les années 1970, issues des Latxa Carra Rubia du pays basque espagnol (Legarra et al., 2014). A l'origine, les éleveurs français ont fait le choix de sélectionner des animaux sans cornes (ce phénotype ségrégeant dans la population). Cependant, à la fin des années 80, 1/3 des agneaux candidats à la sélection étaient cryptorchides (non descente uni ou bilatérale des testicules dans le scrotum, [Figure 15.7](#)) et donc infertiles. Afin de faire face à ce problème, une stratégie avait été mise en place partant du constat que tous les pères des cryptorchides étaient sans cornes (tête plate ou parfois un creux au départ attendu des cornes). Elle consistait donc à ne sélectionner comme reproducteurs mâles que des agneaux cornus issus de mères non cornues. Ces animaux produiraient à leur tour des mâles (possiblement avec des cornes mais non cryptorchides) et des femelles sans cornes pour les élevages. A l'heure actuelle le problème n'est pas vraiment résolu puisque 20% des femelles élites (mères à béliers) sont disqualifiées sur standard en raison de la présence de cornes, et 5% des agneaux candidats sont encore écartés de la sélection car cryptorchides. C'est dans ce contexte que j'ai été amené à réfléchir de façon globale au lien entre MTRDHH4, *RXFP2*, le cornage et la cryptorchidie.

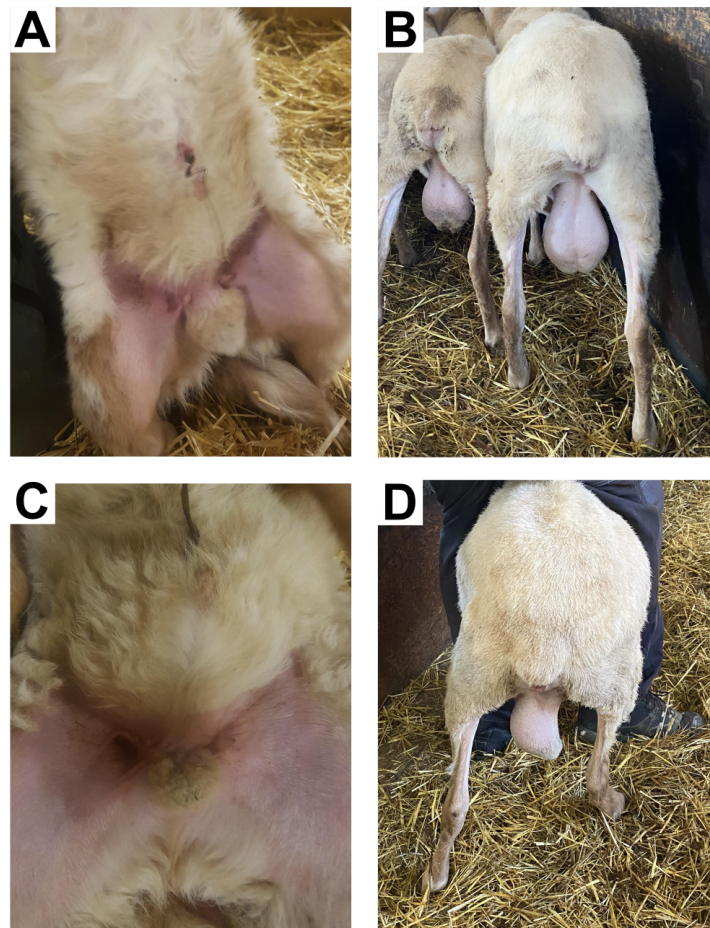


FIGURE 15.7 – Cas de cryptorchidie en race Manech Tête Rousse. (A) Agneau témoin (7 jours), (B) Agneau témoin (7 mois), (C) Agneau cryptorchide bilatéral (7 jours), (D) Agneau cryptorchide unilatéral (7 mois).

15.3.1 Ségrégation de la mutation cornage en Manech Tête Rousse

La présence de cornes permet aux ruminants de se défendre des prédateurs et a un rôle important dans les interactions sexuelles. Cependant, lors de la domestication, les animaux naturellement sans cornes ont été sélectionnés pour les facilités d'élevage. Un déterminisme génétique de la présence/absence de cornes a été identifié par Wiedemar et Drögemüller (2015) sous la forme d'un rétrogène, une insertion anti-sens d'un ARN épissé issu du gène *EEF1A1* (*Eukaryotic Translation Elongation Factor 1 Alpha 1*) dans la partie 3'UTR du gène *RXFP2* (*Relaxin/Insulin-like Family Peptide Receptor 2*) (OMIA 000483-9940).

Cette mutation de type insertion/délétion est connue dans de nombreuses races ovines, responsable du sans cornes plus particulièrement lorsque l'insertion est à l'état homozygote (génotype As/As). La localisation de la mutation est décrite dans la Figure 15.8. Cette insertion ajoute une séquence d'ARN antisens de *EEF1A1* à l'extrémité 3'UTR du transcrite *RXFP2*. L'hypothèse fonctionnelle d'action de cette mutation est une interférence ARN, comme le mécanisme décrit pour la mutation contrôlant le type de toison (Demars et al., 2017). Les transcrits du gène *EEF1A1* (NC_040259.1, OAR8 :282,488-288,942pb) se lieraient à l'extrémité 3' de l'ARNm *RXFP2* muté inséré, donnant un hybride ARN double brin qui serait dégradé, réduisant ainsi les niveaux de production de la protéine *RXFP2*. La dose de *RXFP2* ne serait donc plus suffisante

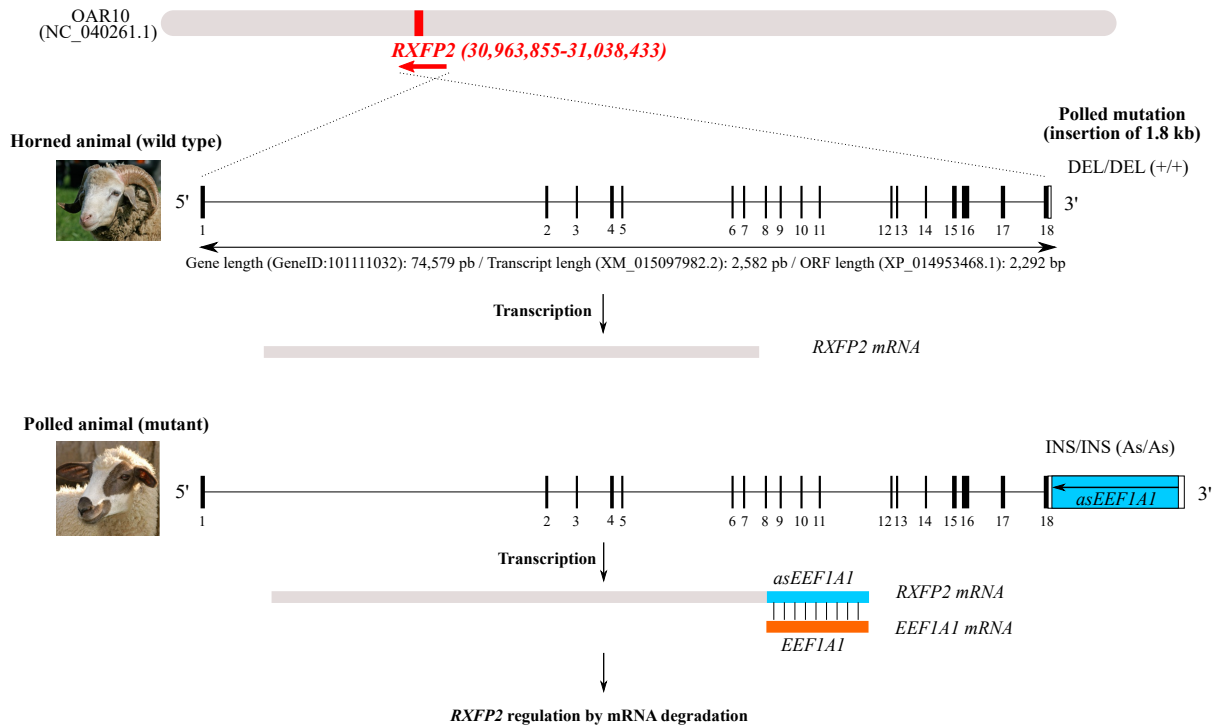


FIGURE 15.8 – Description de la mutation de type insertion/délétion dans le 3'UTR du gène *RFXP2* responsable de l'absence/présence de cornes.

pour le développement initial des cornes.

Afin de déterminer si la mutation ségrège dans les races laitières françaises (Figure 15.9), j'ai extrait les marqueurs de la puce basse densité (*RFXP2_insert_L1* (A/G) et *RFXP2_insert_R1* (A/T)) qui permettent de fournir le génotypage à la mutation présence/absence de cornes. Ainsi, on constate que l'insertion (As) est fixée en Lacaune (race sans cornes) mais n'existe pas dans les races cornues Basco-Béarnaise et Manech tête Noire. Dans la race Corse (race cornue), on observe une faible ségrégation de l'allèle "As" avec une fréquence de 0,01. De manière intéressante, l'allèle "As" sans cornes ségrège dans la population Manech Tête Rousse, pour laquelle on observe des animaux cornus et sans cornes, avec une fréquence allélique de 0,62 (Figure 15.10).

Afin de caractériser l'effet de la mutation, 1 008 mâles (324 béliers adultes du CIA et 684 agneaux génomiques de la cohorte 2021) et 414 brebis Manech Tête Rousse génotypées ont été phénotypées pour le cornage en trois catégories : cornes, absence de cornes et le phénotype intermédiaire présentant des bourgeons de cornes ou chicots. L'analyse génotype/phénotype montre un déterminisme génétique dominant du cornage chez le mâle. En effet 68% des mâles hétérozygotes "+/As" ont soit des cornes développées soit des chicots. Chez la femelle, le déterminisme semble récessif, les deux copies de la délétion sont nécessaires pour avoir des cornes (Figure 15.11). A noter que chez la femelle les cornes sont beaucoup moins développées que chez le mâle (Figure 15.10).

Ainsi, la mutation As dans le 3'UTR du gène *RFXP2* ségrège en race Manech Tête Rousse. L'absence d'insertion As (+/+) conduit à des animaux cornus. A l'état hétérozygote (+/As), on retrouve des animaux cornus, avec chicots ou sans cornes et à l'état homozygote inséré (As/As),

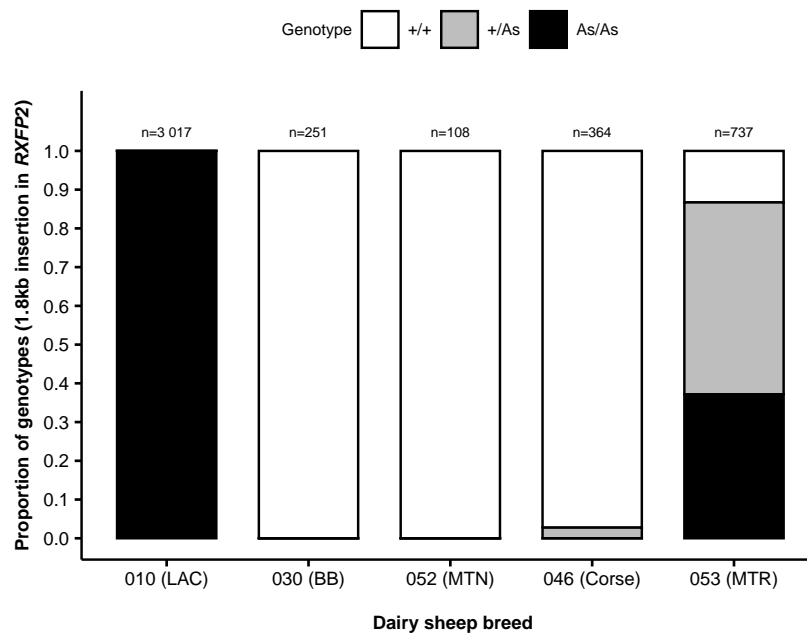


FIGURE 15.9 – Distribution des génotypes de l'insertion "As" de 1,8kb dans *RXFP2* pour la cohorte d'agneaux MTR de la sélection génomique 2021 (n=4 477).

la mutation conduit à l'absence de cornes ([Figure 15.12](#)).



FIGURE 15.10 – Cornage en Manech Tête Rousse. (A) Béliers sans cornes, (B) Béliers cornus, (C, D) Béliers avec chicots (bourgeons de cornes plus ou moins développés), (E,F) Femelles cornues.

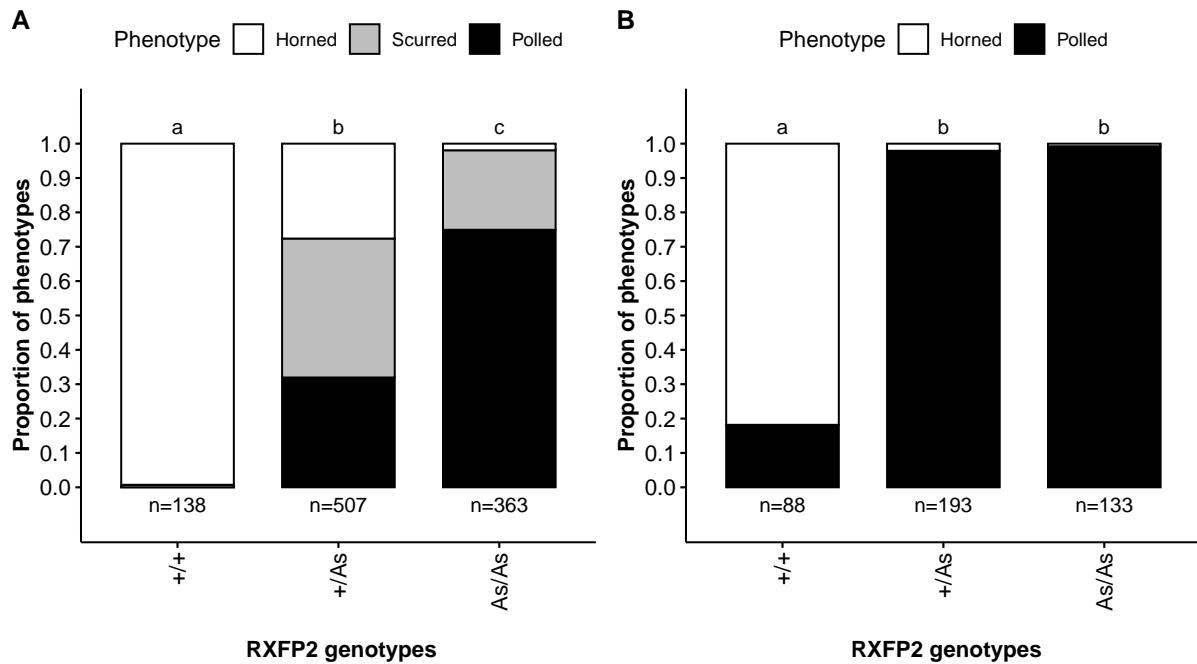


FIGURE 15.11 – Lien entre les génotypes à l'insertion "As" de 1,8kb dans le gène *RXFP2* et les phénotypes de cornage (A) chez les mâles (n=1 008) et (B) chez les femelles (n=414). Horned= animal cornu, Scurred= animal avec chicots (bourgeons de cornes variables de quelques centimètres), Polled= animal sans cornes. Le phénotype des femelles a été établi à partir du fichier de qualification de la race (disqualification lorsque la brebis possède des cornes). Des lettres différentes indiquent des différences statistiquement significatives (Test de Fisher, $P < 0,05$).

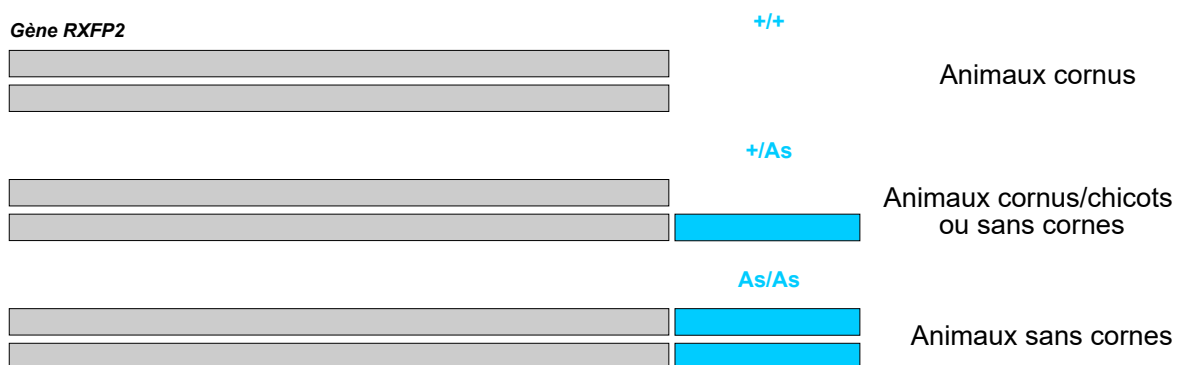


FIGURE 15.12 – Effet de l'insertion "As" de 1,8kb dans le gène *RXFP2* sur le phénotype de cornage en race Manech Tête Rousse.

15.3.2 Etude de l'expression de *RXFP2*

Aucune publication à ce jour n'a étudié finement le mécanisme moléculaire de la mutation. Afin de travailler sur ce sujet, j'ai collaboré avec Aurélien Capitan (INRAE), Julie Rivière (INRAE) et Aurélie Bonnet (Elliance) possédant des échantillons biologiques d'intérêt.

Une première étape s'est focalisée sur l'expression de *RXFP2* en immunohistochimie (IHC) à partir d'échantillons de peau frontale et de bourgeon de corne d'un fœtus témoin de génotype "+/+" à 76 jours post-fécondation. Les échantillons ont été fixés dans du formol à 10% pendant 24 heures et inclus dans la paraffine selon les protocoles histologiques standard. Des sections de cinq micromètres d'épaisseur ont été montées sur des lames adhésives (Klinipath-KP-PRINTER ADHESIVES). L'immunomarquage a été réalisé à l'aide du système de coloration Leica Bond RXm IHC avec le kit Bond Polymer Refined (DS9800) et un anticorps dirigé contre l'extrémité N-terminale de *RXFP2* (<https://www.sigmaaldrich.com/FR/fr/product/sigma/sab2900590>) dilué à 1%. Après révélation, les lames ont été scannées (Pannoramic Scan 150; 3D Histech) pour visualisation (Viewer CaseCenter 2.9; 3D Histech). L'analyse IHC montre une expression très localisée de *RXFP2* dans l'épiderme et certaines cellules des follicules pileux (Figure 15.13).

La seconde étape consistait à étudier l'expression de *RXFP2* dans des tissus de génotype "+/+" et "As/As" par PCR quantitative pour tenter de mettre en évidence un effet dose de l'allèle "As". Ces expérimentations sont en cours de réalisation.

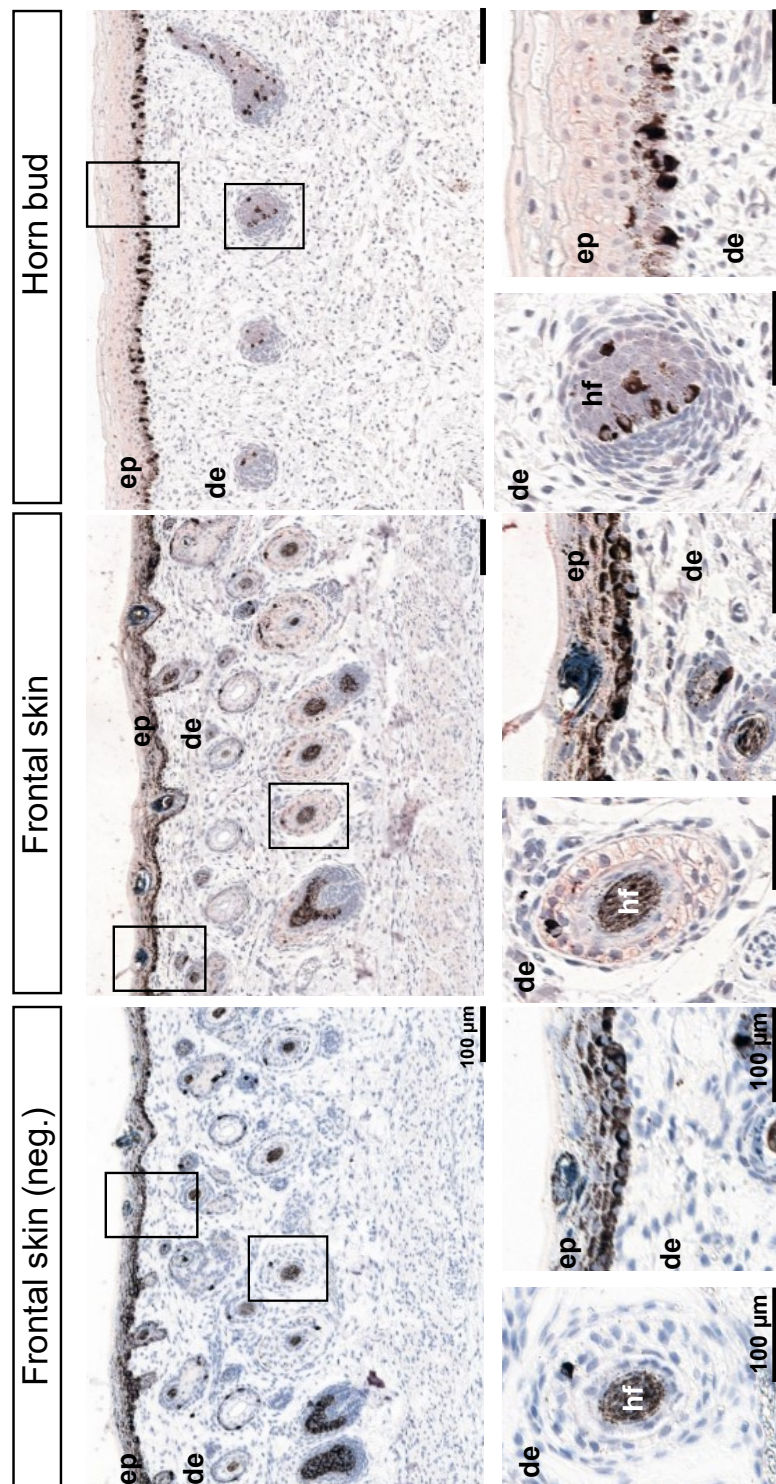


FIGURE 15.13 – Immunomarquage de la protéine RXFP2 dans des coupes tissulaires de peau frontale et de bourgeon de corne d'un fœtus témoin (+/+) à 76 jours post-fécondation. Neg. : contrôle négatif. Les panneaux inférieurs représentent des agrandissements des follicules pileux (à gauche), du derme supérieur et de l'épiderme (à droite). Barres = 100 μ m. ep= épiderme ; de=derme ; hf=follicule pileux. Travail réalisé par Julie Rivière et Aurélien Capitan (INRAE).

15.3.3 Identification d'une mutation causale de la cryptorchidie

Dans la littérature, la cryptorchidie est une anomalie connue dans de nombreuses espèces. Chez l'homme, de nombreuses mutations sont identifiées dans le gène *INSL3* (*insulin-like-3*) (OMIM 219050). Ce gène code pour une hormone appartenant à la superfamille des peptides analogues à l'insuline et est le seul ligand physiologique connu du récepteur RXFP2, dont la signalisation est couplée aux protéines G (GPCR). Le complexe INSL3/RXFP2 a un rôle essentiel dans le développement du gubernaculum pour la descente transabdominale initiale du testicule et dans le maintien d'une bonne santé reproductive chez l'homme (Esteban-Lopez & Agoulnik, 2020). Des mutations localisées dans *RXFP2* sont également causales pour la cryptorchidie chez l'homme (Ayers et al., 2019).

Une mutation dans l'un des deux gènes pourrait donc avoir des conséquences sur la descente des testicules. Ainsi, j'ai extrait les variants ségrégeant dans les séquences tout génome des 22 Manech Tête Rousse disponibles et présents dans les 2 gènes candidats fonctionnels pour expliquer la cryptorchidie, *INSL3* et *RXFP2* (Figure 15.14 A et C). La recherche de variants, avec un impact prédit modéré ou fort, met en évidence 9 mutations candidates (8 mutations faux-sens et une mutation décalage cadre de lecture) dans le gène *RXFP2* mais aucun variant candidat n'est localisé dans le gène *INSL3* (Table 15.1).

TABLE 15.1 – Liste des mutations identifiées en Manech Tête Rousse localisées dans le gène *RXFP2*. Seules les mutations prédites avec un impact modéré ou fort sont considérées. Les positions sont données sur la version du génome ovin Oar_rambouillet_v1.0.

Mutation	Annotation	Conséquence fonctionnelle
g.31,038,419G>T	Exon 1, <i>RXFP2</i> (c.15C>A)	MODERATE, faux-sens (p.His5Gln)
g.31,038,408G>A	Exon 1, <i>RXFP2</i> (c.26C>T)	MODERATE, faux-sens (p.Ser9Phe)
g.31,038,397C>T	Exon 1, <i>RXFP2</i> (c.37G>A)	MODERATE, faux-sens (p.Glu13Lys)
g.31,038,357G>A	Exon 1, <i>RXFP2</i> (c.77C>T)	MODERATE, faux-sens (p.Ser26Leu)
g.31,038,351A>G	Exon 1, <i>RXFP2</i> (c.83T>C)	MODERATE, faux-sens (p.Ile28Thr)
g.31,002,724C>G	Exon 4, <i>RXFP2</i> (c.377G>C)	MODERATE, faux-sens (p.Lys126Ser)
g.30,974,959C>T	Exon 14, <i>RXFP2</i> (c.1117G>A)	MODERATE, faux-sens (p.Glu373Lys)
g.30,964,387delC	Exon 18, <i>RXFP2</i> (c.2050delG)	HIGH, frameshift (p.Val684Serfs*8)
g.30,964,378G>A	Exon 18, <i>RXFP2</i> (c.2059C>T)	MODERATE, faux-sens (p.Leu687Phe)

On retrouve le variant g.31,038,357G>A (c.77C>T, p.Ser26Leu) dans l'exon 1 de *RXFP2* initialement associé à MTRDHH4. Mais, compte tenu des conséquences fonctionnelles prédites des mutations, j'ai priorisé la mutation g.30,964,387delC (c.2050delG) dans l'exon 18 du gène *RXFP2*,

car elle crée un décalage du cadre de lecture et conduit à l'apparition prématurée d'un codon stop en position 691 (p.Arg246Alafs*3). Cette mutation entraîne la perte du dernier domaine transmembranaire et de la queue C-terminale du récepteur *RXFP2* (p.Val684Serfs*8) ([Figure 15.14 D et E](#)).

Afin de valider la mutation candidate OAR10 :g.30,964,387delC identifiée dans le gène *RXFP2*, des agneaux atteints de cryptorchidie (bilatérale n=134, unilatérale n=23) et des agneaux témoins (n=53) ont été phénotypés pour la cryptorchidie et le cornage, 7 jours après leur naissance. Des prélèvements sanguins ont été réalisés sur ces animaux pour réaliser les génotypages. Ces animaux viennent compléter les échantillons des 1 008 mâles non cryptorchides par définition puisqu'il s'agit des 324 béliers adultes actifs du CIA et des 684 agneaux génomiques de la cohorte 2021 sélectionnés pour être reproducteurs. Afin de génotyper la mutation sur ces 1 218 mâles, j'ai développé un test moléculaire spécifique de la mutation (voir [Annexe : Table 21.1](#)).

L'étude des effets des deux mutations, insertion 1,8kb (+, As) et c.2050delG (+, DelG), dans le gène *RXFP2* sur le développement des cornes et la descente testiculaire est représentée dans la [Figure 15.15](#). On confirme que l'insertion "As" à l'état homozygote est bien responsable de l'absence de cornes chez le mâle (61% d'animaux sans cornes sont As/As). Cependant, si on associe un allèle DelG, on accentue le phénotype avec plus de 90% de sans cornes, et cela passe à quasiment 100% chez les double-homozygotes As/As_DelG/DelG. L'effet est encore plus flagrant sur les hétérozygotes +/As dont la proportion de sans cornes passe de 14% à 72% en présence d'un allèle DelG (+/As_+/DelG). Ceci indique un effet fonctionnel de la mutation DelG sur le cornage. Étonnement, l'allèle DelG n'est jamais retrouvé sans la présence de l'allèle As. L'allèle DelG semble donc être en déséquilibre de liaison total avec l'insertion "As".

Concernant le phénotype de descente testiculaire ([Figure 15.15 B](#)), on constate que la très grande majorité des cas de cryptorchidie bilatérale est de génotype DelGAs/DelGAs, alors que les mâles As/As sont normaux. Ceci indique que la mutation DelG dans *RXFP2* serait responsable de la cryptorchidie chez les agneaux MTR. En complément, tous les agneaux cryptorchides unilatéralement sont hétérozygotes à cette mutation (+/DelG). L'association des mutations As et DelG dans *RXFP2* expliquerait donc à elle-seule le lien observé par les éleveurs entre absence de cornes et cryptorchidie en race MTR. Dans la cohorte d'agneaux 2021, la mutation c.2050delG ségrége à une fréquence allélique de 0,17.

Afin de mieux caractériser l'effet des deux mutations, 200 béliers du CIA âgés de 1 an et demi, et issus de la cohorte sélection génomique 2021 pour lesquels un phénotypage avait été réalisé à 3 mois, ont été phénotypés pour le diamètre de la base des cornes (avec un pied à coulisse) et la circonférence scrotale (avec un mètre ruban) ([Figure 15.16](#)). On retrouve un effet dose des deux mutations en lien avec le cornage, avec un effet fort de la mutation c.2050delG à l'état hétérozygote qui réduit la circonférence des cornes. Cependant, les deux mutations n'ont aucun effet sur la circonférence scrotale des béliers.

Avec cette étude, j'ai identifié une nouvelle mutation récessive responsable de la cryptorchidie chez les ovins mâles. Cette mutation OAR10 : g.30,964,387delC dans le gène *RXFP2* (c.2050delG), située en amont de la mutation As liée au cornage, semble a minima capable de contrôler également le cornage renforçant l'effet de la mutation As. En parallèle, j'ai génotypé la mutation c.2050delG

chez 413 brebis phénotypées pour le cornage (Figure 15.17). On constate que les brebis cornues sont de génotypes $+/+$ (non porteuses des deux mutations). La présence d'une seule copie de l'allèle "As" suffit pour avoir des brebis sans cornes. Néanmoins, j'ai aussi identifié 13 brebis sans cornes de génotype DelGAs/DelGAs pour lesquelles on imagine un effet sur la reproduction de ces brebis. En effet, INSL3, ligand de RXFP2, est largement exprimé par les cellules ovariennes spécifiquement chez les ruminants et constitue un régulateur majeur de la physiologie de la reproduction femelle (Ivell & Anand-Ivell, 2018). Cependant, l'étude des carrières reproductives de ces brebis ne permet pas de mettre en évidence un impact de la mutation DelG à l'état homozygote chez la femelle.

Finalement, le lien entre les deux mutations, l'insertion de 1,8kb (As) et la délétion c.2050delG, identifiées dans le gène *RXFP2* et l'impact sur les phénotypes de cornage et cryptorchidie est synthétisé dans la Figure 15.18. On fait l'hypothèse que la mutation c.2050delG est apparue dans un haplotype sans cornes contenant l'allèle As. La sélection des animaux sans cornes a tiré la mutation responsable de la cryptorchidie par le phénomène d'auto-stop.

La rétro-insertion de 1,8kb dans *RXFP2*, responsable de l'absence de cornes est une mutation connue dans de nombreuses races ovines. On peut donc supposer que la nouvelle mutation c.2050delG dans *RXFP2*, en déséquilibre de liaison avec l'allèle As, existe également dans d'autres races. Pour répondre à cette question, j'ai génotypé les deux mutations de *RXFP2* sur un panel d'ADN "diversité ovine" comprenant des animaux issus de nombreuses races laitières et allaitantes (Rochus et al., 2018) (Figure 15.19). On observe que la mutation DelG responsable de la cryptorchidie ségrége dans les populations Manech Tête Rousse (FRA), Latxa Cara Rubia (ESP), Latxa Cara Negra Euskadi (ESP), Mérinos d'Arles (FR) et Causse du Lot (FRA) pour lesquelles ségrége aussi l'allèle "As".

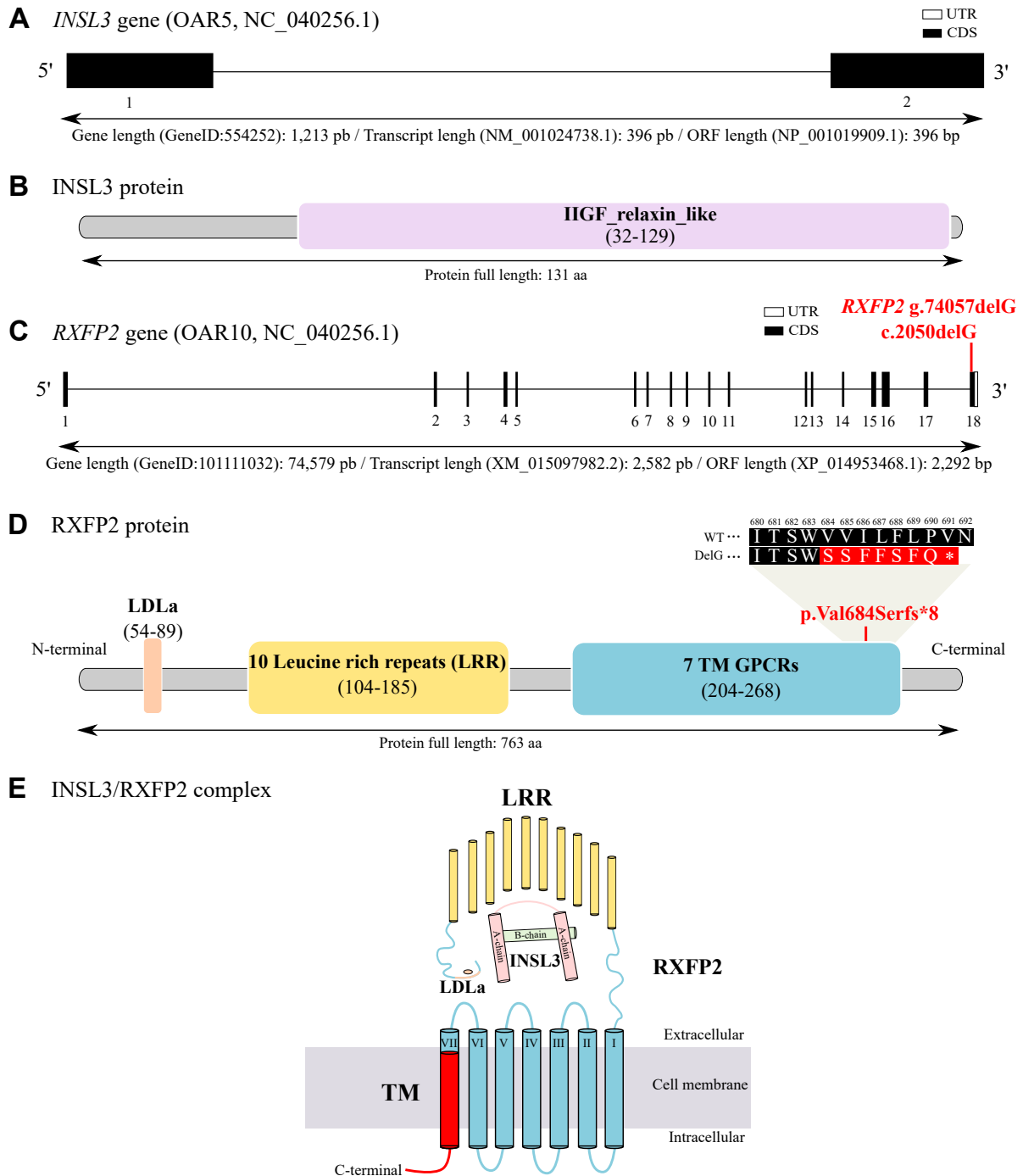


FIGURE 15.14 – Description des gènes *INSL3* et *RXFP2*. (A) Structure du gène *INSL3* (NC_040256.1, OAR5 :5,327,208-5,328,420pb) (B) Protéine INSL3 (NP_001019909.1) composée d'un domaine "IGF_relatin_like" (C) Structure du gène *RXFP2* (NC_040261.1, OAR10 :30,963,855-31,038,433pb) et localisation de la mutation *RXFP2* g.74057delG (c.2050delG) identifiée dans le dernier exon (XM_015097982.2). (UTR : région non traduite; CDS : séquence codante) (D) Le récepteur RXFP2 (XP_014953468.1) est constitué d'un domaine lipoprotéine de basse densité (LDLa), de dix domaines répétés riches en leucine (LRR) et de sept domaines hélicoïdaux transmembranaires (TM). La mutation crée un décalage du cadre de lecture et conduit à l'apparition prématurée d'un codon stop en position 691 (p.Val684Serfs*8) (E) Complexe INSL3 (ligand)/RXFP2 (récepteur), adapté d'après Esteban-Lopez et Agoulnik (2020). La mutation conduit à la perte du domaine transmembranaire VII et de la partie C-terminale (symbolisés en rouge).

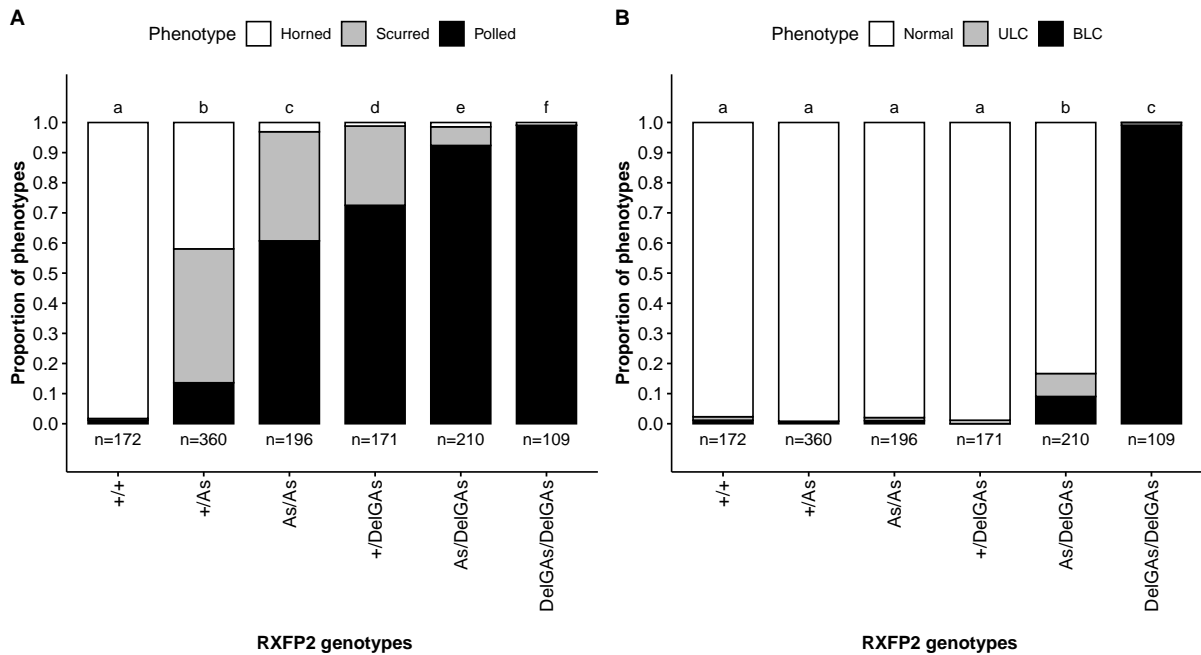


FIGURE 15.15 – Effets des deux mutations, insertion 1,8kb (+, As) et c.2050delG (+, DelG), dans le gène *RXFP2* sur le développement des cornes (A) et la descente testiculaire (B) chez le mâle. Au total, 1 218 mâles MTR composés de 172, 360, 196, 171, 210 et 109 individus par catégorie de génotype (de gauche à droite, respectivement) ont été phénotypés pour les deux caractères entre 7 jours et 5 ans d'âge. BLC : cryptorchidie bilatérale; ULC : cryptorchidie unilatérale. Des lettres différentes indiquent des différences statistiquement significatives (Test de Fisher, $P < 0,05$).

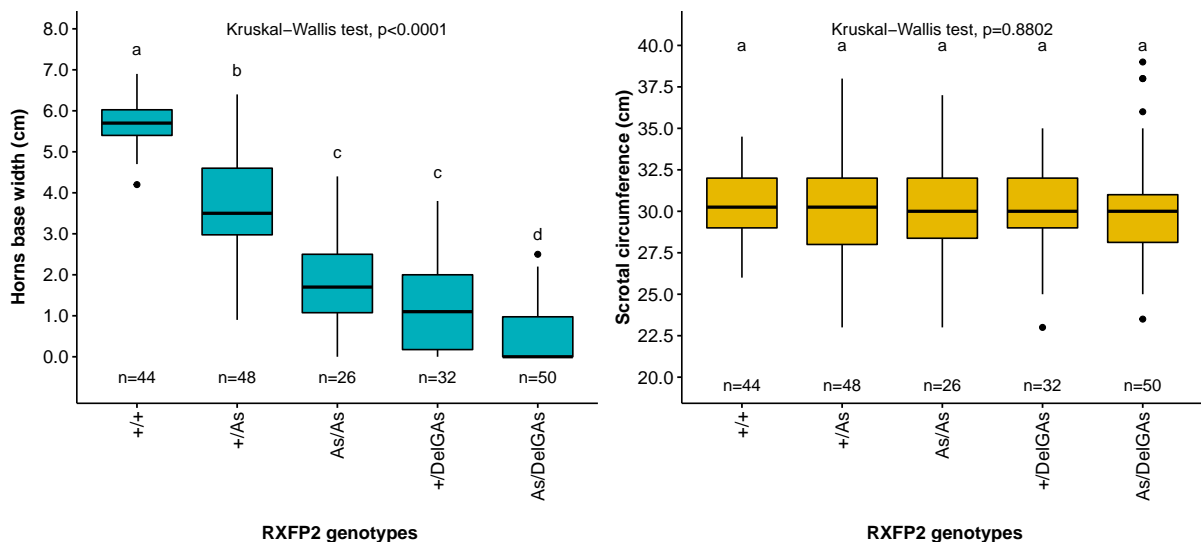


FIGURE 15.16 – Effets des deux mutations insertion 1,8kb (+, As) et c.2050delG (+, DelG) dans le gène *RXFP2* sur le diamètre des cornes (A) et la circonférence scrotale (B) chez le bélier ($n = 200$). Des lettres différentes indiquent des différences statistiquement significatives (Test de Wilcoxon avec comparaison 2 à 2, $P < 0,05$).

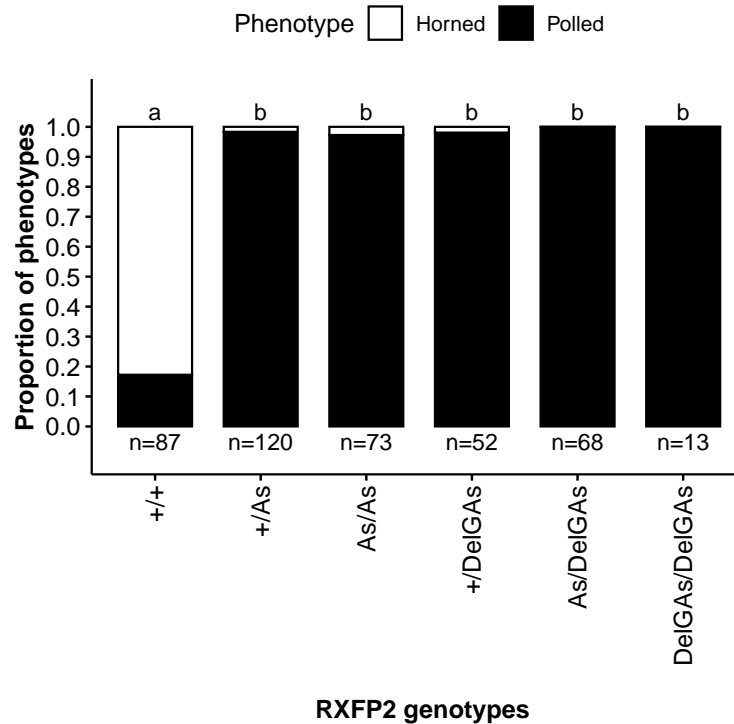


FIGURE 15.17 – Effets des deux mutations insertion 1,8kb (+, As) et c.2050delG (+, DelG) dans le gène *RXFP2* sur le développement de la corne chez les femelles (n=413). Des lettres différentes indiquent des différences statistiquement significatives (Test de Fisher, $P < 0,05$).



FIGURE 15.18 – Représentation des allèles As et DelG du gène *RXFP2* et leurs effets phénotypiques sur le cornage et la cryptorchidie.

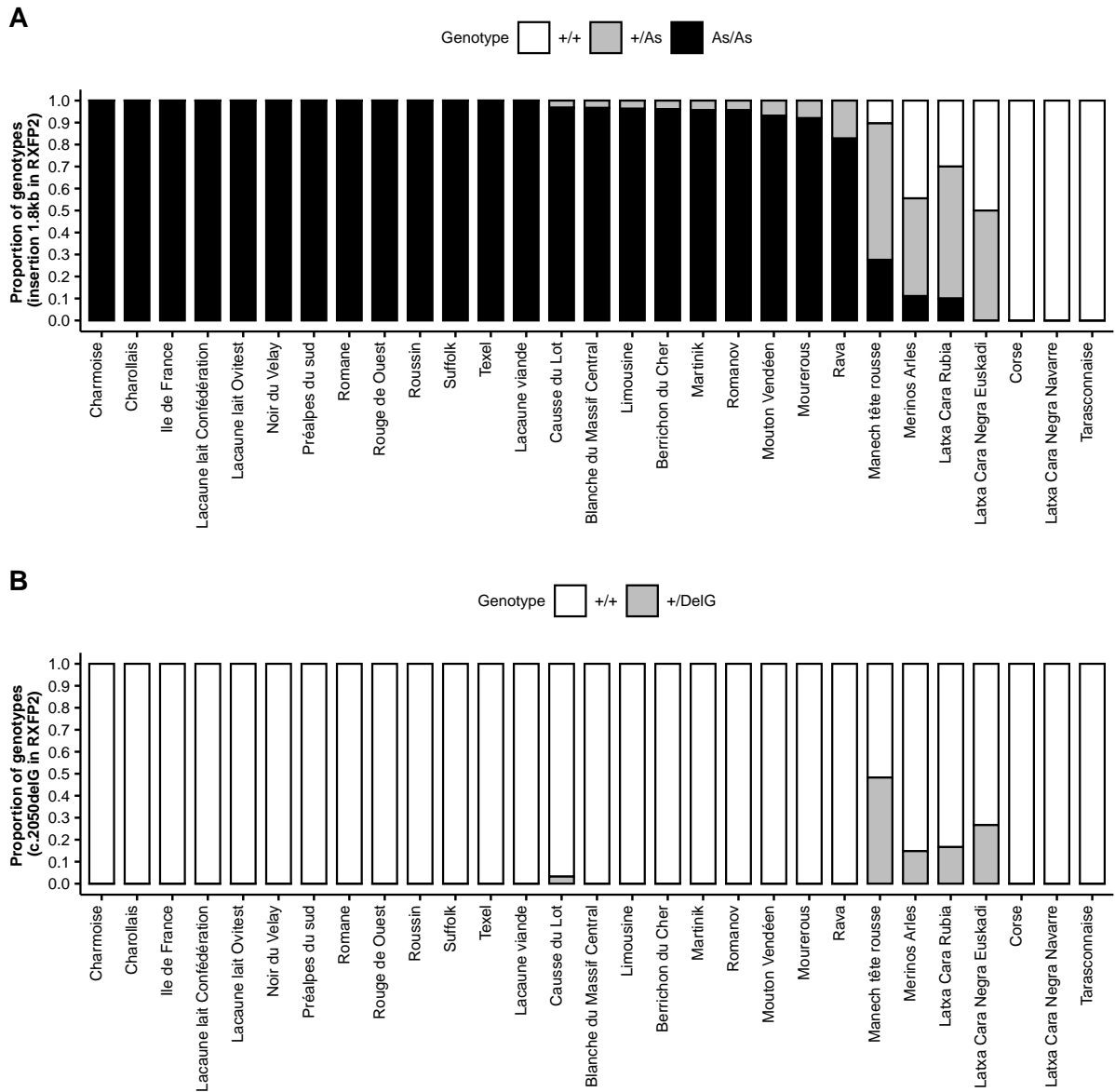


FIGURE 15.19 – Distribution des génotypes de l'insertion de 1,8kb (A) et de la délétion c.2050delG (B) identifiées dans le gène *RXFP2* dans un panel diversité composé de 29 races ovines. Le nombre d'individus par race varie entre 17 et 40.

15.3.4 Lien entre MTRDHH4 et cryptorchidie

Précédemment, nous avons fait l'hypothèse que MTRDHH4, qui contient le gène *RXFP2*, est associé partiellement à l'insertion de 1,8kb dans le gène *RXFP2* en lien avec le cornage. Avec la nouvelle mutation identifiée dans *RXFP2* (c.2050delG) responsable de la cryptorchidie, il est probable que le déficit dans la région de MTRDHH4 soit en fait lié à la cryptorchidie. En effet, les agneaux cryptorchides sont de fait écartés de la sélection (incapacité reproductive future) et ne sont donc pas génotypés. D'ailleurs, l'équilibre des génotypes à cette mutation pour la cohorte d'agneaux 2021 pour la sélection génomique (n=684) ne respecte pas la loi d'Hardy-Weinberg ($\text{Chi}^2 = 27,65$, $\text{df} = 2$, P-valeur = $9,9 \times 10^{-7}$), avec un déficit total d'animaux homozygotes (19 attendus, zéro observé).

Afin de mieux comprendre cette région, j'ai extrait tous les haplotypes (fréquence $\geq 1\%$, déficit > 0) présents dans cette région, nommés MTRDHH4a à MTRDHH5m, triés par P-valeur croissante (Figure 15.20). Initialement lors de la première analyse en Août 2021 (6 845 animaux MTR), j'ai identifié un seul haplotype satisfaisant les critères (fréquence $\geq 1\%$, déficit $\geq 75\%$ et P-Poisson $\leq 1,9 \times 10^{-4}$) et définissant MTRDHH4 (appelé par la suite MTRDHH4b). Cependant, lors de la nouvelle analyse de Septembre 2022 (7 405 animaux MTR), j'ai identifié un nouvel haplotype significativement en déficit dans la même région (nommé MTRDHH4a) pour lequel 11 animaux homozygotes sont attendus alors que 2 sont observés.

L'analyse d'association entre le statut aux 11 haplotypes identifiés dans la région MTRDHH4 et la mutation c.2050delG dans *RXFP2*, montre un fort lien pour 5 haplotypes de la région MTRDHHa, b, g, h et l (Figure 15.21). En effet, tous les animaux hétérozygotes porteurs de ces haplotypes sont hétérozygotes à la mutation. Lorsque l'on regarde les différents marqueurs qui composent ces haplotypes associés, on distingue un haplotype recombinant plus court, composé de 4 marqueurs "AGGC" en position 453 à 456, entourant la mutation c.2050delG (Figure 15.20). On retrouve également cet haplotype plus court "AGGC" dans les haplotypes MTRDHH4c, f et k, possiblement porteurs de formes ancestrales de l'haplotype plus court sans la mutation c.2050delG. L'analyse d'association avec l'insertion de 1,8kb dans *RXFP2* confirme bien que la mutation c.2050delG n'existe que dans les haplotypes associés à l'allèle sans cornes "As" (Figure 15.22). Tous les animaux hétérozygotes aux haplotypes MTRDHH4a, b, g, h ou l ne sont pas porteurs de l'allèle sauvage du cornage.

Cependant, pour les haplotypes MTRDHH4a et b associés à la mutation c.2050delG, on observe 2 animaux homozygotes à l'haplotype pour MTRDHH4a et 1 animal homozygote à l'haplotype pour MTRDHH4b (Figure 15.20). En réalité, ces trois animaux sont des femelles, certainement mères à béliers, ne présentant pas de défauts morphologiques visibles et sans cornes.

En conclusion, la mutation c.2050delG dans *RXFP2*, responsable de la cryptorchidie en auto-stop avec la mutation sans cornes (allèle "As"), est très certainement responsable du déficit en animaux homozygotes identifié par l'haplotype MTRDHH4.

Haplotype	NbAtt	NBObs	NBObs pop	Frq he	Déficit	P-Poisson	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470								
MTRDHH4a	14	2	2	9.7%	86%	1.04E-04	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A							
MTRDHH4b	11	1	1	8.4%	91%	1.45E-04	G	G	A	A	A	A	A	A	A	A	A	A	A	A	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A						
MTRDHH4c	29	18	27	13.7%	38%	1.94E-02	A	A	A	A	A	A	C	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A					
MTRDHH4d	14	9	15	9.9%	37%	9.52E-02	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A				
MTRDHH4e	7	4	6	6.2%	43%	1.75E-01	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A				
MTRDHH4f	26	21	30	12.6%	19%	1.87E-01	G	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A				
MTRDHH4g	1	0	0	3.6%	100%	2.59E-01	G	G	A	A	A	A	A	A	A	A	A	A	A	A	C	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
MTRDHH4h	1	0	0	3.1%	100%	3.23E-01	G	G	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
MTRDHH4i	1	0	0	2.7%	100%	3.61E-01	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
MTRDHH4j	5	4	4	6.6%	24%	4.00E-01	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
MTRDHH4k	14	13	17	9.7%	4%	5.10E-01	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
MTRDHH4l	0	0	0	1.6%	100%	8.46E-01	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
MTRDHH4m	0	0	0	1.8%	100%	8.53E-01	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A

FIGURE 15.20 – Liste des 13 haplotypes localisés dans la région de MTRDHH4. NbAtt= nombre d'animaux homozygotes attendus ; NbObs= nombre d'animaux homozygotes observés ; NbObs pop= nombre d'animaux homozygotes observés dans l'ensemble de la population génotypée (n=7 405) ; Frq he= fréquence des porteurs hétérozygotes. La région MTRDHH4 est définie par 30 marqueurs SNP (marqueurs SNP à la position 441 à 470). La mutation OAR10 :g.30,964,387delC (c.2050delG) dans *RXFP2* est située entre le marqueur 452 (OAR10_29381795.1 ; g.30,850,504A>G) et 453 (OAR10_29538398.1 ; g.31,008,754A>X), localisée par la barre noire.

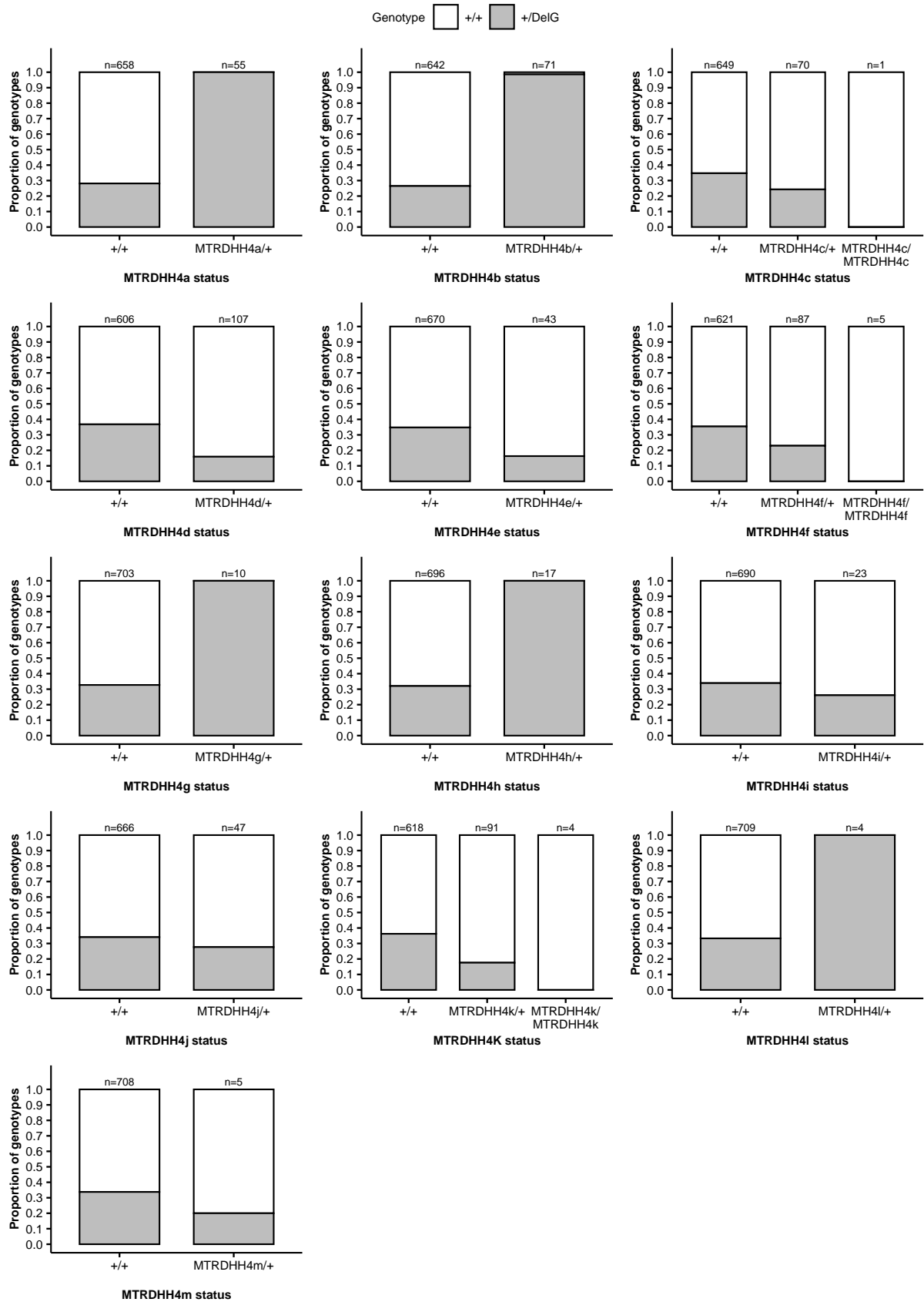


FIGURE 15.21 – Lien entre le statut aux haplotypes identifiés dans la région *MTRDHH4* et la mutation *OAR10* :g.30,964,387delC (c.2050delG) dans *RXFP2* (n=713). Les haplotypes *MTRDHH4a* et *MTRDHH4b* sont deux haplotypes significativement en déficit d’animaux homozygotes.

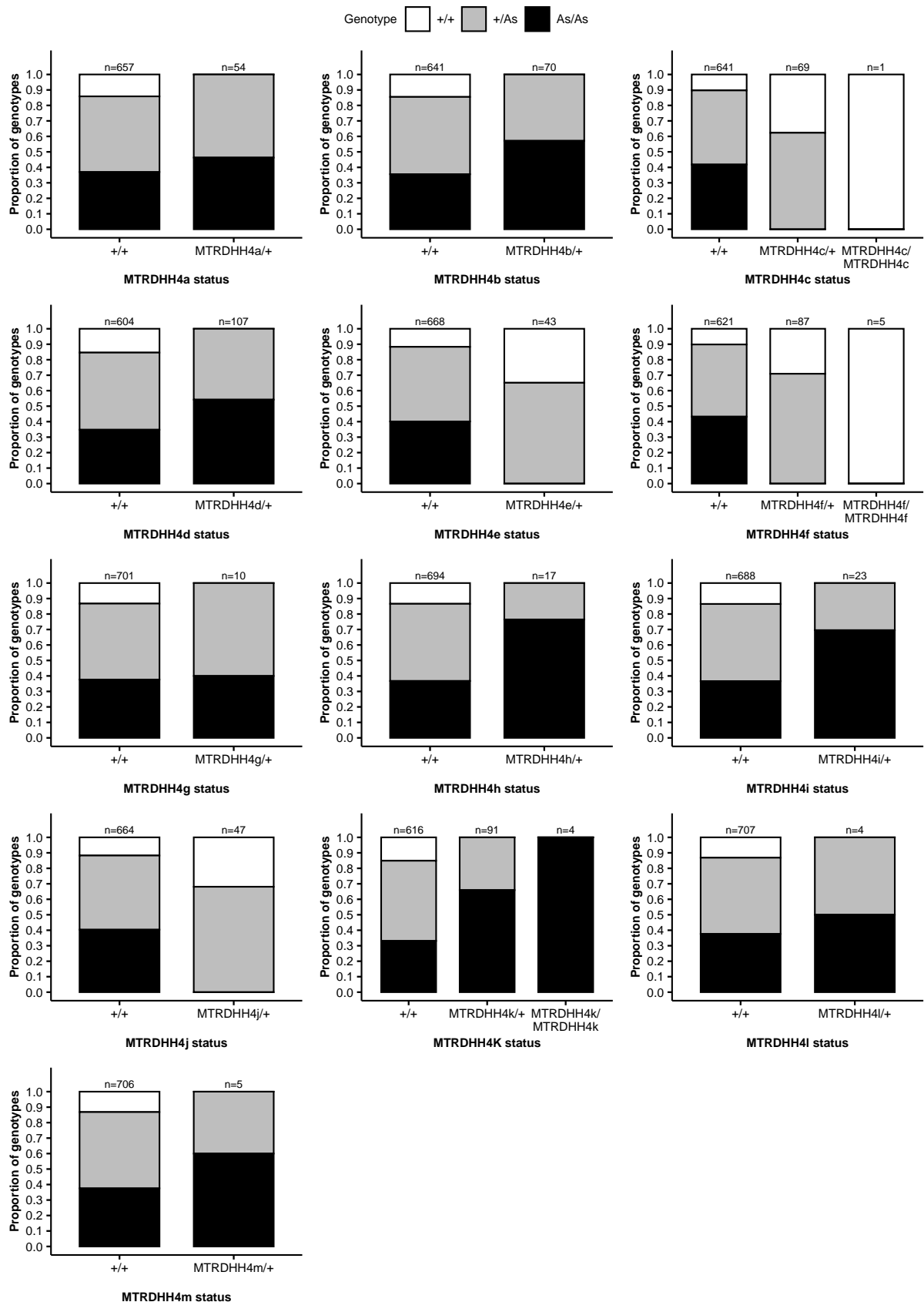


FIGURE 15.22 – Lien entre le statut aux haplotypes identifiés dans la région MTRDHH4 et l'insertion de 1,8kb dans le 3'UTR du gène *RXFP2* (n=711). Les haplotypes MTRDHH4a et MTRDHH4b sont deux haplotypes significativement en déficit d'animaux homozygotes. Le génotypage spécifique de la mutation a été extrait de la puce basse densité pour les agneaux génomiques de la cohorte 2021.

VI

Discussion générale et perspectives

16	Génétique inverse pour la recherche de déficit d'animaux homozygotes chez les petits ruminants	249
16.1	Existence d'haplotypes en déficit d'animaux homozygotes dans les races ovines laitières	
16.2	Réflexion sur la méthode de recherche de déficit	
16.3	Quelles sont les opportunités futures ?	
17	Les mutations dans les haplotypes en déficit	253
17.1	Découverte de nombreuses mutations récessives candidates	
17.2	Limites de la méthode utilisée pour rechercher les mutations	
18	Causalité des mutations	257
18.1	Mutations létales embryonnaires	
18.2	Mutations létales périnatales	
18.3	Mutations en lien avec la morphologie	
18.4	Les mutations non étudiées	
18.5	Pertinence des mutations identifiées en tant que modèles animaux de pathologies	
19	Gestion dans les schémas de sélection	267
19.1	Évolution de la fréquence des mutations	
19.2	Les mutations identifiées sont-elles spécifiques de races ?	
19.3	Comment gérer les anomalies génétiques dans les schémas de sélection ?	

Au cours de ma thèse, j'ai été amené à utiliser différentes approches de génétique quantitative, de génétique moléculaire et de génomique pour mettre en évidence de nouvelles mutations récessives délétères chez les petits ruminants. Des compétences en zootechnie m'ont également permis d'apporter des éléments de causalité pour un certain nombre de mutations. Dans le cadre de ce travail de thèse, j'ai pu atteindre mes trois objectifs principaux initialement fixés. Le premier consistait à étudier la faisabilité d'utiliser l'approche de génétique inverse dans les populations de petits ruminants laitiers engagées dans un programme de sélection génomique. Il s'agissait de détecter des haplotypes en déficit d'animaux homozygotes, supposés abriter des mutations récessives létales. Le deuxième objectif a été d'étudier leurs effets sur les caractères de reproduction (afin de confirmer la létalité) et de production (avantage sélectif des hétérozygotes). Le troisième et dernier objectif consistait à identifier les mutations causales sous-jacentes de ces haplotypes en déficit par l'analyse des séquences de génomes entiers des hétérozygotes et de valider la causalité de quelques mutations. L'ensemble des résultats obtenus pour atteindre ces objectifs sera discuté dans les chapitres suivants.



16. Génétique inverse pour la recherche de déficit d'animaux homozygotes chez les petits ruminants

La recherche de déficit en homozygotes est une méthode initialement développée par VanRaden et al. (2011) pour mettre en évidence des mutations embryonnaires récessives létales jusqu'alors impossibles à détecter par les approches de génétique classique. Cette méthode de génétique inverse est maintenant utilisée dans de nombreuses espèces pour identifier avec succès de nombreux haplotypes en déficit d'animaux homozygotes : en bovins (Résumé dans la [Table 4.1](#)), porcs (Derks et al., 2017 ; Derks, Lopes, et al., 2018), poulets (Derks, Megens, et al., 2018), dindes (Abdalla et al., 2020) et chevaux (Todd et al., 2020). Cependant, aucune analyse dans les populations de petits ruminants n'avait été effectuée avant mon travail de thèse.

16.1 Existence d'haplotypes en déficit d'animaux homozygotes dans les races ovines laitières

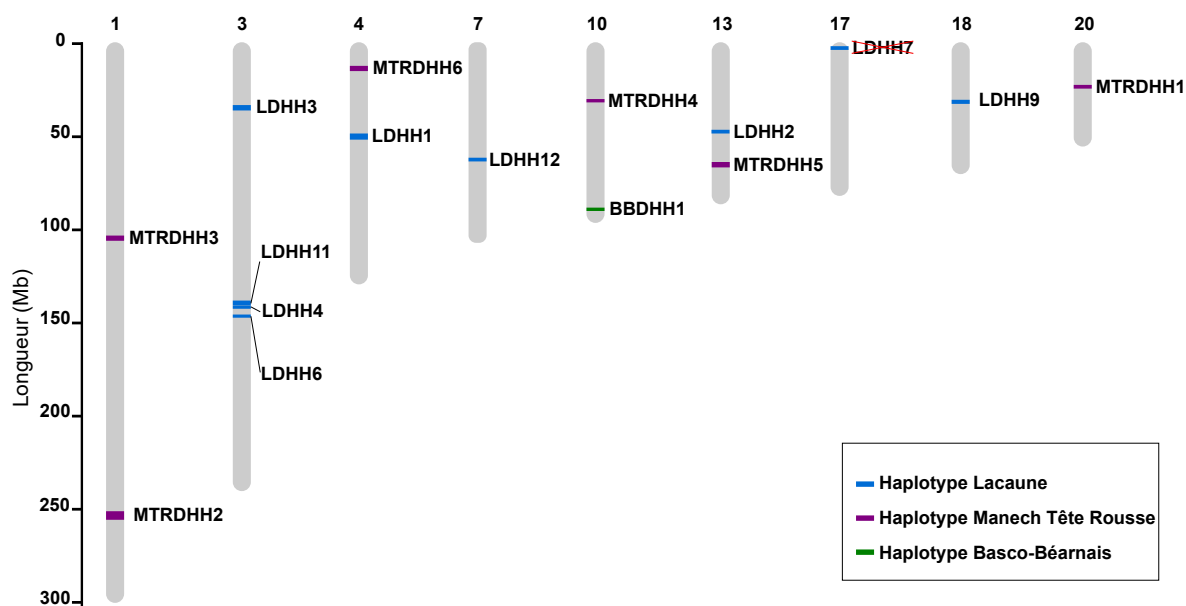


FIGURE 16.1 – Localisation des haplotypes identifiés montrant un déficit significatif d'animaux homozygotes dans les trois populations laitières ovines Lacaune, Manech Tête Rousse et Basco-Béarnaise.

La mise en place de la sélection génomique en ovins (2015 en Lacaune et 2017 en Races Ovines Laitières des Pyrénées, ROLP) et en caprins (2018) permet d'avoir accès aujourd'hui à une grande quantité de génotypes disponibles. La recherche de déficit a été réalisée dans les cinq races ovines (Lacaune, ROLP et Corse) et les deux races caprines (Saanen et Alpine) laitières françaises. Avec cette méthode, j'ai pu mettre en évidence des régions en déficit d'animaux homozygotes en

ovins dans les races Lacaune, la Manech Tête Rousse et Basco-Béarnaise. Ainsi, 14 régions (8 en Lacaune, 5 en Manech Tête Rousse et 1 en Basco-Béarnaise) ont été identifiées au total et sont représentées dans la [Figure 16.1](#). Les haplotypes identifiés sont spécifiques de chaque race puisque les haplotypes signent des régions différentes selon les races.

16.2 Réflexion sur la méthode de recherche de déficit

La disponibilité de grandes populations d'individus génotypés est un pré-requis pour appliquer la méthode de génétique inverse. En France, la mise en place des programmes de sélection génomique spécifiques de race a permis la fourniture de cette grande quantité de génotypes. Sachant que les animaux sont génotypés avant l'âge d'un mois, les déficits d'homozygotes observés permettent de cibler des mutations récessives conduisant à la mort précoce des animaux (stade embryonnaire, foetal, néonatal ou juvénile) avant l'âge du génotypage.

Cependant, l'utilisation de cette approche nécessite d'avoir un nombre de génotypes disponibles suffisant. En effet, les différentes analyses de génétique inverse en bovins ont permis d'estimer les effectifs nécessaires pour identifier des haplotypes ayant des fréquences allant de 1% à 20% ([Table 16.1](#)). En caprins (Saanen et Alpine), aucun haplotype en déficit d'animaux homozygotes n'a été mis en évidence. En effet, même si l'on dispose de 3 025 génotypes en Saanen et 4 003 en Alpine, seulement 65% des animaux génotypés sont informatifs (appartenant à des trios génotypés), limitant la découverte de tels déficits dans ces races pour lesquelles seuls les haplotypes ayant une fréquence supérieure à 10% pourraient être détectés, ce qui ne semble pas être le cas dans les deux populations caprines. Les effectifs d'animaux génotypés sont encore trop faibles en races ovines Manech Tête Noire et Corse pour identifier des régions en déficit.

TABLE 16.1 – **Effectif nécessaire pour mettre en évidence des haplotypes en déficit d'animaux homozygotes.** *D'après Hozé et al. (2020)*

Fréquence de l'haplotype	20%	10%	5%	1%
Effectif nécessaire :	500	2 000	8 000	200 000

Les paramètres d'exécution pour la recherche de déficit en petits ruminants sont similaires à ceux utilisés en bovins (Fritz et al., 2016) avec une fréquence haplotypique $\geq 1\%$ et un déficit $\geq 75\%$ pour cibler des mutations létales. Le choix de la taille de la fenêtre utilisée pour rechercher des haplotypes en déficit d'animaux homozygotes (HHD) est un élément crucial car il ne doit pas être (i) ni trop court car l'haplotype peut exister à une forte fréquence dans la population sans être en déséquilibre de liaison avec la mutation (haplotype non informatif) et (ii) ni trop long car la mutation peut se retrouver associée à plusieurs haplotypes par le biais des recombinaisons. Les analyses sont réalisées avec des génotypes moyenne densité. Avec un marqueur informatif tous les 67kb (génomme ovin de 2,6Gb, nombre de marqueurs informatifs de 38 523 SNP) et une fenêtre de 20 marqueurs, la taille moyenne des HHD est de 1,3Mb, présentant un comportement identique à une mutation létale très généralement de faible fréquence ($<10\%$).

Une différence notable entre notre étude sur les ovins et la littérature sur les bovins est la composition de la population génotypée. En effet, dans les populations ovines laitières, moins de 1% des animaux génotypés ont leurs deux parents génotypés tandis qu'en bovins cet effectif est beaucoup plus élevé variant entre 30% et 50% (Fritz et al., 2018, 2013 ; Mesbah-Uddin et al., 2019). En effet chez les ovins, les mères des candidats à la sélection ne sont pas systématiquement génotypées pour les programmes de sélection génomique, ce qui réduit le pouvoir de détection des HHD dans notre analyse. Ainsi, le calcul du nombre d'animaux homozygotes attendus en ovins repose principalement sur le trio descendant, père et grand-père maternel. Le statut de la mère est prédit par son père et la fréquence de l'haplotype dans la population totale. Le génotypage de femelles pourrait permettre de gagner en puissance en augmentant la fraction homozygote.

Cette méthode, initialement utilisée pour mettre en évidence des mutations létales comme j'ai pu le démontrer au cours de cette thèse, peut également permettre d'identifier des anomalies de morphologie qui conduisent à écarter souvent visuellement des animaux de la sélection. Au cours de ce travail, de tels haplotypes en déficit ont été identifiés en lien avec des déterminismes ou des gènes connus pour la stature/sensibilité aux mammites avec LDHH4 et LDHH11 en lien avec *SOCS2*, ou pour le cornage/cryptorchidie avec MTRDHH4 en lien avec *RXFP2*, ou encore pour la couleur de toison avec MTRDHH5 en lien avec *ASIP* en Manech Tête Rousse. Par des analyses approfondies des régions, j'ai pu mettre en évidence que ces haplotypes sont en fait en déséquilibre de liaison plus ou moins fort avec des mutations sous sélection. Cette relative faible association est certainement liée à la taille des haplotypes qui est incompatible avec la plus forte fréquence de ces mutations. En conséquence, elles se retrouvent associées à plusieurs haplotypes par le biais des recombinaisons. Par exemple, trois haplotypes en Lacaune signent la mutation dans *SOCS2*. Avec les effectifs de génotypages utilisés (un maximum de 35 221 en Lacaune), j'ai pu détecter LDHH5 avec la fréquence la plus faible de 3,7%. Mais cette méthode permet en théorie de détecter des variants encore plus rares. En effet, si l'on considère un haplotype ayant une fréquence à 1%, la fréquence d'homozygotes attendue n'est que de 1/10 000. Pour mettre en évidence un déficit significatif pour de tels variants il faudra attendre d'avoir une à plusieurs centaines de milliers de génotypages comme indiqué dans la [Table 16.1](#).

16.3 Quelles sont les opportunités futures ?

La recherche de HHD peut facilement être réalisée en routine. J'ai pu constater au cours de ma thèse que les effectifs d'animaux génotypés pouvaient évoluer rapidement. Notamment en Lacaune où l'effectif a quasiment doublé en 3 ans puisqu'on passe de 19 102 animaux génotypés en Mars 2019 à 35 221 en Septembre 2022. La recherche de déficit en routine permet de confirmer les haplotypes précédemment identifiés. Comme nous l'avons vu au cours des différentes analyses, cela a permis d'affiner les régions présentant des haplotypes proches en déséquilibre de liaison (haplotypes LDHH4-5, LDHH8-9-10). De plus, l'accumulation des génotypages permet d'identifier de nouvelles régions en déficit d'animaux homozygotes. Ainsi, en Septembre 2022, j'ai pu mettre en évidence un nouvel haplotype en Lacaune présentant un déficit total (9 animaux homozygotes attendus, fréquence de porteurs hétérozygotes de 2,9%) localisé sur OAR7 : 62,547,980-64,671,509 (composé de 25 marqueurs), appelé LDHH12 ([Figure 16.1](#)). De même, en MTR la nouvelle

analyse de Septembre 2022 a permis d'identifier un haplotype en déficit partiel (91%, 11 animaux homozygotes attendus alors qu'un seul animal homozygote est observé), localisé sur OAR4 : 15,022,129-16,701,229 (composé de 27 marqueurs, fréquence de 8,6%). En outre, l'haplotype LDHH7 initialement détecté lors de l'analyse de Mars 2019, n'est plus retrouvé dans l'analyse de Septembre 2022, suggérant qu'il s'agit d'un faux positif. Comme indiqué précédemment, l'accumulation des génotypes permet d'identifier des haplotypes avec des fréquences plus faibles. L'ensemble de ce travail de génétique inverse permettrait même d'identifier des régions abritant des mutations létales avant même leur émergence visible en élevage. Les mutations d'aujourd'hui causeront les anomalies de demain si elles ne sont pas gérées.

La validation de l'utilisation de la méthode en ovins laitiers offre de nouvelles perspectives prometteuses pour identifier de telles régions en ovins allaitants. Deux races allaitantes, Lacaune viande et Blanche du Massif Central, sont en train de mettre en place un programme de sélection génomique permettant d'acquérir de nombreuses données de génotypages dans l'avenir (projet OVIGEN). D'ailleurs, j'ai pu montrer que la mutation létale dans *CCDC65* (LDHH6) découverte en Lacaune lait ségrège aussi dans ces deux populations allaitantes confirmant tout l'intérêt de l'approche de génétique inverse pour ces races (Ben Braiek et al., 2022).

A partir de l'exploitation des caractères de fertilité et de mortinatalité enregistrés en routine dans les schémas de sélection, il a été possible d'aller vers les hypothèses de létalité pour certains des haplotypes, en ciblant des stades plus spécifiques au moment du développement embryonnaire précoce (LDHH1, LDHH2 en Lacaune), tout au long de la gestation (LDHH9 en Lacaune, MTRDHH2 en MTR) ou autour de la naissance (LDHH3, LDHH6 en Lacaune et MTRDHH1 en MTR). Ceci démontre tout l'intérêt des informations phénotypiques. La remontée systématique des causes de réforme dans les bases de données nationales pour les petits ruminants serait un atout indéniable supplémentaire pour l'extension des approches de déficit en homozygotes dans la phase d'élevage des jeunes (de la naissance à la puberté). Ceci permettrait certainement d'identifier de nouvelles anomalies. En effet, pour MTRDHH3, je suspecte que cet haplotype soit lié à un défaut immunitaire pouvant conduire à la mort juvénile des agneaux (mutation dans *FCGR1A*). Cependant, aucune information de survie des agneaux n'est disponible dans les bases de données après la naissance (à quelques jours près). Dans le cadre du projet Effitness en bovins (A. Capitan, INRAE), l'analyse des déficits en homozygotes est conduite de manière conjointe avec l'étude des trajectoires de vie des animaux génotypés. Avec la mise en place de l'observatoire des anomalies génétiques en petits ruminants (projet PRESAGE), la remontée de cas (agneaux morts, présentant des défauts morphologiques), pourra aussi permettre d'identifier de nouvelles anomalies génétiques par une approche de génétique classique (analyse d'association, cartographie par homozygotie).

17. Les mutations dans les haplotypes en déficit

17.1 Découverte de nombreuses mutations récessives candidates

En l'absence de matériel biologique des animaux homozygotes mutés, il était nécessaire de rechercher les mutations chez les animaux porteurs hétérozygotes des haplotypes. Pour cela, j'ai fait des hypothèses fortes pour identifier ces mutations. J'ai extrait tous les SNP et petites insertions/délétions (InDels) situés dans les régions des différents DHH étendues de 1 Mb de chaque côté. Je faisais l'hypothèse que la mutation candidate était en parfait déséquilibre de liaison avec l'haplotype et qu'elle était spécifique de race. Cette hypothèse se traduisait par les filtres suivants : (i) la mutation est un variant par rapport au génome de référence de race Rambouillet (variant non porté par l'animal Rambouillet) ; (ii) présente à l'état hétérozygote chez tous les moutons de la race qui sont porteurs hétérozygotes du DHH (0/1) ; (iii) mais absente, c'est-à-dire à l'état homozygote référence (0/0), chez tous les moutons de la race qui sont non porteurs du DHH et chez les autres races plus lointaines d'un point de vue phylogénétique (0/0). J'ai également considéré en priorité les mutations perte de fonction dans la partie codante (*Loss-of-function*, LoF) connues pour avoir un impact biologique fort sur les protéines (non-sens, site accepteur/donneur, perte/gain codon start/stop, faux-sens). En effet, ce sont les principaux variants responsables d'anomalies génétiques dans de nombreuses espèces (MacArthur et al., 2012 ; Charlier et al., 2016), et comme nous avons pu le voir dans l'inventaire des anomalies génétiques identifiées en petits ruminants repris dans la [Table 2.3](#).

Ainsi dans mon analyse, en retenant de tels variants, j'ai pu identifier 9 mutations causales candidates : 4 mutations faux-sens (*PREB*/*GPN1*/LDHH3, *EDC3*/LDHH9 et *FCGR1A*/MTRDHH3), 2 mutations non-sens (*CCDC65*/LDHH6 et *MMUT*/MTRDHH2), 2 mutations entraînant un décalage du cadre de lecture (*ORC5*/LDHH1 et *SLC33A1*/MTRDHH2, liées à des duplications d'une seule paire de bases) et une causant la perte d'un codon (*IDI1*/LDHH2) ([Figure 17.1](#)). Les fonctions biologiques des gènes portant les mutations candidates sont impliquées dans des fonctions essentielles telles que la réplication (*ORC5*) et la réparation de l'ADN (*GPN1*), le métabolisme énergétique (*IDI1*, *SLC33A1* et *MMUT*), la réponse immunitaire (*FCGR1A*), la motilité ciliaire (*CCDC65*), la régulation de la transcription (*PREB*, *GPN1*), ou encore la dégradation des ARNm (*EDC3*). Il est donc aisé de penser que la perte de fonction de ces gènes entraîne à l'état homozygote des phénotypes délétères voir létaux. L'exploitation des bases de données humaines et souris avec une idée de biologie comparée m'a effectivement permis de confirmer des gènes candidats associés à de la létalité. En effet, les gènes *CCDC65* (LDHH6), *EDC3* (LDHH9), *MMUT* (MTRDHH1) et *SLC33A1* (MTRDHH2) ont pu être mis en évidence par cette approche.

Cependant, l'application stricte des filtres (phénotypes souris létaux et impliqués dans des

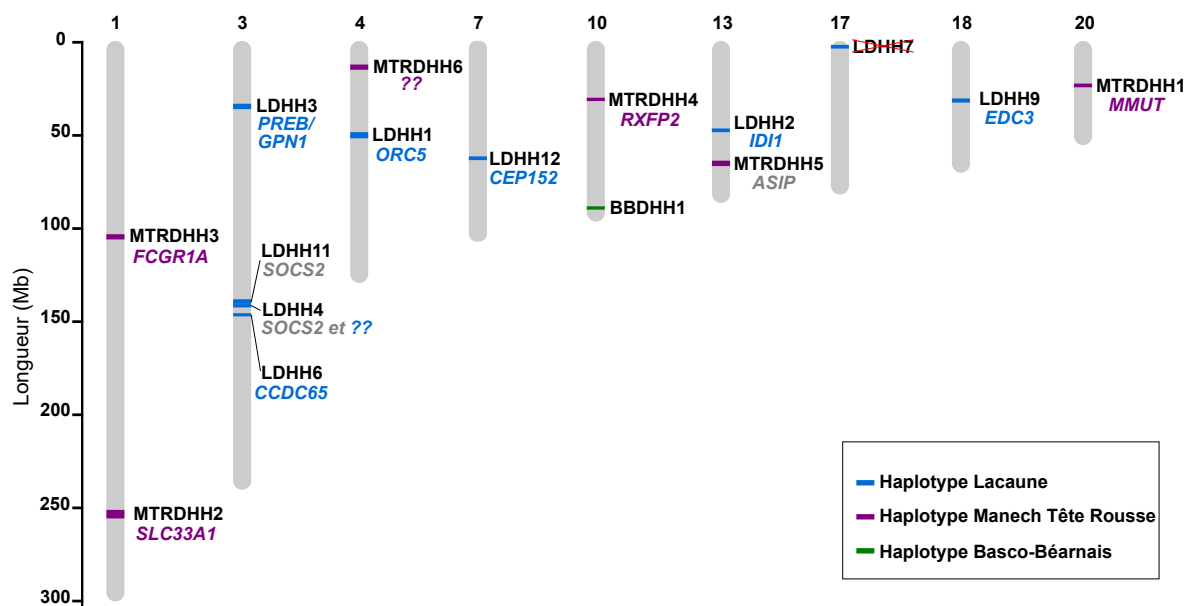


FIGURE 17.1 – Localisation des haplotypes et des gènes contenant les mutations candidates identifiées montrant un déficit significatif d'animaux homozygotes dans les trois populations laitières ovines Lacaune, Manech Tête Rousse et Basco-Béarnaise.

syndromes connus chez l'homme) ne permet pas forcément de retenir le bon gène candidat. En effet, alors que j'ai trouvé des mutations perte de fonction dans les gènes *ORC5*, *GPN1*, *IDI1*, *FCGR1A* ou encore *PREB*, ceux-ci n'apparaissaient pas dans la liste des candidats car soit aucun phénotype délétère n'est décrit chez la souris, soit aucun chez l'homme. C'est d'ailleurs une opportunité pour proposer le mouton comme modèle d'intérêt.

On peut également constater que les mutations causales identifiées sont localisées dans les intervalles stricts des haplotypes et non dans l'intervalle étendue de 1 Mb de chaque côté. L'exemple du CVM (*Complex Vertebral Malformation*) en bovins montre pourtant que la mutation causale peut être localisée assez loin (6 Mb) par rapport à l'haplotype en déficit d'animaux homozygotes (Fritz et al., 2016). Il reste donc pertinent pour les analyses futures de garder un intervalle de confiance assez large pour identifier de nouvelles mutations.

Le succès de l'identification de ces nombreuses mutations a reposé sur la disponibilité d'un grand nombre d'animaux séquencés tout génome. Au laboratoire, on disposait déjà d'un grand nombre d'animaux Lacaune et Manech Tête Rousse séquencés obtenus dans le cadre de projets de recherche antérieurs, et avec des génotypes connus sur les puces SNP permettant de connaître le statut aux haplotypes. L'accessibilité de ces données a permis dès le début de ma thèse, d'identifier des mutations candidates. Seulement 5 animaux supplémentaires ont été séquencés pendant ma thèse (régions LDHH4 et MTRDHH4). En effet, la découverte de nouvelles mutations peut être très rapide. D'ailleurs, lors de l'analyse de Septembre 2022, j'ai identifié deux nouveaux haplotypes (LDHH12 et MTRDHH4) et grâce aux données WGS disponibles, j'ai de nouveau identifié une unique mutation non-sens OAR7 :g.63,254,935C>T (c.3598C>T, p.Gln1200*) dans le gène *CEP152* (*Centrosomal Protein 152*) pour l'haplotype LDHH12. CEP152 est une protéine centrale du centrosome organisant les microtubules des cellules animales qui influencent la forme, la polarité et la motilité, et avec une fonction cruciale dans la division cellulaire (Andersen et

al., 2003). Chez l'homme, l'altération de *CEP152* est associée au syndrome de microcéphalie primaire (OMIM 614852) et au syndrome de Seckel (OMIM 613823). Chez la souris, les embryons homozygotes pour une mutation perte de fonction de *Cep152* présentent un nombre réduit de centrosomes et de cils, une apoptose accrue et une létalité à mi-gestation (Bazzi & Anderson, 2014). Elle se rajoute à la liste des mutations intéressantes à étudier.

17.2 Limites de la méthode utilisée pour rechercher les mutations

Bien que la découverte de mutations ait été fructueuse, il reste encore une région (LDHH4) en déficit total d'animaux homozygotes pour laquelle je n'ai pas mis en évidence de mutation candidate dans le codant. Je fais l'hypothèse que LDHH4 contient une mutation récessive létale (aucun animal homozygote observé dans la population alors que 11 sont attendus) qui est apparue dans un haplotype contenant l'allèle g.139,302,270T avec la mutation *SOCS2* et qu'elle subit un phénomène d'"auto-stop". Ce cas de figure nous permet de discuter des filtres appliqués.

Le premier point de discussion repose sur l'hypothèse que l'haplotype identifié est en parfait déséquilibre de liaison avec la mutation causale. Or ceci n'est pas toujours vrai en cas de pénétrance incomplète. Par exemple Fritz et al. (2018), en recherchant des déficits en homozygotie, ont mis en évidence une mutation dans le gène *SDE2* (intervenant dans la maintenance des télomères). Parmi l'ensemble des variants présents dans l'intervalle de 1,1Mb en déficit d'homozygotie, seul un polymorphisme présente une corrélation supérieure à 0,5 entre le statut des animaux à l'haplotype et le génotype des variants.

Le deuxième point de discussion repose sur l'application de filtres stricts sur l'effet des variants. En effet, je me suis focalisé sur les variants perte de fonction prédits avec un impact fort ou modéré. Cependant, des mutations pouvant impacter les séquences promotrices ou régulatrices et donc l'expression des gènes pourraient être considérées dans un second temps. En effet, 36 variants ont une corrélation de 1 entre leurs allèles et le statut de l'haplotype LDHH4.

Le troisième point de discussion repose sur l'analyse des variants de type SNP et petites insertions/délétions. En effet, l'analyse de variants de structure (non mis en évidence par les logiciels de détection de variants usuels) pourrait permettre d'identifier de nouvelles mutations candidates. Par exemple en bovins, une délétion de 660kb (OAR12 (UMD3.1) : g.2,010,0648_20,763,119del), supprimant quatre gènes, est létale à l'état homozygote (Kadri et al., 2014).

Le quatrième point de discussion repose sur l'annotation des génomes de référence. Par exemple, la région LDHH4 contient des clusters de gènes homéotiques (*HOXC*) et de kératine (*KRT*) pour lesquels des erreurs d'annotations peuvent être présentes et qui biaisent souvent l'alignement correct des séquences courtes des WGS, altérant la détection des variants. Afin de palier à cela, dans le cadre du projet SeqOccin, nous avons fait réaliser des génomes de référence en séquençage longues lectures dans chacune des races étudiées. L'alignement sur ces nouveaux génomes des séquences WGS déjà acquises pourrait permettre d'améliorer la recherche de variants pour ces régions complexes.

18. Causalité des mutations

18.1 Mutations létales embryonnaires

L'étude des accouplements à risque pour les mutations dans *ORC5* et *IDI1* n'a pas permis de confirmer une baisse de fertilité dans les croisements à risque réalisés, n'ayant pas eu accès à un nombre suffisant de brebis Lacaune porteuses de ces mutations. Il faudrait pouvoir réaliser des accouplements à plus grande échelle en élevages commerciaux comme ceux que j'ai pu planifier en Manech Tête Rousse. Néanmoins, ces accouplements ont permis de développer un test moléculaire de gestation (GDMT) basé sur l'expression de gènes stimulés par l'interféron τ . Ce test a été utilisé en élevage Manech Tête Rousse pour mettre en évidence des pertes fœtales pour la mutation dans *SLC33A1*.

Par accouplements orientés entre porteurs de la mutation dans *SCL33A1*, j'ai pu confirmer la baisse de la réussite à l'IA observée pour MTRDHH2. Cette baisse de 12% pourrait être associée à des pertes embryonnaires. Afin d'explorer cette hypothèse, nous avons testé le diagnostic moléculaire de gestation 15 jours après IA (le moment de l'implantation de l'embryon) confronté à l'échographie classique réalisée entre 45 jours et 60 jours post-IA. Ce test moléculaire GDMT, basé sur l'expression dans les cellules sanguines du gène *MX1* connu pour être fortement stimulé par l'interféron τ (Mauffré et al., 2016), a prédit 16 brebis gestantes alors que seulement 12 étaient réellement gestantes. Ceci laisse penser que pour 4 brebis hétérozygotes, des pertes fœtales sont survenues entre 15 jours et 60 jours de gestation, que j'attribue à la létalité des fœtus homozygotes *dupG/dupG*. Le rôle de *SCL33A1* dans la régulation du métabolisme de l'acétyl-CoA pourrait expliquer en grande partie le défaut d'embryogenèse. En effet, l'acétyl-CoA pourrait jouer un rôle majeur dans la régulation de la croissance cellulaire, de la prolifération et de l'apoptose, suggérant que les défaillances métaboliques de l'acétyl-CoA sont responsables de l'échec embryonnaire (Tsuchiya et al., 2014). Cependant, la relation de cause à effet entre la mutation et l'arrêt de développement embryonnaire reste à établir formellement par une expérimentation permettant une production d'embryons génotypés comme proposé dans la [Figure 16.1](#). Celle-ci pourrait également être réalisable pour les mutations dans *ORC5* et *IDI1*. J'ai réalisé des expérimentations préliminaires de productions d'embryons de 7 jours (stade blastocyste) pour les mutations *SLC33A1* et *IDI1* qui ne m'ont pas encore permises de conclure sur un arrêt précoce du développement embryonnaire.

D'autres méthodes de détection précoce de la gestation sont disponibles telles que l'échographie transrectale précoce (Rickard et al., 2017) ou les dosages de biomarqueurs circulants comme la progestérone, la protéine B (PSP-B) ou les glycoprotéines associées à la grossesse (PAG) (Karen et al., 2003). Cependant, ces approches ne sont vraiment efficaces qu'à partir du 28^{ème} jour de

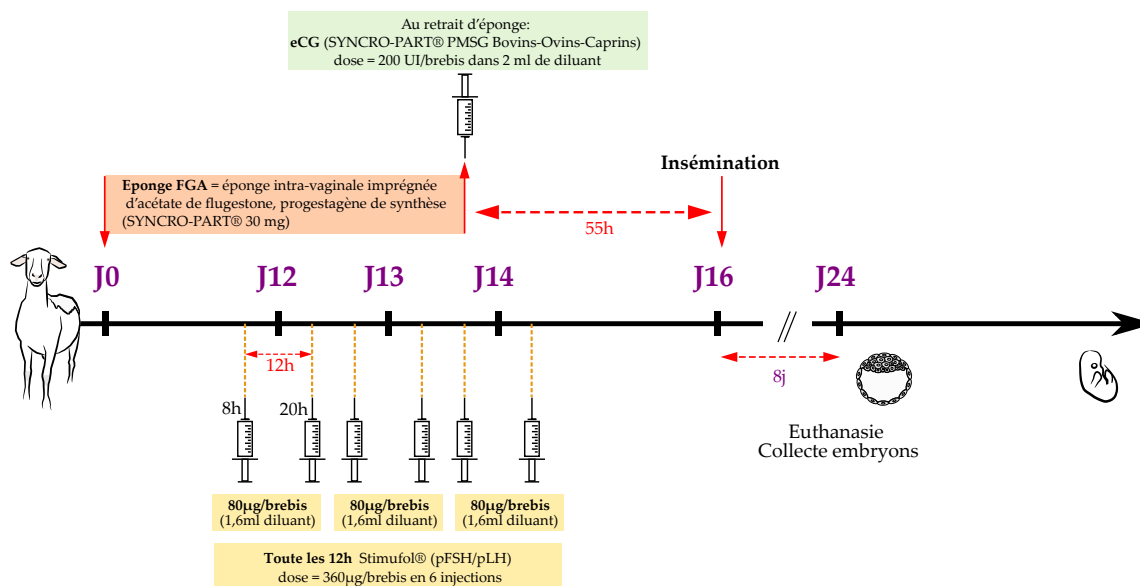


FIGURE 18.1 – Chronogramme de la procédure de production d'embryons chez les ovins pour étudier les mutations dans *IDI1* et *SLC33A1*.

gestation après le moment prévu de retour en œstrus (Mauffré et al., 2016).

Même si le test de diagnostic moléculaire de gestation nous a permis de mettre en évidence des pertes fœtales précoces et offre de bonnes opportunités pour prédire l'état de gestation à J15, son application en élevage est possiblement limitée. En effet, seuls 66% des échantillons d'ARN de cette étude étaient exploitables. Ce problème était principalement dû à la dégradation de l'ARN avec un contrôle limité du stockage de l'échantillon de sang lors de la collecte dans les fermes. Les échantillons devraient être congelés directement dans un tampon adéquat inactivant les RNases. En se concentrant sur la méthode de diagnostic, l'aire (AUC) d'une courbe ROC est un bon estimateur pour évaluer la puissance d'un test de diagnostic (Janssens & Martens, 2020). Dans le cas présent, on obtient une AUC de 0,687 qui est considérée comme acceptable (Swets, 1988). La valeur seuil de 63% pour l'expression relative de *MX1* a été déterminée par la méthode ROC01 au lieu de l'indice de Youden généralement utilisé dans l'analyse ROC (Perkins & Schisterman, 2006). En effet, l'indice de Youden donne un poids égal à la sensibilité et à la spécificité, alors que le point de coupure optimal de la méthode ROC01 privilégie la sensibilité afin d'identifier les brebis avec un GDMT positif et gestantes aux jours 45-60. Initialement, le GDMT était basé sur l'expression de deux gènes, *MX1* et *STAT1*, mais seule l'expression relative de l'ARNm de *MX1* était différente entre les brebis gestantes et vides du groupe témoin. Comme indiqué précédemment, l'interféron τ est principalement produit par le conceptus dans une période comprise entre 14 et 17 jours après l'IA et des échantillons de sang ont été prélevés à J15 comme proposé dans Mauffré et al. (2016) et proche du J14 évalué en Lacaune. Cependant, dans la littérature, *STAT1* montre une expression d'ARNm plus élevée aux jours 16-17 (Spencer et al., 2008) et la surexpression de l'ARNm de *MX1* a pu être maintenue jusqu'au jour 21 dans les cellules mononucléaires du sang périphérique (Yankey et al., 2001) et comme j'ai pu le voir chez les brebis Lacaune (paragraphe 13.1.3). Ainsi, J17 pourrait être préférable pour prélever des échantillons de sang pour une analyse possiblement plus discriminante.

18.2 Mutations létales périnatales

18.2.1 Mutation dans *MMUT*

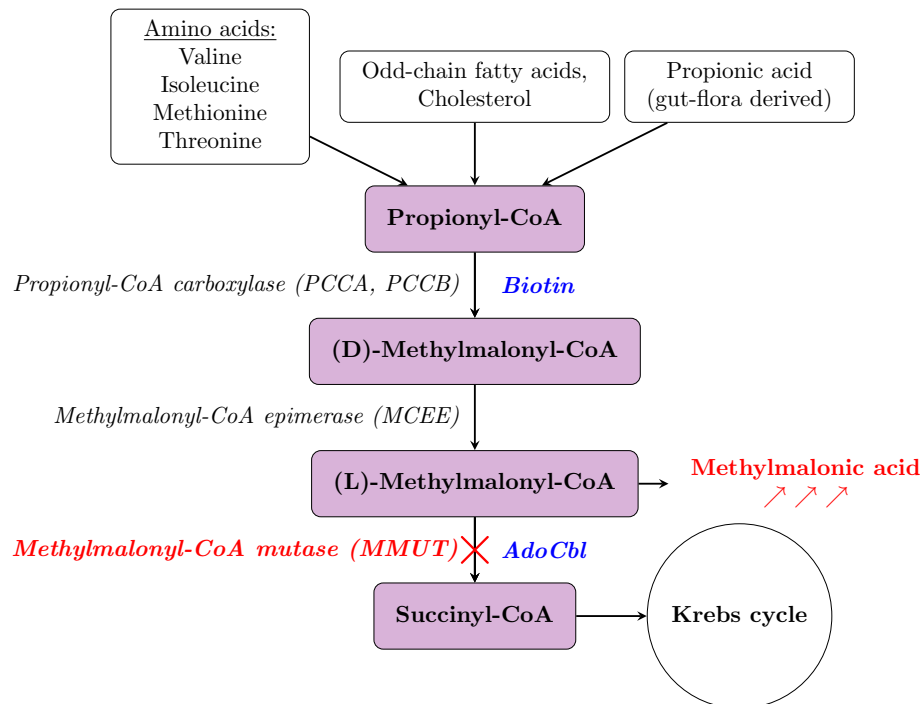


FIGURE 18.2 – Voie métabolique impliquant *MMUT*

Par les croisements à risque entre porteurs hétérozygotes de la mutation dans *MMUT*, j'ai pu prouver le caractère récessif léthal de cette mutation, la mortalité des agneaux homozygotes intervenant dans les premières 24h. Chez l'homme, les variants délétères du gène *MMUT* provoquent une "acidémie méthylmalonique" (OMIM 609058, MMA), un trouble du métabolisme autosomique récessif. *MMUT* joue un rôle dans une voie métabolique (Figure 18.2) partant de la dégradation des acides aminés (valine, isoleucine, méthionine et thréonine), des acides gras à chaîne impaire, du cholestérol et de l'acide propionique en succinyl-CoA par trois enzymes importantes : la propionyl-CoA carboxylase (PCC), la Méthylmalonyl-CoA épimérase (MCE) et la Méthylmalonyl-CoA mutase (*MMUT*) (Baumgartner et al., 2014 ; Deodato et al., 2006). La protéine *MMUT* est une enzyme mitochondriale qui catalyse le L-méthylmalonyl-CoA en succinyl-CoA, un intermédiaire du cycle de Krebs. L'isomérisation du méthylmalonyl-CoA nécessite l'adénylcobalamine (AdoCbl), la forme cofacteur de la vitamine B12 (également appelée cobalamine) (González-Montaña et al., 2020). Lorsque l'enzyme *MMUT* n'est pas fonctionnelle, l'acide méthylmalonique s'accumule dans les fluides corporels (principalement dans le sang et l'urine) (Deodato et al., 2006 ; Lucienne et al., 2020). Chez la souris homozygote knock-out pour *Mmut* l'augmentation d'acide malonique est très rapide dans les 24 premières heures après la naissance (Peters et al., 2003). Malheureusement, je n'ai pas pu observer cette accumulation anormale d'acide méthylmalonique dans les échantillons de sang et d'urine dosés chez les agneaux homozygotes mutés. Comme nous avons prélevé sur des animaux morts pour la plupart, cela

pourrait altérer les résultats. J'ai pourtant mis en évidence par western blot, l'absence de la protéine complète pour les animaux homozygotes dans le foie et le rein. En effet, la séquence épitopique de l'anticorps polyclonal MMUT utilisé était située après la position du codon stop (p.409Gln*) de 451 à 750 acides aminés (homologie de séquence spécifique de 94,3% entre lapin et mouton). La protéine mutante ne contient pas non plus le domaine de liaison B12 (615-733 acides aminés), et elle est donc incapable de convertir le L-méthylmalonyl-CoA en succinyl-CoA.

Bien que les agneaux homozygotes soient phénotypiquement normaux à la naissance, j'émetts l'hypothèse qu'une fois qu'ils commencent à se nourrir du lait maternel, ils ne peuvent pas métaboliser les acides aminés et les acides gras, et l'acide méthylmalonique se retrouve en quantité toxique dans l'organisme. Il faudrait pouvoir collaborer avec des cliniciens spécialistes de ce type de maladie métabolique chez l'homme pour établir un diagnostic clinique précis des conséquences de la mutation dans *MMUT* découverte chez les ovins.

18.2.2 Mutation dans *CCDC65*

Pour la mutation dans *CCDC65*, les accouplements à risque ont aussi permis de prouver le rôle létal de cette mutation pour les agneaux homozygotes. Chez l'homme, de nombreux variants sont identifiés dans le gène *CCDC65*, responsables de Dyskinésie Ciliaire Primitive (PCD), une maladie génétique à transmission autosomique récessive (CILD27, OMIM 615504). La protéine est impliquée dans l'assemblage du complexe régulateur axonémique nexine-dynéine (Bower et al., 2018).

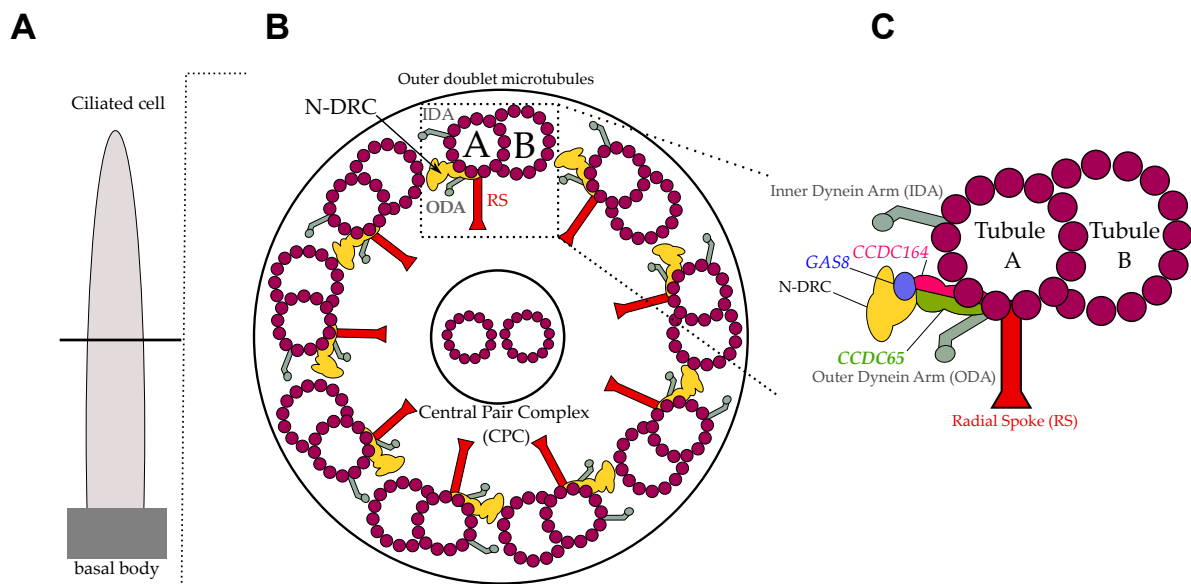


FIGURE 18.3 – **Structure ciliaire et organisation des axonèmes** (A) Représentation schématique d'une cellule ciliée; (B) Représentation schématique d'une coupe transversale d'axonème; (C) Représentation schématique d'un doublet externe de microtubules avec la localisation de la protéine *CCDC65* du complexe régulateur Nexin-Dynein (N-DRC), qui est une sous-unité du N-DRC. *Adapté de* (Morohoshi et al., 2020; Gui et al., 2019; Horani et al., 2013; Poprzeczko et al., 2019; Kurkowiak et al., 2015).

L'axonème est la partie motrice axiale des cils composé de neuf doublets de microtubules situés en périphérie et d'un complexe de paires centrales (Heuser et al., 2009). Chaque doublet

externe est composé de deux tubules (A et B) et est associé à plusieurs complexes protéiques : bras de dynéine externe et interne (ODAs et IDAs), Nexin-Dynein Regulatory Complex (N-DRC) et Radial Spoke (RS) (Heuser et al., 2009 ; Gui et al., 2019 ; J. Lin et al., 2011 ; Bower et al., 2013 ; Morohoshi et al., 2020). En particulier, le N-DRC est important pour le glissement entre les doublets de microtubules externes adjacents afin de permettre la motilité des cils. Le complexe est composé d'au moins 11 sous-unités protéiques présentes chez *C. reinhardtii* et conservées chez les mammifères (Heuser et al., 2009 ; Gui et al., 2019 ; J. Lin et al., 2011 ; Bower et al., 2018). Parmi elles, des variants dans l'un des trois gènes : *CCDC164* (OMIM 615294 (Wirschell et al., 2013)), *CCDC65* (OMIM 615504 (Austin-Tse et al., 2013 ; Horani et al., 2013)) ou *GAS8* (OMIM 616726 (Olbrich et al., 2015)) ont été identifiés comme étant à l'origine de dyskinésie ciliaire. Bower et al. (2018) ont démontré que l'absence de sous-unité CCDC65 déstabilisait l'association avec CCDC164 et GAS8. Au niveau pulmonaire, l'altération de la N-DRC a des conséquences sur les battements de cils et la clairance du mucus des voies respiratoires. Chez l'homme, les patients atteints de PCD en lien avec des mutations dans *CCDC65* présentent des maladies infectieuses récurrentes de la sphère ORL (oreille nez gorge) comme la bronchite, la sinusite et/ou l'otite (Austin-Tse et al., 2013 ; Horani et al., 2013). Ces profils cliniques sont en accord avec nos observations faites chez des agneaux homozygotes pour le variant c.331G>T dans *CCDC65*. Ils présentaient en effet une détresse respiratoire dès la naissance, associée à un retard de croissance (McRae et al., 2016). Les autopsies ont également révélé une large hépatisation des lobes pulmonaires ressemblant à une pneumonie infectieuse. J'émetts l'hypothèse qu'en raison de la mutation, la protéine CCDC65 tronquée entraîne la déstabilisation du N-DRC, un battement anormal des cils et altère la clairance mucociliaire entraînant une obstruction des voies respiratoires par du mucus chargé d'agents pathogènes (bactéries, mycoplasmes, virus). La charge en agents pathogènes pourrait dépendre des conditions sanitaires d'élevage, de l'immunité individuelle et du traitement vétérinaire appliqué, ce qui peut expliquer que la létalité puisse survenir entre les stades périnatal et juvénile. Concernant les béliers hétérozygotes, l'analyse des données de production de semences ne met pas en évidence un effet de la mutation sur la quantité et la qualité des semences (volume, concentration et motilité).

18.3 Mutations en lien avec la morphologie

Avec la méthode de recherche de déficit d'animaux homozygotes, j'ai pu en mettre en évidence des haplotypes en lien avec des caractères que j'ai regroupés sous le terme "morphologie" impliqués dans la stature et sensibilité aux mammites (*SOCS2*/LDHH4/LDHH11), la coloration de la toison (*ASIP*/MTRDHH5) et le cornage/cryptorchidie (*RXFP2*/MTRDHH4).

J'ai pu mettre en évidence que l'allèle de sensibilité aux mammites g.139,302,270T dans *SOCS2* était en lien avec les haplotypes LDHH4 et LDHH11. A ce jour, cette mutation qui influence aussi la stature et la production laitière, n'est pas connue pour être létale à l'état homozygote. Je fais l'hypothèse que LDHH4 contient une mutation récessive létale (aucun animal homozygote observé dans la population alors que 11 sont attendus) qui est apparue dans un haplotype contenant l'allèle muté T de *SOCS2* et subit un phénomène d'"auto-stop" avec la mutation *SOCS2* en sélection. C'est certainement ce signal que M. Petit avait identifié lors de la première analyse sur

un nombre plus réduit de géotypages en Lacaune (Petit et al., 2015), voir [Chapitre 5 - Situation du sujet](#)). La première fois que j'ai cherché à identifier la mutation candidate de LDHH4, j'ai pu mettre en évidence une mutation faux-sens (OAR3 :g.142,735,640C>T ; p.Gly299Gln) dans le gène *MYG1/C3H12orf10* (*MYG1 Exonuclease/Chromosome 12 Open Reading Frame 10*), présent à l'état hétérozygote chez un seul animal LDHH4/+, mais cette mutation n'a pas été confirmée lors du séquençage d'un second animal (LDHH4/+ mais non porteur de la mutation candidate). Une nouvelle mutation reste donc à identifier pour LDHH4. Lorsqu'elle sera identifiée, il sera intéressant de rechercher des animaux recombinants, hétérozygotes à la mutation "létale" et non porteurs de l'allèle muté T, pour envisager son étude et ses effets sans biais avec *SOCS2*.

Le lien entre MTRDHH5 et la délétion de 5pb connue dans *ASIP* explique en partie le déficit observé. En effet, dans l'ensemble de la population géotypée, 14 animaux homozygotes à MTRDHH5 sont observés dont 12 sont des femelles et 10 sont disqualifiées pour être de couleur noire. Cependant, on retrouve des béliers Del/Del en CIA mais leur toison n'est pas noire. Par contre, ces derniers présentent une tâche rousse autour du cou ([Figure 18.4](#)). Cela pourrait suggérer qu'une autre mutation localisée dans un autre gène de coloration donne ce phénotype, comme *MC1R* (Våge et al., 1999) ou *TYRP1* (Beraldi et al., 2006). Le travail de cette région reste donc à poursuivre.



FIGURE 18.4 – Bélier de géotype Del/Del (délétion de 5pb dans *ASIP*) présent au CIA avec une tâche rousse dans le cou.

Même si MTRDHH4 est en lien partiel avec la mutation sans cornes (allèle "As"), l'étude de cet haplotype a permis de mettre en évidence la mutation c.2050delG dans le gène *RXFP2* responsable de la cryptorchidie et donc du déficit identifié par contre-sélection des jeunes agneaux présentant cette anomalie. Cette découverte a été facilitée par la remontée de cas au travers du projet PRESAGE d'observatoire des anomalies. Finalement, j'ai identifié une nouvelle mutation dans *RXFP2*, responsable de la cryptorchidie, mais qui accentue aussi le phénotype sans cornes. Comme les mutations cornes/sans cornes et cryptorchidie sont très proches (1 kb d'écart), la chance de trouver un animal recombinant entre les deux mutations est très faible. De ce fait, aucun animal est +/+ au cornage et delG/delG pour la cryptorchidie pouvant nous assurer du phénotype propre de la mutation c.2050delG. Afin de confirmer l'effet propre de la mutation cryptorchidie,

il pourrait être envisageable de créer des animaux delG/delG par édition de génome dans des haplotypes non porteurs de la mutation "As" cornage. De plus, j'ai pu voir que cette mutation ne semble pas impacter la carrière reproductive des brebis homozygotes. Cependant, compte tenu du rôle de la signalisation INSL3/RXFP2 dans le fonctionnement ovarien (Esteban-Lopez & Agoulnik, 2020), je formule l'hypothèse que la mutation de *RXFP2* pourrait peut-être jouer sur la réserve ovarienne de follicules en croissance. Pour tester cette hypothèse, il serait pertinent d'étudier les profils endocriniens et les cycles de reproduction des brebis homozygotes, en particulier au regard de la sécrétion de l'hormone anti-Mullérienne AMH, marqueur endocrinien de cette réserve chez la femme et les ruminants (Monniaux et al., 2014).

18.4 Les mutations non étudiées

L'étude d'un nombre insuffisant d'accouplements à risque pour les mutations dans *ORC5* et *IDI1* n'a pas permis de confirmer une baisse de fertilité. Il faudrait également pouvoir réaliser des accouplements à plus grande échelle en élevages commerciaux Lacaune comme ceux que j'ai pu planifier en Manech Tête Rousse (diagnostic moléculaire de gestation, échographie J45, suivi des mises-bas et génotypages des agneaux).

De plus, toutes les mutations identifiées pendant mon travail de thèse n'ont pas pu être travaillées (*EDC3/LDHH9*; *PREB/GPN1/LDHH3*; *FCGR1A/MTDHH3*). Pour étudier ces mutations, il serait également nécessaire d'effectuer des croisements à risque et d'effectuer un suivi des mises-bas et des agneaux. Pour la mutation *EDC3* localisée dans *LDHH9*, elle est supposée conduire à des avortements tout au long de la gestation, comme suggéré par une baisse de réussite à l'IA et une hausse du taux de mortinatalité lors de l'étude populationnelle en race Lacaune. Son étude pourrait donc être conduite comme pour la mutation dans *SLC33A1* en MTR. Pour valider la causalité des mutations dans *PREB* et *GPN1*, il serait d'abord judicieux de génotyper à très grande échelle les deux mutations pour déterminer laquelle est la plus associée à l'haplotype *LDHH3*. La réalisation d'accouplements à risque permettrait là encore de caractériser la mortalité agneaux observée dans les analyses populationnelles. D'un point de vue moléculaire, ces deux gènes sont des facteurs modulant la transcription et pourraient être étudiés en mettant en place des tests fonctionnels *in vitro* avec des rapporteurs d'activité transcriptionnelle, du type luciférase. Au cours de ma thèse, j'ai tenté de développer un test fonctionnel pour *PREB* (facteur de transcription capable de se fixer sur le promoteur de la prolactine) par la mise au point d'un essai luciférase en clonant (*i*) le promoteur du gène *PRL* (*Prolactin*) ovin dans un vecteur contenant le gène de la luciférase et (*ii*) en clonant le cDNA ovin de *PREB* (formes sauvage et mutée) dans un vecteur d'expression. Cependant, ces expérimentations n'ont pas permis de valider fonctionnellement l'effet de la mutation. La stratégie d'accouplements à risque reste également pertinente pour étudier la causalité de la mutation dans *FCGR1A/MTRDHH3*, possiblement associée à des problèmes immunitaires. Cela permettrait de générer des agneaux homozygotes que l'on pourrait suivre sur une période plus longue et sur lesquels nous pourrions réaliser un challenge immunitaire avec des injections de type lipopolysaccharide (LPS) (Henry et al., 2009; Dillingh et al., 2014) pour révéler l'effet de la mutation.

Enfin, pour déterminer la stratégie à adopter pour étudier la nouvelle mutation identifiée en

Septembre 2022 dans le gène *CEP152* du nouvel haplotype LDHH12, il est nécessaire d'analyser les données populationnelles pour les caractères de fertilité et mortinatalité en Lacaune. Ces résultats permettront de confirmer ou non la létalité, et permettront de poser des hypothèses comme je l'ai fait avec toutes les autres mutations.

18.5 Pertinence des mutations identifiées en tant que modèles animaux de pathologies

Depuis des siècles, l'homme a tiré parti de ses similitudes avec les animaux (notamment les mammifères) pour mieux comprendre son anatomie, sa physiologie et ses pathologies (Barré-Sinoussi & Montagutelli, 2015). Les souris de laboratoire ont été développées en tant que souches consanguines qui ont une composition génétique très homogène pour augmenter la reproductibilité des résultats et la puissance statistique des expériences. En biologie moderne, la recherche se concentre principalement sur un nombre limité d'espèces sélectionnées pour leur taux de reproduction élevé, leur faible intervalle entre générations et leur facilité d'élevage en laboratoire. L'étude de ces espèces « modèles » a considérablement amélioré notre compréhension des bases moléculaires de nombreux phénotypes, et en particulier des troubles génétiques humains, avec la caractérisation de milliers de gènes/mutations au cours des 40 dernières années.

Grâce aux connaissances apportées par les modèles murins et les pathologies avec déterminisme génétique connus chez l'homme, et la synthèse de ces connaissances dans des bases de données (MGI, OMIM), il est possible d'établir des hypothèses sur les phénotypes sous-jacents causés par des mutations à effet fort. En effet, dans mon travail de thèse, j'ai pu mettre en évidence des mutations dans des gènes déjà connus pour être associés à de la létalité chez des modèles de souris KO et également associées à des troubles génétiques connus chez l'homme (*IDI1*, *EDC3*, *SLC33A1*, *CCDC65* et *MMUT*). Notamment dans le cas de mutations supposées être létales à l'état embryonnaire, le modèle murin s'avère pertinent puisqu'il permet de caractériser le moment de la létalité plus facilement (manipulation génétique aisée, courte durée de gestation, cycle reproductif court). En effet, la production d'embryons en ovins est une étape délicate (superovulation, insémination et collecte d'embryons) et n'a pas permis dans cette thèse d'aboutir à la validation de la causalité des mutations *SLC33A1* et *IDI1* lors des expérimentations préliminaires (voir [paragraphe 18.1](#)).

Toutefois, le modèle murin peut présenter des limites à l'étude de pathologies humaines puisque les descriptions cliniques varient souvent d'une souche à l'autre (Barré-Sinoussi & Montagutelli, 2015). L'absence de variabilité génétique et les conditions d'élevage standardisées limitent les chances de détecter des pathologies à pénétrance incomplète et à expressivité variable alors qu'elles sont nombreuses chez l'homme, favorisées par divers facteurs de prédispositions génétiques ou environnementaux (Bourneuf et al., 2017). A titre d'exemple, l'étude de la mutation naturelle dans le gène *CCDC65* présent en race Lacaune a permis une description clinique précise des individus atteints et de tester des traitements adaptés (administration de glucocorticoïdes et antibiotiques) pour lesquels j'ai pu mettre en évidence une pénétrance incomplète (létalité du stade périnatal à juvénile) expliquée par des environnements différents (élevage, centre d'IA) associés à des charges de pathogènes variables. De même, afin d'étudier la mutation dans

MMUT, le modèle ovin pourrait être un très bon modèle pour tester de nouveaux traitements (vaccin, complémentation alimentaire). En 2022, la société Moderna Inc. a lancé un essai de vaccin à ARNm pour l'acidémie méthylmalonique. Cet ARNm (encapsulé dans une nanoparticule lipidique) code une protéine *MMUT* fonctionnelle et restaure l'activité métabolique de l'enzyme (<https://www.technologynetworks.com/biopharma/news/moderna-launches-trial-of-mrna-vaccine-for-methylmalonic-acidemia-352951>). Le modèle ovin pourrait donc être pertinent pour étudier à plus grande échelle ce vaccin (identité de 90% entre ARNm de *MMUT* ovins et humains). A ce jour, le modèle ovin est utilisé pour étudier plus d'une cinquantaine de pathologies (Banstola & Reynolds, 2022). Cependant, l'utilisation d'un tel modèle impose des considérations éthiques plus importantes et un coût d'élevage plus onéreux que les modèles classiques de laboratoire.

Parmi les gènes que j'ai mis en évidence, plusieurs n'ont pas de modèle de souris KO (*GPN1*) ni de profil clinique pour des pathologies humaines rapportées (*ORC5*, *PREB*, *GPN1* et *FCGR1A*). Dans ce cas de figure, le modèle ovin pourrait se montrer pertinent pour étudier de nouvelles pathologies.

Les bovins, ovins et caprins, en particulier, ont une fécondité effective plus élevée (grâce à l'insémination animale), une meilleure diversité génétique et une plus grande similitude avec les humains en termes de physiologie, de stature et de longévité que les souris. De plus, grâce à l'avènement des technologies à haut débit, il est désormais possible d'avoir accès à l'ensemble des variations génétiques naturelles de ces populations d'élevage et cela offre des perspectives intéressantes pour améliorer notre compréhension de la fonction des gènes chez les mammifères et des conséquences pathologiques de leur mutation. L'étude de ces nouveaux modèles animaux est une opportunité de proposer des gènes candidats pour des pathologies non encore élucidées chez l'homme (en général, le nombre de patients atteints est relativement faible pour étudier les pathologies).

19. Gestion dans les schémas de sélection

19.1 Évolution de la fréquence des mutations

Les fréquences des haplotypes identifiés et des mutations causales sous-jacentes varient entre 3% et 16% dans les populations Lacaune et MTR. Une fois apparu, un variant est soumis à différentes forces évolutives qui vont influencer l'évolution de sa fréquence.

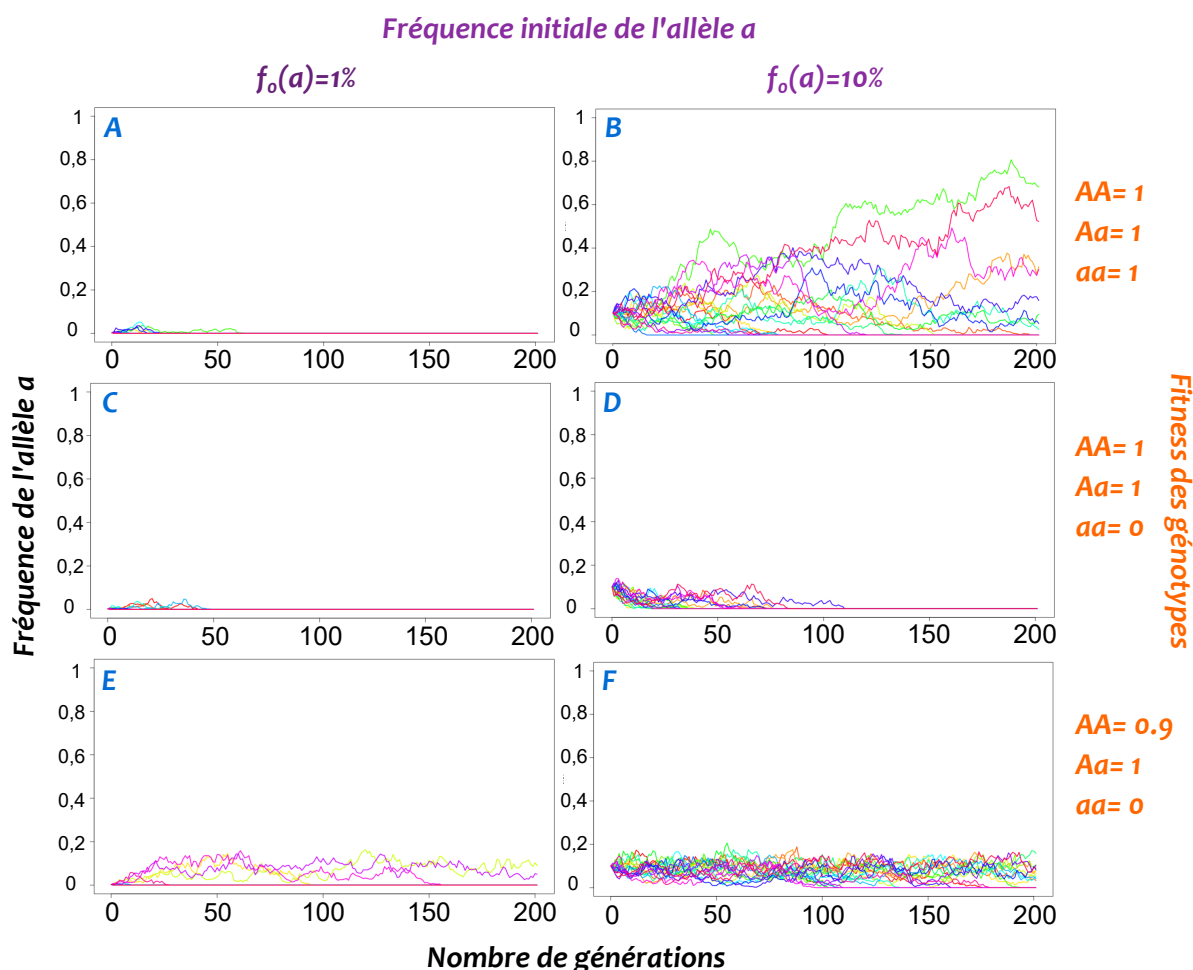


FIGURE 19.1 – Simulation de la fréquence d'un allèle délétère dans une population de **taille efficace $N_e=150$** au cours du temps sur 200 générations. (A, C, E) Fréquence initiale de l'allèle délétère à 1%. (B, D, F) Fréquence initiale de l'allèle délétère à 10%. On considère un locus biallélique (Aa) avec A l'allèle sauvage et a l'allèle délétère. (A, B) Dans une situation classique, les fitness des trois génotypes sont égales. (B, C) Situation où l'allèle a est délétère à l'état homozygote et les hétérozygotes ne présentent pas d'avantage sélectif. L'ensemble des simulations a été réalisé avec l'interface web <https://phytools.shinyapps.io/drift-selection/> obtenu avec le package R learnPopGen (Revell, 2019). Le nombre de simulations réalisées est de 200.

Les variants responsables d'anomalies génétiques ont des fréquences faibles à modérées (1-10%) et peuvent rapidement être éliminés en quelques générations par effet de dérive génétique (Georges et al., 2019). Afin d'étudier l'évolution de la fréquence d'une mutation délétère dans les populations de petits ruminants, j'ai réalisé des simulations sur 200 générations prenant en compte la taille efficace de la population (fixée à $N_e=150$ en petits ruminants), la fréquence initiale de la mutation (1% et 10%) et les fitness des génotypes (Figure 19.1). Par dérive génétique, un variant peut être éliminé ou diffusé largement dans la population (jusqu'à potentiellement être fixé). Ce phénomène est d'autant plus important lorsque la fréquence de la mutation est très faible et la taille efficace est petite (Figure 19.1 A et B). Cependant, lorsque l'allèle est délétère sans avoir un avantage sélectif, il peut se maintenir à une fréquence très faible dans la population pendant de nombreuses générations avant d'être éliminé (entre 50 et 100 générations en fonction de la fréquence initiale de l'allèle délétère) (Figure 19.1 C et D).

L'autre phénomène pouvant expliquer le maintien de variant délétère dans les populations est la sélection. En effet, certaines mutations donnent un avantage phénotypique aux individus à l'état hétérozygote. Avec une meilleure fitness (capacité reproductive), les reproducteurs vont plus facilement diffuser dans les schémas de sélection. En analysant les caractères en sélection, certains haplotypes offrent un potentiel avantage sélectif des béliers porteurs hétérozygotes, notamment pour la production laitière augmentée de leurs filles (LDHH3, LDHH9 et LDHH11 ; MTRDHH2). Cependant, ces différences sont très faibles (correspond à un an de progrès génétique) et ne permettent pas d'expliquer totalement la fréquence de ces haplotypes dans les populations. Cela explique qu'une mutation délétère peut se retrouver à une fréquence stable (entre 5-15%) dans une population (Figure 19.1 E et F).

Finalement, les fréquences observées semblent être expliquées par la dérive génétique. Cependant, même si l'allèle muté ne présente pas d'avantage sélectif, il peut être maintenu à une certaine fréquence par le phénomène d'auto-stop. En effet, deux exemples issus de ce travail de thèse illustrent ce phénomène. Le premier concerne l'haplotype LDHH4, supposé contenir une mutation récessive létale, qui porte l'allèle muté c.286>T dans le gène *SOCS2* possédant un effet pléiotrope qui conduit à augmenter le gabarit, la croissance et la production laitière des animaux hétérozygotes et homozygotes mutés par rapport aux animaux non porteurs (Rupp et al., 2015). Le deuxième exemple est la mutation cryptorchidie (c.2050DelG) qui est apparue dans un haplotype sans cornes (allèle "As"). Dans la race Manech Tête Rousse, les éleveurs ont fait le choix de sélectionner des animaux sans cornes, sans se douter qu'ils allaient faire émerger la mutation cryptorchidie.

19.2 Les mutations identifiées sont-elles spécifiques de races ?

Lorsque l'on identifie de nouveaux variants dans une population, il est intéressant de génotyper une large cohorte d'animaux de races différentes pour étudier la ségrégation de cette nouvelle mutation. Au cours de ma thèse, j'ai pu réaliser ce travail pour les mutations létales dans les gènes *CCDC65* en Lacaune et *MMUT*, *SLC33A1* et *FCGR1A* en Manech Tête Rousse et la mutation responsable de la cryptorchidie dans le gène *RXFP2*.

L'analyse de diversité a révélé la ségrégation du variant c.331G>T (*CCDC65*) non seulement

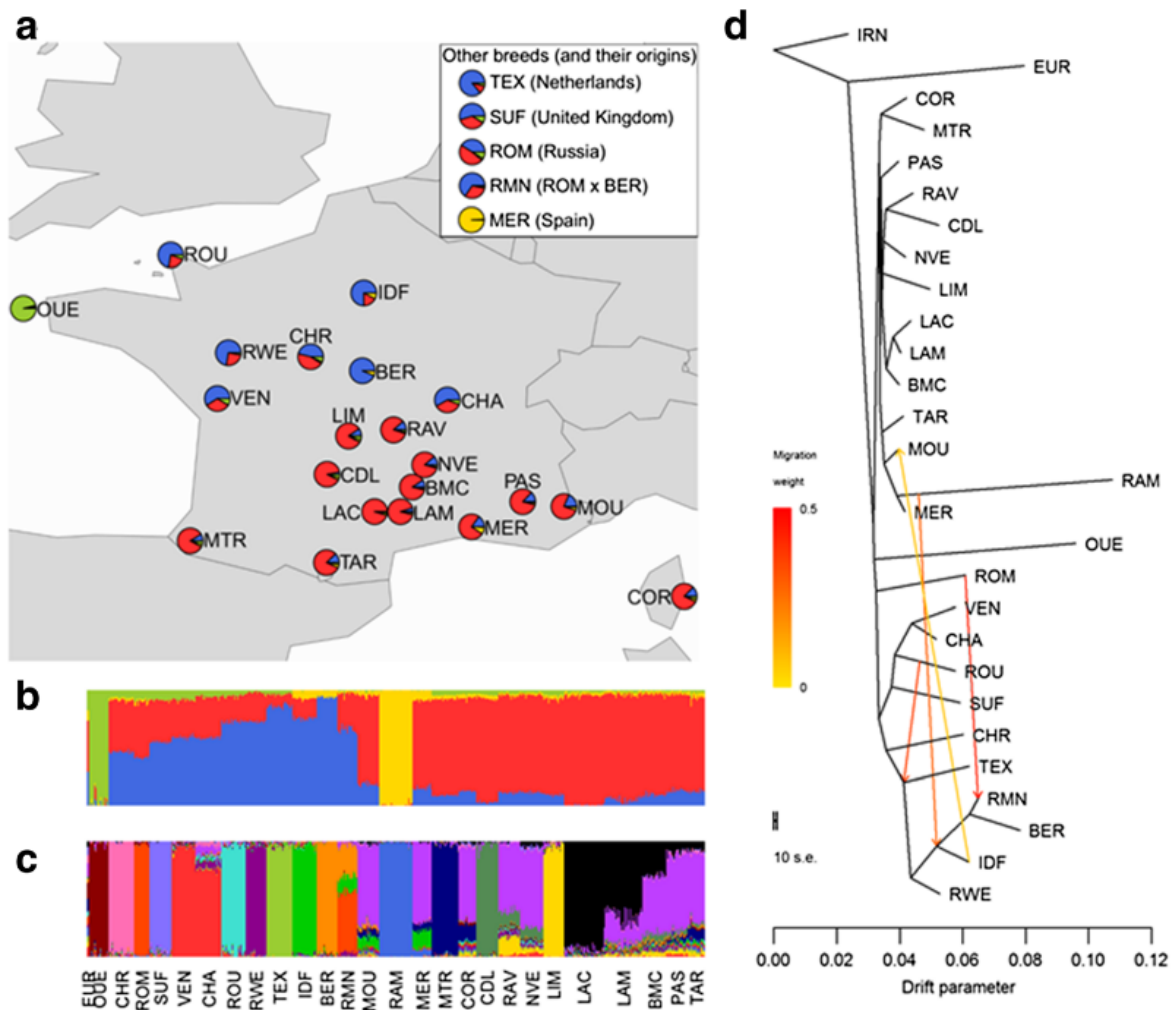


FIGURE 19.2 – **Structure de la population des ovins français.** (a) Coefficients d'ascendance individuels moyens au sein de la race pour quatre populations ancestrales. Les races sont cartographiées selon leur origine géographique. Coefficients d'ascendance individuels pour (b) quatre populations ancestrales et (c) 16 populations ancestrales. (d) Arbre de population à vraisemblance maximale avec quatre événements de migration. *Extrait de* (Rochus et al., 2018).

dans la population Lacaune laitière (LAC) mais aussi dans les populations Lacaune viande (LAM) et Blanche du Massif Central (BMC). Le cheptel ovin Lacaune a une histoire complexe avec la création de deux lignées, l'une à viande et l'autre à vocation laitière, avec quatre schémas de sélection indépendants appartenant à deux entreprises de sélection. Curieusement, la fréquence des porteurs de LDHH6 est 2,5 fois plus élevée dans une entreprise de sélection que dans l'autre (0,16 contre 0,06). En élevage, les allèles délétères pourraient être associés à de meilleures performances lorsqu'ils sont hétérozygotes pour expliquer leur maintien dans les populations (J. Cole et al., 2016; Kadri et al., 2014; Hedrick, 2015). Cependant, aucun avantage hétérozygote n'a été identifié sur les caractères sélectionnés de production laitière dans la population Lacaune pour les porteurs hétérozygotes LDHH6 et les deux entreprises de sélection ont des objectifs de sélection similaires. Ainsi, la différence observée dans la fréquence de l'allèle c.331T dans *CCDC65* entre les deux entreprises de sélection pourrait s'expliquer par la dérive génétique et la diffusion massive de béliers porteurs. La ségrégation du variant dans *CCDC65* dans les souches Lacaune laitières et à viande indique que l'événement de mutation est antérieur à la création des deux lignées. De plus,

le variant ségrége également dans la race BMC. Les analyses de la structure des populations ont révélé que la Lacaune et la BMC, toutes deux originaires du Massif Central, partageaient la même origine au sein des populations ovines du sud de l'Europe, ce qui peut expliquer la ségrégation d'allèles identiques dans les deux populations (Figure 19.2). La présence du variant est également identifiée chez la Lacaune suisse, qui pourrait s'expliquer par des importations de reproducteurs Lacaune français. Cependant, avec le nombre de porteurs détectés dans notre analyse (10 sur 37), il est important de vérifier la fréquence des mutations dans l'ensemble de la population suisse Lacaune car le risque de générer des agneaux homozygotes est assez élevé.

Les analyses des fréquences des haplotypes identifiés en Lacaune au sein des deux entreprises de sélection suggèrent que l'ensemble de ces haplotypes, bien que les fréquences soient différentes, contient des mutations apparues avant la création des deux schémas de sélection (voir paragraphe 10.1). Seulement quelques animaux du schéma Confédération (période 2017-2022) sont porteurs des haplotypes LDHH3, LDHH4 et LDHH9, suggérant aujourd'hui que ces mutations ne ségrégent plus que dans le schéma Ovitest.

Les trois mutations récessives létales (*MMUT*/*MTRDHH1*, *SLC33A1*/*MTRDHH2* et *FCGR1A*/*MTRDHH3*) et la mutation de cryptorchidie identifiées en Manech Tête Rousse sont toutes les quatre identifiées en race Latxa Cara Rubia. Cette ségrégation n'est pas anodine puisque les Manech ont été importées du pays basque espagnol dans les années 1970 et sont très apparentées génétiquement aux races Latxa espagnoles d'aujourd'hui (Legarra et al., 2014).

Le génotypage d'individus plus anciens (disponibles dans les centres de ressources biologiques) permettrait de vérifier ces partages d'allèles. Ce travail a été réalisé pour la race mouton vendéen concernant la mutation localisée dans le gène *ITGB4* (*Integrin Subunit Beta 4*) associée au syndrome d'épidermolyse bulleuse jonctionnelle (OMIA 001948-9940).

Le fait que certaines de ces mutations responsables d'anomalies puissent ségréger dans plusieurs races va m'obliger dans les prochaines analyses de recherche de mutations d'être moins strict sur les filtres utilisés concernant les non porteurs des haplotypes en déficit dans les races phylogénétiquement proches de la race de découverte des haplotypes.

19.3 Comment gérer les anomalies génétiques dans les schémas de sélection ?

La prise en compte des anomalies génétiques dans les schémas de sélection peut se faire dans le choix des reproducteurs et/ou dans la stratégie de gestion des accouplements.

Une première méthode de gestion dans le choix des reproducteurs repose sur la prise en compte des anomalies génétiques dans l'objectif de sélection, un poids économique est donné à chaque anomalie. L'impact économique dans la population d'une anomalie génétique dépend de l'âge d'apparition de l'anomalie et de sa fréquence (Boichard, Grohs, Michot, et al., 2016). Le coût économique engendré par l'anomalie est d'autant plus important lorsqu'elle ségrége dans la population à une fréquence élevée (Hozé et al., 2018). Cette méthode a été développée en bovins (Boichard, Grohs, Michot, et al., 2016) et est transposable en petits ruminants (Raoul et al., 2022). Une deuxième méthode de gestion dans le choix des reproducteurs consiste à faire évoluer la méthode de Sélection à Parenté Minimum Génomique (SPMG, méthode d'*Optimum Contribution Selection*) basée sur la gestion combinée du progrès génétique et de la variabilité génétique (J-J

Colleau, INRAE, GABI). Cette méthode SPMG permet d'optimiser la contribution des meilleurs pères et mères élites et de les accoupler entre eux en minimisant les parentés entre reproducteurs. Cette méthode est utilisée en caprins (applicable car utilisation de la semence congelée, les accouplements sont programmés plusieurs mois avant l'IA) et a été améliorée pour prendre en compte les anomalies génétiques à partir de 2023 (Palhière et al., 2022). La SPMG intégrant les anomalies génétiques n'affecte pas le progrès génétique, mais permet des augmentations/réductions de fréquences alléliques notables sur l'ensemble des mutations considérées. Son application en ovins reste limitée par l'utilisation de doses en semence fraîche.

Les anomalies génétiques peuvent également être contrôlées en optimisant les stratégies d'accouplement pour éviter de produire des porteurs homozygotes (J. B. Cole, 2015). Trois principaux modes de gestion peuvent être mis en place : (i) la contre-sélection de l'anomalie : on ne rentre plus de béliers porteurs de l'anomalie en CIA, (ii) on évite les accouplements à risque en se basant sur les génotypes des mâles (seuls les béliers collectifs sont génotypés) en interdisant les accouplements entre mâles porteurs et filles de mâles porteurs, (iii) on génotype les femelles (non réalisé en ovins pour le moment) pour permettre d'éviter les accouplements à risque. Des outils de gestion des anomalies commencent à exister en bovins. Par exemple, l'outil pANO, qui a été développé en race Montbéliarde (Brochard et al., 2018), a permis de diminuer le nombre d'accouplements à risque programmés de 25% à 1,04% en prenant en compte 4 anomalies différentes.

L'intégration de ces mutations dans les schémas de sélection, nécessite d'avoir accès aux génotypages des animaux. Pour cela, l'ensemble des mutations découvertes pendant ma thèse seront mises sur la puce de génotypage lors de la prochaine mise à jour de celle-ci.

La découverte des mutations peut même conduire à revoir l'organisation des schémas de sélection. Grâce à mes travaux, qui ont permis d'élucider le déterminisme génétique commun à l'apparition des phénotypes "cryptorchide" et "cornu" en MTR, il est envisagé de reconsidérer les stratégies de sélection pour optimiser leur éradication.

VII

Productions scientifiques

20	Articles scientifiques	275
21	Communication lors de congrès	377

20. Articles scientifiques

Article 1 (publié) - Ben Braiek et al., 2021 - *Genetics Selection Evolution* - **Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep.** doi : 10.1186/s12711-021-00634-1.

Article 2 (publié) - Ben Braiek et al., 2022 - *Genes* - **A Nonsense Variant in *CCDC65* Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep.** doi : 10.3390/genes13010045.

Article 3 - Ben Braiek et al., 2023a - bioRxiv, rédigé pour *Plos Genetics* - **Homozygous haplotype deficiency in Manech Tête Rouse dairy sheep revealed a nonsense variant in *MMUT* gene affecting newborn lamb viability.** doi : 10.1101/2023.03.10.531894.




Article 4 - Ben Braiek et al., 2023b - bioRxiv, rédigé pour *Animal genetics* - **A single base pair duplication in *SLC33A1* gene causes fetal losses and neonatal lethality in Manech Tête Rouse dairy sheep.** doi : 10.1101/2023.03.13.532360.

RESEARCH ARTICLE

Open Access



Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep

Maxime Ben Braiek¹ , Stéphane Fabre^{1*} , Chris Hozé², Jean-Michel Astruc³ and Carole Moreno-Romieux^{1†} 

Abstract

Background: Homozygous recessive deleterious mutations can cause embryo/fetal or neonatal lethality, or genetic defects that affect female fertility and animal welfare. In livestock populations under selection, the frequency of such lethal mutations may increase due to inbreeding, genetic drift, and/or the positive pleiotropic effects of heterozygous carriers on selected traits.

Results: By scanning the genome of 19,102 Lacaune sheep using 50 k single nucleotide polymorphism (SNP) phased genotypes and pedigree data, we identified 11 Lacaune deficient homozygous haplotypes (LDHH1 to LDHH11) showing a highly significant deficit of homozygous animals ranging from 79 to 100%. These haplotypes located on chromosomes 3, 4, 13, 17 and 18, spanned regions from 1.2 to 3.0 Mb long with a frequency of heterozygous carriers between 3.7 and 12.1%. When we compared at-risk matings (between carrier rams and daughters of carrier rams) and safe matings, seven of the 11 haplotypes were associated with a significant alteration of two fertility traits, a reduced success of artificial insemination (LDHH1, 2, 8 and 9), and/or an increased stillbirth rate (LDHH3, 6, 8, 9, and 10). The 11 haplotypes were also tested for a putative selective advantage of heterozygous carrier rams based on their daughter yield deviation for six dairy traits (milk, fat and protein yields, fat and protein contents and lactation somatic cell score). LDHH1, 3, 4, 5, 7, 9 and 11 were associated with positive effects on at least one selected dairy trait, in particular milk yield. For each haplotype, the most probable candidate genes were identified based on their roles in lethality of mouse knock-out models and in mammalian genetic disorders.

Conclusions: Based on a reverse genetic strategy, we identified at least 11 haplotypes with homozygous deficiency segregating in French Lacaune dairy sheep. This strategy represents a first tool to limit at-risk matings in the Lacaune dairy selection scheme. We assume that most of the identified LDHH are in strong linkage disequilibrium with a recessive lethal mutation that affects embryonic or juvenile survival in sheep but is yet to be identified.

Background

Most of the individuals in a population are likely to be heterozygous for several loss-of-function mutations. When these mutations are homozygous, they very often lead to early embryo death (embryonic lethal mutations), or to developmental defects that affect fetuses and subsequently young individuals with varying degrees of severity [1]. Advances in genomic approaches and

*Correspondence: stephane.fabre@inrae.fr

†Stéphane Fabre and Carole Moreno-Romieux contributed equally to this work

¹ GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France

Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

whole-genome sequencing in humans or in species of agronomic interest have shown that an individual can carry about a hundred of these mutations [2, 3].

In livestock under selection, the effective size of populations (N_e), which is used as an indicator of genetic diversity, is limited ($N_e \sim 100\text{--}300$) compared with that of the human population ($N_e \sim 10,000$); and as a result, the number of reproducers is rather small, particularly given the widespread use of artificial insemination (AI) [2, 4]. Even if genetic diversity and inbreeding parameters are managed, selection programs provide a fairly favorable context for the emergence of homozygous individuals with genetic defects that increase in frequency with inbreeding and/or overuse of certain sires and can finally jeopardize fertility in the whole population [5]. This has been observed in cattle where about 1% of the embryos die due to their homozygosity at one of the 10 identified lethal embryonic mutations [1]. In addition, the frequency of recessive lethal alleles could also increase in a population if they are associated with heterozygous advantages due to positive pleiotropic effects on selected production traits such as milk production in dairy cattle [6, 7], although in the homozygous state they are responsible for embryonic losses. Identification of these causal mutations has become a major issue with the emergence of genetic defects with obvious consequences on animal welfare and also have major economic implications. Indeed, in France these disorders cause losses that range from 50 to 100 million euros per year in cattle populations when their impact on fertility (about 5% decrease), loss of calves, and veterinary procedures are included in the calculation [8].

In recent decades, several genomic tools have been developed to help improve fertility in dairy cattle [9]. Among these tools, two methods have enabled the identification and characterization of recessive genetic defects and lethal mutations that affect fertility. First, homozygosity-mapping is an efficient way to map genetic defects based on a case/control approach using only a few biological samples (e.g. DNA or tissues) from affected and non-affected live animals [10]. However, embryonic and fetal lethal mutations, which are more frequently associated with fertility, have not been identified using this approach due to the difficulty to obtain biological samples. These mutations are more efficiently detected by a reverse genetic screen approach using large sets of single nucleotide polymorphism (SNP) chip genotyped animals and fertility records, such as those provided by genomic selection. In cattle, the original works of VanRaden et al. [11] and Fritz et al. [12] were based on the identification of haplotypes for which homozygous carrier animals are absent or show a more significant homozygous haplotype deficiency (HHD) than expected. Their strategy

used phased 50 k SNP genotypes from trios (offspring, sire, dam or maternal grand-sire), and the search for statistically significant HHD based on sliding windows of 20 to 100 SNPs. The underlying hypothesis is based on the linkage disequilibrium between these haplotypes and deleterious recessive mutations located nearby. This reverse genetic screen strategy has led to the identification of HHD regions that harbor 14 causal mutations in seven dairy cattle breeds. Among these, 11 HHD are associated with embryonic lethal mutations in Holstein [11–18], Jersey [19], Fleckvieh [20], Montbéliarde [12, 21], and Normande [22], and three are associated with juvenile mortality in Ayrshire [23], Brown Swiss [24], and Fleckvieh [20]. With the recent increased use of genomic selection, the accumulation of genotyping data has enabled the identification of recessive lethal mutations by reverse genetic screening in other species such as pig and chicken [25, 26]. However, to date there are no such studies in sheep.

Compared to cattle, the management of genetic diversity in dairy sheep takes advantage of their more local selection and breed management and of the use of a wider range of rams to produce fresh semen during a short reproductive campaign (May to August) [27]. For example, the efficiency of the management of genetic diversity in Lacaune dairy sheep was explained by an effective population size of 336 [28]. However, since the implementation of genomic selection in 2015, the number of Lacaune rams that enter the AI program was reduced to balance the cost of genotyping [27]. Thus, the widespread use of a limited number of AI rams could favor the emergence of recessive alleles and possibly embryonic or fetal lethal mutations that affect fertility.

In order to discover such mutations, a reverse genetic screen method was applied to the large genome-wide SNP dataset available from a genomic selection program in Lacaune dairy sheep. The specific objectives of this study were to identify haplotypes with a deficit of homozygous animals, to test the hypothesis of a negative impact of these haplotypes on fertility traits in the case of at-risk matings, to test their putative pleiotropic effects on milk production traits, and to propose candidate genes that could harbor the causal mutations.

Methods

Animal and genotyping data

The genotyped dairy Lacaune animals [$n = 19,102$ born between 1996 and 2019 (see Additional file 1: Figure S1)] were obtained from the selection schemes of two breeding companies, OVITEST (Saint-Léon, France) and the *Confédération Générale de Roquefort* (Millau, France). Table 1 lists the details on all the animals used in the study (mainly rams) that were genotyped either

Table 1 Description of genotyped animals

Year of birth	≤ 2014	2015–2016	≥ 2017		Total
Background	Research programs	Genomic selection	Genomic selection		
Number of animals	6587 rams 1517 ewes	3986 rams	7012 rams		19,102
SNP chip	MD	MD	LD (n = 6502)	MD (n = 510)	LD (n = 6502) MD (n = 12,600)
Genotyping age (months)	> 12	1–5	1–5	8–12	

MD medium density (50 k), LD low density (15 k)

on a medium-density (MD) SNP chip (Illumina Ovine SNP50 BeadChip, 54,241 SNPs, n = 12,600 genotyped animals between 1 and 12 months of age, born since 1996) or a low-density (LD) SNP chip (SheepLD v.1, 15,000 SNPs, n = 6502 genotyped animals between 1 and 5 months of age, born since 2017).

Both SNP chips (LD and MD) purchased from Illumina Inc. (San Diego, USA) were used to genotype the animals at Labogena (<http://www.labogena.fr/>) or Aveyron Labo (<http://www.aveyron-labo.com/>) genotyping facilities. Genotype data were obtained within the framework of different research programs before 2015, and subsequently from the ongoing Lacaune dairy sheep genomic selection program [29]. The pedigree of the genotyped animals was extracted from the official French livestock data system (*Systèmes Nationaux d'Information Génétique*, France Génétique Elevage, Paris, France).

Genotype quality control, imputation and phasing

The quality control of each SNP was based on three criteria: (i) a call frequency higher than 97% (% of genotyped animals for each SNP), (ii) a minor allele frequency higher than 1%, and (iii) accordance with the Hardy–Weinberg equilibrium ($P > 10^{-5}$). LD to MD genotype imputation and phasing of all genotypes were implemented with the *FImpute v2.2* software [30]; the LD and MD chips had 11,342 common SNPs. The accuracy of LD to MD imputation of Lacaune genotypes was previously assessed and showed a concordance rate per animal of 99.05%, a concordance rate per SNP of 99.12%, and a squared Pearson's correlation coefficient of 94.95% between imputed and observed SNP genotypes [31]. For subsequent identification of HHD, 38,696 SNPs on the 26 autosomal sheep chromosomes were available and mapped to the *Ovis aries* genome assembly Oar_v2.0 ([32], GigaDB, Oar_v2.0 coordinates are available at <https://doi.org/10.5524/100023>), the version used for the current genomic evaluation.

Detection of homozygous haplotype deficiency

Based on phased MD genotype data, the ovine genome was scanned using a sliding window approach of 20 consecutive SNPs to identify HHD on the 26 sheep autosomes by comparing the observed and expected number of homozygous animals using the method developed by Fritz et al. [12] in French dairy cattle, and adapted to ovine data as described below.

For each window of 20 consecutive SNPs, the frequencies of all observed haplotypes were calculated from the maternal phase, which is associated with a greater diversity of haplotypes. Only haplotypes with a frequency higher than 1% were selected. The choice of a sliding window of 20 consecutive SNPs, representing approximately 1.0 to 1.5 Mb (50 k SNP chip with an informative SNP every 60 kb, 3 Gb genome), and of a haplotype frequency higher than 1% was based on a previous simulation to estimate the frequency of recessive lethal mutations in breeding populations [2]. With an effective population size ranging from 100 to 500, as in Lacaune sheep ($N_e = 336$ [28]), the frequency of recessive lethal mutations is expected to range from 1 to 3%.

For each selected haplotype k , the number of observed homozygous animals ($N_{Obs}(k)$) was compared to the expected number of homozygous animals ($N_{Exp}(k)$). The number of expected homozygous animals $N_{Exp}(k)$ was estimated using the within-trio transmission probability with the formula described in Fritz et al. [12]. Two types of trios were considered, 38 progeny-sire-dam trios (transmission probability of haplotype k is estimated based on the genotyped sire and dam) and 15,530 progeny-sire-maternal grandsire trios (transmission probability of haplotype k is estimated based on the genotyped sire and maternal grandsire).

The probability of observing “ q ” homozygotes with an expectation “ λ ” was estimated using the Poisson distribution and calculated with the *ppois* function ($ppois(q = N_{Obs}(k), \lambda = N_{Exp}(k))$) in the RStudio software (Version 1.1.456), as previously described in Mesbah-Uddin et al. [22]. Each k haplotype was

assumed to be significantly deficient in homozygotes when the P-value was lower than 1.9×10^{-4} . This threshold was obtained by a Bonferroni correction for multiple testing at a 5‰ level of significance assuming that the number of independent tests was equal to the number of chromosomes ($n=26$). Among the significant haplotypes, only those with a severe deficiency that ranged from 75 to 100%, were retained as HHD ($(N_{Exp}(k) - N_{Obs}(k))/N_{Exp}(k) \geq 0.75$).

When the significant HHD of 20 SNPs were consecutive (i.e., shifted from the previous one by a single SNP) and showed the same minimum number of homozygous animals, they were clustered together to define a larger haplotype (all these primary HHD are in total linkage disequilibrium with each other), which we refer to as 'Lacaune deficient homozygous haplotype' (LDHH) (see Additional file 2: Tables S1 and S2). The homozygous, heterozygous, and non-carrier status of each haplotype constituting an LDHH region was then determined for each animal in the studied population ($n=19,102$).

Linkage disequilibrium was estimated between two LDHH regions on the same chromosome by the r^2 coefficient measure that was introduced by Hill and Robertson [33]. For each LDHH region, a bi-allelic locus was defined as allele 1, i.e., the detected LDHH showing a deficit in homozygotes, and as allele 2, i.e., all other haplotypes identified in the same region. The coordinates of the SNPs included in each haplotype were obtained for the ovine genome assembly Oar_v2.0 and were repositioned on the genome assembly Oar_v3.1 [32] (available from GenBank, GCA_000298735.1) for further genetic analyses.

Analysis of fertility and dairy production traits

Analysis of fertility traits

Trait records of Lacaune matings between 2006 and 2018 were obtained from the national database. We studied a first set of two fertility traits, i.e. artificial insemination success (AIS) and stillbirth rate (SBR). Among all the records, we focused only on matings between ewes with a genotyped sire and a genotyped ram, the sire and ram both having a known status at each LDHH ($n=1,155,835$ matings). AIS was coded "1" for success and "0" for failure based on lambing date according to the gestation length starting from the day of AI (147 ± 5 days). SBR was determined only in the AI success group, and coded "1" if there was at least one stillbirth in the litter or "0" if all lambs were born alive ($n=804,577$ matings). Four different types of mating are possible for each LDHH: (i) non-carrier ram \times ewe from a non-carrier sire, (ii) non-carrier ram \times ewe from a carrier sire, (iii) carrier ram \times ewe from a non-carrier sire, and (iv) carrier ram \times ewe from a carrier sire. Mating type (iv) was

considered as at-risk mating, and the cluster of mating types (i), (ii) and (iii) were considered as safe mating. A logistic threshold binary model with a logit link function was used to compare AIS and SBR between at-risk and safe matings, using the GLIMMIX procedure in the SAS software (version 9.4; SAS Institute Inc., Cary, NC). The model used is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{e}$, where \mathbf{y} is a vector of "0" or "1" coding for AIS or SBR by considering the corresponding observations Z ($Z=\{0,1\}$ for AIS, $Z=\{0,1,2+\}$ for SBR) and a threshold for each variable, so that $\mathbf{y} = 1$ if $Z \geq 1$ or $\mathbf{y} = 0$ otherwise; \mathbf{X} is the incidence matrix of fixed effects; $\boldsymbol{\beta}$ is a vector of fixed effects; \mathbf{Z} is an incidence matrix of random effects; $\boldsymbol{\gamma}$ is a vector of random effects, and \mathbf{e} is a vector of residual error effects. The fixed effects for AIS and SBR were mating type (safe or at-risk), month of AI (March to September), and lactation number (L1, L2, L3 and L4+). For SBR only, prolificacy of the ewe (1, 2, 3+ lambs/litter) was added as a fixed effect. For AIS and SBR, the random effect was interaction herd \times year ($n=470$ herds between 2006 and 2018). Traits were considered to differ significantly when the fixed effect mating type had a P-value lower than 6.3×10^{-4} . This threshold was obtained by Bonferroni correction for multiple testing at a 1% level of significance by correcting the number of independent tests assumed to be the number of significant LDHH regions multiplied by the two independent fertility traits studied.

Analysis of milk parameters

Daughter yield deviations (DYD) for milk parameters from genotyped sires with known status at each LDHH were computed from genetic evaluations (GenEval, Jouyen-Josas, France). The DYD corresponds to the average performance of the daughters of each sire, corrected for environmental effects and the average genetic value of the mothers [29]. The six parameters studied were milk yield (MY), fat (FC) and protein (PC) contents, fat (FY=MY \times FC) and protein (PY=MY \times PC) yields, and lactation somatic cell score (LSCS). LSCS corresponds to the average SCS per lactation, i.e. the log-transformation of test-day somatic cell count (SCC) defined by $SCS = \log_2\left(\frac{SCC}{10000}\right) + 3$ [34–36]. To compare all the traits on the same scale, each DYD was divided by its genetic standard deviation and referred to as standardized DYD (sDYD). Only genotyped rams with records from at least 20 daughters were included in the analysis in order to obtain sufficiently accurate DYD values ($n \sim 5400$ rams). Each trait was tested by variance analysis comparing LDHH carrier and non-carrier rams using the GLM procedure in the SAS software (version 9.4; SAS Institute Inc., Cary, NC). The fixed effect model is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where \mathbf{y} is a vector of sDYD data for each trait; \mathbf{X} is an incidence matrix of fixed effects; $\boldsymbol{\beta}$ is a vector of the fixed

effects and \mathbf{e} is a vector of residual error effects. The fixed effects are the genetic status (carrier, non-carrier) and year of birth (2000 to 2016) to correct for annual genetic gain. Traits differ significantly between carrier and non-carrier rams when the effect of the genetic status is significant, i.e. with a P-value lower than 3.1×10^{-4} . This threshold was obtained by Bonferroni correction for multiple testing at a 1% level of significance by correcting the number of independent tests assumed to be the number of significant LDHH regions multiplied by the four studied traits (MY, FC, PC and LSCS).

Identification of positional and functional candidate genes

The coordinates of each LDHH region were obtained from the ovine genome assembly Oar_v3.1. and extended 1 Mb upstream and 1 Mb downstream to obtain gene annotations from the Ensembl release 99 using the Biomart tool [37] (accessed on 20/03/2020; see Additional file 3: Table S3). Annotations and genome organization within these regions were found to be the same as those of the most recent ovine genome assembly Rambouillet v1.0 (NCBI *Ovis aries* Annotation Release 103, 2019-02-06, GCF_002742125.1).

Gene information (gene description, biological process, and molecular function) was extracted from several databases (NCBI Entrez: <https://www.ncbi.nlm.nih.gov/Web/Search/entrezfs.html>; MGI: www.informatics.jax.org; IMPC: <https://www.mousephenotype.org>). The list of identified genes was then sorted to identify the most relevant candidate genes according to (i) their known implication in lethal phenotypes in knockout/loss-of-function

mouse models based on viability information in the IMPC database and mortality/aging (embryonic, prenatal, perinatal, neonatal, postnatal, preweaning, premature death) information in the MGI database, and (ii) their association with abortion/death/autosomal recessive disorders from Online Mendelian Inheritance in Man (OMIM) (<https://omim.org>) and OMIA: Online Mendelian Inheritance in Animal (<https://omia.org>).

Results

Identification of HHD in Lacaune dairy sheep

By screening the genome of 16,346 animals (belonging to trios) from real/imputed 50 k genotyping data, we detected 266 highly significant HHD of 20 consecutive SNPs, each with a frequency higher than 1% and a deficit of homozygous animals higher than 75%. The location of these haplotypes along the ovine genome is shown as a Manhattan plot (see Additional file 4: Figure S2). As explained in “Methods” section, when significant HHD of 20 SNPs were consecutive (shifted by a single SNP) and showed the same minimum number of homozygous animals, they were clustered to define 11 larger haplotypes containing 23 to 48 SNPs and named LDHH (Table 2).

Among these haplotypes, five LDHH presented a complete deficit of observed homozygous animals (LDHH1 to 5) and six LDHH presented a partial deficit ranging from 79 to 96% of the expected number of homozygous animals (LDHH6 to 11). The length of the identified haplotypes ranged from 1.2 to 3.0 Mb on the ovine genome v3.1. LDHH3, 4, 5, 6 and 11 are located on *Ovis aries* (OAR) chromosome 3. Only LDHH4 (24

Table 2 List of Lacaune deficient homozygous haplotypes

Haplotype	OAR	Number of markers ^a	Position ^b (Mb)	Carrier frequency ^c (%)	Number of homozygotes			
					Exp ^d	Obs ^e	Deficit (%)	Poisson P-value
LDHH1	4	46/42	43.4–46.3	6.7	21	0	100	7.6×10^{-10}
LDHH2	13	28/26	44.8–46.8	6.2	17	0	100	4.1×10^{-8}
LDHH3	3	48/39	32.0–34.9	4.3	10	0	100	7.0×10^{-5}
LDHH4	3	24/19	132.4–133.9	3.8	10	0	100	5.5×10^{-5}
LDHH5	3	29/23	131.1–132.7	3.7	9	0	100	1.9×10^{-4}
LDHH6	3	27/21	136.2–137.4	12.1	72	3	96	3.5×10^{-27}
LDHH7	17	29/27	0.0–1.6	4.7	12	1	92	9.9×10^{-5}
LDHH8	18	23/20	25.7–27.5	5.5	14	2	86	9.4×10^{-5}
LDHH9	18	37/33	31.3–33.5	4.4	13	2	85	2.2×10^{-4}
LDHH10	18	26/23	33.0–34.6	5.7	17	3	82	4.1×10^{-5}
LDHH11	3	28/28	128.9–131.1	7.2	19	4	79	3.8×10^{-5}

^a Number of LDHH markers refers to ovine genome reference assembly Oar_v2.0/Oar_v3.1 and listed in Additional file 2: Table S3

^b Position on ovine genome assembly Oar_v3.1

^c Frequency of carriers in the entire genotyping population (n = 19,102)

^d Expected

^e Observed

SNPs) and LDHH5 (29 SNPs) are in high linkage disequilibrium (71%). Although they share nine SNPs at their ends, LDHH4 and LDHH5 were not originally clustered together since some consecutive 20-SNP HHD between the two LDHH did not follow our clustering rule (they were neither significant nor had the same minimum number of homozygous animals). LDHH8 is located on OAR18, 3.7 Mb from LDHH9 and 5.4 Mb from LDHH10, and shows a moderate linkage disequilibrium of 55% and 40% with these two haplotypes, respectively. Likewise, LDHH9 (37 SNPs) and LDHH10 (26 SNPs) are in high linkage disequilibrium (72%), share 11 SNPs, but were not clustered together for the same reason as LDHH4 and 5. Other LDHH are located on OAR4 (LDHH1), OAR13 (LDHH2) and OAR17 (LDHH7). Consequently, the 11 LDHH identified are most probably associated with only eight independent causal mutations in the eight following genomic regions on OAR3 (LDHH3, 32.0–34.9 Mb; LDHH11, 128.9–131.1 Mb; LDHH4–5, 131.1–133.9 Mb and LDHH6, 136.2–137.4 Mb), OAR4 (LDHH1, 43.4–46.3 Mb), OAR13 (LDHH2, 44.8–46.8 Mb), OAR17 (LDHH7, 0–1.6 Mb) and OAR18 (LDHH8–9–10, 25.7–34.6 Mb). When calculated based

on the whole genotyped population, the frequency of carriers ranged from 3.7 to 6.7% for LDHH with a total deficit of homozygotes, and from 4.4 to 12.1% for LDHH with a partial deficit.

Impact of LDHH on the success rate of AI and on stillbirth rate

To check for a putative effect of the 11 LDHH on embryonic, fetal and/or perinatal lethality in the dairy Lacaune population, two fertility-associated traits that could reflect the consequences of these mutations were analyzed: AI success (AIS: 1,155,835 matings) and stillbirth rate (SBR: 804,577 matings) (Fig. 1).

AIS could be a good proxy for embryonic losses during the first weeks after AI. In our study population, the average AIS was 69.6%. Interestingly, at-risk matings for four LDHH (1, 2, 8 and 9) had a significant negative effect on AIS leading to a decreased success rate of 2.2% for LDHH1 ($P=3.2\times 10^{-6}$) up to 3.2% for LDHH2 ($P=2.6\times 10^{-9}$).

In parallel, evaluating the stillbirth rate could be a useful way to identify mutations that cause perinatal lethality. In our study population, 5.1% of litters included

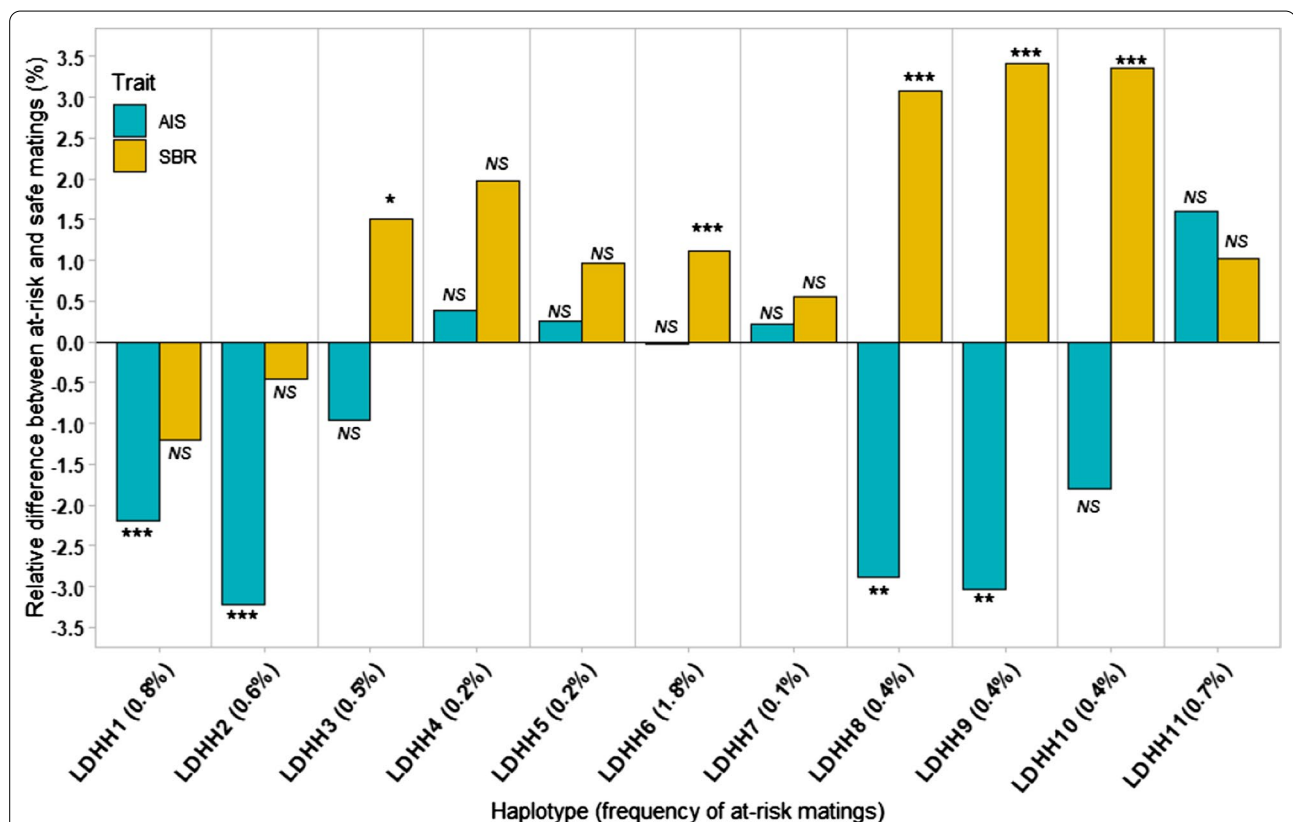


Fig. 1 Effects of LDHH on the success rate of artificial insemination and on the stillbirth rate in at-risk matings compared to safe matings. AIS artificial insemination success, SBR stillbirth rate. For each LDHH, the frequency of at-risk mating is shown in parentheses. Significant effects are indicated by the corrected P-value for multiple tests with a threshold set at $\alpha=1\%$: * $P<6.3\times 10^{-4}$; ** $P<6.3\times 10^{-5}$ and *** $P<6.3\times 10^{-6}$

at least one stillbirth. Five significant LDHH (3, 6, 8, 9 and 10) were associated with a 1.1% increase in SBR for LDHH6 ($P=5.9\times 10^{-7}$) up to 3.4% for LDHH9 ($P=8.5\times 10^{-12}$) in at-risk compared to safe matings.

Pleiotropic effects of LDHH on milk production traits

For the 11 LDHH regions detected in dairy Lacaune, we tested whether a putative selective advantage existed at the heterozygous state for six dairy traits routinely included in genomic evaluations. sDYD of five milk production traits (milk, fat and protein yields, and fat and protein contents) and lactation somatic cell score (a proxy for mastitis) were compared between carrier and non-carrier rams of each LDHH ($n\sim 5400$ genotyped rams with DYD). Figure 2 shows the relative differences in sDYD calculated between heterozygous and non-carrier rams. For all milk production traits, positive values indicate an improvement of the selected traits, whereas for the lactation somatic cell score, a positive value indicates a deterioration of udder health in the progeny of heterozygous rams.

Among the 11 haplotypes, seven had a significant effect on sDYD for at least one trait. Daughters of LDHH1 carrier rams had a higher protein content (sDYD +0.11), whereas those of LDHH3 carriers had a significantly lower protein yield (sDYD -0.12) and protein content (sDYD -0.13). Moreover, LDHH3 was significantly associated with a lower somatic cell score in milk (sDYD -0.10). Both LDHH4 and LDHH5 carriers showed a significant increase in milk production (sDYD +0.22 and +0.21, respectively) and in protein yield (sDYD +0.14 and +0.13, respectively) but had negative effects on fat (sDYD -0.19 and -0.21, respectively) and protein content (sDYD -0.20 in both). LDHH7 and LDHH9 carriers showed a higher sDYD for milk production with an increase of +0.17 and +0.11, respectively. LDHH11 was also associated with a +0.12 increase in sDYD for milk production, and a wider effect spectrum, a positive effect on fat and protein yields sDYD (with +0.09 and +0.14, respectively), but a negative effect on LSCS (+0.12) and fat content sDYD (-0.12).

Candidate genes from LDHH regions

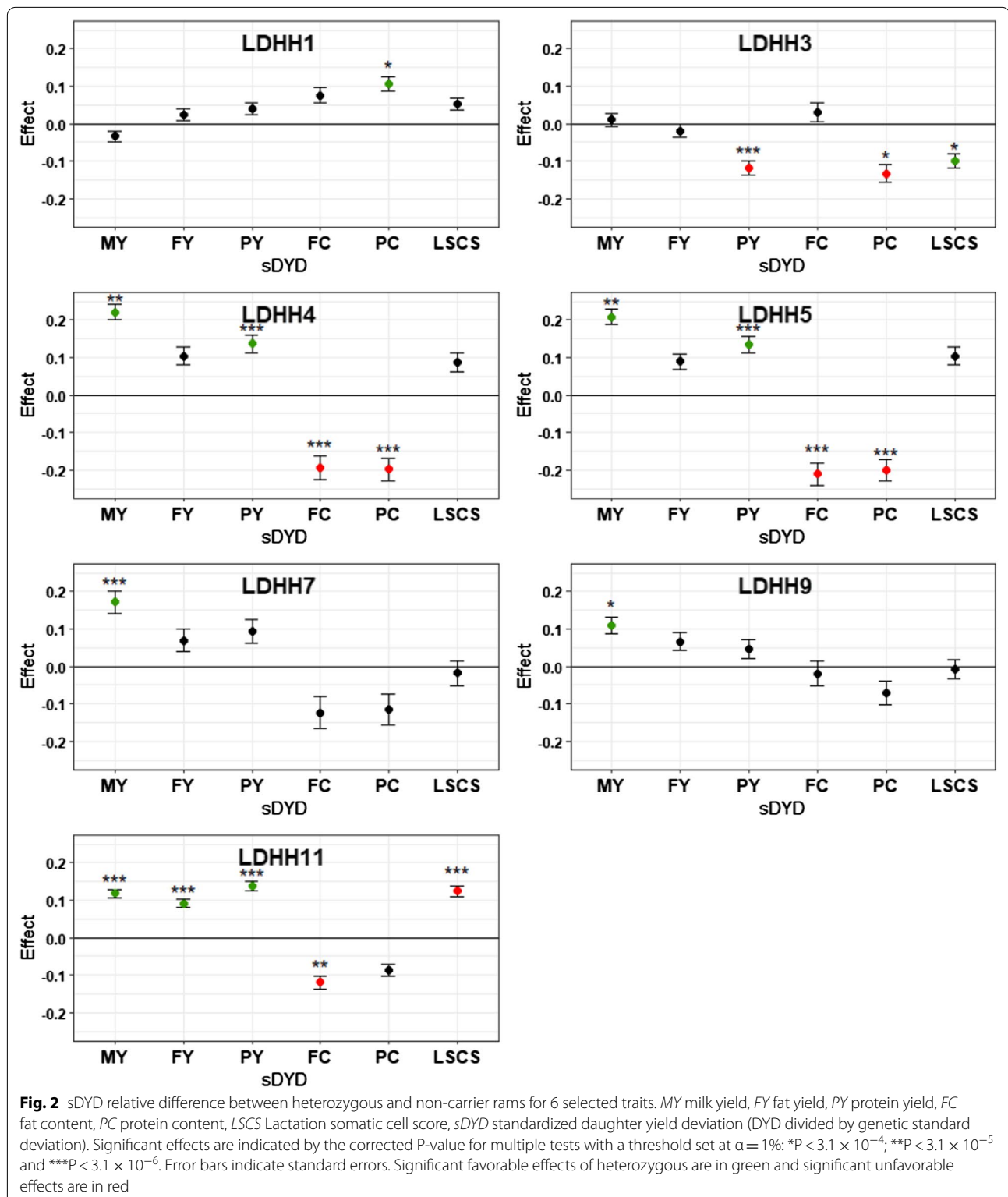
Among the 340 protein coding genes identified in the 11 LDHH regions extended by 1 Mb on each side (see Additional file 3: Table S3), 59 are associated with lethality when invalidated in mice (all homozygous mutant alleles lead to lethality, complete penetrance) and 40 show sub-viable phenotypes (higher deficit of homozygous mutant animals than expected according to Mendelian inheritance, incomplete penetrance). Among these genes, several are listed in the OMIM and OMA databases as being associated with mammalian autosomal recessive

disorders (Table 3). Most of these 99 genes are involved in nucleic acid synthesis/DNA replication/transcription (*PCNA*, *CAD*, *HOXC13*, and *ARNT2*), or encode structural or signaling proteins (*MAGI2*, *RELN*, *PRNP*, *FERMT1*, *IFT172*, *CEP83*, *KRT8*, *WNT1*, *CCDC65*, and *TLL1*) or are involved in basal metabolic processes (*PMPCB*, *ID11*, *POMC*, *HADH-A/B*, *EIF2B4*, *SCN8A*, *PFKM*, *FAH*, *MPI*, *CYP1A2*, *STRA6*, and *CRADD*). Knock-out mouse models of these genes have mainly revealed defects that affect offspring soon after birth (perinatal to weaning stages) and jeopardize the survival of the young.

Discussion

We identified eight independent genomic regions (named LDHH) in Lacaune dairy sheep showing complete or partial deficit of homozygous haplotypes, which suggests that several independent recessive deleterious mutations segregate in this population. To successfully identify HHD in dairy sheep, we used the same criteria as those used in dairy cattle, i.e. sliding windows of 20 consecutive SNPs and considered significant haplotypes with a frequency higher than 1% [12]. Using about 20,000 related animals genotyped with the 50 k SNP chip, the number of HHD identified and the carrier frequencies (between 4 and 12%) in dairy sheep are in line with the estimations reported in cattle [17]. Detection of HHD segregating at a lower frequency would have required more animals, but this will be possible in the near future thanks to the ongoing accumulation of genotyping data. Even if the length of these haplotypes depends on population structure, genome size, or the density of the SNP chip used, the length (from 1.2 to 3.0 Mb) of the LDHH identified here is in line with previously reported observations in other livestock with haplotypes spanning regions of less than 5 Mb [1]. One notable difference between our study on sheep and the literature on cattle is that less than 1% of the genotyped Lacaune animals have their two parents genotyped. Indeed, in Holstein and Normande cattle for which HHD were identified, this number reaches 30%. However, in sheep, dams are not routinely genotyped for genomic selection programs, which reduces the HHD detection power of our analysis since, in the trios included here, most of the dam genotypes are estimated by the status of the maternal grandsire and the frequency of the haplotype in the population.

In our population, we estimated that the frequencies of LDHH carriers ranged from 3.7 to 12.1%, which is in line with those of previous studies on cattle and pig (recently reviewed in Georges et al. [1]). Considering that these haplotypes are associated with deleterious mutations in the homozygous state, such frequencies could be considered as being quite high compared to the estimated



frequency in humans, which is less than 1% [2]. In live-stock populations with a small effective size ($N_e = 336$ in dairy Lacaune [28]), allele frequencies fluctuate randomly

from one generation to the next due to small sampling of reproducers. Thus, the frequency of a deleterious recessive mutation can increase sharply by genetic drift from

Table 3 Most probable candidate genes affecting viability in mouse knockout models and associated with mammalian autosomal recessive disorders

Haplotype	Gene name ^a	Gene description	Mammalian recessive disorder ^b
LDHH1	<i>MAGI2</i> [†]	Membrane associated guanylate kinase, WW and PDZ domain containing 2	Nephrotic syndrome
	<i>PMPCB</i> [†]	Peptidase, mitochondrial processing beta subunit	Multiple mitochondrial dysfunctions syndrome
	<i>RELN</i> ^{†*}	Reelin	Lissencephaly and cerebellar hypoplasia
LDHH2	<i>IDI1</i> [†]	Isopentenyl-diphosphate delta isomerase 1	Zellweger syndrome and neonatal adrenoleukodystrophy
	<i>PRNP</i> [†]	Major prion protein	Spongiform encephalopathy
	<i>PCNA</i> ^{†*}	Proliferating cell nuclear antigen	Ataxia-telangiectasia-like disorder
	<i>FERMT1</i> [†]	Fermitin family member 1	Kindler syndrome
LDHH3	<i>POMC</i> [*]	Proopiomelanocortin	Obesity, adrenal insufficiency
	<i>HADHA</i> [†] , <i>HADHB</i> [†]	Hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit alpha/beta	Long-chain 3-hydroxyacyl-CoA dehydrogenase deficiency, trifunctional protein deficiency with myopathy and neuropathy
	<i>CAD</i> [†]	Carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase	Abortion; epileptic encephalopathy, early infantile
	<i>EIF2B4</i> [†]	Eukaryotic translation initiation factor 2B subunit delta	Leukoencephalopathy, ovarioleukodystrophy
	<i>IFT172</i> [†]	Intraflagellar transport 172	Retinitis pigmentosa, short-rib thoracic dysplasia ± polydactyly
LDHH4–5	<i>HOXC13</i> [*]	Homeobox C13	Ectodermal dysplasia
	<i>KRT8</i> ^{†*}	Keratin 8	Cryptogenic cirrhosis
	<i>SCN8A</i> ^{†*}	Sodium voltage-gated channel alpha subunit 8	Spinocerebellar ataxia
LDHH6	<i>WNT1</i> ^{†*}	Wnt family member 1	Osteogenesis imperfecta, type XV
	<i>CCDC65</i> [*]	Coiled-coil domain containing 65	Primary ciliary dyskinesia
	<i>PFKM</i> [*]	Phosphofructokinase, muscle	Glycogen storage disease VII
LDHH7	<i>TLL1</i> [*]	Tolloid like 1	Heart malformation
LDHH8–9–10	<i>FAH</i> [†]	Fumarylacetoacetate hydrolase	Tyrosinemia
	<i>ARNT2</i> ^{†*}	Aryl hydrocarbon receptor nuclear translocator 2	Webb-Dattani syndrome
	<i>MPI</i> [†]	Mannose phosphate isomerase	Congenital disorder of glycosylation, type Ib
	<i>CYP1A2</i> [*]	Cytochrome P450 family 1 subfamily A member 2	Metabolizer of a cognitive enhancer
	<i>STRA6</i> [†]	Stimulated by retinoic acid 6	Microphthalmia
LDHH11	<i>CRADD</i> [†]	CASP2 and RIPK1 domain containing adaptor with death domain	Mental retardation, with variant lissencephaly
	<i>CEP83</i> [†]	Centrosomal protein 83	Nephronophthisis

^a Homozygous lethal[†] (complete penetrance) and homozygous sub-viable* (incomplete penetrance) genes affecting developmental stages reported in knockout databases (IMPC and/or MGI)

^b Mammalian autosomal recessive disorders reported in OMIM and/or OMIA databases

an initial value close to 0 up to several percent, as we observed in Lacaune (from 2 to 6%), due to the spread of very influential carriers and their progeny [5]. The Lacaune sheep population has a complex history with the creation of two lines, one for meat and one for dairy purposes, and four independent selection schemes. Based on this particular population structure, further simulation studies would be useful to estimate the effect of drift on the frequency of LDHH carriers. Apart from genetic drift, maintaining or increasing the frequency of deleterious alleles in a livestock population can be accomplished by balancing selection when deleterious alleles provide a heterozygous advantage on selected traits [5].

Several examples of homozygous deleterious mutations have been reported with a heterozygous advantage in livestock [38], such as the 660-kb deletion (spanning four genes) identified in Nordic Red cattle that leads to higher milk production in heterozygous carriers [7], the 2-pb deletion in the *MRC2* gene leading to the Crooked Tail syndrome in Belgian Blue cattle, but enhances muscular development in heterozygous carriers [39], or the 212-kb deletion affecting the *BBS9* and *BMPER* genes associated with higher growth rates in pig [40]. In the current study, seven of the 11 LDHH (i.e. six of the eight independent regions) are associated with positive effects on at least one milk trait (milk, protein and fat yields, protein and

fat contents and LSCS) in the heterozygous state. The observed difference between carrier and non-carrier rams corresponds mainly to 1 or 2 years of genetic gain for the given traits [34]. For example, the observed difference in milk yield corresponds to +4.0 L up to +8.1 L for LDHH9 and LDHH4 carriers, respectively. These results are relatively similar to the effect observed for lethal mutations detected in US dairy cattle [6]. Surprisingly, LDHH6 which is the most frequent haplotype in our study (carrier frequency of 12.1%) had no pleiotropic effect on milk traits, which was also the case of the less frequent LDHH2, LDHH8 and LDHH10. Apart from the hypotheses of genetic drift or association with selected traits, the high frequency observed for LDHH6 could be explained by balancing selection of other traits that are obviously not implemented in the selection scheme, such as morphological phenotypes that match breed criteria (stature, fleece type and color, hornless).

Based on linkage disequilibrium, the 11 LDHH identified delimit only eight independent genomic regions probably associated with eight causal mutations. Within these regions, we searched for candidate genes that may be affected by these mutations in accordance with the characteristics of the genotyped population, the deficit of homozygous animals and the observed impact on AIS and SBR (assuming lethality from the embryonic to the juvenile stage). We reviewed mouse knockout models and report genetic disorders that are listed in human and animal databases (Table 3). Knowledge based on bovine studies shows that recessive lethal alleles are caused by loss-of-function mutations (stop-gain, frameshift, missense, splice site and deletion) affecting genes mainly involved in cell division, DNA replication, transcription, RNA processing, or coding for structural proteins or in essential metabolic processes [2, 41]. Based on AIS and SBR analyses, we were able to classify LDHH into four patterns, each with a hypothesized associated effect. The first pattern groups LDHH1 on OAR4 and LDHH2 regions on OAR13. Both haplotypes have a significant impact on AIS that we associate with early embryonic loss. Accordingly, we found that *PMPCB* and *PCNA* are strong candidate genes in the LDHH1 and LDHH2 regions, respectively. Indeed, homozygous invalidation of both genes in mouse causes embryonic lethality around the implantation stage [MGI:1920328, MGI:97503]. The second pattern corresponds to LDHH that have an impact on both AIS and SBR and that we associate with embryo/fetal loss throughout the gestation period, i.e. LDHH8, 9 and 10 on OAR18. In this case, we suspect a unique lethal mutation that affects one of the four following candidate genes *FAH*, *ARNT2*, *MPI* and *STRA6*, which are reported to be homozygous lethal in knock-out mouse models (*Fah* [MGI:95482], *Arnt2* [MGI:107188],

Mpi [MGI:97075] [42] and *Strat6* [MGI:107742]). The third pattern groups LDHH3 and LDHH6 on OAR3 and affects only SBR. We assume that these LDHH harbor lethal mutations with effects that occur at the end of gestation and/or very soon after birth. Within the LDHH3 region in complete homozygous deficiency, the *CAD* gene is the most obvious candidate. Indeed, invalidation of the *Cad* gene in mouse causes preweaning lethality with complete penetrance [MGI:1916969], and a missense mutation in *CAD* (g.72399397T>C; p.Tyr452Cys) associated with the NH7 homozygous deficient haplotype [OMIA 002201-9913] causes late abortion during the last months of gestation in Normande cattle [22]. In the LDHH6 region, three genes *Wnt1*, *Ccdc65*, and *Pfkm*, are associated with perinatal, neonatal or preweaning lethality in mouse knock-out models [MGI:98953, MGI:2146001, MGI:97548]. Among these, a mutation in *CCDC65* encoding a nexin-dynein regulatory complex, is reported to cause “ciliary dyskinesia, primary 27” [CILD27, OMIM #615504], which is a neonatal respiratory distress syndrome in humans [43]. Interestingly, a respiratory syndrome is also reported for the homozygous deficient Braunvieh haplotype 2 (BH2) [OMIA 001939-9913] associated with perinatal and juvenile mortality in cattle due to a mutation in *TUBD1* (*tubulin delta 1*) that disorganizes the microtubules in airway cilia [24]. Finally, the fourth pattern groups LDHH4–5, LDHH7 and LDHH11 for which we failed to produce evidence for an alteration of fertility traits when comparing at-risk and safe matings. This could be due to the relatively low frequency of at-risk matings in our dataset for LDHH4–5 (0.2%) and LDHH7 (0.1%), which made it impossible to reach statistical significance. Alternatively, these haplotypes could also host mutations that affect lamb survival in a later growth period, as observed for heart malformation in mouse [44] and in humans (“atrial septal defect 6” [ASD6, OMIM #613087]; [45]) associated with the *TLL1* gene located in the LDHH7 region. The last hypothesis for these LDHH in the fourth pattern concerns an effect related to morphological defects or breed standards that is possibly counter-selected in selection schemes.

Particular attention should also be focused on OAR3 and the region spanned by LDHH4–5 and LDHH11. These haplotypes are located in the region of the p.R96C mutation (OAR3, g.129722200C>T, genome assembly v3.1) that has been previously detected in the *suppressor of cytokine signaling 2* (*SOCS2*) gene in Lacaune dairy sheep and is associated with sensitivity to mastitis [46]. Interestingly, in our study, these LDHH, and particularly LDHH11 that includes *SOCS2*, had positive effects on milk production, on fat and protein yields, but negative effects on LSCS and fat content, as already observed for the C variant of the p.R96C mutation. Since 2017,

genotypes for OAR3:129722200C>T are available for all animals genotyped on the LD chip when candidate rams are selected to enter the breeding center. Although LDHH4 and LDHH5 are not in linkage disequilibrium with LDHH11, all three are associated with the *SOCS2* mutation. Indeed, 69% of the LDHH4, 92% of the LDHH5 and 97% of the LDHH11 heterozygous carriers were C/T for the *SOCS2* mutation but only 16%, 18% and 23% of the heterozygotes for the *SOCS2* mutation carried one of these haplotypes, respectively (see Additional file 5: Table S4). This could be explained by the fact that the *SOCS2* mutation is carried by two different haplotypes, one in strong linkage disequilibrium with LDHH4–5 and the other with LDHH11. Thus, linkage disequilibrium with the *SOCS2* mutation may explain the observed effect of LDHH4–5 and LDHH11 on milk traits. However, OAR3:129722200C>T in *SOCS2* is not a candidate lethal mutation associated with these haplotypes that could explain their deficit of homozygotes. We hypothesize that a recessive lethal allele located in one of those haplotypes is in partial linkage disequilibrium with the mutant allele of *SOCS2*. Maintaining frequencies of LDHH4–5 (4%) and LDHH11 (7%) in the Lacaune population would then not be explained by a heterozygous advantage of deleterious mutations but by genetic hitchhiking due to *SOCS2* and its association with selected dairy traits [5].

Conclusions

In this study, we demonstrated the possible segregation of recessive lethal alleles in a sheep population by applying a reverse genetic screen approach on a large genotyping dataset and searching for homozygous haplotype deficiency. Among the 11 LDHH genomic regions detected, we hypothesize that at least eight independent recessive mutations cause early embryonic loss, peri/neonatal lethality or severe health defects in young lambs. We identified the most obvious candidate genes that are assumed to be altered by these mutations, and provide strong working hypotheses to identify them by whole-genome sequencing of heterozygous carriers of the different LDHH. Identification of causal mutations, and of the corresponding altered genes, is important to accurately identify the phenotype they control and to improve our knowledge of the fundamental mechanisms underlying the phenotype. As already observed for haplotypes associated with deleterious recessive mutations, particularly in dairy cattle, most of the LDHH evidenced in dairy sheep were associated with decreased fertility, but also had positive pleiotropic effects on milk production. Management of these haplotypes—or of their causal mutations once they are discovered—in the Lacaune dairy sheep

selection scheme through reasoned mating of carrier rams and putative carrier ewes could improve overall fertility and lamb viability. Moreover, if these mutations segregate more widely in other ovine populations, the consequences would apply more broadly to sheep breeding in general.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-021-00634-1>.

Additional file 1: Figure S1. Distribution of genotyped animals. Figure S1 shows the number of genotyped animals according to sex and year of birth.

Additional file 2: Table S1. Clustering HHD in LDHH regions. Table S1 shows all significant haplotypes of 20 markers (i.e., 266 HHD with frequency > 1%, P-value < 1.9×10^{-4} and deficit $\geq 75\%$). As described in the “Methods” section, the 266 HHD could be associated with 11 LDHH regions which are shown in green. **Table S2.** SNPs defining the LDHH regions. Table S2 gives the position of each SNP within LDHH regions according to the sheep reference genome v2.0 and v3.1, and the phased alleles of each deficient haplotype.

Additional file 3: Table S3. Positional candidate genes within LDHH regions. Table S3 summarizes all the protein coding genes that are located within the 11 LDHH regions extended by 1 Mb on each side. Genomic coordinates refer to the sheep reference genome v3.1. When available, gene information and association with autosomal recessive disorders in mammals are reported for each protein coding gene based on the following databases: NCBI Entrez: <https://www.ncbi.nlm.nih.gov/Web/Search/entrezfs.html>; MGI: www.informatics.jax.org; IMPC: <https://www.mousephenotype.org>; OMIM: Online Mendelian Inheritance in Man (<https://omim.org>) and OMA: Online Mendelian Inheritance in Animal (<https://omia.org>).

Additional file 4: Figure S2. Manhattan plot of HHD. Each point represents one haplotype of 20 markers with a frequency > 1% in the maternal phase. The red line represents the P-value threshold (1.9×10^{-4}) to consider a haplotype significantly deficient in homozygotes. Only HHD with a deficit in homozygotes $\geq 75\%$ were selected and resulted in the identification of 266 significant HHD (represented by green dots).

Additional file 5: Table S4. Contingency table between LDHH4, 5 and 11 status and genotypes for the *SOCS2* mutation OAR3:129722200C>T. Ram genotypes at both *SOCS2* OAR3:129722200C>T and LDHH loci.

Acknowledgements

The authors thank the Phenofinlait, SheepSNPQTL, Roquefort’In, Degeram and GOLD project consortia for genomic selection implementation in French dairy Lacaune. We are grateful to the UPRA Lacaune, CNBL, FGE and Valogène who made the genotyping data available for this study. We thank also Sébastien Fritz (Allièze) for helpful discussions.

Authors’ contributions

MB performed the analyses, interpreted the results, and drafted the manuscript. CH designed the script to identify homozygous haplotype deficiency, MB and CMR adapted the script in sheep. JMA managed access to genotype and phenotype data. CMR and SF conceived and designed the research. CMR and SF supervised the analyses, helped interpret the results, wrote, reviewed and edited the final manuscript. All authors read and approved the final manuscript.

Funding

This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Grant Agreement No. 772787 (SMARTER). MB was supported by a Ph.D. grant for the HOMLET program co-funded by APIS-GENE and *Région Occitanie*.

Availability of data and materials

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France. ²Allice, 149 rue de Bercy, 75595 Paris, France. ³Institut de l'Élevage, 24 chemin de Borde-Rouge, 31326 Castanet-Tolosan, France.

Received: 12 October 2020 Accepted: 19 April 2021

Published online: 01 May 2021

References

- Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. *Nat Rev Genet*. 2019;20:135–56.
- Charlier C, Li W, Harland C, Littlejohn M, Coppieters W, Creagh F, et al. NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res*. 2016;26:1333–41.
- MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. *Science*. 2012;335:823–8.
- Danchin-Burge C, Danvy S, Laloë D, Verrier E. Création d'un observatoire de la VARIabilité génétique des RUMInants et des Equidés (VARUME). *Innov Agron*. 2017;55:235–45.
- Boichard D, Grohs C, Danchin-Burge C, Capitan A. Genetic defects: definition, origin, transmission and evolution, and mode of action. *INRAE Prod Anim*. 2016;29:297–306.
- Cole JB, Null DJ, VanRaden PM. Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. *J Dairy Sci*. 2016;99:7274–88.
- Kadri NK, Sahana G, Charlier C, Iso-Touru T, Guldbandsen B, Karim L, et al. A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock. *PLoS Genet*. 2014;10:e1004049.
- Hozé C, Fritz S, Baur A, Grohs C, Danchin-Burge C, Boichard D. Accounting for genetic conditions in breeding objectives in dairy cattle. In: Proceedings of the 24ème Rencontres Autour des Recherches sur les Ruminants: 5–6 December, Paris; 2018.
- Capitan A, Michot P, Baur A, Saintilan R, Hozé C, Valour D, et al. Genetic tools to improve reproduction traits in dairy cattle. *Reprod Fertil Dev*. 2015;27:14–21.
- Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, et al. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat Genet*. 2008;40:449–54.
- VanRaden PM, Olson KM, Null DJ, Hutchison JL. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J Dairy Sci*. 2011;94:6153–61.
- Fritz S, Capitan A, Djari A, Rodriguez SC, Barbat A, Baur A, et al. Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in *GART*, *SHBG* and *SLC37A2*. *PLoS One*. 2013;8:e65550.
- Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet*. 2014;46:858–65.
- McClure MC, Bickhart D, Null D, VanRaden P, Xu L, Wiggins G, et al. Bovine exome sequence analysis and targeted SNP genotyping of recessive fertility defects BH1, HH2, and HH3 reveal a putative causative mutation in *SMC2* for HH3. *PLoS ONE*. 2014;9:e92769.
- Adams HA, Sonstegard TS, VanRaden PM, Null DJ, Van Tassell CP, Larkin DM, et al. Identification of a nonsense mutation in *APAF1* that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *J Dairy Sci*. 2016;99:6693–701.
- Schütz E, Wehrhahn C, Wanjek M, Bortfeld R, Wemheuer WE, Beck J, et al. The Holstein Friesian lethal haplotype 5 (HH5) results from a complete deletion of *TBF1M* and cholesterol deficiency (CDH) from an ERV-(LTR) insertion into the coding region of *APOB*. *PLoS One*. 2016;11:e0154602.
- Hozé C, Escouflaire C, Mesbah-Uddin M, Barbat A, Boussaha M, Deloche MC, et al. Short communication: a splice site mutation in *CENPU* is associated with recessive embryonic lethality in Holstein cattle. *J Dairy Sci*. 2020;103:607–12.
- Fritz S, Hozé C, Rebours E, Barbat A, Bizard M, Chamberlain A, et al. An initiator codon mutation in *SDE2* causes recessive embryonic lethality in Holstein cattle. *J Dairy Sci*. 2018;101:6220–31.
- Sonstegard TS, Cole JB, VanRaden PM, Van Tassell CP, Null DJ, Schroeder SG, et al. Identification of a nonsense mutation in *CWC15* associated with decreased reproductive efficiency in Jersey cattle. *PLoS One*. 2013;8:54872.
- Pausch H, Schwarzenbacher H, Burgstaller J, Flisikowski K, Wurmser C, Jansen S, et al. Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. *BMC Genomics*. 2015;16:312.
- Michot P, Fritz S, Barbat A, Boussaha M, Deloche M-C, Grohs C, et al. A missense mutation in *PFAS* (*phosphoribosylformylglycinamide synthase*) is likely causal for embryonic lethality associated with the MH1 haplotype in Montbéliarde dairy cattle. *J Dairy Sci*. 2017;100:8176–87.
- Mesbah-Uddin M, Hozé C, Michot P, Barbat A, Lefebvre R, Boussaha M, et al. A missense mutation (p.Tyr452Cys) in the *CAD* gene compromises reproductive success in French Normande cattle. *J Dairy Sci*. 2019;102:6340–56.
- Venhoranta H, Pausch H, Flisikowski K, Wurmser C, Taponen J, Rautala H, et al. In frame exon skipping in *UBE3B* is associated with developmental disorders and increased mortality in cattle. *BMC Genomics*. 2014;15:890.
- Schwarzenbacher H, Burgstaller J, Seefried FR, Wurmser C, Hilbe M, Jung S, et al. A missense mutation in *TUBD1* is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. *BMC Genomics*. 2016;17:400.
- Derks MFL, Megens H-J, Bosse M, Lopes MS, Harlizius B, Groenen MAM. A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics*. 2017;18:858.
- Derks MFL, Megens H-J, Bosse M, Visscher J, Peeters K, Bink MCAM, et al. A survey of functional genomic variation in domesticated chickens. *Genet Sel Evol*. 2018;50:17.
- Buisson D, Astruc J-M, Barillet F. Overview and perspectives of the management of genetic diversity in French dairy sheep. *INRAE Prod Anim*. 2018;31:1–12.
- IDELE. VARUME. 2016. <http://idele.fr/presse/publication/idelesolr/recomends/varume-resultats-2016-1.html/>. Accessed 2 Apr 2021.
- Astruc J-M, Baloche G, Buisson D, Labatut J, Lagriffoul G, Larroque H, et al. Genomic selection in French dairy sheep. *INRAE Prod Anim*. 2016;29:41–55.
- Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 2014;15:478.
- Larroque H, Chassier M, Saintilan R, Astruc J-M. Imputation accuracy from a low density SNP panel in 5 dairy sheep breeds in France. In: Proceedings of the 68th annual meeting of the European association for animal production: 28 August–1 September, Tallinn; 2017.
- Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*. 2014;344:1168–73.
- Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet*. 1968;38:226–31.
- Barillet F, Lagriffoul G, Marnet P-G, Larroque H, Rupp R, Portes D, et al. Breeding objectives and reasoned strategy of implementation at the population level for French dairy sheep breeds. *INRAE Prod Anim*. 2016;29:19–40.

35. IDELE. Comment sont évalués génétiquement les ovins laitiers en France. 2016. http://idele.fr/no_cache/recherche/publication/idelesolr/recom mends/comment-evaluer-genetiquement-les-ovins-laitiers-en-france.html. Accessed 2 Apr 2021.
36. Rupp R, Lagriffoul G, Astruc JM, Barillet F. Genetic parameters for milk somatic cell scores and relationships with production traits in French Lacaune dairy sheep. *J Dairy Sci.* 2003;86:1476–81.
37. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. *Nucleic Acids Res.* 2020;48:D682–8.
38. Hedrick PW. Heterozygote advantage: the effect of artificial selection in livestock and pets. *J Hered.* 2015;106:141–54.
39. Fasquelle C, Sartelet A, Li W, Dive M, Tamma N, Michaux C, et al. Balancing selection of a frame-shift mutation in the *MRC2* gene accounts for the outbreak of the crooked tail syndrome in Belgian Blue cattle. *PLoS Genet.* 2009;5:e1000666.
40. Derks MFL, Lopes MS, Bosse M, Madsen O, Dibbits B, Harlizius B, et al. Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS Genet.* 2018;14:e1007661.
41. Fritz S, Michot P, Hozé C, Grohs C, Boussaha M, Boichard D, et al. Anticipate the emergence of genetic defects from genomics data. *INRAE Prod Anim.* 2016;29:339–50.
42. DeRossi C, Bode L, Eklund EA, Zhang F, Davis JA, Westphal V, et al. Ablation of mouse phosphomannose isomerase (*Mpi*) causes mannose 6-phosphate accumulation, toxicity, and embryonic lethality. *J Biol Chem.* 2006;281:5916–27.
43. Austin-Tse C, Halbritter J, Zariwala MA, Gilberti RM, Gee HY, Hellman N, et al. Zebrafish ciliopathy screen plus human mutational analysis identifies *C21orf59* and *CCDC65* defects as causing primary ciliary dyskinesia. *Am J Hum Genet.* 2013;93:672–86.
44. Clark TG, Conway SJ, Scott IC, Labosky PA, Winnier G, Bundy J, et al. The mammalian tolloid-like 1 gene, *Tll1*, is necessary for normal septation and positioning of the heart. *Development.* 1999;126:2631–42.
45. Stańczak P, Witecka J, Szydło A, Gutmajster E, Lisik M, Auguściak-Duma A, et al. Mutations in mammalian *tolloid-like 1* gene detected in adult patients with ASD. *Eur J Hum Genet.* 2009;17:344–51.
46. Rupp R, Senin P, Sarry J, Allain C, Tasca C, Ligat L, et al. A point mutation in suppressor of cytokine signalling 2 (*Socs2*) increases the susceptibility to inflammation of the mammary gland while associated with higher body weight and size and higher milk production in a sheep model. *PLoS Genet.* 2015;11:e1005629.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



Article

A Nonsense Variant in *CCDC65* Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep

Maxime Ben Braiek ¹, Carole Moreno-Romieux ¹, Charlotte Allain ², Philippe Bardou ³, Arnaud Bordes ¹, Frédéric Debat ¹, Cord Drögemüller ⁴ , Florence Plisson-Petit ¹, David Portes ², Julien Sarry ¹, Némuel Tadi ¹, Florent Woloszyn ¹ and Stéphane Fabre ^{1,*} 

- ¹ GenPhySE, Université de Toulouse, Institut National de Recherche Pour L'agriculture, L'alimentation et L'environnement (INRAE), Ecole Nationale Vétérinaire de Toulouse (ENVT), 31326 Castanet-Tolosan, France; maxime.ben-braiek@inrae.fr (M.B.B.); carole.moreno-romieux@inrae.fr (C.M.-R.); arnaud.bordes@inrae.fr (A.B.); frederic.debat@inrae.fr (F.D.); florence.plisson-petit@inrae.fr (F.P.-P.); julien.sarry@inrae.fr (J.S.); nemuel.tadi@inrae.fr (N.T.); florent.woloszyn@inrae.fr (F.W.)
- ² UE Domaine de La Fage, Institut National de Recherche Pour L'agriculture, L'alimentation et L'environnement (INRAE), 12250 Saint-Jean et Saint-Paul, France; charlotte.allain@inrae.fr (C.A.); david.portes@inrae.fr (D.P.)
- ³ Sigeneae, Institut National de Recherche Pour L'agriculture, L'alimentation et L'environnement (INRAE), 31326 Castanet-Tolosan, France; philippe.bardou@inrae.fr
- ⁴ Institute of Genetics, Vetsuisse Faculty, University of Bern, 3012 Bern, Switzerland; cord.droegemueller@vetsuisse.unibe.ch
- * Correspondence: stephane.fabre@inrae.fr



Citation: Ben Braiek, M.; Moreno-Romieux, C.; Allain, C.; Bardou, P.; Bordes, A.; Debat, F.; Drögemüller, C.; Plisson-Petit, F.; Portes, D.; Sarry, J.; et al. A Nonsense Variant in *CCDC65* Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep. *Genes* **2022**, *13*, 45. <https://doi.org/10.3390/genes13010045>

Academic Editor:
Beatriz Gutierrez-Gil

Received: 17 November 2021

Accepted: 20 December 2021

Published: 24 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: We recently demonstrated that the Lacaune deficient homozygous haplotype 6 (LDHH6) potentially hosts a recessive perinatal lethal mutation in Lacaune dairy sheep mapped on OAR3. In the present study, we have analyzed the whole-genome sequences of two Lacaune ram heterozygous carriers of LDHH6. After variant calling and filtering against the variants of 86 non-carrier rams, we have identified a single nucleotide variant (SNV) in the two LDHH6 carriers whose variant allele induced a premature stop codon (p.Glu111*) in the *Coiled-Coil Domain Containing 65* (*CCDC65*) gene. *CCDC65* is involved in the assembly of the nexin-dynein regulatory complex for the formation of microtubules in ciliated cells. In order to identify the phenotype in homozygous sheep, we generated at-risk matings ($n = 17$) between rams and ewes heterozygous for the candidate variant in *CCDC65*. A total of 16 lambs were born alive with five genotyped as homozygous carriers. The homozygous lambs suffered from respiratory problems, and four of them died within the first month of life. At necropsy, we observed a broad hepatization of lung lobes possibly induced by infectious pneumonia. The management of this lethal recessive allele (frequency of 0.06) through reasoned mating in the Lacaune sheep selection schemes could reduce lamb mortality by 2%.

Keywords: homozygous haplotype deficiency; LDHH6; *CCDC65*; juvenile lethal mutation; genetic disorder; primary cilia dyskinesia (PCD); ciliopathy; development; sheep

1. Introduction

Health and animal welfare are major concerns in livestock populations with important economic repercussions. During the last few decades, mortality in the first month of life remains high around 15–20% in sheep and most of this mortality occurs within the first three days of postnatal life [1,2]. The risk factors for lamb mortality are mainly due to parameters depending on the mother (nutrition, mothering ability, health status, parturition conditions), the lamb itself (low birth weight, vigor, colostrum intake, congenital malformations) and/or their environment (extreme weather, predation, infectious diseases) [1,3].

Although adequate farm environment and management practices could favor lamb viability (reviewed in Dwyer et al. [1]), it would be possible to act on the genetic aspect of

lamb survival. However, the QTL approach for mapping genetic variants/loci affecting this trait is difficult due to low heritability ($h^2 < 0.1$) and multifactorial nature of lamb survival or lamb birthweight (i.e., best proxy for lamb survival) [1,4–6]. Nevertheless, thanks to the use of high throughput genomic tools (single nucleotide polymorphism arrays, whole-genome sequences), the study of genetic disorders and analysis of associated data have successfully led the identification of many causal variants affecting the viability of young animals [7].

Nowadays, two main approaches are used to identify and characterize causal variants associated with genetic disorders. The first one is the homozygosity-mapping based on a case-control approach using only few biological samples from affected and non-affected animals [8–10]. In this pioneering use of homozygosity-mapping using large number of SNP array in cattle, Charlier et al. have identified three causal variants responsible for *congenital muscular dystony* types 1 [OMIA 001450-9913] and 2 [OMIA 001451-9913] in Belgian Blue cattle and *ichthyosis fetalis* [OMIA 002238-9913] in Italian Chianina cattle [8]. In sheep, this approach was used successfully for the first time in the *microphthalmia* [OMIA 000649-9940] in Texel [11]. Since the last few decades, this method is efficient to map genetic defects and their associated variants as shown by the growing number of discovered causal variants referenced in the OMIA: Online Mendelian Inheritance in Animal database (<https://www.omia.org/>; accessed 2 September 2021). As example, with this approach, several pathogenic variants were identified causing *junctional epidermolysis bullosa* in cattle [12,13] and sheep [14–16] and the affected animals died shortly after birth due to skin disorders (injuries on limb extremities, mucous membranes). However, when mutations affect viability of developing embryos during gestation, phenotype or biological samples are not available. Thus, these mutations are more efficiently detectable by a second approach using reverse genetic screens. This approach is based on the availability of large sets of SNP array genotyped animals provided by genomic selection for example [17,18]. Reverse screens identified haplotypes for which homozygous carrier animals are absent or significantly less observed when compared to the expected number based on haplotype frequency. In cattle, the original works of VanRaden et al. [17] and Fritz et al. [18] using reverse genetic screen followed by whole-genome sequence analyses have successfully identified candidate causal variants in linkage disequilibrium with haplotypes with significant partial or total deficit in homozygous animals. Going on with this approach, numerous embryonic lethal mutations were identified in Holstein and Jersey breeds altering *APAF1* [19], *SMC2* [20,21], *GART* [18], *TFB1M* [22], *SDE2* [23], *CENPU* [24] and *CWC15* [25] genes. Additionally, neonatal/juvenile lethal mutations associated with homozygous haplotype deficiency were also identified in Ayrshire and Braunvieh breeds affecting *UBE3B* [20] and *TUBD1* [26] genes.

Recently, with the implementation of genomic selection in small ruminants [27] and the availability of large genotyping datasets, we have identified homozygous deficient haplotypes segregating in the Lacaune dairy sheep population by a reverse genetic screen method [28]. We have identified 11 haplotypes with significant deficit in homozygous animals, ranging from 79 to 100%. Some of them were associated with fertility loss at artificial insemination, and/or with increased stillbirth rate. We therefore assumed that these haplotypes are likely to host recessive mutations causing embryonic/fetal or perinatal lethality, respectively. Among these haplotypes, LDHH6 (Lacaune deficient homozygous haplotype 6) [OMIA 002342-9940] was the most frequent (12.1% of heterozygous carriers, located on OAR3:146.2–147.9 Mb on Oar_rambouillet_v1.0) and presented a homozygous deficit of 96% (three homozygous animals were observed, whereas 72 animals were expected; $p = 3.5 \times 10^{-27}$). LDHH6 showed an increased stillbirth rate in at-risk matings between carrier rams and daughters of carrier rams when compared to safe matings. We have previously reported several candidate genes in the LDHH6 region (*WNT1*, *CCDC65* and *PFKM* genes), but the underlying causative variant is still unknown [28].

In the present study, we have taken advantage of whole-genome sequence data to identify the causal variant associated with the LDHH6 haplotype, studied the segregation

of the variant in several sheep breeds and validated the lethal effect of the proposed causal variant by generating at-risk matings.

2. Materials and Methods

2.1. Sequencing Data

For variant calling, publicly available data of 88 ovine short read Illumina HiSeq whole genome sequences (WGS) from 14 breeds generated in various INRAE and Teagasc research projects were used. Among them, 24 WGS were obtained from dairy Lacaune sheep, and also genotyped with the OvineSNP50 Beadchip from Illumina in the framework of Lacaune dairy sheep genomic selection program [27]. A description of the different breeds and all the accession numbers of sequencing raw data are available in Supplementary Table S1.

2.2. WGS Variant Calling and Annotation

Reads mapping and variants calling were performed using Nextflow v20.10.0 and Sarek pipeline v2.6.1 [29]. Genome Analysis Toolkit (GATK) best practices were followed as implemented in the Sarek pipeline. In this project, the following steps were performed: aligning the reads with BWA v0.7.17-r1188 [30] against the ovine genome assembly Oar_rambouillet_v1.0 (GCF_002742125.1), marking duplicate reads (MarkDuplicates), base quality recalibration (BQSR) and calling germline small variants (HaplotypeCaller in GVCF mode) with GATK v4.1.7.0 [31], annotation of small variants with SnpEff v4.3t [32] and quality control with MultiQC v1.8 [33].

2.3. Identification of Candidate Causal Variants

Based on OvineSNP50 Beadchip genotyping, two Lacaune animals were detected as LDHH6 heterozygous carriers (others as non-carriers) among the 24 sequenced Lacaune genomes [28]. All SNPs, small insertion and deletion variants located within the LDHH6 region extended by 1 Mb from each side were extracted from OAR3 (Oar_rambouillet_v1.0; NC_040254.1:145,243,481-148,946,399pb) using SnpSift Filter [32]. The selection filter of candidate polymorphisms was as follows: (i) variant allele compared to the Rambouillet reference genome; (ii) in the heterozygote state in the 2 heterozygous LDHH6-carriers Lacaune sheep; and (iii) in the reference homozygous state in all LDHH6-non carriers Lacaune sheep and in other non-related breeds. These variants were checked manually using the Integrative Genomics Viewer (IGV) [34] to confirm the bioinformatics variant calling prior to further investigation.

2.4. Biological Samples and CCDC65-Specific Genotyping Assay

Blood samples (3 ml) were collected by jugular vein puncture with the Venoject system containing EDTA (Terumo, Tokyo, Japan) and directly stored at 4 °C or −20 °C depending on further use. Ear samples were obtained with a tissue sampling unit (TSU, Allflex Europe, Vitré, France) taking an ear punch (1 mm³) directly placed in the TSU storage buffer at 4 °C. Ear biopsies were placed twice consecutively in 180 µL of 50 mM NaOH, heated 10 min at 95 °C, neutralized with 20 µL of 1 M HCl, and then vortexed during 10 s. Part of the blood samples was used for extraction of genomic DNA as described in Bodin et al. [35]. All other samples were used for direct genotyping without DNA purification on whole blood or on neutralized NaOH treatment solution of ear biopsies [36].

Among the candidate polymorphisms within the LDHH6 region, the potential CCDC65 causal variant (Oar_rambouillet_v1.0, NC_040254.1:g.147,207,999C > A; XM_004006389.4:c.521G > T) was genotyped either by RFLP or PACE (PCR allele competitive extension) analysis. RFLP was resolved on 2% agarose gel using BsaJI restriction enzyme (New England Biolabs, Ipswich, MA, USA) after either a first step of Terra PCR Direct Polymerase Mix amplification (Takara Bio, Kusatsu, Japan) using 1 µL of total blood (or 5 µL ear biopsy solution), or a PCR using GoTaq Flexi DNA Polymerase (Promega, Madison, WI, USA) with 50 ng of purified DNA. The following amplification primers (forward 5'-GAGCTGCGTGTGTAAGATGA-3' and reverse 5'-CCTCCAGCTCCATGTTGTAA-

3') were designed using Primer3Plus software [37]. PCR for RFLP was performed on an ABI2720 thermocycler (Applied Biosystems, Waltham, MA, USA) with the following conditions: 5 min at 94 °C, 35 cycles of 30 s at 94 °C, 30 s at 58 °C and 30 s at 72 °C, followed by 5 min final extension at 72 °C. Fluorescent PACE analysis was done with 15 ng of purified DNA using the PACE-IR 2x Genotyping Master mix (3CR Bioscience) in the presence of 12 µM of a mix of extended allele specific forward primers (5'-GAAGGTGACCAAGTTCATGCTGGACCTGTCAGAAGCCGAGG-3' and 5'-GAAGGTGACCAAGTTCATGCTGGACCTGTCAGAAGCCGAGT-3') and 30 µM of common reverse primer (5'-AGGGCGTGGGCGTGCTGCT-3') in a final volume of 10 µL. The touch-down PCR amplification condition was 15 min at 94 °C for the hot-start activation, 10 cycles of 20 s at 94 °C, 54–62 °C for 60 s (dropping 0.8 °C per cycle), then 36 cycles of 20 s at 94 °C and 60 s at 54 °C performed on an ABI9700 thermocycler followed by a final point read of the fluorescence on an ABI QuantStudio 6 real-time PCR system and using the QuantStudio software 1.3 (Applied Biosystems). The accuracy of the genotyping was validated by Sanger sequencing on few samples.

The presence of the *CCDC65* variant was checked in a DNA set of the 2021 cohort of 2952 Lacaune male lamb candidates for genomic selection. DNA was extracted by Labogena (<http://www.labogena.fr>, accessed on 2 September 2021) for low-density SNP chip genotyping (SheepLD v.3, on behalf the dairy Lacaune breed industry) and LDHH6 status was established as previously described [28]. A DNA diversity panel of 872 animals from 25 French sheep breeds [38] and 8 Swiss sheep breeds [39] was also specifically genotyped for the *CCDC65* variant.

2.5. Generation of Homozygous Lambs

Dairy Lacaune ewes ($n = 245$) from two INRAE experimental flocks (Langlade and La Fage, agreement numbers: D3142901 and A312031, respectively) were genotyped for the *CCDC65* variant. Heterozygous ewes ($n = 17$) were artificially inseminated (AI) with fresh semen from heterozygous carrier rams ($n = 3$) selected among the genotyped animals for genomic selection with a known status at the LDHH6 haplotype [28] and further genotyped as heterozygous at the *CCDC65* locus. An ultrasound diagnosis of gestation was realized between 45 and 60 days after AI. Lambs were weighted at birth and 15 days after birth. Ear biopsies from newborn lambs (TSU Allflex) were collected for *CCDC65* genotyping at the same time of the mandatory electronic ear tag identification.

2.6. Ethics Statement

All experimental procedures were approved (approval numbers 01171.02 and 752056.00) 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.

3. Results

3.1. Screening of WGS Data Identifies a Nonsense Variant in *CCDC65* Gene Associated with LDHH6

To identify the putative causal variant hosted by the LDHH6 haplotype, we considered biallelic polymorphisms (SNPs and InDels) proven from 88 ovine WGS containing 24 Lacaune sequences and, among them, two LDHH6 heterozygous carriers. Variant search analysis and annotations were deliberately limited to the LDHH6 region (OAR3: 146,243,481–147,946,399 pb) extended by 1 Mb from each side. In this 3.7 Mb region, we detected 53,632 polymorphisms with a quality score > 30. After filtering, we identified 11 SNPs only present in a heterozygous state in the genome of the two LDHH6 heterozygous carriers (Table 1).

Table 1. Candidate SNPs, locations and functional annotations.

Position	Ref/Alt	Quality Score	Location Annotation	Functional Consequence ^a
145,928,967	C/G	347.3	Intergenic, downstream of <i>FAM186A</i>	Modifier
146,173,708	G/A	396.0	Intergenic, upstream of <i>ASIC1</i>	Modifier
146,566,556	C/T	133.1	Intronic, <i>FAM186B</i>	Modifier
146,718,479	G/A	272.7	Intronic, <i>SPATS2</i>	Modifier
146,809,812	T/C	604.4	Intronic, <i>DNAJC22</i>	Modifier
147,207,999	C/A	506.7	Exonic, <i>CCDC65</i> (c.521G > T)	High, stop-gain (p.Glu111*)
147,345,297	G/A	192.3	Intergenic, upstream of <i>TEX49</i>	Modifier
148,189,184	A/G	596.7	Intergenic	Modifier
148,212,194	A/C	181.7	Intergenic	Modifier
148,417,713	A/G	545.8	Intergenic, downstream of <i>ZNF641</i>	Modifier
148,904,267	G/A	439.1	Intronic, <i>HDAC7</i>	Modifier

^a variant annotation and effect predicted by SnpEff [32].

Among those SNPs, only one was predicted to highly alter gene function. This SNV (NC_040254.1:g.147,207,999C > A; XM_004006389.4:c.521G > T) leads to a nonsense variant located in exon 3 of the *Coiled-Coil Domain Containing 65* (*CCDC65*) gene (Figure 1a,b). The variant should result in a premature stop codon and a truncated protein in position 111 (XP_004006438.1:p.Glu111*) while the entire translated *CCDC65* protein is composed of 498 amino acids (Figure 1c). The variant localizes in the first Coiled-Coil (CC) domain in a highly conserved region of the protein between mammal species and even in algae (*Chlamydomonas reinhardtii*) where the role of *CCDC65* was originally studied [40–43] (Supplementary Figure S1). The resulting truncated protein has kept the major part of the NYD-SP28 domain shared with *CCDC164* also part of the N-DRC [44] but has lost two CC domains known to be implicated in protein–protein interaction [42,43].

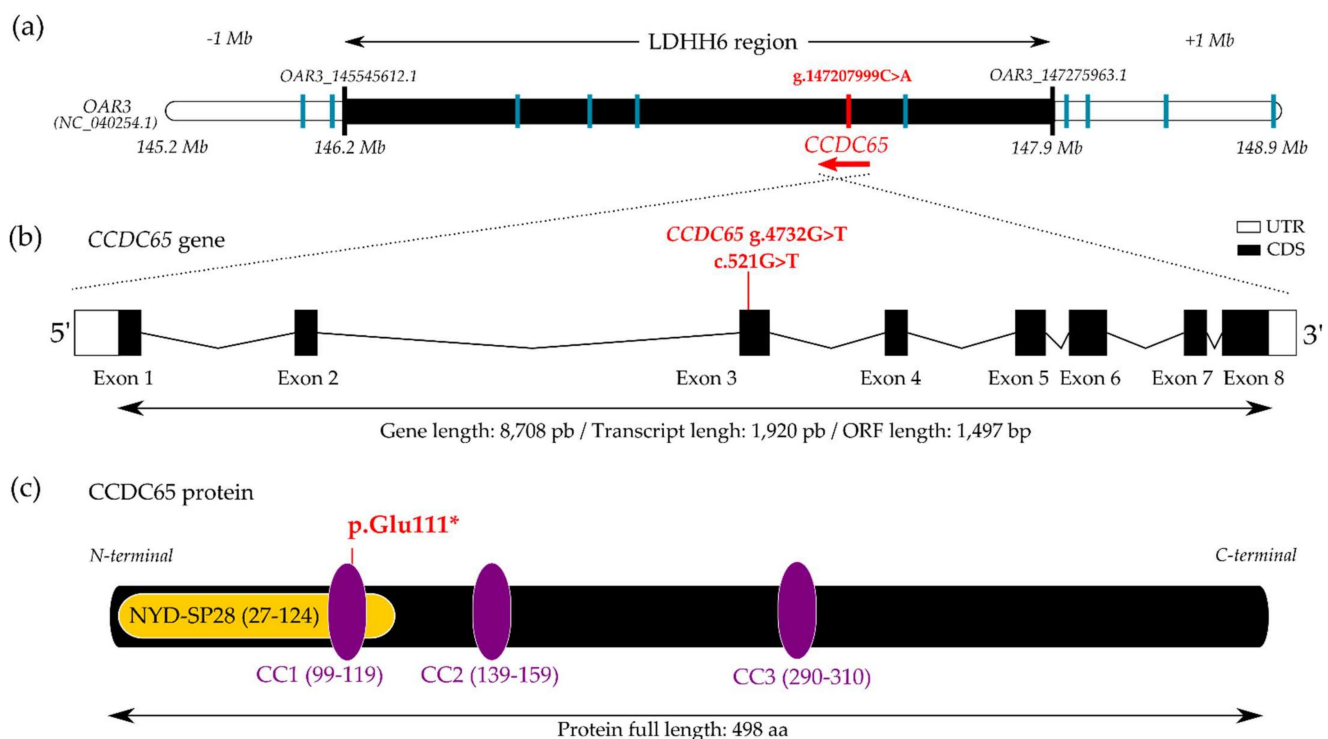


Figure 1. Nonsense variant in *CCDC65* gene associated with the LDHH6 lethal haplotype. (a) LDHH6 haplotype (NC_040254.1:OAR3:146,243,481pb–147,946,399pb) extended by 1 Mb from each side. The black bar indicates the limits of LDHH6 haplotype with the first and last markers (Illumina OvineSNP50 OAR3_145545612.1 and OAR3_147275963.1). Vertical lines indicate the positions of the candidate causal SNV in *CCDC65* (red) and 10 other SNVs fully associated with LDHH6 (blue);

(b) *CCDC65* gene structure and localization of the g.4732G > T (GeneID:101104220)/c.521G > T (XM_004006389.4) variant in the third exon (UTR: untranslated region, CDS: coding sequence); (c) schematic representation of *CCDC65* protein (XP_004006438.1) with one N-terminal 'NYD-SP28' (Pfam PF14772) and three coiled-coil (CC) domains (UniProtKB-W5QCQ9), and position of the premature C-terminal residue, pGlu111.

3.2. Variant Association with *LDHH6* and Population Estimation of Allele Frequency

In order to provide additional evidence in favor of the *CCDC65* polymorphism as the causal variant, we have genotyped the c.521G > T SNP in a cohort of 2952 Lacaune lambs with known status at the *LDHH6* locus (Supplementary Figure S2). The contingency table indicated a clear association between the *LDHH6* status and the nonsense variant in *CCDC65* (Table 2, Fischer's exact test $p < 0.0001$, without the homozygous carrier individual). Based on this genotyping, the c.521T variant allele frequency was calculated at 7%, and, as expected, the distribution of genotypes was not consistent with the Hardy–Weinberg equilibrium (Chi-square test, $p < 0.001$).

Table 2. Contingency table between *LDHH6* status and genotype at c.521G > T in *CCDC65*.

Genotype	+/+	LDHH6/+	LDHH6/LDHH6	Total
G/G	2540	3	0	2543
G/T	9	399	0	408
T/T	0	0	1	1
Total	2549	402	1	2952

+/+: non-carriers; LDHH6/+ : heterozygous carriers and LDHH6/LDHH6: homozygous carriers.

Since 2017, all Lacaune male lamb candidates for genomic selection were genotyped on low-density sheep (SheepLD chip) between one and five months of age, with a good representation of the genetic diversity in the selection schemes. In the Lacaune dairy breed, the population consists of two subpopulations with separate selection schemes conducted by two breeding companies. Overall, *LDHH6* heterozygous carrier frequency observed is stable in the population between 2017 and 2021. Nevertheless, the frequency of carriers was two-fold higher for breeding company 1 compared to breeding company 2 (Figure 2).

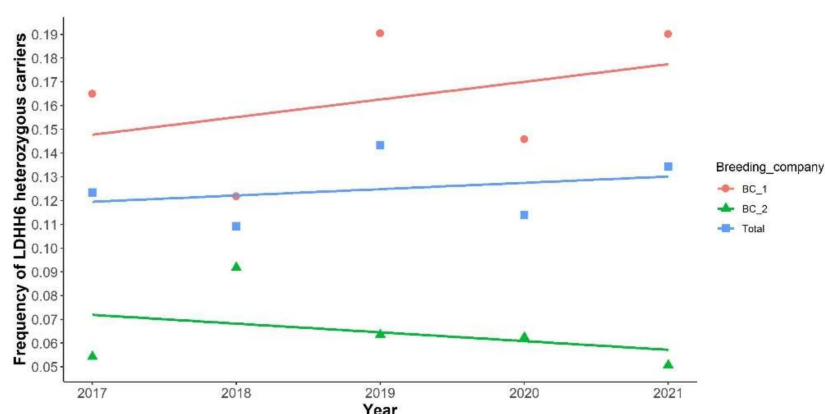


Figure 2. Evolution of *LDHH6* heterozygous carrier frequency between 2017 and 2021 in dairy Lacaune male lambs. The frequency is indicated either for all candidates to genomic selection (Total) or depending on the breeding company (BC_1 or 2).

3.3. Generation of At-Risk Matings to Obtain Homozygote Lambs

The genotyping of the *CCDC65* variant in two experimental flocks of dairy Lacaune ewes ($n = 245$) enabled us to identify 17 heterozygous carriers. These ewes were inseminated with fresh semen from heterozygous carrier rams to generate at-risk matings. Forty-five days after AI, 11 ewes were diagnosed as pregnant by ultrasonography. This corresponded to an AI success of 65%, in line with the average AI success of 69% recorded in the whole

dairy Lacaune population [28]. Gestation length was recorded as normal (146 ± 2 days), and no abortion or stillborn was observed. At the end of gestation, 16 lambs were born with 7 males and 9 females. For each lamb, an ear punch was collected for genotyping of the c.521G > T *CCDC65* variant. Five lambs were genotyped homozygous non-carriers, 6 were heterozygous and 5 were homozygous carriers, consistent with the Hardy–Weinberg equilibrium (Chi-square test, $p = 0.317$). All lambs were weighted at birth (males: 4.9 ± 1.0 kg, females: 3.9 ± 0.6 kg), and no significant difference of birthweight was observed between genotypes. A second weighing was carried out at day 15 and average daily gain (ADG) was calculated on the 0–15 day period. Based on Wilcoxon’s non-parametric test, the homozygous carrier (T/T) showed a significant lower ADG compared to the other lambs (G/G or G/T) (Figure 3).

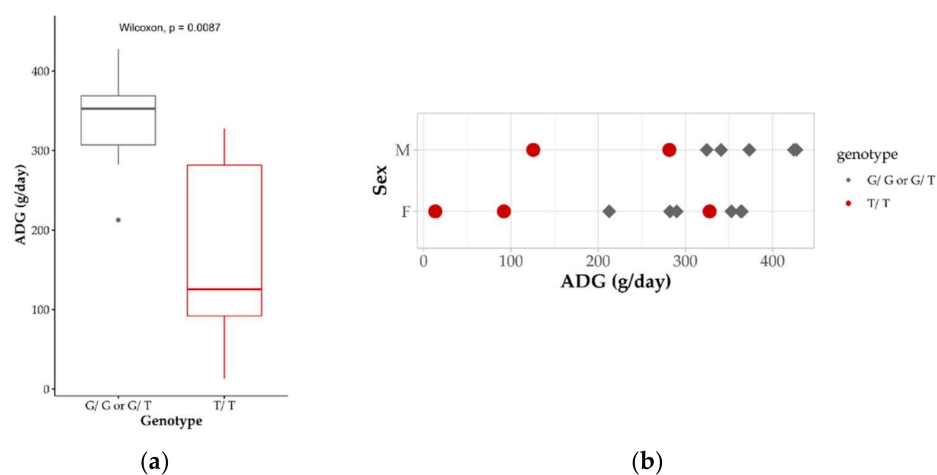


Figure 3. *CCDC65* genotype effect on average daily gain during the 0–15 day period. Mean ADG (g/day) according to *CCDC65* c.521G > T genotype (a) and ADG distribution by sex (M: male; F: female) and genotype (b).

Clinical examination revealed respiratory problems, such as tachypnea and a runny nose for 4 over 5 T/T lambs from the first days of life. Despite an appropriate veterinary treatment for suspected pulmonary infectious diseases (based on glucocorticoids and antibiotics administration), the recurrent respiratory distress resulted in the declining of general body condition (as attested by the lower ADG) leading to the natural death or euthanasia of the affected lambs between 15 and 25 days after birth. Only one T/T female lamb with the highest ADG (327.8 g/day during the 0–15 day period) has exceeded the weaning age with light respiratory syndrome and was sold to a sheep fattener.

Additionally, the previous analysis of the genomic selection cohort has allowed to identify a homozygous animal for the *CCDC65* variant (Table 2). This male lamb was present in a breeding center, and we obtained the information that he received an adapted veterinary treatment several times due to respiratory problems. The lamb finally died at the age of five and a half months. Overall, these observations have confirmed the hypothesis of a deleterious recessive *CCDC65* variant almost perfectly associated with LDHH6.

An autopsy was performed on the four deceased T/T lambs. A broad hepatization of lungs probably linked to an infectious pneumonia was particularly observed (Figure 4).

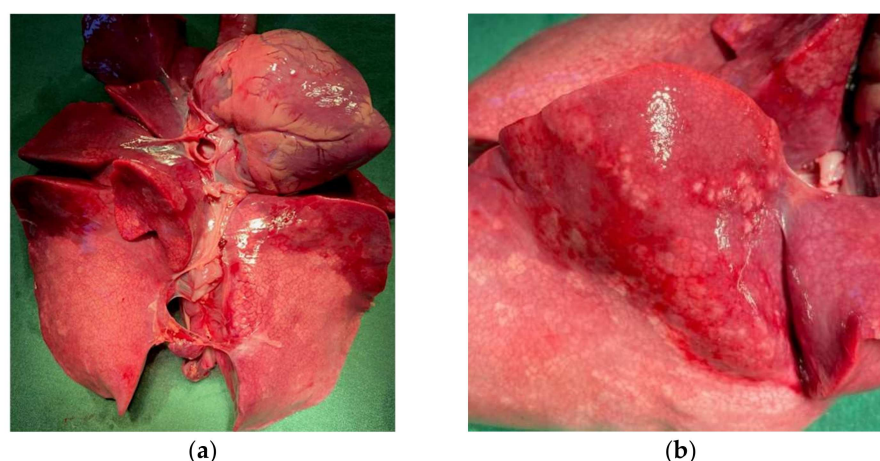


Figure 4. *CCDC65* homozygous affected lamb with lung lesions. (a) heart-pulmonary system of an affected homozygous lamb; (b) magnification of hepatized lobes.

3.4. Occurrence of the *CCDC65* Variant in Further Populations

An ovine DNA diversity panel of 25 French and eight Swiss sheep breeds representing 872 animals was checked for the segregation of the c.521G > T variant in *CCDC65* (Table 3). Initially shown in the French Lacaune dairy sheep, *CCDC65* genotyping has also revealed its occurrence in the French Lacaune meat strain and in a Lacaune dairy population reared in Switzerland. Moreover, two heterozygous animals were shown in the Blanche du Massif Central (BMC) breed. All the other tested French and Swiss breeds were non-carriers of the T allele.

Table 3. *CCDC65* c.521G > T genotype distribution from a DNA diversity panel of French (FR) and Swiss (CH) ovine breeds.

Breed	Total	Genotype		Breed	Total	Genotype	
		G/G	G/T			G/G	G/T
Berrichon du Cher (FR)	30	30		Mouton Vendéen (FR)	30	30	
Blanche du Massif Central (FR)	31	29	2	Noir du Velay (FR)	28	28	
Causse du Lot (FR)	32	32		Préalpes du sud (FR)	27	27	
Charmoise (FR)	31	31		Rava (FR)	29	29	
Charollais (FR)	30	30		Romane (FR)	30	30	
Corse (FR)	30	30		Romanov (FR)	26	26	
East Friesian (CH)	18	18		Rouge de l'Ouest (FR)	30	30	
Engadine Red (CH)	3	3		Roussin (FR)	30	30	
Ile de France (FR)	28	28		Saaser Mutte (CH)	10	10	
Lacaune (meat) (FR)	45	43	2	Suffolk (FR)	29	29	
Lacaune (milk) (FR)	40	37	3	Swiss Mirror (CH)	11	11	
Lacaune (milk) (CH)	37	27	10	Swiss White Alpine (CH)	14	14	
Limousine (FR)	30	30		Tarasconnaise (FR)	33	33	
Manech tête rousse (FR)	29	29		Texel (FR)	27	27	
Martinik (FR)	23	23		Valais Blacknose (CH)	14	14	
Merinos d'Arles (FR)	27	27		Valais Red (CH)	13	13	
Mourerous (FR)	27	27					
				Total	872	855	17

4. Discussion

We have previously reported that LDHH6 [OMIA 002342-9940] was the most significant haplotype in deficit of homozygous animals in French Lacaune dairy sheep. Analyses based on at-risk matings between LDHH6 heterozygous carriers have shown an increase in stillbirth rate, suggesting that this haplotype harbored a harmful recessive mutation [28]. Using sheep whole-genome sequences, the present study identified a causal variant as

the g.147,207,999C > A substitution on ovine chromosome 3 linked to the LDHH6 haplotype, and this SNV corresponds to the c.521G > T nonsense variation in the *CCDC65* gene introducing a premature stop codon (p.Glu111*).

In the LDHH6 region, we have previously highlighted three candidate genes, *WNT1*, *CCDC65*, and *PFKM*, whose knock-out models in mice fitted well with perinatal, neonatal or preweaning lethality [28]. Interestingly, variant analysis from the whole genome sequence of LDHH6 heterozygous carriers has revealed candidate variants in ovine *CCDC65*, but not in *WNT1* or *PFKM* or in close vicinity of these two genes. Moreover, among all candidate polymorphisms, the SNP detected in the exon 3 of *CCDC65* was the most obvious candidate since it was the only one predicted to strongly alter the protein function by creating a premature stop codon allowing to synthesize only the first N-terminal quarter of the protein.

In humans, pathogenic variants in *CCDC65* cause Primary Cilia Dyskinesia (PCD), a genetic disorder with autosomal recessive inheritance (CILD27, OMIM#615504). The protein is involved in the assembly of the axonemal nexin-dynein regulatory complex [45]. Axoneme is the axial motor part of cilia composed of nine doublets of microtubules located in the periphery and one central pair complex [40]. Each outer doublet is composed of two tubules (A and B) and is associated with several complexes: outer and inner dynein arms (ODAs and IDs), Nexin-Dynein Regulatory Complex (N-DRC) and Radial Spoke (RS) [40–43,46]. In particular, the N-DRC is important for the sliding between adjacent outer doublet microtubules to allow cilia motility. The complex is composed of at least 11 subunits shown in *C. reinhardtii* and conserved among mammals [40–43]. Among them, variants in one of the three genes: *CCDC164* (OMIM #615294 [47]), *CCDC65* (OMIM #615504 [44,48]) or *GAS8* (OMIM #616726 [49]) have been reported to cause PCD. Bower et al. have demonstrated that the absence of *CCDC65* subunit destabilized the association with *CCDC164* and *GAS8* [45]. Indeed, *CCDC65* forms a base plate for N-DRC. Alteration in N-DRC have big consequences for cilia beating leading to the respiratory problems. These three subunits are important for cilia and flagella motility and mucus clearance to eject pathogenic organisms from respiratory tracts. In human, loss-of-function mutations in *CCDC65* genes in PCD patients have been identified and are associated with recurrent infectious diseases of the ENT (ear nose throat) sphere as bronchitis, sinusitis, and/or otitis [44,48]. The above clinical profiles are well in line with our observations made in lambs homozygous for the c.521G > T variant in *CCDC65*. Homozygous lambs have recurrent neonatal respiratory problems. At the autopsy, we particularly observed a broad hepatization of lung lobes resembling infectious pneumonia also associated with growth delay [50]. We hypothesize that, due to the loss of Coiled-Coil domains, the truncated *CCDC65* leads to the destabilization of the N-DRC, abnormal cilia beating and impairs the mucociliary clearance resulting in airways obstruction by mucus loaded with pathogens (bacteria, mycoplasma, virus). Pathogens load could depend on the breeding sanitary conditions, individual immunity and veterinary treatment applied, which can explain that the lethality can occur between perinatal and juvenile stages. This is to be connected to the previous observation of a partial homozygous deficit for LDHH6 at the age of genotyping (72 expected and 3 observed) [28]. Indeed, Lacaune candidate rams for genomic selection are genotyped on average at 3 months of age using LD SNP chip indicating that some LDHH6/*CCDC65* homozygous lambs have exceeded this age as observed for a female born from our at-risk mating, and a male in a breeding center exceeding 5 months of age.

Interestingly, a respiratory syndrome is also reported in cattle for the homozygous deficient Braunvieh haplotype 2 (BH2) [OMIA 001939-9913] associated with perinatal and juvenile mortality due to the missense variant (p.His210Arg) in *TUBD1* (*tubulin delta 1*) that also disorganizes the microtubules in airway cilia [26]. Additionally, a stop-gain variant in *CCDC39* (p.Arg96*) [OMIA 001540-9615], also essential for the assembly of inner dynein arms [51,52], was found in Bobtail dogs associated with PCD suffering from chronic airway diseases [51].

Based on the sheep gene atlas (<http://biogps.org/sheepatlas/>; accessed 2 September 2021), *CCDC65* is highly expressed in tissues and organs hosting ciliated or flagellated cells

as lungs, fallopian tube and testis [53]. The high expression in testis raises the question on the impact of the *CCDC65* variant on sperm motility, knowing that the first characterization of the protein was made in human sperm tail [54] and several mutations causing PCD lead to sperm defects [55]. The three homozygous male lambs we had access to died before puberty, preventing us from getting the semen. We then evaluated the impact of the mutation at the heterozygous state based on semen quality records from artificial insemination center. Semen quality is controlled for each ejaculate according to volume, concentration and sperm motility [56]. However, no significant effect of the mutation was demonstrated for the parameters recorded (Supplementary Table S2).

By genotyping the complete 2021 cohort of candidate lambs for genomic selection, we have validated an almost perfect association between LDHH6 status and alleles at the *CCDC65* variant (Table 2). Only 12 among 2952 animals do not match with the expectations (described in Supplementary Figure S3). Indeed, nine LDHH6 non-carriers were heterozygous carriers for the variant allele. Specific focus on the haplotypes of these animals have revealed shorter recombinant versions of the LDHH6 haplotypes (between 4 and 26 markers out of 27 markers composing the LDHH6 haplotype). We also identify three LDHH6 heterozygous carriers which do not carry the *CCDC65* variant. This discrepancy could be attributed to errors from SNP array genotyping, phasing and/or imputation.

Diversity analysis revealed the segregation of the *CCDC65* c.521G > T variant not only in the Lacaune dairy population but also in the Lacaune meat and Blanche du Massif Central (BMC) populations. Additionally, we found the *CCDC65* variant in the Ensembl variant database (rs1085624756) compiling the whole genome sequences from the International Sheep Genome Consortium (448 animals from 58 breeds all over the world). The rs1085624756 variant is only present in a Lacaune dairy animal (LRLACU000000000084). The Lacaune sheep population has a complex history with the creation of two lines, one for meat and one for dairy purposes, with four independent selection schemes depending on two breeding companies. Intriguingly, the LDHH6 carrier frequency is 2.5 fold higher in one breeding company compared to the other one (0.16 vs. 0.06). In livestock, deleterious alleles could be associated with better performances when heterozygous to explain their maintaining in the populations [7,57,58]. However, no heterozygous advantage has been identified on milk selected traits in the Lacaune population for LDHH6 heterozygous carriers [28], and the two breeding companies have similar selection objectives on milk traits. Thus, the difference observed in the *CCDC65* allele frequency between the two breeding companies could be explained by genetic drift and the use of very influential carriers. The segregation of the *CCDC65* variant in both dairy and meat strains of Lacaune sheep indicates that the mutation event predates the creation of the two lineages. Moreover, the variant is also segregating in the BMC breed. Population structure analyses revealed that Lacaune and BMC, both originating from the Massif Central in France, shared the same origin within the Southern European sheep populations that may explain the segregation of identical alleles in both populations [38]. Interestingly, these two breeds also share another important variant in the *BMP15* gene controlling prolificacy [36]. The presence of the variant in Swiss Lacaune could be explained by imports of French Lacaune reproducers. However, with the number of carriers detected in our analysis (10 over 37), it is important to check the mutation frequency in the overall Swiss Lacaune population because the risk of generating homozygous lambs is quite high.

5. Conclusions

This is the first report of the identification of a recessive lethal variant in sheep through a reverse genetic screen method using highthroughput genotyping completed by whole genome sequencing of few informative animals. The present study identifies a causal recessive nonsense variant in *CCDC65* gene linked to the homozygous deficient haplotype named LDHH6 and associated with rearing loss in Lacaune dairy sheep. The loss-of-function mutation in *CCDC65* is likely to cause a *primary cilia dyskinesia* syndrome similar to *CCDC65*-related ciliopathies in humans. Homozygous affected lambs suffer from res-

piratory defects predisposing infectious pneumonia with a fatal ending at a young age. Based on the LDHH6/CCDC65 allele frequency of around 0.06 and a mortality lamb rate at 15% (recorded on 804,577 matings, described in [28]), we estimate that 1/50 (2%) of dead lambs is homozygous for the CCDC65 variant. Consequently, management of this causal variant in the Lacaune sheep selection schemes through reasoned mating of carrier rams and ewes is important to improve health, welfare and overall lamb viability.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes13010045/s1>, Table S1: Accession numbers of the 88 whole-genome sequences used in the analysis, Table S2. Semen parameters of dairy Lacaune rams from artificial insemination center, Figure S1. CCDC65 protein multiple sequence alignment, Figure S2. Representative genotyping results for c.521G > T variant in CCDC65, Figure S3. Haplotypes of the LDHH6 region for 12 animals showing mismatch with the CCDC65 SNV.

Author Contributions: Conceptualization, M.B.B., S.F. and C.M.-R.; methodology, M.B.B., S.F. and C.M.-R.; software, P.B.; formal analysis, M.B.B., S.F., C.M.-R. and P.B.; investigation, M.B.B., S.F., C.M.-R., A.B., F.D., F.P.-P., J.S., N.T., F.W., C.A., D.P. and C.D.; resources, M.B.B., S.F., C.M.-R., A.B., F.D., N.T., C.A., D.P., P.B. and C.D.; writing—original draft preparation, M.B.B.; writing—review and editing, S.F., C.M.-R., P.B. and C.D.; visualization, M.B.B. and S.F.; supervision, S.F. and C.M.-R.; funding acquisition, S.F. and C.M.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research has received funding from the European Union’s Horizon 2020 research and innovation program under the Grant Agreement No. 772787 (SMARTER) and PRESAGE project (CASDAR No. 20ART1532777, the responsibility of the French ministry of Agriculture and Food cannot be engaged). M.B.B. was supported by a Ph.D. grant for the HOMLET program co-funded by APIS-GENE and Région Occitanie.

Institutional Review Board Statement: Experimental animals were reared in two INRAE farms with agreement numbers D3142901 (Langlade) and A312031 (La Fage) delivered by the French Departmental Directorate of Population Protections 31. All experimental procedures on animals were approved (approval numbers 01171.02 and 752056.00) 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.

Data Availability Statement: The WGS data used in this study are publicly available, accession numbers are described in Table S1.

Acknowledgments: We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing help and/or computing and/or storage resources. The authors thank the Lacaune breeding companies who provided access to private genomic data and biological samples.

Conflicts of Interest: C.D. is the guest editor of this special issue of *Genes (Genetics and Breeding of Small Ruminants)* but has not in any way been involved in or interacted with the journal’s review process or editorial decision-making. The authors declare no other conflict of interest.

References

1. Dwyer, C.M.; Conington, J.; Corbiere, F.; Holmøy, I.H.; Muri, K.; Nowak, R.; Rooke, J.; Vipond, J.; Gautier, J.-M. Invited Review: Improving Neonatal Survival in Small Ruminants: Science into Practice. *Animal* **2016**, *10*, 449–459. [[CrossRef](#)] [[PubMed](#)]
2. Dwyer, C.M. Genetic and Physiological Determinants of Maternal Behavior and Lamb Survival: Implications for Low-Input Sheep Management. *J. Anim. Sci.* **2008**, *86*, E246–E258. [[CrossRef](#)] [[PubMed](#)]
3. Flinn, T.; Kleemann, D.O.; Swinbourne, A.M.; Kelly, J.M.; Weaver, A.C.; Walker, S.K.; Gatford, K.L.; Kind, K.L.; van Wettere, W.H.E.J. Neonatal Lamb Mortality: Major Risk Factors and the Potential Ameliorative Role of Melatonin. *J. Anim. Sci. Biotechnol.* **2020**, *11*, 107. [[CrossRef](#)]
4. Cloete, S.W.P.; Misztal, I.; Olivier, J.J. Genetic Parameters and Trends for Lamb Survival and Birth Weight in a Merino Flock Divergently Selected for Multiple Rearing Ability. *J. Anim. Sci.* **2009**, *87*, 2196–2208. [[CrossRef](#)] [[PubMed](#)]
5. Hatcher, S.; Atkins, K.D.; Safari, E. Lamb Survival in Australian Merino Sheep: A Genetic Analysis. *J. Anim. Sci.* **2010**, *88*, 3198–3205. [[CrossRef](#)]

6. Riggio, V.; Finocchiaro, R.; Bishop, S.C. Genetic Parameters for Early Lamb Survival and Growth in Scottish Blackface Sheep. *J. Anim. Sci.* **2008**, *86*, 1758–1764. [[CrossRef](#)] [[PubMed](#)]
7. Cole, J.B.; Null, D.J.; VanRaden, P.M. Phenotypic and Genetic Effects of Recessive Haplotypes on Yield, Longevity, and Fertility. *J. Dairy Sci.* **2016**, *99*, 7274–7288. [[CrossRef](#)]
8. Charlier, C.; Coppieters, W.; Rollin, F.; Desmecht, D.; Agerholm, J.S.; Cambisano, N.; Carta, E.; Dardano, S.; Dive, M.; Fasquelle, C.; et al. Highly Effective SNP-Based Association Mapping and Management of Recessive Defects in Livestock. *Nat. Genet.* **2008**, *40*, 449–454. [[CrossRef](#)] [[PubMed](#)]
9. Ytournal, F.; Gilbert, H.; Boichard, D. Comment affiner la localisation d'un QTL ? *INRAE Prod. Anim.* **2008**, *21*, 147–158. [[CrossRef](#)]
10. Georges, M.; Charlier, C.; Hayes, B. Harnessing Genomic Information for Livestock Improvement. *Nat. Rev. Genet.* **2019**, *20*, 135–156. [[CrossRef](#)]
11. Becker, D.; Tetens, J.; Brunner, A.; Bürstel, D.; Ganter, M.; Kijas, J.; Drögemüller, C. Microphthalmia in Texel Sheep Is Associated with a Missense Mutation in the Paired-Like Homeodomain 3 (PITX3) Gene. *PLoS ONE* **2010**, *5*, e8689. [[CrossRef](#)] [[PubMed](#)]
12. Michot, P.; Fantini, O.; Braque, R.; Allais-Bonnet, A.; Saintilan, R.; Grohs, C.; Barbieri, J.; Genestout, L.; Danchin-Burge, C.; Gourreau, J.-M.; et al. Whole-Genome Sequencing Identifies a Homozygous Deletion Encompassing Exons 17 to 23 of the Integrin Beta 4 Gene in a Charolais Calf with Junctional Epidermolysis Bullosa. *Genet. Sel. Evol.* **2015**, *47*, 37. [[CrossRef](#)] [[PubMed](#)]
13. Peters, M.; Reber, I.; Jagannathan, V.; Raddatz, B.; Wohlsein, P.; Drögemüller, C. DNA-Based Diagnosis of Rare Diseases in Veterinary Medicine: A 4.4 Kb Deletion of ITGB4 Is Associated with Epidermolysis Bullosa in Charolais Cattle. *BMC Vet. Res.* **2015**, *11*, 48. [[CrossRef](#)] [[PubMed](#)]
14. Fabre, S.; Chantepie, L.; Plisson-Petit, F.; Sarry, J.; Woloszyn, F.; Genet, C.; Drouilhet, L.; Tosser-Klopp, G. A Novel Homozygous Nonsense Mutation in ITGB4 Gene Causes Epidermolysis Bullosa in Mouton Vendéen Sheep. *Anim. Genet.* **2020**, *52*, 138–139. [[CrossRef](#)] [[PubMed](#)]
15. Suárez-Vega, A.; Gutiérrez-Gil, B.; Benavides, J.; Perez, V.; Tosser-Klopp, G.; Klopp, C.; Keenel, S.J.; Arranz, J.J. Combining GWAS and RNA-Seq Approaches for Detection of the Causal Mutation for Hereditary Junctional Epidermolysis Bullosa in Sheep. *PLoS ONE* **2015**, *10*, e0126416. [[CrossRef](#)]
16. Letko, A.; Dijkman, R.; Strugnell, B.; Häfliger, I.M.; Paris, J.M.; Henderson, K.; Geraghty, T.; Orr, H.; Scholes, S.; Drögemüller, C. Deleterious AGXT Missense Variant Associated with Type 1 Primary Hyperoxaluria (PH1) in Zwartbles Sheep. *Genes* **2020**, *11*, 1147. [[CrossRef](#)] [[PubMed](#)]
17. VanRaden, P.M.; Olson, K.M.; Null, D.J.; Hutchison, J.L. Harmful Recessive Effects on Fertility Detected by Absence of Homozygous Haplotypes. *J. Dairy Sci.* **2011**, *94*, 6153–6161. [[CrossRef](#)] [[PubMed](#)]
18. Fritz, S.; Capitan, A.; Djari, A.; Rodriguez, S.C.; Barbat, A.; Baur, A.; Grohs, C.; Weiss, B.; Boussaha, M.; Esquerré, D.; et al. Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2. *PLoS ONE* **2013**, *8*, e65550. [[CrossRef](#)]
19. Adams, H.A.; Sonstegard, T.S.; VanRaden, P.M.; Null, D.J.; Van Tassell, C.P.; Larkin, D.M.; Lewin, H.A. Identification of a Nonsense Mutation in APAF1 That Is Likely Causal for a Decrease in Reproductive Efficiency in Holstein Dairy Cattle. *J. Dairy Sci.* **2016**, *99*, 6693–6701. [[CrossRef](#)] [[PubMed](#)]
20. Daetwyler, H.D.; Capitan, A.; Pausch, H.; Stothard, P.; van Binsbergen, R.; Brøndum, R.F.; Liao, X.; Djari, A.; Rodriguez, S.C.; Grohs, C.; et al. Whole-Genome Sequencing of 234 Bulls Facilitates Mapping of Monogenic and Complex Traits in Cattle. *Nat. Genet.* **2014**, *46*, 858–865. [[CrossRef](#)]
21. McClure, M.C.; Bickhart, D.; Null, D.; VanRaden, P.; Xu, L.; Wiggins, G.; Liu, G.; Schroeder, S.; Glasscock, J.; Armstrong, J.; et al. Bovine Exome Sequence Analysis and Targeted SNP Genotyping of Recessive Fertility Defects BH1, HH2, and HH3 Reveal a Putative Causative Mutation in SMC2 for HH3. *PLoS ONE* **2014**, *9*, e92769. [[CrossRef](#)]
22. Schütz, E.; Wehrhahn, C.; Wanjek, M.; Bortfeld, R.; Wemheuer, W.E.; Beck, J.; Brenig, B. The Holstein Friesian Lethal Haplotype 5 (HH5) Results from a Complete Deletion of TBF1M and Cholesterol Deficiency (CDH) from an ERV-(LTR) Insertion into the Coding Region of APOB. *PLoS ONE* **2016**, *11*, e0154602. [[CrossRef](#)]
23. Fritz, S.; Hozé, C.; Rebours, E.; Barbat, A.; Bizard, M.; Chamberlain, A.; Escouflaire, C.; Jagt, C.V.; Boussaha, M.; Grohs, C.; et al. An Initiator Codon Mutation in SDE2 Causes Recessive Embryonic Lethality in Holstein Cattle. *J. Dairy Sci.* **2018**, *101*, 6220–6231. [[CrossRef](#)]
24. Hozé, C.; Escouflaire, C.; Mesbah-Uddin, M.; Barbat, A.; Boussaha, M.; Deloche, M.C.; Boichard, D.; Fritz, S.; Capitan, A. Short Communication: A Splice Site Mutation in CENPU Is Associated with Recessive Embryonic Lethality in Holstein Cattle. *J. Dairy Sci.* **2020**, *103*, 607–612. [[CrossRef](#)] [[PubMed](#)]
25. Sonstegard, T.S.; Cole, J.B.; VanRaden, P.M.; Van Tassell, C.P.; Null, D.J.; Schroeder, S.G.; Bickhart, D.; McClure, M.C. Identification of a Nonsense Mutation in CWC15 Associated with Decreased Reproductive Efficiency in Jersey Cattle. *PLoS ONE* **2013**, *8*, e54872. [[CrossRef](#)]
26. Schwarzenbacher, H.; Burgstaller, J.; Seefried, F.R.; Wurmser, C.; Hilbe, M.; Jung, S.; Fuerst, C.; Dinhopf, N.; Weissenböck, H.; Fuerst-Waltl, B.; et al. A Missense Mutation in TUBD1 Is Associated with High Juvenile Mortality in Braunvieh and Fleckvieh Cattle. *BMC Genom.* **2016**, *17*, 400. [[CrossRef](#)] [[PubMed](#)]
27. Astruc, J.-M.; Baloche, G.; Buisson, D.; Labatut, J.; Lagriffoul, G.; Larroque, H.; Robert-Granie, C.; Legarra, A.; Barillet, F. Genomic selection in French dairy sheep. *INRAE Prod. Anim.* **2016**, *29*, 41–55.

28. Ben Braiek, M.; Fabre, S.; Hozé, C.; Astruc, J.-M.; Moreno-Romieux, C. Identification of Homozygous Haplotypes Carrying Putative Recessive Lethal Mutations That Compromise Fertility Traits in French Lacaune Dairy Sheep. *Genet. Sel. Evol.* **2021**, *53*, 41. [[CrossRef](#)]
29. Garcia, M.; Juhos, S.; Larsson, M.; Olason, P.I.; Martin, M.; Eisfeldt, J.; DiLorenzo, S.; Sandgren, J.; Ståhl, T.D.D.; Ewels, P.; et al. Sarek: A Portable Workflow for Whole-Genome Sequencing Analysis of Germline and Somatic Variants. *F1000Research* **2020**, *9*, 63. [[CrossRef](#)]
30. Li, H.; Durbin, R. Fast and Accurate Long-Read Alignment with Burrows–Wheeler Transform. *Bioinformatics* **2010**, *26*, 589–595. [[CrossRef](#)]
31. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)]
32. Cingolani, P.; Patel, V.M.; Coon, M.; Nguyen, T.; Land, S.J.; Ruden, D.M.; Lu, X. Using *Drosophila Melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet.* **2012**, *3*, 35. [[CrossRef](#)] [[PubMed](#)]
33. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics* **2016**, *32*, 3047–3048. [[CrossRef](#)] [[PubMed](#)]
34. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration. *Brief. Bioinform.* **2013**, *14*, 178–192. [[CrossRef](#)]
35. Bodin, L.; Di Pasquale, E.; Fabre, S.; Bontoux, M.; Monget, P.; Persani, L.; Mulsant, P. 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. [[CrossRef](#)] [[PubMed](#)]
36. Chantepie, L.; Bodin, L.; Sarry, J.; Woloszyn, F.; Plisson-Petit, F.; Ruesche, J.; Drouilhet, L.; Fabre, S. Genome-Wide Identification of a Regulatory Mutation in BMP15 Controlling Prolificacy in Sheep. *Front. Genet.* **2020**, *11*, 585. [[CrossRef](#)]
37. Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J.A.M. Primer3Plus, an Enhanced Web Interface to Primer3. *Nucleic Acids Res.* **2007**, *35*, W71–W74. [[CrossRef](#)]
38. Rochus, C.M.; Tortereau, F.; Plisson-Petit, F.; Restoux, G.; Moreno-Romieux, C.; Tosser-Klopp, G.; Servin, B. Revealing the Selection History of Adaptive Loci Using Genome-Wide Scans for Selection: An Example from Domestic Sheep. *BMC Genom.* **2018**, *19*, 71. [[CrossRef](#)]
39. Signer-Hasler, H.; Burren, A.; Ammann, P.; Drögemüller, C.; Flury, C. Runs of Homozygosity and Signatures of Selection: A Comparison among Eight Local Swiss Sheep Breeds. *Anim. Genet.* **2019**, *50*, 512–525. [[CrossRef](#)] [[PubMed](#)]
40. Heuser, T.; Raytchev, M.; Krell, J.; Porter, M.E.; Nicastro, D. The Dynein Regulatory Complex Is the Nexin Link and a Major Regulatory Node in Cilia and Flagella. *J. Cell Biol.* **2009**, *187*, 921–933. [[CrossRef](#)] [[PubMed](#)]
41. Gui, L.; Song, K.; Tritschler, D.; Bower, R.; Yan, S.; Dai, A.; Augspurger, K.; Sakizadeh, J.; Grzemska, M.; Ni, T.; et al. Scaffold Subunits Support Associated Subunit Assembly in the *Chlamydomonas* Ciliary Nexin–Dynein Regulatory Complex. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 23152–23162. [[CrossRef](#)]
42. Lin, J.; Tritschler, D.; Song, K.; Barber, C.F.; Cobb, J.S.; Porter, M.E.; Nicastro, D. Building Blocks of the Nexin–Dynein Regulatory Complex in *Chlamydomonas* Flagella. *J. Biol. Chem.* **2011**, *286*, 29175–29191. [[CrossRef](#)] [[PubMed](#)]
43. Bower, R.; Tritschler, D.; VanderWaal, K.; Perrone, C.A.; Mueller, J.; Fox, L.; Sale, W.S.; Porter, M.E. The N-DRC Forms a Conserved Biochemical Complex That Maintains Outer Doublet Alignment and Limits Microtubule Sliding in Motile Axonemes. *Mol. Biol. Cell* **2013**, *24*, 1134–1152. [[CrossRef](#)]
44. Austin-Tse, C.; Halbritter, J.; Zariwala, M.A.; Gilberti, R.M.; Gee, H.Y.; Hellman, N.; Pathak, N.; Liu, Y.; Panizzi, J.R.; Patel-King, R.S.; et al. Zebrafish Ciliopathy Screen Plus Human Mutational Analysis Identifies C21orf59 and CCDC65 Defects as Causing Primary Ciliary Dyskinesia. *Am. J. Hum. Genet.* **2013**, *93*, 672–686. [[CrossRef](#)] [[PubMed](#)]
45. Bower, R.; Tritschler, D.; Mills, K.V.; Heuser, T.; Nicastro, D.; Porter, M.E. DRC2/CCDC65 Is a Central Hub for Assembly of the Nexin–Dynein Regulatory Complex and Other Regulators of Ciliary and Flagellar Motility. *Mol. Biol. Cell* **2018**, *29*, 137–153. [[CrossRef](#)]
46. Morohoshi, A.; Miyata, H.; Shimada, K.; Nozawa, K.; Matsumura, T.; Yanase, R.; Shiba, K.; Inaba, K.; Ikawa, M. Nexin–Dynein Regulatory Complex Component DRC7 but Not FBXL13 Is Required for Sperm Flagellum Formation and Male Fertility in Mice. *PLoS Genet.* **2020**, *16*, e1008585. [[CrossRef](#)] [[PubMed](#)]
47. Wirschell, M.; Olbrich, H.; Werner, C.; Tritschler, D.; Bower, R.; Sale, W.S.; Loges, N.T.; Pennekamp, P.; Lindberg, S.; Stenram, U.; et al. The Nexin–Dynein Regulatory Complex Subunit DRC1 Is Essential for Motile Cilia Function in Algae and Humans. *Nat. Genet.* **2013**, *45*, 262–268. [[CrossRef](#)]
48. Horani, A.; Brody, S.L.; Ferkol, T.W.; Shoseyov, D.; Wasserman, M.G.; Ta-shma, A.; Wilson, K.S.; Bayly, P.V.; Amirav, I.; Cohen-Cymberek, M.; et al. CCDC65 Mutation Causes Primary Ciliary Dyskinesia with Normal Ultrastructure and Hyperkinetic Cilia. *PLoS ONE* **2013**, *8*, e72299. [[CrossRef](#)] [[PubMed](#)]
49. Olbrich, H.; Cremers, C.; Loges, N.T.; Werner, C.; Nielsen, K.G.; Marthin, J.K.; Philipsen, M.; Wallmeier, J.; Pennekamp, P.; Menchen, T.; et al. Loss-of-Function GAS8 Mutations Cause Primary Ciliary Dyskinesia and Disrupt the Nexin–Dynein Regulatory Complex. *Am. J. Hum. Genet.* **2015**, *97*, 546–554. [[CrossRef](#)]
50. McRae, K.M.; Baird, H.J.; Dodds, K.G.; Bixley, M.J.; Clarke, S.M. Incidence and Heritability of Ovine Pneumonia, and the Relationship with Production Traits in New Zealand Sheep. *Small Rumin. Res.* **2016**, *145*, 136–141. [[CrossRef](#)]

51. Merveille, A.-C.; Davis, E.E.; Becker-Heck, A.; Legendre, M.; Amirav, I.; Bataille, G.; Belmont, J.; Beydon, N.; Billen, F.; Clément, A.; et al. CCDC39 Is Required for Assembly of Inner Dynein Arms and the Dynein Regulatory Complex and for Normal Ciliary Motility in Humans and Dogs. *Nat. Genet.* **2011**, *43*, 72–78. [[CrossRef](#)]
52. Poprzeczko, M.; Bicka, M.; Farahat, H.; Bazan, R.; Osinka, A.; Fabczak, H.; Joachimiak, E.; Wloga, D. Rare Human Diseases: Model Organisms in Deciphering the Molecular Basis of Primary Ciliary Dyskinesia. *Cells* **2019**, *8*, 1614. [[CrossRef](#)] [[PubMed](#)]
53. Clark, E.L.; Bush, S.J.; McCulloch, M.E.B.; Farquhar, I.L.; Young, R.; Lefevre, L.; Pridans, C.; Tsang, H.G.; Wu, C.; Afrasiabi, C.; et al. A High Resolution Atlas of Gene Expression in the Domestic Sheep (*Ovis Aries*). *PLoS Genet.* **2017**, *13*, e1006997. [[CrossRef](#)] [[PubMed](#)]
54. Zheng, Y.; Zhang, J.; Wang, L.; Zhou, Z.; Xu, M.; Li, J.; Sha, J. Cloning and Characterization of a Novel Sperm Tail Protein, NYD-SP28. *Int. J. Mol. Med.* **2006**, *18*, 1119–1125. [[CrossRef](#)] [[PubMed](#)]
55. Sironen, A.; Shoemark, A.; Patel, M.; Loebinger, M.R.; Mitchison, H.M. Sperm Defects in Primary Ciliary Dyskinesia and Related Causes of Male Infertility. *Cell. Mol. Life Sci.* **2020**, *77*, 2029–2048. [[CrossRef](#)] [[PubMed](#)]
56. David, I.; Druart, X.; Lagriffoul, G.; Manfredi, E.; Robert-Granié, C.; Bodin, L. Genetic and Environmental Effects on Semen Traits in Lacaune and Manech Tête Rousse AI Rams. *Genet. Sel. Evol.* **2007**, *39*, 405. [[CrossRef](#)]
57. Kadri, N.K.; Sahana, G.; Charlier, C.; Iso-Touru, T.; Guldbandsen, B.; Karim, L.; Nielsen, U.S.; Panitz, F.; Aamand, G.P.; Schulman, N.; et al. A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. *PLoS Genet.* **2014**, *10*, e1004049. [[CrossRef](#)] [[PubMed](#)]
58. Hedrick, P.W. Heterozygote Advantage: The Effect of Artificial Selection in Livestock and Pets. *J. Hered.* **2015**, *106*, 141–154. [[CrossRef](#)]

Homozygous haplotype deficiency in Manech Tête Rousse dairy sheep revealed a nonsense variant in *MMUT* gene affecting newborn lamb viability

Maxime Ben Braiek¹, Carole Moreno-Romieux^{1§}, Céline André², Jean-Michel Astruc³, Philippe Bardou⁴, Arnaud Bordes¹, Frédéric Debat¹, Francis Fidelle², Chris Hozé⁵, Florence Plisson-Petit¹, François Rivemale¹, Julien Sarry¹, Némuel Tadi¹, Florent Woloszyn¹ and Stéphane Fabre^{1*}

¹GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet-Tolosan, France

²CDEO, Quartier Ahetzia, 64130, Ordiarp, France

³Institut de l'Élevage, 24 chemin de Borde-Rouge, F-31321, Castanet-Tolosan, France

⁴Sigenae, INRAE, 31326 Castanet-Tolosan, France

⁵Eliance, 75595 Paris, 149 rue de Bercy, France

§ *in memoriam*

*Corresponding author

Email: stephane.fabre@inrae.fr (SF)

Maxime BEN BRAIEK, ORCID iD <https://orcid.org/0000-0003-4770-0867>

Stéphane FABRE, ORCID iD <https://orcid.org/0000-0001-7350-9500>

Abstract

Recessive deleterious variants are known to segregate in livestock populations as in human, and some may cause lethality when homozygous. By scanning the genome of 6,845 Manech Tête Rousse dairy sheep using phased 50k SNP genotypes and pedigree data, we searched for homozygous haplotype deficiency. Five Manech Tête Rousse deficient homozygous haplotypes (MTRDHH1 to 5) were identified with a homozygous deficiency ranging from 84% to 100%. These haplotypes are located on OAR1 (MTRDHH2 and 3), OAR10 (MTRDHH4), OAR13 (MTRDHH5) and OAR20 (MTRDHH1), and have frequencies ranging from 7.8% to 16.6%. When comparing at-risk mating between DHH carriers to safe mating between non-carriers, two DHH (MTRDHH1 and 2) showed significant effects on decreasing artificial insemination success and/or increasing stillbirth rate. We particularly investigated the MTRDHH1 haplotype highly increasing stillbirth rate, and we identified a single nucleotide variant inducing a premature stop codon (p.Gln409*) in the *MMUT* gene (*methylmalonyl-CoA mutase*) by using a whole genome sequencing (WGS) approach. We generated homozygous lambs for the *MMUT* mutation by oriented mating, and most of them died within the first 24h after birth without any obvious clinical defect. RT-qPCR and western blotting performed on post-mortem liver and kidney biological samples showed a decreased expression of *MMUT* mRNA in the liver and absence of a full-length *MMUT* protein in mutated homozygous lambs. In parallel, MTRDHH4 and MTRDHH5 showed partial association with variants in *RXFP2* and *ASIP* genes, respectively, already known to control horned/polled and coat color phenotypes in sheep, two morphological traits accounting in the MTR breed standard. Further investigations are needed to identify the supposed recessive deleterious variant hosted by MTRDHH2 and MTRDHH3. Anyway, an appropriate management of these haplotypes/variants in the MTR dairy sheep selection program should increase the overall fertility and lamb survival.

Author Summary

In this article, we used reverse genetics screen in ovine using large genotype data available in the framework of genomic selection program in Manech Tête Rousse dairy sheep. We identified five genomic regions with a highly significant deficit in homozygous animal. These regions are thus supposed to host recessive deleterious mutations. In one of these genomic regions, we identified a nonsense mutation in *MMUT* that alters the functioning of this essential gene of cell metabolism, causing perinatal mortality of homozygous lambs. In this work, we also identified other regions possibly associated with morphological appearance part of the breed standard such as polledness and coat color. Increasing knowledge in these genomic regions will help the future genetic management of the Manech Tête Rousse breed, particularly to reduce lamb mortality.

Introduction

In livestock, genetic selection has largely improved production traits over the past decades, but the last one has seen the emergence of new technological tools allowing to implement the genomic selection that further enhanced the genetic progress [1]. The availability of high-density single nucleotide polymorphism (SNP) chip and the improvement of knowledge of genomes (genome assembly and gene annotations) have allowed to fine-map genomic regions and identify causal variants associated with production traits [1–4]. Despite a successful selection on these traits, undesired decline of fertility was observed [5]. Although the environment explains a large part of performance in ruminants (heritability < 0.05 [6,7]), genetic studies have made it possible to correct this trend and improve fertility traits [8]. These studies have shown Mendelian monogenic disorders as one of the causes of fertility failure.

Nowadays two main approaches were broadly developed to identify recessive deleterious variants. The first approach is a “top-down” strategy based on case-control analysis performing genome-wide association [9] when biological samples from affected animals are available. In this method, distinctive phenotypes between non-affected and affected animals are essential. Subsequently, homozygosity mapping approach could be performed to detect homozygous regions in affected animals supposed to host the causal variant, further determined by whole genome sequencing (WGS) data [10,11]. In livestock, Charlier et al. [10] have used this approach for the first time and successfully detected three causal variants in cattle breeds located in *ATP2A1*, *SLC6A5* and *ABCA12* genes responsible for “congenital muscular dystony” types 1 (OMIA 001450-9913) and 2 (OMIA 001451-9913) and “ichthyosis fetalis” (OMIA 002238-9913), respectively. However, this approach showed some limits when biological samples and descriptive phenotypes are not available. To raise this drawback, a second approach called “bottom-up” or reverse genetic screen strategy was developed to specifically identify recessive lethal variants. This strategy, initially developed by VanRaden et al. [12], is based on the exploitation of large number of genotyped animals easily available from genomic selection datasets to detect haplotypes showing deficit in homozygous animals, with a significant deviation from the Hardy-Weinberg equilibrium. Initially, this method was developed to detect embryonic lethal variant but the generalization of the method can also fine-map deleterious variants leading to neonatal or juvenile lethality and morphological disorders. This reverse genetic screen has successfully identified numerous deficient homozygous haplotypes in several species, cattle [12–27], pigs [28,29], chicken [30], turkey [31] and horses [32], and the whole genome sequencing that followed has revealed the associated causative variants hosted by these regions [13,14,18,19,21,24–27,33–41]. We recently, and for the first time, validated this approach in sheep with the identification of 8 independent deficient homozygous haplotypes in the Lacaune dairy breed [42]. Thereafter, focusing on the Lacaune Deficient Homozygous Haplotype 6 (LDHH6, OMIA 002342-9940), we identified a nonsense variant in *CCDC65* gene causing juvenile mortality associated with respiratory distress when homozygous [43].

In the present work, still leveraging genomic selection data, we searched for lethal variants using a reverse genetic screen in Manech Tête Rousse (MTR) dairy sheep. This breed is raised in the French Basque Country and represents the second most important breed by its population size (~450,000 ewes) in France [44]. Genetic variability is well controlled in MTR breed with an increase by +0.42% of inbreeding per generation over 2000-2009. The effective population size ranges from 118 to 154 according to the estimation methods [44,45]. As for other dairy sheep breeds in France, genomic selection in MTR was implemented in 2017 [46].

The aim of this study was to identify deficient homozygous haplotypes by reverse genetic screen using large amount of genotyping data available in MTR dairy sheep, to test the hypothesis of negative impacts on fertility traits in at-risk matings, to propose relevant candidate genes located in these regions that could host recessive deleterious variants. We particularly focus on one region to identify the associated causal variant from WGS data and manage at-risk mating between carriers to determine the associated phenotype.

Results

Identification of deficient homozygous haplotypes in Manech Tête Rouse dairy sheep

Using a reverse genetic screen strategy based on 5,271 genotyped animals belonging to trios, we have detected 150 highly significant Homozygous Haplotype Deficiency (HHD) of 20 SNP markers (listed in S1A Table). These 150 HHD were clustered to five independent regions called “Manech Tête Rouse Deficient Homozygous Haplotype” (MTRDHH). Three haplotypes showed a total deficit in homozygous animals (MTRDHH1, 2 and 3), whereas two haplotypes, MTRDHH4 and 5, only showed a partial deficit (84% and 91%) with 1 and 8 homozygous animals genotyped while 11 and 49 were expected, respectively (Table 1, Fig 1, S1A Table). The complete description of MTRDHH SNP markers (SNP name, SNP allele and position on sheep reference genomes Oar_v3.1, Oar_Rambouillet_v1.0 and ARS-UI_Ramb_v2.0) is available in S1B Table. The different MTRDHH were located on OAR20 (MTRDHH1), OAR1 (MTRDHH2 and 3), OAR10 (MTRDHH4) and OAR13 (MTRDHH5), and their length ranged from 1.1 to 4.6 Mb on Oar_rambouillet_v1.0. The observed frequencies of heterozygous carriers were between 7.8% and 16.6%. MTRDHH2 and MTRDHH3 both located on OAR1 were not in linkage disequilibrium. Consequently, the five MTRDHH identified are likely to harbor five independent variants causal of the observed homozygous deficiency.

Table 1. List of Manech Tête Rouse deficient homozygous haplotypes.

Haplotype	OAR	^a Number of markers	^b Position (Mb)	^c Heterozygous carrier frequency (%)	Number of homozygotes			
					^d Exp	^e Obs	Deficit	Poisson P-value
MTRDHH1	20	32	23.0-25.0	9.7%	13	0	100%	2.9×10^{-6}
MTRDHH2	1	66	251.9-256.4	8.7%	10	0	100%	3.8×10^{-5}
MTRDHH3	1	39	103.8-106.6	7.8%	9	0	100%	9.6×10^{-5}
MTRDHH4	10	26	30.5-31.5	8.7%	11	1	91%	1.5×10^{-4}
MTRDHH5	13	29	64.3-67.2	16.6%	49	8	84%	5.3×10^{-13}

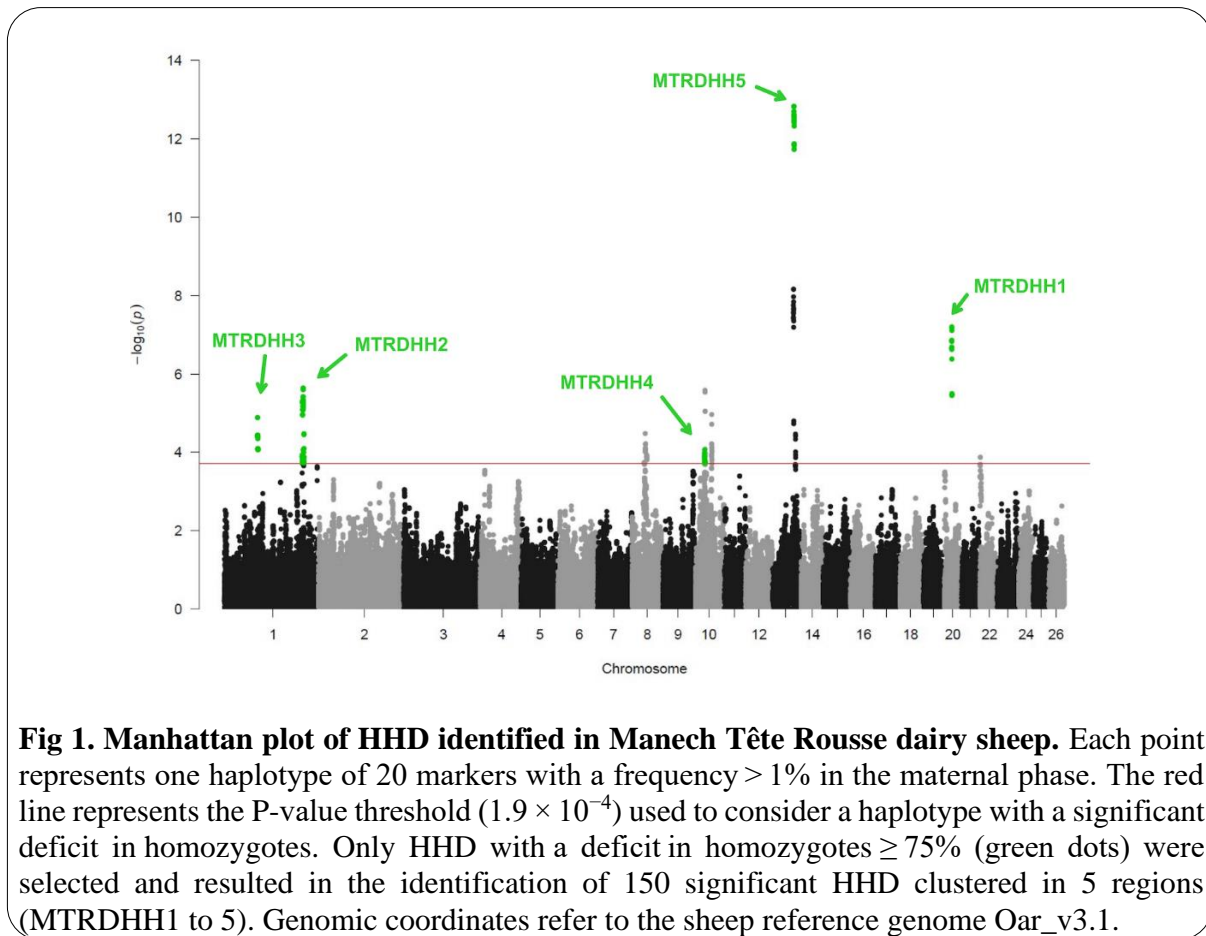
^aMTRDHH haplotypes and SNP markers composition are details in S1 Table

^bPosition on ovine genome assembly Oar_rambouillet_v1.0

^cFrequency of carriers in the entire genotyping population (n=6,845)

^dExpected

^eObserved



Impact of MTRDHH on fertility traits

In order to identify a putative lethal effect of the five MTRDHH, two recorded fertility traits were analyzed: artificial insemination success, a proxy for embryonic loss (AIS: 330,844 matings) and stillbirth rate associated with perinatal lethality (SBR: 201,637 matings) (Fig 2). The average AIS of the population was 60.9%. When comparing at-risk and safe matings, only MTRDHH2 showed a significant decrease of -3.3% of AIS ($P=3.5 \times 10^{-4}$) in at-risk matings. The average SBR of the population was 7.5%. As described in Fig 2, MTRDHH1 and 2 showed a huge increase in SBR with +7.5% ($P=4.0 \times 10^{-24}$) and +4.3% ($P=1.3 \times 10^{-6}$) in at-risk matings compared to safe matings, respectively. The three other haplotypes showed no significant impact on the fertility traits studied.

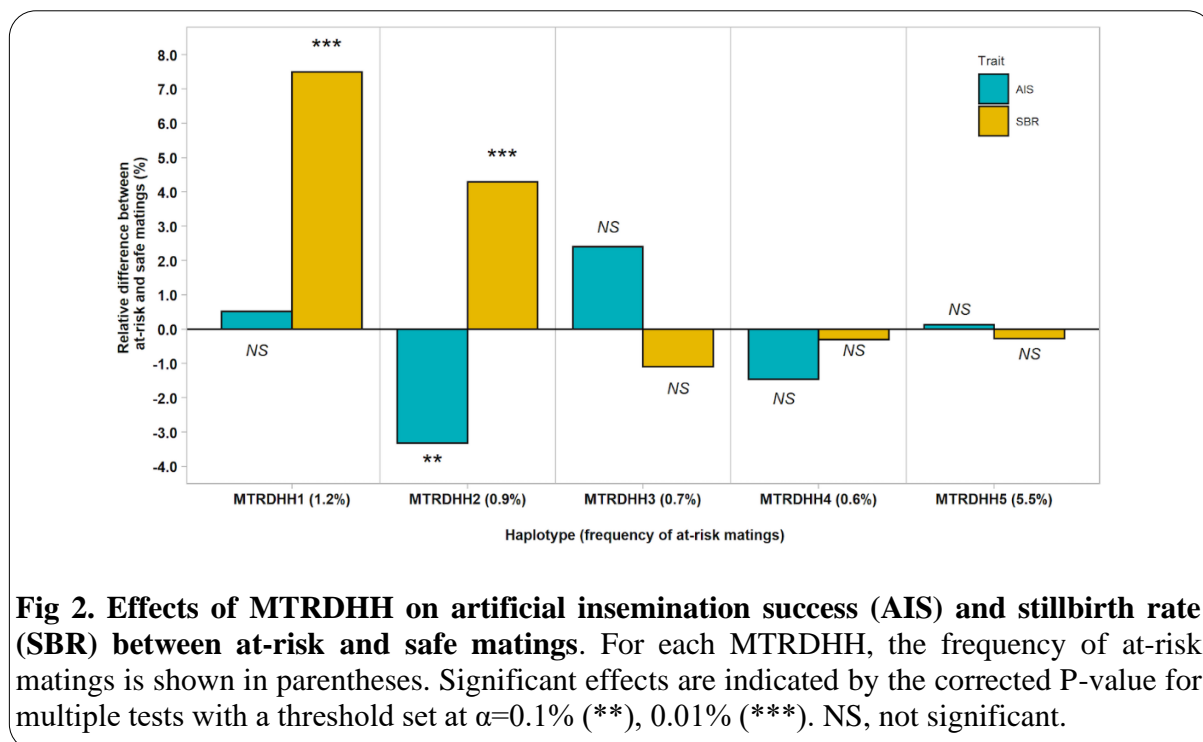


Fig 2. Effects of MTRDHH on artificial insemination success (AIS) and stillbirth rate (SBR) between at-risk and safe matings. For each MTRDHH, the frequency of at-risk matings is shown in parentheses. Significant effects are indicated by the corrected P-value for multiple tests with a threshold set at $\alpha=0.1\%$ (**), 0.01% (***). NS, not significant.

Pleiotropic effects of MTRDHH on milk production traits

Six dairy traits are routinely included in genomic evaluation of the French dairy sheep. Thus, standardized daughter yield deviation (sDYD) of five milk production traits (milk, fat and protein yields, and fat and protein contents) and lactation somatic cell score (a proxy for udder health) were compared between carrier and non-carrier rams for each of the 5 MTRDHH evidenced (Fig 3). Among the five haplotypes, three were associated with significant effect on sDYD. Daughters of MTRDHH2 carrier rams showed a significant increase in milk production (sDYD $+0.06$, $P=6.5 \times 10^{-3}$) but a decrease in protein content (sDYD -0.11 , $P=4.7 \times 10^{-4}$). For MTRDHH4, there was a significant increase in lactation somatic cell score (sDYD -0.13 , $P=2.4 \times 10^{-3}$), and daughters of MTRDHH5 carrier rams showed higher fat yield (sDYD -0.06 , $P=9.6 \times 10^{-3}$). Additionally, the total merit genomic index, named ISOLg, was extracted from each male lamb of the 2021 genomic selection cohort to estimate the MTRDHH impact on the genetic gain of the selected traits. No significant difference was observed on ISOLg between heterozygous carrier and non-carrier lambs for each MTRDHH (S1 Fig).

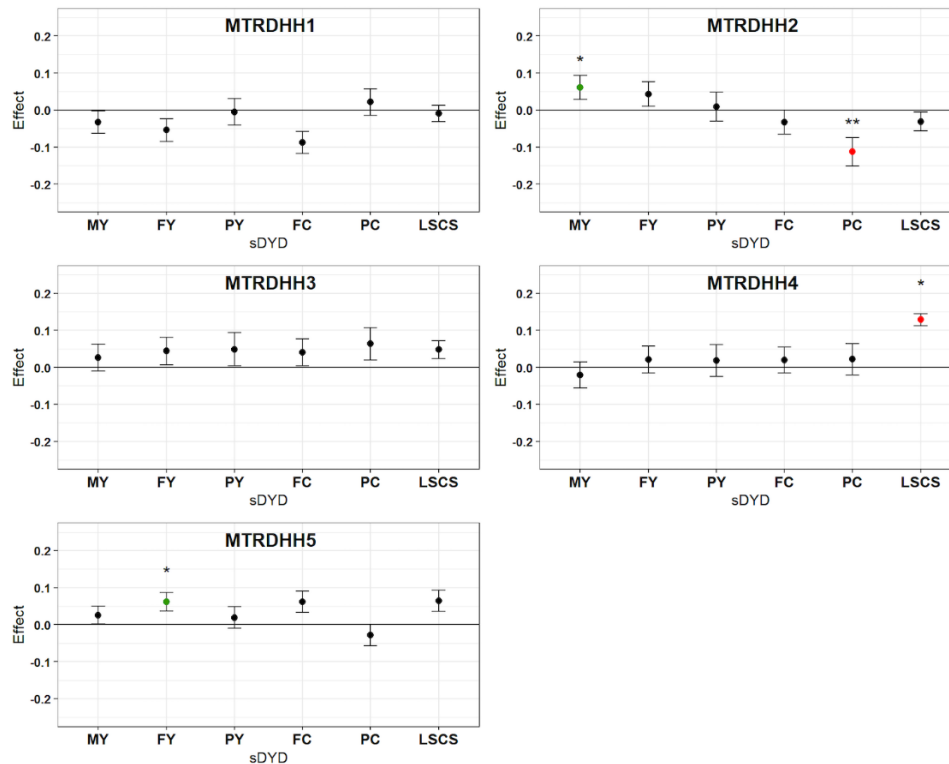
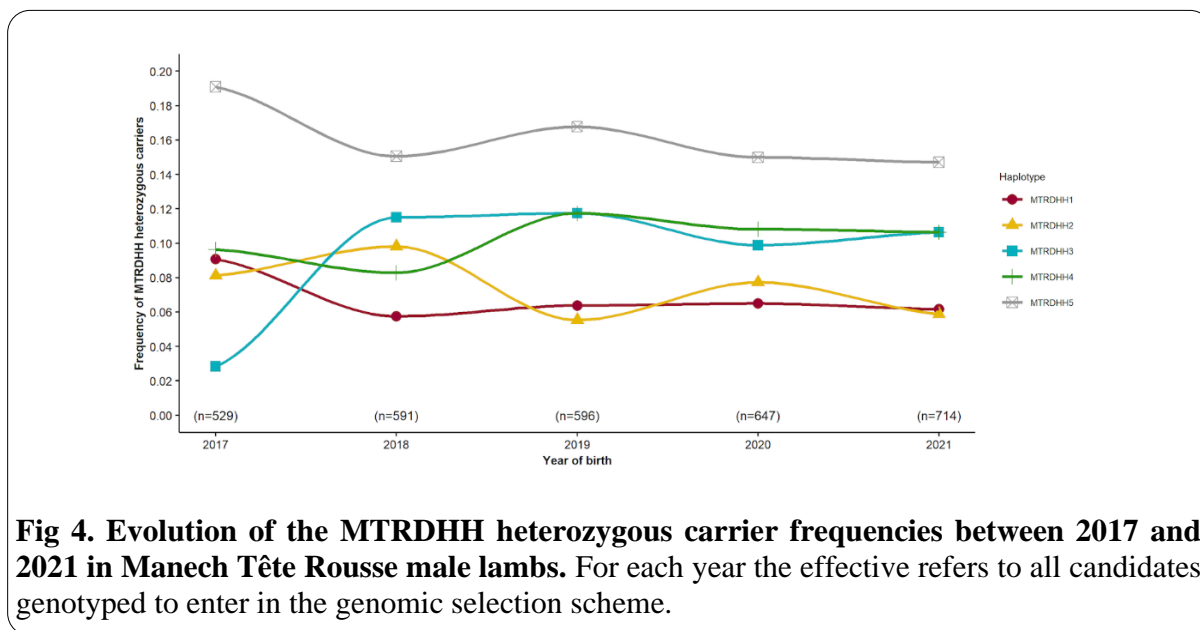


Fig 3. sDYD relative difference between heterozygous and non-carrier rams for 6 traits under selection. MY: milk yield, FY: fat yield, PY: protein yield, FC: fat content, PC: protein content, LSCS: lactation somatic cell score, sDYD standardized daughter yield deviation (DYD divided by genetic standard deviation). sDYD relative difference value is obtained from lsmeans estimate according to mating class. Significant effects are indicated by the corrected P-value for multiple tests with a threshold set at $\alpha=5\%$ (*), 0.1% (**). Error bars indicate standard errors. Significant favorable effects of heterozygous are in green while significant unfavorable effects are in red.

Evolution of the MTRDHH frequencies in the population

Since the implementation of genomic selection in 2017, all candidate rams coming from elite mating were genotyped on low density SNP chip at 7 days old, representing the genetic diversity disseminated by AI in the selection scheme. As shown in Fig 4, the frequencies of the MTRDHH heterozygous carriers were quite stable during the last five years around 6.8, 7.4, 9.3, 10.2 and 16.1% for MTRDHH1, 2, 3, 4 and 5, respectively. Nevertheless, we can notice a spectacular increase in the frequency of MTRDHH3 from 2.8% to 10.9% when comparing 2017 and 2018.



Candidate genes located in MTRHH regions

Within the five MTRDHH genomic regions extended by 1Mb on each side, 408 protein coding genes are annotated (S2 Table). When available, information on mouse phenotypes (including lethal phenotypes) and association with mammalian genetic disorders were extracted for each gene using MGI, IMPC, OMIM and OMIA databases. Among the 408 genes, we highlighted 64 genes involved in lethal phenotypes in knock-out mice, and 45 genes associated with human genetic disorders. Twenty-three relevant candidate genes were identified by the intersection of both information (Fig 5). In addition, 7 genes are known to be associated with genetic disorders or morphological traits in livestock (*GJA5*, *ITGA10*, *ADAMTLS4*, *RXFP2*, *KIF3B*, *ASIP* and *CEP250*). Overall, these candidate genes are involved in essential functions such as transcription (*POLR3GL*, *SF3B4*, *PRPF3*, *ASXL1* and *DNMT3B*), cell division (*POGZ*, *BRCA2*, *PIGU* and *CEP250*), basal metabolic processes (*MMUT*, *SLC33A1*, *TARS2* and *AHCY*), cell structure and signaling (*CD2AP*, *PKHD1*, *GJA5*, *ITGA10*, *ECM1*, *GJA5*, *ITGA10*, *ECM1*, *PRUNE1*, *RXFP2*, *KL*, *POFUT1*, *KIF3B*, *ASIP* and *GSS*), and DNA/protein binding (*TFAP2B* and *PEX11B*).

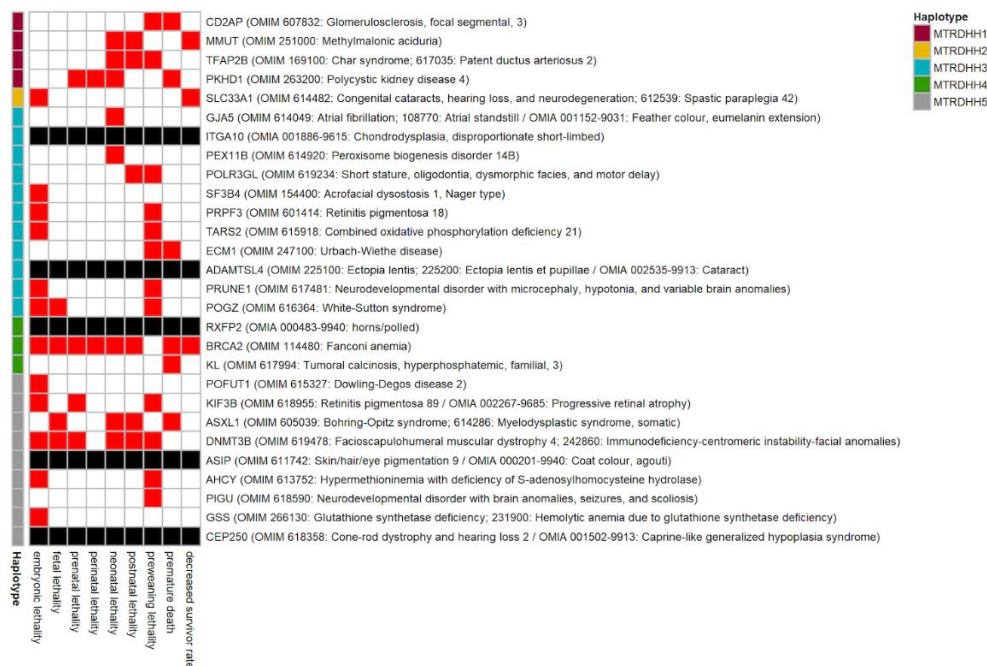
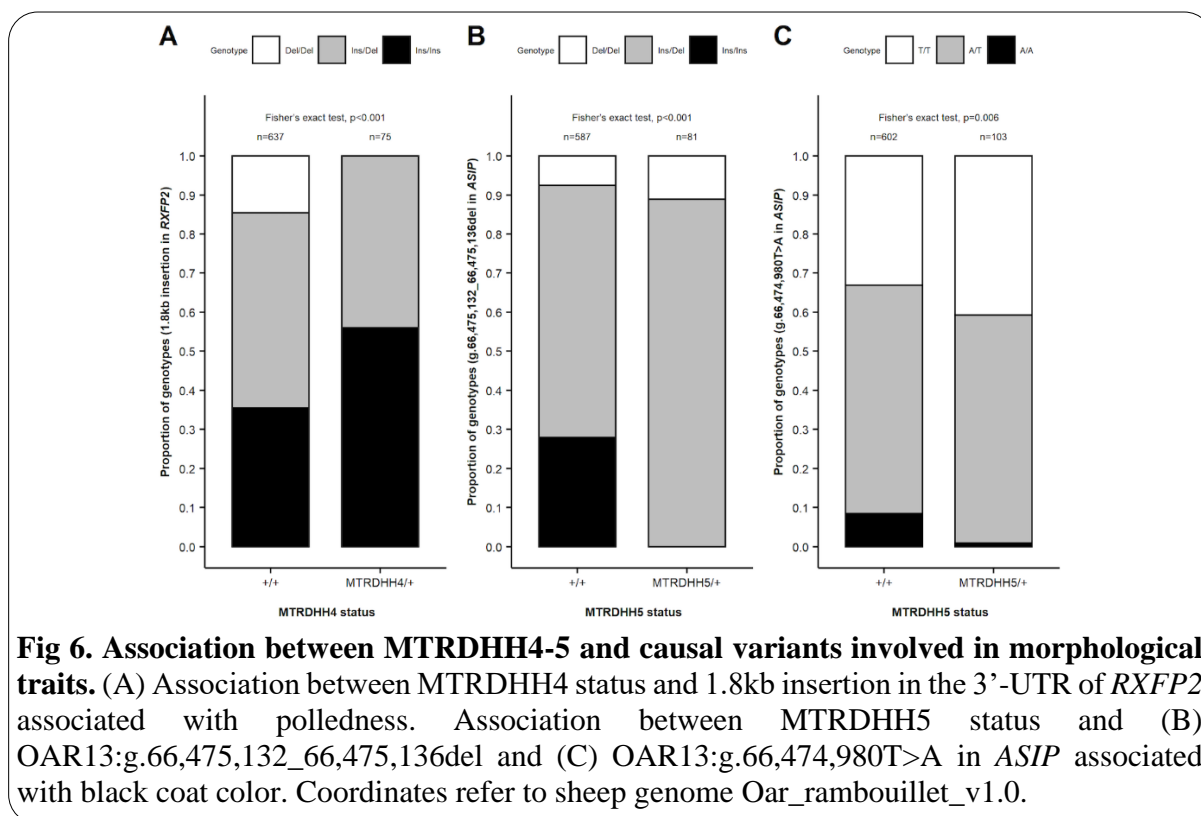


Fig 5. Relevant candidate genes located in MTRDHH implicated in lethal mouse phenotype and/or associated with mammalian genetic disorders. Red boxes indicate time of known lethality stages in mouse when the gene is knocked-out from MGI/IMPC databases. When the gene is associated with mammalian genetic disorders, the OMIM and/or OMIA trait phenotypes are described in parentheses. Black boxes indicate a gene implicated in animal genetic disorder (OMIA) but with no lethal phenotype observed in mouse.

Known variants in *RXFP2* and *ASIP* genes are partially associated with MTRDHH4 and MTRDHH5

The above list of candidate genes particularly highlighted *RXFP2* (within MTRDHH4) and *ASIP* (within MTRDHH5) genes well-known to impact sheep morphological traits such as polledness [47] and black coat color [48,49], respectively. Interestingly, horned female and black males are not desired in accordance with the MTR breed standards. Using the 22 WGS of MTR animals, we searched for the 1.8kb insertion in the 3'-UTR of *RXFP2* associated with polledness (OMIA 000483–9940, [47]) and the different variants affecting *ASIP* leading to recessive black coat color (OMIA 000201-9940); OAR13:g.66,475,132_66,475,136del and g.66,474,980T>A, Oar_rambouillet_v1.0) [48,49]. S2 Fig shows the segregation of these variant among the 22 sequenced animals and their relationship with the status of MTRDHH4 and 5, but with no obvious association. This being made on a reduced number of animals, we specifically genotyped a larger set of animals (n=714 male lambs born in 2021) for the 1.8kb insertion in *RXFP2* and the two variants in *ASIP*. Using this cohort, we evidenced a partial association between MTRDHH4 and the 1.8kb insertion in the 3'-UTR of *RXFP2*, none heterozygous MTRDHH4/+ being Del/Del corresponding to the horned phenotype (Fig 6). Concerning *ASIP* variants, MTRDHH5 was also partially associated with the 5pb deletion. Indeed, 89% of the MTRDHH5 heterozygous carriers were heterozygous for the 5pb deletion and none Ins/Ins animal was MTRDHH5 carrier. In contrast, we identified almost the same proportion of genotypes for g.66,474,980T>A in MTRDHH5 carriers and non-carriers (+/+), suggesting that MTRDHH5 was not linked to this SNV (Fig 6C).



Identification of a nonsense variant in *MMUT* gene associated with MTRDHH1

With the most important impact on stillbirth rate increased by +7.5% in at-risk matings (Fig 2), we particularly focused on MTRDHH1 that may represent a putative recessive lethal haplotype. In order to identify the MTRDHH1 causal mutation, we have considered biallelic variants (SNP and InDels) for 100 ovine WGS containing the 22 Manech Tête Rousse dairy sheep, and among them two heterozygous carriers of the MTRDHH1 haplotype. Within the MTRDHH1 region extended by 1Mb on each side, 78,019 variants were called with a quality score >30, call rate >95% and only four candidate variants had a perfect correlation ($r^2=1$) between biallelic variant genotypes and MTRDHH1 status (Table 2, Fig 7A). Among those candidate variants, we identified two small insertions, one intergenic SNV, and one nonsense (stop-gain) SNV located in the *Methylmalonyl-CoA Mutase (MMUT)* gene. This latter SNV (NC_040271.1: g.23,776,347G>A; XM_004018875.4: c.1225C>T; Fig 7B, C) in *MMUT* is predicted to create a premature stop codon at position 409 encoded by exon 6 (XP_004018923.1:p.Gln409*) whereas the full protein length is composed of 750 amino acids (Fig 7D). The variant would disrupt the methylmalonic coenzyme-A mutase domain and would result in the loss of the vitamin B12 binding domain.

Table 2. Candidate variants located in MTRDHH1.

Position	Ref/Alt	Quality score	Location Annotation	Functional Consequence ^a
23,436,234	G/GTCACA	385.8	Intergenic	Modifier
23,436,236	T/TTTGTG	385.8	Intergenic	Modifier
23,776,347	G/A	146.0	Exonic, <i>MMUT</i> (c.1225C>T)	High, stop-gain (p.Gln409*)
23,969,676	C/T	370.3	Intergenic	Modifier

^aVariant annotation and effect predicted by SnpEff [50].

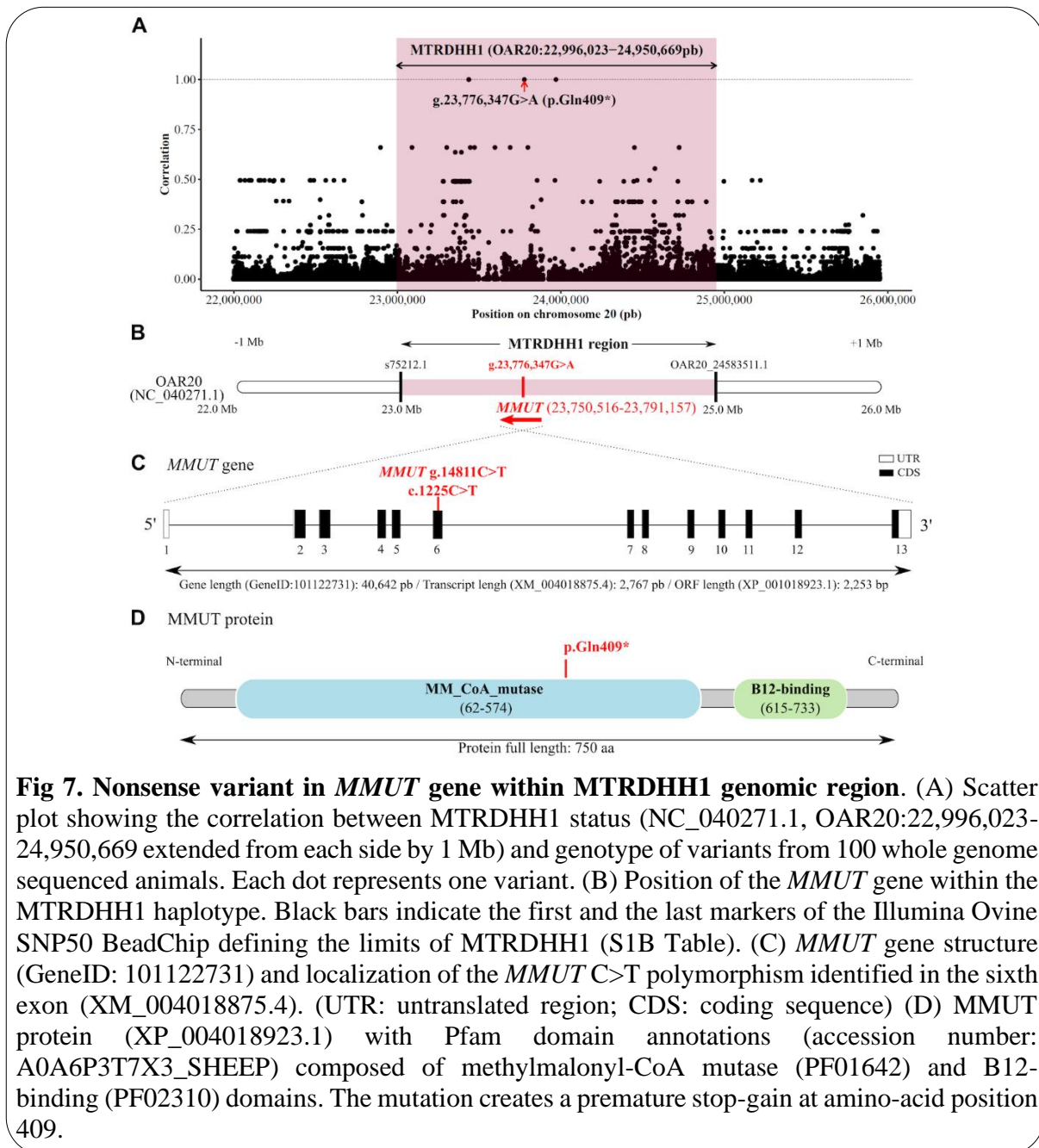


Fig 7. Nonsense variant in *MMUT* gene within *MTRDHH1* genomic region. (A) Scatter plot showing the correlation between *MTRDHH1* status (NC_040271.1, OAR20:22,996,023-24,950,669 extended from each side by 1 Mb) and genotype of variants from 100 whole genome sequenced animals. Each dot represents one variant. (B) Position of the *MMUT* gene within the *MTRDHH1* haplotype. Black bars indicate the first and the last markers of the Illumina Ovine SNP50 BeadChip defining the limits of *MTRDHH1* (S1B Table). (C) *MMUT* gene structure (GeneID: 101122731) and localization of the *MMUT* C>T polymorphism identified in the sixth exon (XM_004018875.4). (UTR: untranslated region; CDS: coding sequence) (D) *MMUT* protein (XP_004018923.1) with Pfam domain annotations (accession number: A0A6P3T7X3_SHEEP) composed of methylmalonyl-CoA mutase (PF01642) and B12-binding (PF02310) domains. The mutation creates a premature stop-gain at amino-acid position 409.

In order to validate the association between the *MMUT* variant and *MTRDHH1*, we genotyped the cohort of male lambs born in 2021 (n=714) with a specific genotyping test for the *MMUT* g.23,776,347G>A SNV. The A variant allele frequency was 3.8%. All these animals have a known status at the *MTRDHH1* locus and the contingency table indicates a clear association between the *MTRDHH1* status and *MMUT* variant genotypes (Fig 8, Fischer's exact test $p < 0.0001$). However, 15 animals showed discrepancy between *MMUT* and *MTRDHH1* genotypes, supposed to be in perfect linkage disequilibrium. A specific focus on haplotypes carried by these animals in the *MTRDHH1* region from marker 1 (s75212.1) to marker 32 (OAR20_24583511.1) showed that the 14 animals heterozygous for the variant exhibited shorter recombinant versions of the *MTRDHH1* haplotype (S3 Fig). Nonetheless, one animal was heterozygous for *MTRDHH1* but did not carry the *MMUT* variant.

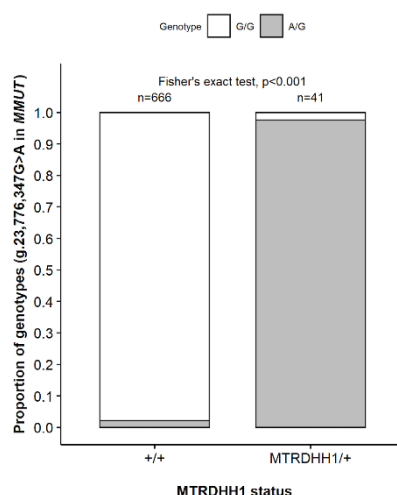


Fig 8. Association of *MMUT* SNV genotypes with *MTRDHH1* status. +/+ : non-carriers; *MTRDHH1*/+ : heterozygous carriers and *MTRDHH1*/*MTRDHH1* : homozygous carriers (Fisher's exact test, $p < 0.001$, without the homozygous *MTRDHH1* carriers).

Occurrence of the *MMUT* SNV in an ovine diversity panel

An ovine diversity panel composed of 25 French sheep breeds, including MTR, and 3 Latxa Spanish sheep breeds related to MTR was used to genotype the *MMUT* g.23,776,347G>A SNV (Table 3). As expected, some MTR animals ($n=5$) from this panel were evidenced as heterozygous carriers, and the variant was also detected in one animal of the Spanish Latxa Cara Rubia population. All the other animals tested did not carry the polymorphism.

Table 3. *MMUT* SNV genotype distribution from a DNA diversity panel of French (FR) and Spanish (ES) ovine breeds.

Breed	Total	Genotype		Breed	Total	Genotype	
		G/G	A/G			G/G	A/G
Berrichon du Cher (FR)	30	30		Martinik (FR)	23	23	
Blanche du Massif Central (FR)	31	31		Merinos d'Arles (FR)	27	27	
Causse du Lot (FR)	32	32		Mourerous (FR)	26	26	
Charmoise (FR)	31	31		Mouton Vendéen (FR)	30	30	
Charollais (FR)	30	30		Noir du Velay (FR)	28	28	
Corse (FR)	30	30		Préalpes du sud (FR)	27	27	
Ile de France (FR)	28	28		Rava (FR)	29	29	
Lacaune (Meat) (FR)	45	45		Romane (FR)	30	30	
Lacaune (Milk) (FR)	40	40		Romanov (FR)	26	26	
Latxa Cara Negra Euskadi (ES)	30	30		Rouge de l'Ouest (FR)	30	30	
Latxa Cara Negra Navarra (ES)	40	40		Roussin (FR)	30	30	
Latxa Cara Rubia (ES)	30	29	1	Suffolk (FR)	29	29	
Limousine (FR)	30	30		Tarasconnaise (FR)	33	33	
Manech Tête Rousse (FR)	29	24	5	Texel (FR)	27	27	
				Total	851	845	6

Viability of homozygous lambs for the *MMUT* variant

To validate the impact of the *MMUT* A variant allele we have managed an oriented mating between heterozygous carriers to generate homozygous lambs. Blood samples were collected from 181 MTR ewes, daughters of MTRDHH1 carrier sires, in 6 private farms. The *MMUT* SNV specific genotyping identified 82 heterozygous ewes. Among these ewes, 73 were raised in the 6 private farms (Experiment 1) and 9 were moved into an INRAE experimental farm (Experiment 2). All ewes were artificially inseminated with fresh semen from *MMUT* g.23776347G/A heterozygous rams. Forty-five days after AI, 44 and 5 were diagnosed as pregnant by ultrasonography in Experiment 1 and 2, respectively. This corresponds to an AIS of 59.8% in accordance with the average AIS of 60.9% determined previously in the whole population. In experiment 1, only 37 among the 44 pregnant ewes were monitored after gestation diagnosis and resulted in the birth of 59 lambs (mean prolificacy of 1.6, litter size ranging from 1 to 3) with a gestation length between 139 and 159 days. In experiment 2, the 5 pregnant ewes gave birth to 13 lambs (mean prolificacy of 2.6, litter size ranging from 2 to 4) with a gestation length between 151 and 157 days. No abortion during the five months of gestation was observed. Finally, 72 lambs (52% males and 48% females) were born and an ear punch was collected for genotyping the *MMUT* SNV (Table 4). The distribution of genotypes did not differ between the two experiments (Fisher's exact test, $p=0.686$). In total, 21 lambs were genotyped homozygous carriers (A/A), 29 heterozygous carriers (A/G) and 21 homozygous non-carriers (G/G). All lambs were monitored during the 0-30 days period until weaning. Twenty-five lambs died during this period representing a huge mortality rate of 34.7% (Fig 9). Contingency table between lamb genotypes (A/A, G/A, G/G) and viability (alive or dead) indicated a higher mortality rate for homozygous A/A lambs (Table 4, Fisher's exact test, $p<0.001$). Indeed, the A/A dead lambs accounted for 78% of the whole lamb mortality. The death of A/A homozygous lambs occurred very soon after birth within the first 24 hours. Clinical examination of dead lambs did not allow us to identify specific symptoms. Two homozygous A/A lambs have passed the weaning age (around 4 weeks). Additionally, in Experiment 2, the 13 lambs were weighted at birth (males: 4.0 ± 0.9 kg, females: 2.9 ± 1.4 kg) and A/A lambs had significantly lower birth weight compared to the other genotypes, regardless gender (Wilcoxon's non parametric test, $p=0.019$) (Fig 10).

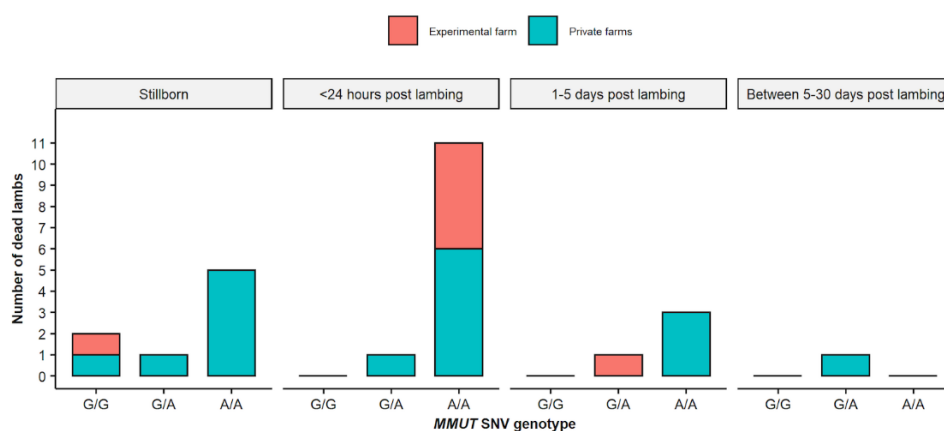


Fig 9. Time distribution of dead lambs in the pre-weaning period. Bar charts depending on *MMUT* SNV genotype and lambing place.

Table 4. Genotyping results of lambs generated in at-risk matings in private and experimental farms according to the sex and *MMUT* SNV genotype.

<i>g.23,776,347G>A</i> genotype in <i>MMUT</i>					
	G/G	G/A	A/A	-/-	Total
Experiment 1:					
Private farms (n=6)					
Male	9	9	10	1	29
Female	8	16	5		29
Undetermined			1		1
Total	17	25	16	1	59
Experiment 2:					
Experimental farm (n=1)					
Male	4	2	2		8
Female	0	2	3		5
Total	4	4	5		13
All					
Male	13	11	12	1	37
Female	8	18	8	0	34
Undetermined			1		1
Total	21 (19^a/2[†])	29 (25^a/4[†])	21 (2^a/19[†])	1* (0^a/1[†])	72

*The ear punch was not available for this lamb.

^a Number of alive lambs.

[†] Number of dead lambs.

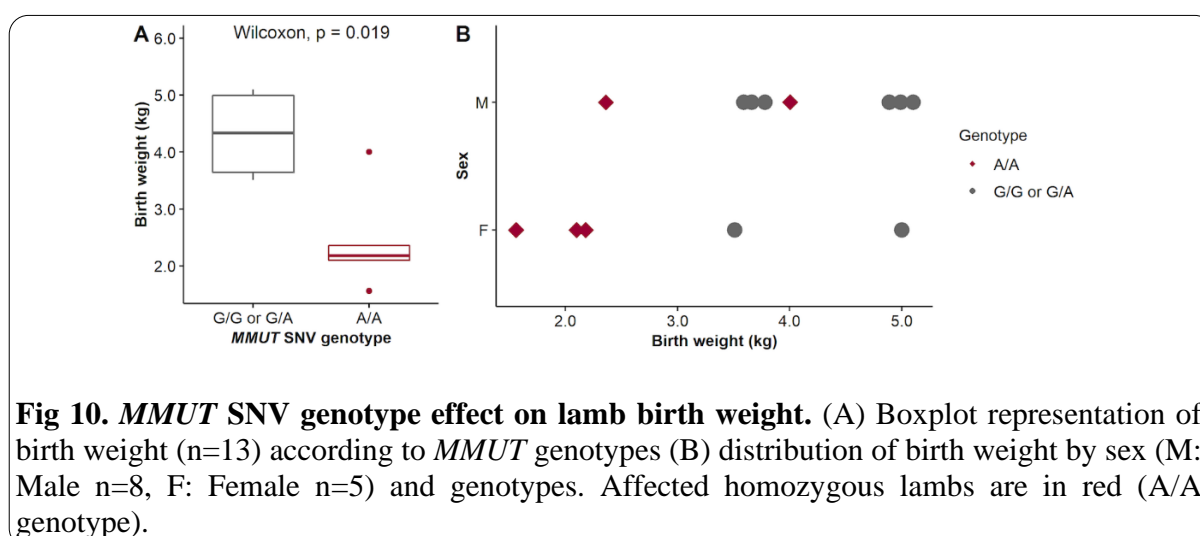


Fig 10. *MMUT* SNV genotype effect on lamb birth weight. (A) Boxplot representation of birth weight (n=13) according to *MMUT* genotypes (B) distribution of birth weight by sex (M: Male n=8, F: Female n=5) and genotypes. Affected homozygous lambs are in red (A/A genotype).

MMUT protein expression and activity

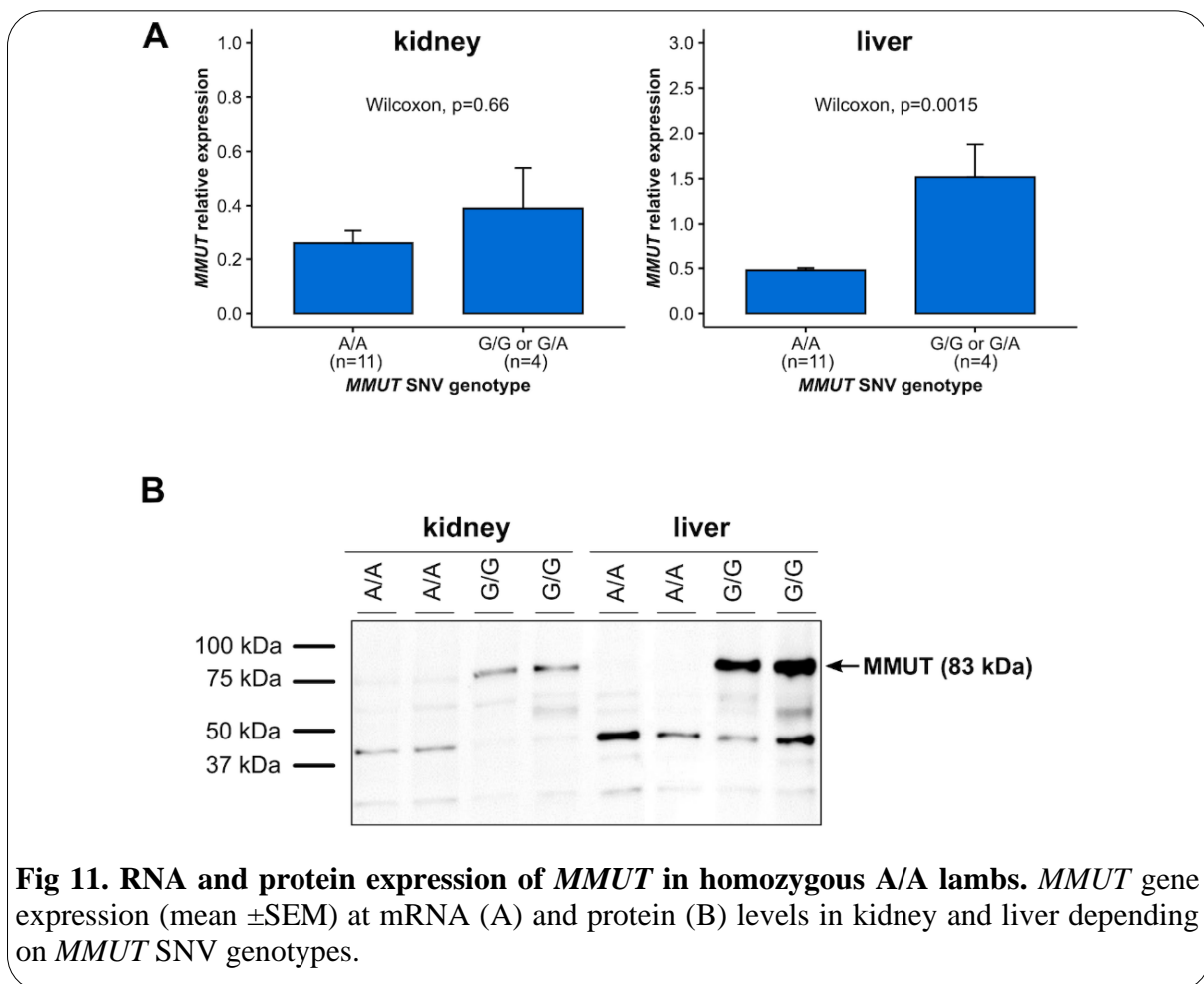
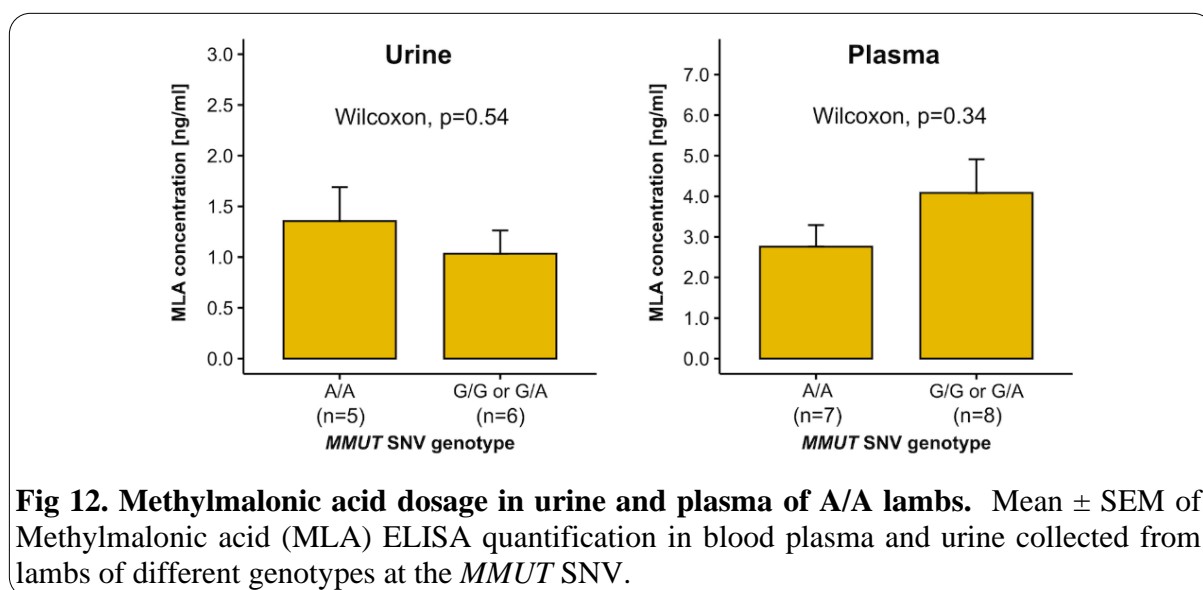


Fig 11. RNA and protein expression of *MMUT* in homozygous A/A lambs. *MMUT* gene expression (mean \pm SEM) at mRNA (A) and protein (B) levels in kidney and liver depending on *MMUT* SNV genotypes.

Based on the sheep gene atlas (<http://biogps.org/sheepatlas/>; accessed 17 February 2022), *MMUT* is shown to be highly expressed in kidney and liver [51]. In order to evidence a putative nonsense-mediated mRNA decay (NMD) due to the nonsense variant in *MMUT*, we evaluated the *MMUT* mRNA relative expression in kidney and liver by qPCR (Fig 11A). We evidenced a significant reduction of *MMUT* expression in liver (Wilcoxon's non parametric test, $p=0.0015$) but not in kidney (Wilcoxon's non-parametric test, $p=0.66$). We also assessed the *MMUT* protein expression from liver and kidney protein extracts collected from two homozygous A/A and two homozygous G/G dead lambs. As expected, using Western blotting, the wild type protein was expressed in kidney and liver whereas the mutated protein was not detected at least as a full-length form in both tissues (Fig 11B). We also tried to evaluate the accumulation of methylmalonic acid quantified by ELISA in urine and blood of A/A lambs collected post-mortem or soon after birth in alive animals (Fig 12) but no significant difference was observed compared to G/G or G/A lambs (Wilcoxon's non-parametric test, $p=0.54$ in urine, Wilcoxon's non-parametric test, $p=0.34$ in plasma).



Discussion

Using a reverse genetic screen in MTR population, we successfully identified five genomic regions (named MTRDHH1 to 5) with a significant deficit of homozygous animals ranging from 84 to 100%. Compared to our previous analysis in Lacaune dairy sheep with 8 independent haplotypes, we identified less deficient haplotypes in MTR possibly due to a lower number of genotyped animals (19,102 Lacaune vs 6,845 MTR) [42]. In the MTR population, we estimated the frequencies of MTRDHH heterozygous carriers between 7.8% and 16.6%, and thus allele frequency between 3.9 and 8.3%. This is in line with an allele frequency of 5% expected from the analysis of a population of 6,000 genotyped animals [24].

We tested a putative impact of each MTRDHH on production traits to search for a selective advantage at heterozygous state. The positive effects of these DHH on selected traits were quite low and only significant for MTRDHH2 carriers on milk yield and for MTRDHH5 on fat yield. It is thus unlikely that selective advantage explains the observed DHH frequencies, or the rapid increase in MTRDHH3 frequency between 2017 and 2018. However, many examples of balancing selection for deleterious alleles have been described in livestock [52]. For example in sheep, a missense variant in *FGFR3* was associated with enhanced skeletal growth and meat yield when heterozygous while it induced chondrodysplasia (spider lamb syndrome) when homozygous (OMIA 001703-9940) [53].

The populational analysis of AIS and SBR recorded on more than 300,000 matings, allowed us to do a first sorting of the different MTRDHH based on their supposed deleterious impact on early gestation (AIS), around the time of birth (SBR), or on postnatal viability or morphological phenotypes when AIS and SBR were not altered. Accordingly, we classified the five haplotypes into three groups and we highlighted potential candidate genes according to their implication in lethal phenotypes in mouse and/or more generally involved in inherited mammalian disorders.

The first group is composed of MTRDHH3 (OAR1), MTRDHH4 (OAR10) and MTRDHH5 (OAR13) which showed no significant impact on fertility traits. We assume that these haplotypes host deleterious variants leading to postnatal lethality or morphological disorders. Within these three regions, 18 candidate genes are of interest for their implication in neonatal to juvenile lethality or associated with morphological defects counter selected at the time of candidate lamb genotyping. Candidate genes associated with postnatal lethality mainly

affect metabolism (*TARS2* in MTRDHH3, *KL* in MTRDHH4) or DNA repair (*BRCA2* in MTRDHH4). Some candidate genes not necessarily lead to lethality when altered, but may affect animal welfare with the alteration of vision (*PRPF3*, *ADAMTSL4* in MTRDHH3 and *KIF3B* in MTRDHH5), neurological disorders (*PRUNE1*, *POGZ* in MTRDHH3 and *ASXL1*, *PIGU*, *AHCY* in MTRDHH5) or morphological/stature defects (*ITGA10*, *POLR3GL*, *ECM1* in MTRDHH2 and *DNMT3B*, *CEP250* in MTRDHH5). However, by searching causal variants in the 22 WGS of MTR animals, we failed to detect candidate mutations in the genes listed above. In contrast, we were puzzled by the presence of *RXFP2* (MTRDHH4) and *ASIP* (MTRDHH5) as positional candidate genes already known to host variants controlling the horned/polled phenotype (OMIA 000483–9940) [47] and black coat color (OMIA 000201–9940) [48,49] in sheep, respectively. Interestingly, in MTR selection scheme, horned females and black animals do not fit with breed standard, and thus are not desirable. This may lead excluding these animals from genotyping which fully express their phenotype at the homozygous state. Accordingly, based on specific SNP markers already present on the LD SNP chip used, we were able to significantly associate MTRDHH5 with the 5pb deletion in *ASIP* leading to black coat color when homozygous. In contrast, the deficit observed for MTRDHH4 is significantly associated with the Ins allele of *RXFP2* leading to the desirable polledness trait. This intriguing observation suggests a hitchhiking phenomenon explaining the deficit [54], with the presence of a deleterious variant in linkage disequilibrium with the Ins allele of *RXFP2*.

The second group with MTRDHH2 (OAR1) associated with significant negative effects on AIS and SBR. Then, we hypothesized that the causative variant hosted by MTRDHH2 could induce embryo/fetal losses throughout the gestation period and until birth. In this region, only *SLC33A1* (*Solute Carrier Family 33 Member 1*) gene has both impact on embryonic lethality and decrease survival rate when knocked-out in mouse (MGI:1332247) and thus appears as an obvious candidate gene. In addition, variants in *SLC33A1* (OMIM 603690) are known to cause “Congenital cataracts, hearing loss, and neurodegeneration” and “Spastic paraplegia 42” phenotypes [55,56]. A study is ongoing to evidence a candidate causal variant affecting the *SLC33A1* gene in linkage disequilibrium with MTRDHH2.

MTRDHH1 (OAR20) was the only haplotype with a strong negative impact exclusively on SBR suggesting that this haplotype hosts a lethal variant affecting the perinatal period. Within the MTRDHH1 region, we identified *MMUT* (*Methylmalonyl-CoA Mutase*), *TFAP2B* (*Transcription Factor AP-2 Beta*) and *PKHD1* (*PKHD1 Ciliary IPT Domain Containing Fibrocystin/Polyductin*) as obvious candidate genes, all resulting in neonatal and/or postnatal lethality when knocked-out in mouse. *MMUT* and *PKHD1* are involved in metabolic disorders such as “Methylmalonic aciduria” (OMIM 251000) and “Polycystic kidney disease” (OMIM 263200), respectively. *TFAP2B* is involved in bone defects and heart failure (OMIM 169100). Using WGS data from 22 MTR animals and among them, two MTRDHH1 heterozygous carriers, we were able to identify four candidate polymorphisms within MTRDHH1, but only one appeared as a strong functional candidate, a SNV (G>A) located in the *MMUT* gene at position g.23,776,347 on OAR20. This SNV leads to a nonsense variation XM_004018875.4: c.1225C>T introducing a premature stop-codon (p.Gln409*). The genotyping of g.23,776,347G>A in 714 animals with a known status at MTRDHH1 indicated an almost perfect association. The only 15 discordant animals were largely explained by shorter recombinant version of the MTRDHH1 haplotype (between 2 to 31 markers surrounding the SNV). Only one MTRDHH1 heterozygous carrier did not carry the *MMUT* variant. This discrepancy could be attributed to errors from SNP array genotyping, phasing and/or imputation. Immunoblotting realized on liver and kidney proteins extracts from homozygous carrier lambs has confirmed the predicted impact of the candidate SNV on the protein. The anti-MMUT antibody has revealed a band at 83kDa as expected for the full-length MMUT polypeptide in wild-type extract. In contrast, due to the lack of the antigenic epitope (from aa

451 to 750) in the p.Gln409* truncated form, the MMUT band was not detected in homozygous carriers proving that this SNV has an effect on the functional expression of the *MMUT* gene, reinforced by a NMD phenomenon detected in liver. In order to confirm the perinatal lethality supposed for MTRDHH1 increasing SBR by 7.5%, we have managed at-risk mating between heterozygous carriers of the p.Gln409* variant. In this experiment, 84% of homozygous lambs died within the first 24 hours after birth fitting perfectly with the hypothesis and explaining the complete deficit of homozygous carriers of MTRDHH1 in the DHH analysis, due to lethality before the time of genotyping. We also evidenced that homozygous newborn lambs had a lower birth weight, an observation which could be compared to postnatal growth retardation in *Mmut* knock-out mice (MGI:5527455).

In human, numerous pathogenic variants in *MMUT* cause “Methylmalonic aciduria” (OMIM 609058, MMA, [57]), an autosomal recessive metabolism disorder. MMUT is part of a metabolic pathway starting from the degradation of amino acids (Valine, Isoleucine, Methionine and Threonine), odd-chain fatty acids, cholesterol and propionic acid to succinyl-CoA by three main enzymes: Propionyl-CoA carboxylase (PCC), Methylmalonyl-CoA epimerase (MCE) and Methylmalonyl-CoA mutase (MMUT) [58,59]. The MMUT protein is a mitochondrial enzyme that catalyze the L-methylmalonyl-CoA to succinyl-CoA, an intermediate of Krebs cycle. The isomerization of methylmalonyl-CoA requires adenosylcobalamin (AdoCbl), the cofactor form of vitamin B12 (also known as Cobalamin) [60]. In sheep, the mutant protein (p.409Gln*) does not carry the B12 binding domain (615-733 amino acids), suggesting that the AdoCbl cofactor is unable to act in the conversion of L-methylmalonyl-CoA to succinyl-CoA. When the MMUT enzyme is not functional, methylmalonic acid (MLA) accumulates in body fluids, mainly in blood and urine [59,61]. However, we failed in evidencing such MLA accumulation in plasma or urine of homozygous A/A lambs, while this was observed in homozygous knock-out mouse 24h after birth [62]. In our study, many of our urine and blood samples were collected at necropsy only after natural death of the lambs without time control, possibly affecting the results. The methylmalonic aciduria in this sheep genetic model will need further clinical investigation.

All the above elements clearly indicate that the deficit of homozygous MTRDHH4 is due to a loss-of-function mutation in the *MMUT* gene altering an essential metabolic pathway. Many reverse genetic screen approaches have also evidenced the association of DHH with mutations in genes implied in metabolism [63]. Particularly in bovine, many variants were evidenced affecting metabolic processes in several breeds such as Braunvieh (BH24/*CPT1C*, lipid metabolism) [19], Holstein (HH4/*GART*, nucleotide metabolism) [13], Montbéliarde (MH1/*PFAS*, nucleotide metabolism; MH2/*SLC37A2*, glucose metabolism) [13,40], Normande (NH7/*CAD*, nucleotide metabolism) [26], and Simmental (SH8/*CYP2B6*, respiratory chain) [27].

Diversity analysis have also revealed the segregation of *MMUT* SNV variant in Spanish Latxa Cara Rubia (LCR) breed. French MTR and Spanish LCR are very close populations. Since 1970s, many exchanges have occurred between these population across the border, first from Spain to France during the seventies, then the reverse since the nineties [64]. The *MMUT* SNV evidenced in this study is not shared by other French sheep breeds and more largely by the individuals from the International Sheep Genome Consortium (dataset composed of 453 animals from 38 breeds all over the world, <https://www.sheepmap.org/>). However, searching this dataset, we found another SNV located in ovine *MMUT* gene (rs1093255812) leading to a premature stop-gain (ENSOART00020022357.1: p.Trp43*). This variant was identified at heterozygous state in the New Zealand Coopworth breed, and as in MTR, it could have the same deleterious impact on lamb viability.

Conclusion

In this study, we firstly identified in the MTR dairy sheep the segregation of five independent haplotypes possibly hosting five recessive deleterious alleles of which at least three have negative impact on fertility traits. Among them, we evidenced the *MMUT* c.1225T (p.409Gln*) variant associated with the MTRDHH1 haplotype causing early lamb mortality when homozygous. This could provide an excellent animal model for the study of methylmalonic aciduria occurring in human with the same recessive genetic determinism affecting *MMUT*. The MTRDHH2 and 3 represent promising haplotypes to discover other recessive lethal mutations in the near future. This reverse genetic study has also allowed us to hypothesize that two of these haplotypes, MTRDHH4 and 5, do not associate with lethal mutations but with mutations possible affecting morphological traits. The causal mutations are still to be identified since the already known mutations in *RXFP2* and *ASIP* do not fit perfectly with the segregation of these two haplotypes. Anyway, an appropriate management of these haplotypes/variants in the MTR dairy sheep selection program should increase the overall fertility and lamb survival, and may help for selection of morphological breed standards.

Materials and methods

Animal and genotyping data

The total dataset is composed of 6,845 genotyped Manech Tête Rousse animals (82% males and 18% females) born between 1993 and 2021 (data description in S4 Fig). Genotyping was performed at the Labogena facility (<http://www.labogena.fr/>) in the framework of the French dairy sheep genomic selection [65]. Both low density SNP chip (SheepLD; n=2,956) and medium density SNP chip (MD, Ovine SNP50 BeadChip; n=3,889) were purchased from Illumina Inc. (San Diego, USA) (Table 5). Pedigree information was extracted from the official national database SIEOL (*Système d'Information en Elevage Ovin Laitier*, France).

Table 5. Description of genotyped animals

Year of birth	< 2017	≥ 2017		Total
Background	Research programs	Genomic selection & Research programs		
Number of animals	2,533 rams 692 ewes	3,077 rams 543 ewes		6,845
SNP chip	MD (n=3,225)	LD (n = 2,956)	MD (n = 664)	LD (n = 2,956) MD (n =3,889)
Genotyping age	>12 months	~15 days	8-12 months	

MD: medium density (50k), LD: low density (15k)

Genotype quality control, imputation and phasing

Quality control for each SNP was carried out following the French genomic evaluation pipeline based on (i) call frequency >97%, (ii) minor allele frequency >1%, (iii) respect Hardy-Weinberg equilibrium ($P > 10^{-5}$). Genotypes were phased and imputed from LD to MD using *FImpute* v2.2 [66]. The accuracy of LD to MD imputation in MTR was previously assessed and resulted in a concordance rate per animal of 98.86%, a concordance rate per SNP of 98.95% and a squared Pearson's correlation coefficient of 93.57% between imputed and observed SNP genotypes [67]. After quality control, the 38,523 remaining autosomal SNPs were mapped onto the *Ovis aries* genome assembly Oar v3.1 (current version used in the French genomic evaluation) [68]. These SNP were also located on the genome assembly Oar_rambouillet_v1.0 (GCF_002742125.1). Genomic coordinates of both version of the sheep genome are available at <https://doi.org/10.6084/m9.figshare.8424935.v2> (<https://www.sheepmap.org/>).

Detection of homozygous haplotype deficiency

Following the method used previously to detect deficit in homozygous haplotypes in sheep [42], we screened the genome of the 5,271 genotyped animals belonging to trios (77 offspring had both parents genotyped, and 4,799 offspring had both sire and maternal grandsire genotyped). Briefly, the method consists in (i) screening the genome using a sliding window of 20 SNP markers, (ii) selecting all 20 SNP haplotypes with frequency >1% on maternal phase, (iii) comparing the observed number $N_{Obs}(k)$ to the expected number $N_{Exp}(k)$ of homozygous offspring for each haplotype k using within-trios transmission probability and further considered haplotypes with $P\text{-poisson} < 1.9 \times 10^{-4}$, (iv) retaining deficit between 75% and 100% defined as $(N_{Exp}(k) - N_{Obs}(k))/N_{Exp}(k)$. Finally, consecutive windows with these same parameters were clustered to define larger region called “Manech Tête Rousse Deficient Homozygous Haplotype” (MTRDHH). For each MTRDHH, the status (homozygous non-carriers, heterozygous and homozygous carriers) was thereafter determined for all 6,845

genotyped animals available. Linkage disequilibrium was estimated between two MTRDHH regions located on the same chromosome by the r^2 coefficient measure as described [42].

Analysis of fertility traits

Mating trait records of MTR between 2006 and 2019 were obtained from the national database SIEOL. Only artificial insemination success (AIS) and stillbirth rate (SBR) records from mating where both sire and maternal grand sire were genotyped (i.e. had a known status at each MTRDHH) were analyzed. AIS was coded “1” for success and “0” for failure based on lambing date according to the gestation length starting from the day of AI (151 ± 7 days; $n=330,844$ mating records). SBR was determined only in the AI success group, and coded “1” if there was at least one stillbirth in the litter or “0” if all lambs were born alive ($n=201,637$ mating records). We considered “at-risk mating” a mating between a carrier ram and a ewe from a carrier sire. We considered “safe mating” when the other combinations occurred: (i) non-carrier ram \times ewe from a non-carrier sire, (ii) non-carrier ram \times ewe from a carrier sire, (iii) carrier ram \times ewe from a non-carrier sire. A logistic threshold binary model with a logit link function was used to compare AIS and SBR between at-risk and safe mating (lsmeans estimate), using the GLIMMIX procedure in the SAS software (version 9.4; SAS Institute Inc., Cary, NC). The fixed effects for AIS and SBR were mating type (safe or at-risk), season of AI (spring or summer), and lactation number (L1, L2, L3 and L4+). For SBR only, prolificacy of the ewe (1, 2, 3+ lambs/litter) was added as a fixed effect. For AIS and SBR, the random effect was the interaction herd \times year ($n=313$ herds between 2006 and 2019). Traits were considered to differ significantly when the mating type fixed effect had a P-value lower than 1.0×10^{-2} after Bonferroni correction for multiple testing with a level of significance α at 5%. This threshold was obtained by dividing the level of significance α by the number of tests corresponding to the number of independent haplotypes ($n=5$).

Analysis of milk parameters and total merit genomic index (ISOLg)

Daughter yield deviations (DYD) for milk traits from genotyped sires with known status at each MTRDHH were computed from official genetic evaluation (GenEval, Jouy-en-Josas, France). The DYD corresponds to the average performance of the daughters of each sire, corrected for environmental effects and the average genetic value of the dam [65]. The six traits studied were milk yield (MY), fat (FC) and protein (PC) contents, fat (FY = MY \times FC) and protein (PY = MY \times PC) yields, and lactation somatic cell score (LSCS) as described [42]. To compare all the traits on the same scale, each DYD was divided by its genetic standard deviation to obtain standardized DYD (sDYD). Only genotyped rams with records from at least 20 daughters were included in the analysis in order to obtain sufficiently accurate DYD values ($n \sim 2570$ rams). Each trait was tested by variance analysis comparing MTRDHH carrier and non-carrier rams using the GLM procedure in the SAS software (version 9.4; SAS Institute Inc., Cary, NC). The fixed effects were the genetic status (carrier, non-carrier) and year of birth (2000 to 2016) to correct for annual genetic gain. Traits were considered to differ significantly when the genetic status fixed effect had a P-value lower than 1.0×10^{-2} after Bonferroni correction for multiple testing with a level of significance α at 5%. This threshold was obtained as explained above for fertility traits ($n=5$ independent haplotypes).

Total merit genomic index in dairy sheep (called “ISOLg”, *Index Synthétique des Ovins Laitiers*) was extracted from the official genomic evaluation (GenEval, Jouy-en-Josas, France) for all the 714 genomic candidate lambs born in 2021. ISOLg is determined by a combination of four selected traits: MY, FC, PC and LSCS. For each MTRDHH, ISOLg from heterozygous carrier and non-carrier lambs were compared with a Wilcoxon non-parametric test under the null hypothesis with a risk of $\alpha=5\%$ using “wilcox.test” function in R software (version 4.1.3, R Core Team, 2022).

Identification of positional and functional candidate genes

All annotated genes located in the MTRDHH region extended by 1 Mb from each side were extracted from the ovine genome Oar_rambouillet_v1.0 (OAR1: NC_040252.1, OAR10: NC_040261.1, OAR13: NC_040264.1 and OAR20: NC_040271.1) using CLC export annotation function (QIAGEN CLC Main Workbench 7.9). Genes with a known knock-out phenotype in mouse including mortality and aging (embryonic, prenatal, perinatal, neonatal, postnatal, preweaning, premature death and decreased survival rate) or associated with mammalian genetic disorders were sorted using “biomaRt” R package (version 2.52.0, <https://doi.org/doi:10.18129/B9.bioc.biomaRt>) extracted from Mouse Genome Informatics (MGI, <http://www.informatics.jax.org>), International Mouse Phenotyping Consortium (IMPC, <https://www.mousephenotype.org>), Online Mendelian Inheritance in Man (OMIM, <https://omim.org>) and Online Mendelian Inheritance in Animal (OMIA, <https://omia.org>) databases (last accession on 27 May 2022). Relevant candidate genes were presented as a heatmap using “pheatmap” R package (version 1.0.12).

Whole genome sequencing data

Publicly available data of 100 ovine short-read Illumina HiSeq/NovaSeq whole genome sequences (WGS) from 14 breeds generated in various INRAE and Teagasc research projects were used for variant calling. Among them, 22 WGS were obtained from MTR dairy sheep also genotyped with the MD SNP chip. A description of the different breeds and the accession numbers of sequencing raw data are available in S3 Table.

WGS variant calling and filtering

Reads mapping, variant calling and functional annotation were performed using Nextflow v20.10.0 and Sarek v2.6.1 pipelines for the 100 short-read WGS as previously described [43]. Regions of interest were extracted using SnpSift Filter [50]. Candidate variants were filtered based on the correlation between haplotype status (homozygous non-carriers, heterozygous and homozygous carriers encoded as 0, 1 and 2, respectively) and allele dosage for bi-allelic variants (also encoded 0, 1 and 2) using `geno--r2` command of VCFtools [69].

Specific variant genotyping assays

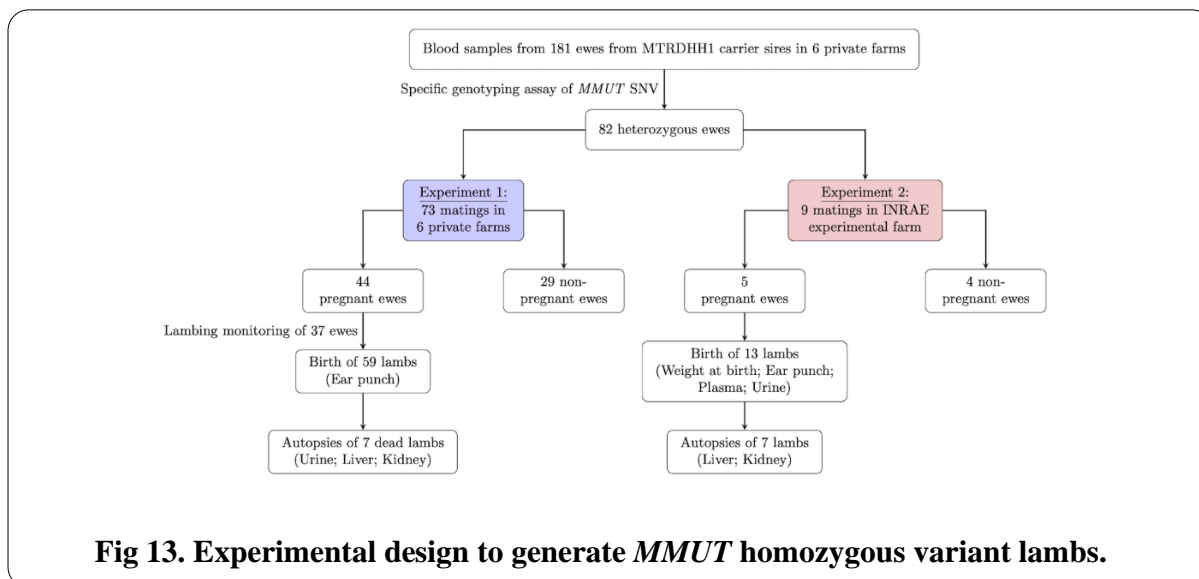
Genotyping at the *RXFP2* (1.8 kb InDel) and *ASIP* (5pb InDel) loci were directly obtained from the LD SNP chip based on the specific markers described in S4 Table. The NC_040264.1:g.66,474,980T>A in the last exon of *ASIP* and the SNP NC_040271.1:g.23,776,347G>A in *MMUT* were both genotyped by PACE (PCR allele competitive extension) analysis. Fluorescent PACE analysis was done with 15 ng of purified DNA using the PACE-IR 2x Genotyping Master mix (3CR Bioscience) in the presence of 12 μ M of a mix of extended allele specific forward primers and 30 μ M of common reverse primers in a final volume of 10 μ L (primer sequences described in S5 Table). The touch-down PCR amplification condition was 15 min at 94°C for the hot-start activation, 10 cycles of 20 s at 94°C, 54–62°C for 60 s (dropping 0.8°C per cycle), then 36 cycles of 20 s at 94°C and 60 s at 54°C performed on an ABI9700 thermocycler followed by a final point read of the fluorescence on an ABI QuantStudio 6 real-time PCR system and using the QuantStudio software 1.3 (Applied Biosystems).

The presence of the *RXFP2*, *ASIP* and *MMUT* variants was checked in a DNA set of the 2021 cohort of 714 MTR male lamb candidates for genomic selection. DNA was extracted by Labogena (Jouy-en-Josas, France) on behalf of the MTR breed industry. A DNA diversity panel

of 851 animals from 25 French sheep breeds [70] and 3 Spanish sheep breeds was also specifically genotyped for the *MMUT* variant.

Generation of homozygous lambs

Blood samples (3 ml) were first collected from 181 ewes, daughters of MTRDHH1 carrier sires, located in 6 private farms by jugular vein puncture with the Venoject system containing EDTA (Terumo, Tokyo, Japan) and directly stored at -20°C . Among them, 82 ewes were genotyped as heterozygous carriers of the *MMUT* variant and were selected to be inseminated with *MMUT* variant heterozygous rams (at-risk matings). Experiment 1 (n=73 ewes) was performed in 6 private farms in Pays-Basque (France) and Experiment 2 (n=9 ewes) was performed at the INRAE experimental farm of Langlade under the agreement number E31429001 (Pompertuzat, France). Experimental design is described in Fig 13. An ultrasound diagnosis of gestation was realized between 45 and 60 days after AI. Gestations were followed and each lamb was monitored from birth to weaning. Ear biopsies (1 mm^3) from the 72 lambs born in both experiments were obtained with a tissue sampling unit (TSU, Allflex Europe, Vitré, France) and directly placed in the TSU storage buffer at 4°C . Ear biopsies were placed twice consecutively in $180\text{ }\mu\text{L}$ of 50 mM NaOH , heated 10 min at 95°C , neutralized with $20\text{ }\mu\text{L}$ of 1 M Tris-HCl , and then vortexed during 10s. All neutralized samples were used for direct genotyping without DNA purification as described [71]. In Experiment 2, all lambs were weighted at birth. Biological samples (plasma, urine, liver and kidney) were collected on animals as described in Fig 13 and frozen at -80°C until use.



Immunoblot analysis

Frozen kidney and liver tissues were crushed in liquid nitrogen using a Mixer Mill during 30 seconds at 30 Hz (MM400, Retsch technology), and 12 mg of tissue powder were lysed with $500\text{ }\mu\text{L}$ of RIPA solution (Ref#R0278, Sigma-Aldrich). The protein extracts were centrifuged for 20 minutes at $16\text{ }000g$ and 4°C , and protein concentration in the supernatant was determined using BCA protein assay Kit (Ref#K812-1000, Biovision). Each protein sample ($45\text{ }\mu\text{g}$) was denatured and reduced in Laemmli buffer ($62.5\text{ mM TRIS pH }6.8$, 2% SDS; 10% glycerol) containing 5% β -mercaptoethanol before SDS-PAGE on a 4 - 15% polyacrylamide gel (Bio-

Rad). Proteins were transferred onto a nitrocellulose membrane blocked with AdvanBlock-Chemi blocking solution (Advansta) during 1 hour at room temperature. After washing in PBS-0.1% Tween 20, the membrane was incubated overnight à 4°C with a rabbit polyclonal anti-MMUT primary antibody (MUT Rabbit pAb, Ref#A3969, ABclonal), followed (after washing) by 1h with a rabbit polyclonal anti-Actin primary antibody (Ref#A2066, Sigma-Aldrich) both at 1/1000 in blocking solution. Revelation of the primary antibodies was performed by incubation with goat anti-rabbit Horseradish peroxidase-conjugated secondary antibody (Ref#A0545, Sigma-Aldrich) at 1/10000 ratio in blocking solution for 1 hour at room temperature, followed by enhanced chemiluminescence detection (WesternBright Quantum HRP substrate, Advansta) on a ChemiDoc Touch low-light camera (Bio-Rad) in automatic mode.

RNA extraction, reverse transcription and quantitative PCR

Total RNA was extracted from 80 mg of frozen kidney (n=15) and liver tissue powders (n=15) in 1mL Trizol reagent (Invitrogen, #Ref 15596-018) and isolated from Nucleospin® RNA II kit (Macherey-Nagel, #Ref 740955.50) according to the manufacturer's protocol and including DNaseI digestion treatment. The RNAs were quantified by spectrophotometry (NanoDrop® ND-8000 spectrophotometer, ThermoFischer) and stored at -80°C. Reverse transcription was carried out from 1µg of total RNA in solution with anchored oligo(dT) T22V (1µL at 10µM), random oligo-dN9 (1µL at 10µM) and dNTPs (2µL at 10 mM) in a reaction volume of 10 µL. This mixture was incubated at 65°C for 5 min in an ABI2700 thermocycler (Applied Biosystems) then ramped down to 4°C. A second reaction mixture (8 µL/reaction) containing the reaction buffer (5µL of First strand Buffer 5X, Invitrogen, France), DTT (Dithiothreitol, 1µL at 0.1M), Rnasine (1µL, 40 units/µL, Promega, France) and Superscript II reverse transcriptase (1µL, 200 units/µL, Invitrogen, France) was added to the denatured RNA solution (final volume reaction of 18µL) then incubated for 50 minutes at 42°C and placed for 15 minutes at 70°C. The complementary DNA (cDNA) solution obtained was directly diluted at 1:5 ratio and stored at -20°C. For each pair of primers, amplification efficiency was evaluated by $E = e^{-1/\alpha}$ which α is the slope of a linear curve obtained from cDNA serial dilution (1:5 to 1:80) and corresponding Ct (cycle threshold) values. Quantitative PCR (qPCR) was performed using 3µL of cDNA at 1:20 ratio, 5µL of SYBR Green real-time PCR Master Mix 2X (Applied Biosystems) and 2µL of primers at 3µM in a total reaction volume of 10µL on qPCR was realized on a QuantStudio 6 Flex Real-Time PCR system (ThermoFisher). Each sample was tested in duplicate. RNA transcript abundance was quantified using the $\Delta\Delta Ct$ method corrected by four reference genes (*GAPDH*, *YWHAZ*, *RPL19* and *SDHA*) and a calibrator sample. Primers were design using Beacon Designer™ 8 (Premier Biosoft). The list of qPCR primer sequences, amplification length and amplification efficiency used is available in S5 Table.

Methylmalonic acid dosage assay

Dosage of methylmalonic acid (MLA) was performed on 11 urine and 15 plasma samples collected on lambs. MLA dosage was performed using Sheep Methylmalonic Acid Elisa kit (MyBioSource, Ref# MBS7266308) following the manufacture's protocol starting with 100µL of samples. The incubation phase with specific antibody was performed overnight at 4°C. The optical density at 450 nm was determined on a GloMax®-Multi Detection System (Promega).

Ethic statement

The experimental procedures on animals were approved (approval numbers APAFIS#30615-2021032318054889 v5) 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.

Acknowledgments

We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing help and/or computing and/or storage resources. The authors thank the Manech Tête Rousse breeding company who provided access to private genomic data and biological samples. We thank also the breeders involved in the project and Soline Szymczak for providing help for genotyping. We are grateful to NEIKER for providing samples for Spanish ovine diversity panel.

Funding

This research has received funding from the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 772787 (SMARTER) and PRESAGE project (CASDAR n°20ART1532777, the responsibility of the French ministry of Agriculture and Food cannot be engaged). MB was supported by a Ph.D. grant for the HOMLET program co-funded by APIS-GENE and Région Occitanie.

References

1. Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. *Nature Reviews Genetics*. 2019;20: 135–156. doi:10.1038/s41576-018-0082-2
2. Rupp R, Mucha S, Larroque H, McEwan J, Conington J. Genomic application in sheep and goat breeding. *Animal Frontiers*. 2016;6: 39–44. doi:10.2527/af.2016-0006
3. Eggen A. The development and application of genomic selection as a new breeding paradigm. *Animal Frontiers*. 2012;2: 10–15. doi:10.2527/af.2011-0027
4. Chamberlain AJ, Cheng HH, Giuffra E, Kuehn C, Tuggle CK, Zhou H. Editorial: Functional Annotation of Animal Genomes. *Frontiers in Genetics*. 2021;12. Available: <https://www.frontiersin.org/article/10.3389/fgene.2021.768626>
5. Pryce JE, Royal MD, Garnsworthy PC, Mao IL. Fertility in the high-producing dairy cow. *Livestock Production Science*. 2004;86: 125–135. doi:10.1016/S0301-6226(03)00145-3
6. Berry DP, Wall E, Pryce JE. Genetics and genomics of reproductive performance in dairy and beef cattle. *Animal*. 2014;8: 105–121. doi:10.1017/S1751731114000743
7. David I, Astruc JM, Lagriffoul G, Manfredi E, Robert-Granié C, Bodin L. Genetic Correlation Between Female Fertility and Milk Yield in Lacaune Sheep. *Journal of Dairy Science*. 2008;91: 4047–4052. doi:10.3168/jds.2008-1113
8. Capitan A, Michot P, Baur A, Saintilan R, Hozé C, Valour D, et al. Genetic tools to improve reproduction traits in dairy cattle. *Reprod Fertil Dev*. 2015;27: 14–21. doi:10.1071/RD14379
9. Duchesne A, Grohs C, Michot P, Bertaud M, Boichard D, Floriot S, et al. From phenotype to the causal mutation: recessive defects. *INRAE Productions Animales*. 2016;29: 319–328. doi:10.20870/productions-animales.2016.29.5.3000
10. Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, et al. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat Genet*. 2008;40: 449–454. doi:10.1038/ng.96
11. Lander ES, Botstein D. Homozygosity Mapping: A Way to Map Human Recessive Traits with the DNA of Inbred Children. *Science*. 1987;236: 1567–1570. doi:10.1126/science.2884728
12. VanRaden PM, Olson KM, Null DJ, Hutchison JL. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy Science*. 2011;94: 6153–6161. doi:10.3168/jds.2011-4624

13. Fritz S, Capitan A, Djari A, Rodriguez SC, Barbat A, Baur A, et al. Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2. *PLOS ONE*. 2013;8: e65550. doi:10.1371/journal.pone.0065550
14. Pausch H, Schwarzenbacher H, Burgstaller J, Flisikowski K, Wurmser C, Jansen S, et al. Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. *BMC Genomics*. 2015;16: 312. doi:10.1186/s12864-015-1483-7
15. Jenko J. Analysis of a large dataset reveals haplotypes carrying putatively recessive lethal and semi-lethal alleles with pleiotropic effects on economically important traits in beef cattle. 2019; 14.
16. Hoff JL, Decker JE, Schnabel RD, Taylor JF. Candidate lethal haplotypes and causal mutations in Angus cattle. *BMC Genomics*. 2017;18: 799. doi:10.1186/s12864-017-4196-2
17. Cooper TA, Wiggans GR, Null DJ, Hutchison JL, Cole JB. Genomic evaluation, breed identification, and discovery of a haplotype affecting fertility for Ayrshire dairy cattle. *Journal of Dairy Science*. 2014;97: 3878–3882. doi:10.3168/jds.2013-7427
18. Null D, Hutchison J, Bickhart D, VanRaden P, Cole J. Discovery of a haplotype affecting fertility in Ayrshire dairy cattle and identification of a putative causal variant. *Journal of Dairy Science*. 2017;100(Suppl. 2): 199(abstr. 205).
19. Häfliger IM, Seefried FR, Spengeler M, Drögemüller C. Mining massive genomic data of two Swiss Braunvieh cattle populations reveals six novel candidate variants that impair reproductive success. *Genetics Selection Evolution*. 2021;53: 95. doi:10.1186/s12711-021-00686-3
20. Sahana G, Nielsen US, Aamand GP, Lund MS, Guldbbrandtsen B. Novel Harmful Recessive Haplotypes Identified for Fertility Traits in Nordic Holstein Cattle. *PLoS ONE*. 2013;8: e82909. doi:10.1371/journal.pone.0082909
21. McClure MC, Bickhart D, Null D, VanRaden P, Xu L, Wiggans G, et al. Bovine Exome Sequence Analysis and Targeted SNP Genotyping of Recessive Fertility Defects BH1, HH2, and HH3 Reveal a Putative Causative Mutation in SMC2 for HH3. *PLOS ONE*. 2014;9: e92769. doi:10.1371/journal.pone.0092769
22. Cooper TA, Wiggans GR, VanRaden PM, Hutchison JL, Cole JB, Null DJ. Genomic evaluation of Ayrshire dairy cattle and new haplotypes affecting fertility and stillbirth in Holstein, Brown Swiss and Ayrshire breeds. 2013; ADSA-ASAS Joint Annual Meeting :poster T206.
23. Fritz S, Hozé C, Rebours E, Barbat A, Bizard M, Chamberlain A, et al. An initiator codon mutation in SDE2 causes recessive embryonic lethality in Holstein cattle. *Journal of Dairy Science*. 2018;101: 6220–6231. doi:10.3168/jds.2017-14119
24. Hozé C, Fritz S, Baur A, Grohs C, Danchin-Burge C, Boichard D. Accounting for genetic conditions in breeding objectives in dairy cattle. In *Proceedings of the 24ème Rencontres Autour des Recherches Sur les Ruminants: 5-6 December*. Paris; 2018.
25. Häfliger IM, Spengeler M, Seefried FR, Drögemüller C. Four novel candidate causal variants for deficient homozygous haplotypes in Holstein cattle. *Sci Rep*. 2022;12: 5435. doi:10.1038/s41598-022-09403-6
26. Mesbah-Uddin M, Hozé C, Michot P, Barbat A, Lefebvre R, Boussaha M, et al. A missense mutation (p.Tyr452Cys) in the CAD gene compromises reproductive success in French Normande cattle. *Journal of Dairy Science*. 2019;102: 6340–6356. doi:10.3168/jds.2018-16100

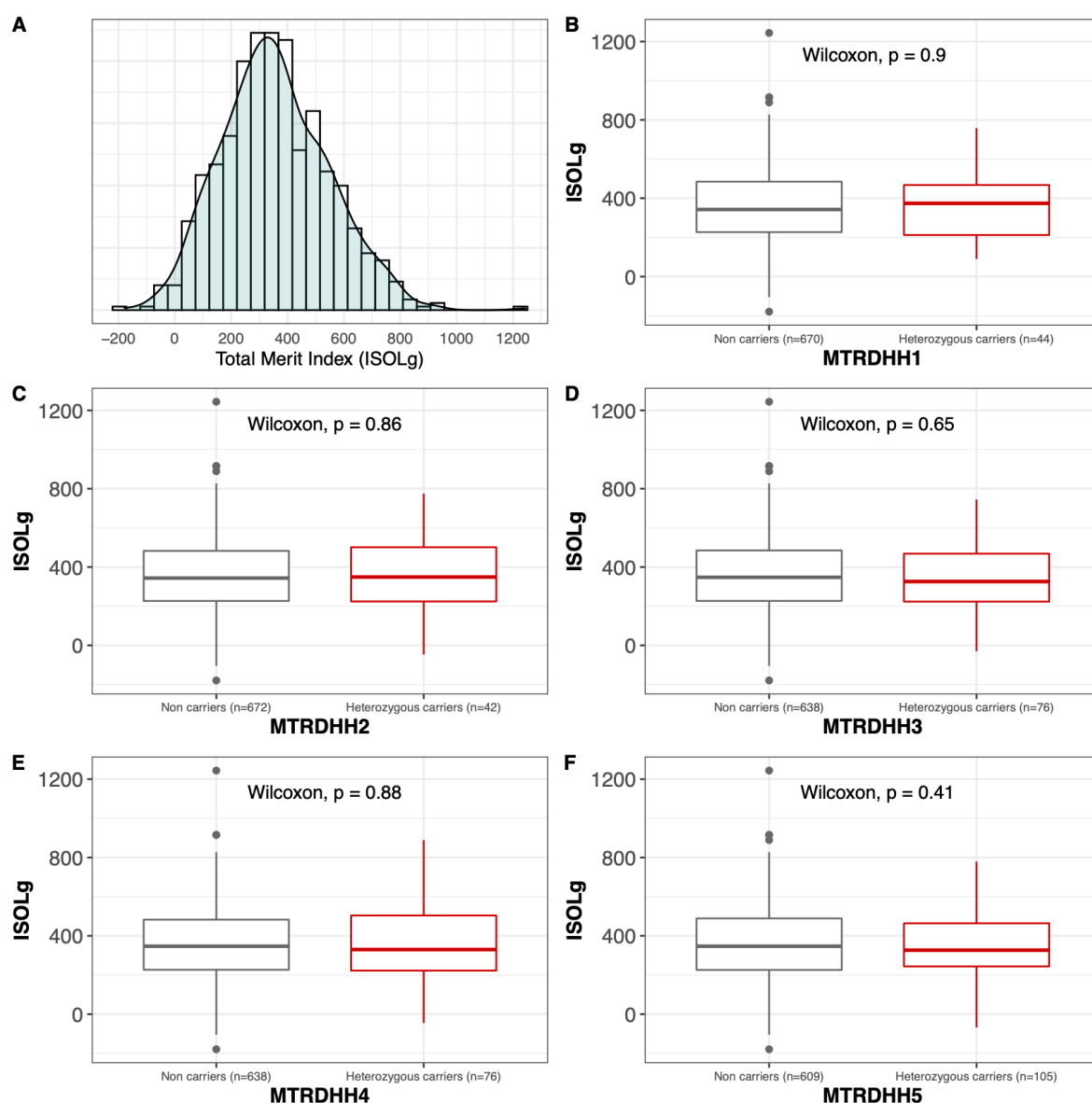
27. Häfliger IM, Seefried FR, Drögemüller C. Reverse Genetic Screen for Deleterious Recessive Variants in the Local Simmental Cattle Population of Switzerland. *Animals*. 2021;11: 3535. doi:10.3390/ani11123535
28. Derks MFL, Megens H-J, Bosse M, Lopes MS, Harlizius B, Groenen MAM. A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics*. 2017;18: 858. doi:10.1186/s12864-017-4278-1
29. Derks MFL, Lopes MS, Bosse M, Madsen O, Dibbits B, Harlizius B, et al. Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. Leeb T, editor. *PLoS Genet*. 2018;14: e1007661. doi:10.1371/journal.pgen.1007661
30. Derks MFL, Megens H-J, Bosse M, Visscher J, Peeters K, Bink MCAM, et al. A survey of functional genomic variation in domesticated chickens. *Genet Sel Evol*. 2018;50: 17. doi:10.1186/s12711-018-0390-1
31. Abdalla EA, Id-Lahoucine S, Cánovas A, Casellas J, Schenkel FS, Wood BJ, et al. Discovering lethal alleles across the turkey genome using a transmission ratio distortion approach. *Animal Genetics*. 2020;51: 876–889. doi:10.1111/age.13003
32. Todd ET, Thomson PC, Hamilton NA, Ang RA, Lindgren G, Viklund Å, et al. A genome-wide scan for candidate lethal variants in Thoroughbred horses. *Sci Rep*. 2020;10: 13153. doi:10.1038/s41598-020-68946-8
33. Venhoranta H, Pausch H, Flisikowski K, Wurmser C, Taponen J, Rautala H, et al. In frame exon skipping in UBE3B is associated with developmental disorders and increased mortality in cattle. *BMC Genomics*. 2014;15: 890. doi:10.1186/1471-2164-15-890
34. Schwarzenbacher H, Burgstaller J, Seefried FR, Wurmser C, Hilbe M, Jung S, et al. A missense mutation in TUBD1 is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. *BMC Genomics*. 2016;17: 400. doi:10.1186/s12864-016-2742-y
35. Burgstaller J, Url A, Pausch H, Schwarzenbacher H, Egerbacher M, Wittek T. Clinical and biochemical signs in Fleckvieh cattle with genetically confirmed Fanconi-Bickel syndrome (cattle homozygous for Fleckvieh haplotype 2). *Berl Munch Tierarztl Wochenschr*. 2016;129: 132–137.
36. Adams HA, Sonstegard TS, VanRaden PM, Null DJ, Van Tassell CP, Larkin DM, et al. Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *Journal of Dairy Science*. 2016;99: 6693–6701. doi:10.3168/jds.2015-10517
37. Yang Y, Si J, Lv X, Dai D, Liu L, Tang S, et al. Integrated analysis of whole genome and transcriptome sequencing reveals a frameshift mutation associated with recessive embryonic lethality in Holstein cattle. *Animal Genetics*. 2021;n/a. doi:10.1111/age.13160
38. Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*. 2014;46: 858–865. doi:10.1038/ng.3034
39. Schütz E, Wehrhahn C, Wanjek M, Bortfeld R, Wemheuer WE, Beck J, et al. The Holstein Friesian Lethal Haplotype 5 (HH5) Results from a Complete Deletion of TBF1M and Cholesterol Deficiency (CDH) from an ERV-(LTR) Insertion into the Coding Region of APOB. *PLOS ONE*. 2016;11: e0154602. doi:10.1371/journal.pone.0154602
40. Michot P, Fritz S, Barbat A, Boussaha M, Deloche M-C, Grohs C, et al. A missense mutation in PFAS (phosphoribosylformylglycinamide synthase) is likely causal for embryonic lethality associated with the MH1 haplotype in Montbéliarde dairy cattle. *Journal of Dairy Science*. 2017;100: 8176–8187. doi:10.3168/jds.2017-12579
41. Sonstegard TS, Cole JB, VanRaden PM, Van Tassell CP, Null DJ, Schroeder SG, et al. Identification of a Nonsense Mutation in CWC15 Associated with Decreased Reproductive Efficiency in Jersey Cattle. *PLOS ONE*. 2013;8: 54872. doi:10.1371/journal.pone.0054872

42. Ben Braiek M, Fabre S, Hozé C, Astruc J-M, Moreno-Romieux C. Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep. *Genetics Selection Evolution*. 2021;53: 41. doi:10.1186/s12711-021-00634-1
43. Ben Braiek M, Moreno-Romieux C, Allain C, Bardou P, Bordes A, Debat F, et al. A Nonsense Variant in CCDC65 Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep. *Genes*. 2022;13: 45. doi:10.3390/genes13010045
44. Buisson D, Astruc J-M, Barillet F. Overview and perspectives of the management of genetic diversity in French dairy sheep. *INRAE Productions Animales*. 2018;31: 1–12. doi:10.20870/productions-animales.2018.31.1.2202
45. IDELE. VARUME. In: idele.fr [Internet]. 2016 [cited 14 Jun 2023]. Available: <https://idele.fr/detail-article/indicateurs-de-variabilite-genetique-races-ovines-laitieres-edition-2016>
46. Astruc J-M, Baloché G, Buisson D, Labatut J, Lagriffoul G, Larroque H, et al. Genomic selection in French dairy sheep. *Inra Prod Anim*. 2016;29: 41–55.
47. Wiedemar N, Drögemüller C. A 1.8-kb insertion in the 3'-UTR of RXFP2 is associated with polledness in sheep. *Animal Genetics*. 2015;46: 457–461. doi:https://doi.org/10.1111/age.12309
48. Smit MA, Shay TL, Beever JE, Notter DR, Cockett NE. Identification of an agouti-like locus in sheep. *Animal Genetics*. 2002;33: 383–385. doi:10.1046/j.1365-2052.2002.00896_5.x
49. Norris BJ, Whan VA. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Res*. 2008;18: 1282–1293. doi:10.1101/gr.072090.107
50. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, et al. Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Frontiers in Genetics*. 2012;3: 35. doi:10.3389/fgene.2012.00035
51. Clark EL, Bush SJ, McCulloch MEB, Farquhar IL, Young R, Lefevre L, et al. A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLOS Genetics*. 2017;13: e1006997. doi:10.1371/journal.pgen.1006997
52. Derks MFL, Steensma M. Review: Balancing Selection for Deleterious Alleles in Livestock. *Frontiers in Genetics*. 2021;12. Available: <https://www.frontiersin.org/article/10.3389/fgene.2021.761728>
53. Beever JE, Smit MA, Meyers SN, Hadfield TS, Bottema C, Albrechtsen J, et al. A single-base change in the tyrosine kinase II domain of ovine FGFR3 causes hereditary chondrodysplasia in sheep. *Animal Genetics*. 2006;37: 66–71. doi:10.1111/j.1365-2052.2005.01398.x
54. Boichard D, Grohs C, Danchin-Burge C, Capitan A. Genetic defects: definition, origin, transmission and evolution, and mode of action. *INRAE Productions Animales*. 2016;29: 297–306. doi:10.20870/productions-animales.2016.29.5.2997
55. Lin P, Li J, Liu Q, Mao F, Li J, Qiu R, et al. A Missense Mutation in SLC33A1, which Encodes the Acetyl-CoA Transporter, Causes Autosomal-Dominant Spastic Paraplegia (SPG42). *The American Journal of Human Genetics*. 2008;83: 752–759. doi:10.1016/j.ajhg.2008.11.003
56. Huppke P, Brendel C, Kalscheuer V, Korenke GC, Marquardt I, Freisinger P, et al. Mutations in SLC33A1 Cause a Lethal Autosomal-Recessive Disorder with Congenital Cataracts, Hearing Loss, and Low Serum Copper and Ceruloplasmin. *The American Journal of Human Genetics*. 2012;90: 61–68. doi:10.1016/j.ajhg.2011.11.030

57. Acquaviva C, Benoist J-F, Pereira S, Callebaut I, Koskas T, Porquet D, et al. Molecular basis of methylmalonyl-CoA mutase apoenzyme defect in 40 European patients affected by mut^o and mut⁻ forms of methylmalonic acidemia: Identification of 29 novel mutations in the MUT gene. *Human Mutation*. 2005;25: 167–176. doi:10.1002/humu.20128
58. Baumgartner MR, Hörster F, Dionisi-Vici C, Haliloglu G, Karall D, Chapman KA, et al. Proposed guidelines for the diagnosis and management of methylmalonic and propionic acidemia. *Orphanet J Rare Dis*. 2014;9: 130. doi:10.1186/s13023-014-0130-8
59. Deodato F, Boenzi S, Santorelli FM, Dionisi-Vici C. Methylmalonic and propionic aciduria. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*. 2006;142C: 104–112. doi:10.1002/ajmg.c.30090
60. González-Montaña J-R, Escalera-Valente F, Alonso AJ, Lomillos JM, Robles R, Alonso ME. Relationship between Vitamin B12 and Cobalt Metabolism in Domestic Ruminant: An Update. *Animals*. 2020;10: 1855. doi:10.3390/ani10101855
61. Lucienne M, Aguilar-Pimentel JA, Amarie OV, Becker L, Calzada-Wack J, da Silva-Buttkus P, et al. In-depth phenotyping reveals common and novel disease symptoms in a hemizygous knock-in mouse model (Mut-ko/ki) of mut-type methylmalonic aciduria. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. 2020;1866: 165622. doi:10.1016/j.bbadis.2019.165622
62. Peters H, Nefedov M, Sarsero J, Pitt J, Fowler KJ, Gazeas S, et al. A Knock-out Mouse Model for Methylmalonic Aciduria Resulting in Neonatal Lethality*. *Journal of Biological Chemistry*. 2003;278: 52909–52913. doi:10.1074/jbc.M310533200
63. Charlier C, Li W, Harland C, Littlejohn M, Coppieters W, Creagh F, et al. NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res*. 2016;26: 1333–1341. doi:10.1101/gr.207076.116
64. Legarra A, Baloche G, Barillet F, Astruc JM, Soulas C, Aguerre X, et al. Within- and across-breed genomic predictions and genomic relationships for Western Pyrenees dairy sheep breeds Latxa, Manech, and Basco-Béarnaise. *Journal of Dairy Science*. 2014;97: 3200–3212. doi:10.3168/jds.2013-7745
65. Astruc J-M, Baloche G, Buisson D, Labatut J, Lagriffoul G, Larroque H, et al. Genomic selection in French dairy sheep. *Inra Prod Anim*. 2016;29: 41–55.
66. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*. 2014;15: 478. doi:10.1186/1471-2164-15-478
67. Larroque H, Astruc J-M. Integration of a low density chip in dairy sheep genomic selection. 2017 p. 27.
68. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, et al. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*. 2014;344: 1168–1173. doi:10.1126/science.1252806
69. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27: 2156–2158. doi:10.1093/bioinformatics/btr330
70. 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: 71. doi:10.1186/s12864-018-4447-x
71. Chantepie L, Bodin L, Sarry J, Woloszyn F, Plisson-Petit F, Ruesche J, et al. Genome-Wide Identification of a Regulatory Mutation in BMP15 Controlling Prolificacy in Sheep. *Front Genet*. 2020;11. doi:10.3389/fgene.2020.00585

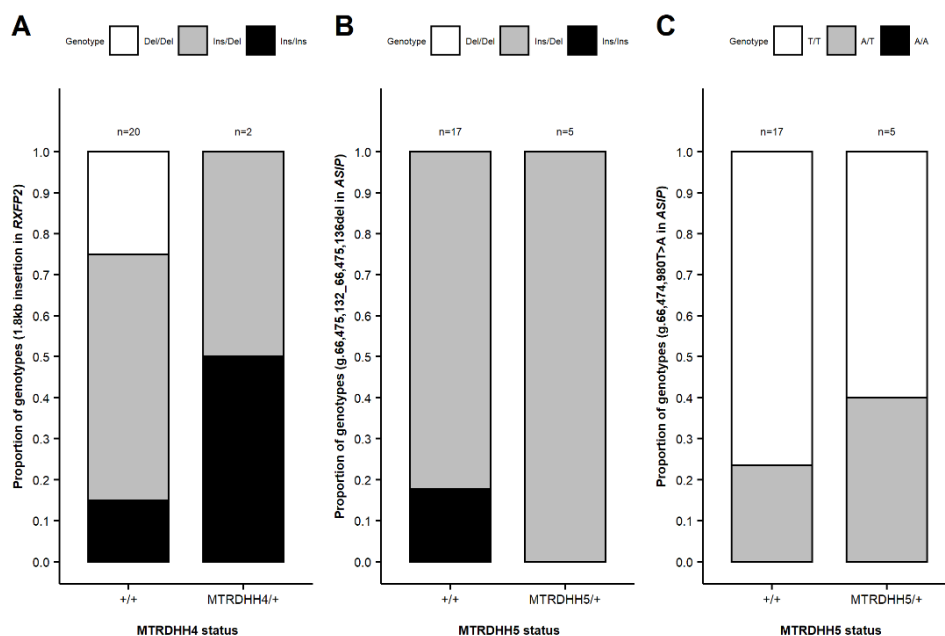
Supporting information captions

S1A Table. Clustering HHD into MTRDHH regions. Table shows all significant haplotypes of 20 markers (150 HHD with frequency > 1%, P-value < 1.9×10^{-4} and deficit $\geq 75\%$). As described in the Materials and methods section, the 150 HHD could be clustered into 5 MTRDHH regions. **S1B Table. SNPs defining the MTRDHH regions.** Table gives the position of each SNP within MTRDHH regions according to the sheep reference genomes Oar_v3.1, Oar_rambouillet_v1.0 and ARS-UI_Ramb_v2.0, and the phased alleles of each deficient haplotype.



S1 Fig. Total merit genomic index (ISOLg) for the 2021 cohort genomic lambs (n=714). (A) Distribution of ISOLg in MTR dairy sheep. The ISOLg is determined by a combination of four selected traits: MY, FC, PC and LSCS. Comparison of ISOLg according to each DHH status, (B) MTRDHH1, (C) MTRDHH2, (D) MTRDHH3, (E) MTRDHH4, (F) MTRDHH5.

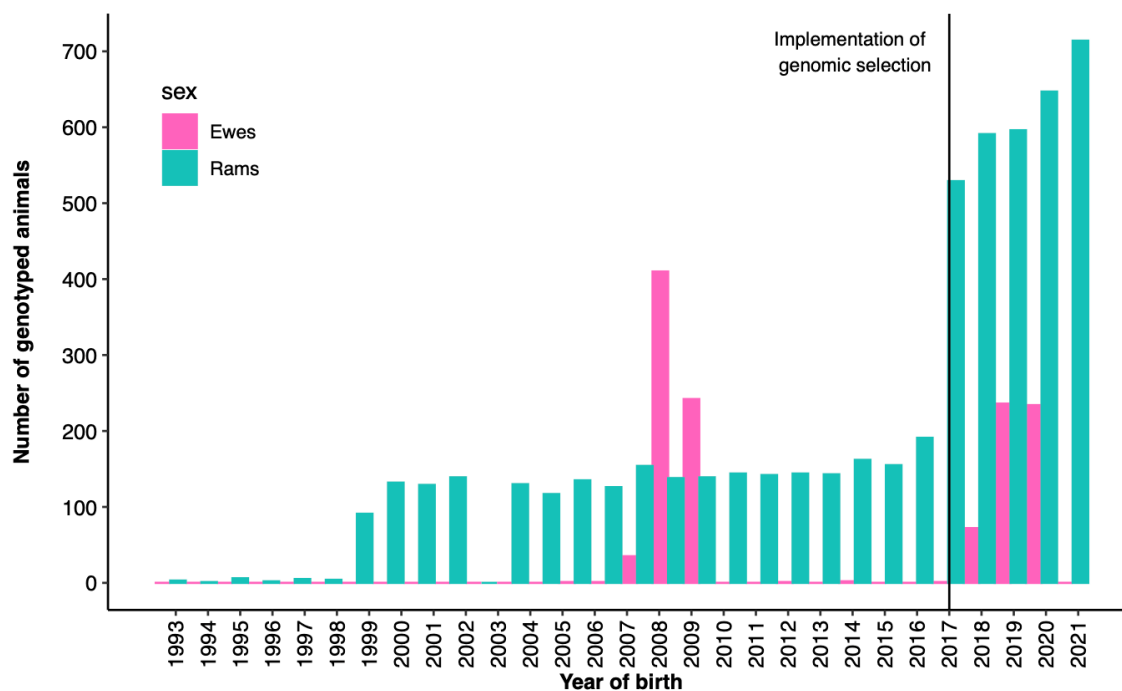
S2 Table. List of the 408 protein coding genes located in the five MTRDHH extended by 1 Mb on each side. Information on mouse phenotypes and association with mammalian disorders were extracted for each gene using several databases: MGI: www.informatics.jax.org; IMPC: <https://www.mousephenotype.org>; OMIM: Online Mendelian Inheritance in Man (<https://omim.org>) and OMIA: Online Mendelian Inheritance in Animal (<https://omia.org>). For mouse phenotypes associated with lethality, these were separated by developmental stages and encoded 0 (no lethality) or 1 if the gene is involved in lethality at the given stage.



S2 Fig. Association between MTRDHH4-5 and causal variants involved in morphological traits from the 22 sequenced MTR animals. (A) Association between MTRDHH4 status and 1.8kb insertion in the 3'-UTR of *RXFP2* associated with polledness. Association between MTRDHH5 status and (B) OAR13:g.66,475,132_66,475,136del and (C) OAR13:g.66,474,980T>A in *ASIP* associated with black coat color. Coordinates refer to sheep genome Oar_rambouillet_v1.0.

Animal	MTRDHH1 haplotype (32 markers)		MMUT genotype
	MTRDHH1 status	MMUT SNV (G>A)	
Animal 1	+/+	G/A	s75212.1 [A/G]
Animal 2	+/+	G/A	OAR20_22661712_X.1 [A/G]
Animal 3	+/+	G/A	OAR20_22784700.1 [A/G]
Animal 4	+/+	G/A	OAR20_22851804.1 [A/C]
Animal 5	+/+	G/A	s39376.1 [G/A]
Animal 6	+/+	G/A	OAR20_22932011_X.1 [A/G]
Animal 7	+/+	G/A	OAR20_22996638.1 [G/A]
Animal 8	+/+	G/A	OAR20_23028179.1 [G/A]
Animal 9	+/+	G/A	OAR20_23063796.1 [A/C]
Animal 10	+/+	G/A	OAR20_23099257_X.1 [G/A]
Animal 11	+/+	G/A	OAR20_23116747.1 [A/G]
Animal 12	+/+	G/A	s66452.1 [A/G]
Animal 13	+/+	G/A	OAR20_23240800.1 [A/G]
Animal 14	+/+	G/A	MMUT SNV (G>A)
Animal 15	MTRDHH1/+	G/C	s04959.1 [G/A]
			s12178.1 [A/G]
			OAR20_23408992.1 [A/C]
			OAR20_23455049.1 [A/G]
			OAR20_23481589_X.1 [G/A]
			OAR20_23552034_X.1 [A/G]
			OAR20_23606688.1 [A/G]
			OAR20_23662826.1 [A/C]
			OAR20_23756997.1 [A/G]
			OAR20_23854908.1 [G/A]
			OAR20_23878963.1 [A/G]
			OAR20_23898925.1 [A/G]
			OAR20_24029500.1 [G/A]
			OAR20_24071930.1 [A/C]
			OAR20_24089242_X.1 [G/A]
			OAR20_24307087.1 [G/A]
			OAR20_24357620.1 [A/G]
			OAR20_24553232.1 [G/A]
			OAR20_24583511.1 [A/G]

S3 Fig. MTRDHH1 recombinant haplotypes from 15 animals showing mismatch between MTRDHH1 status and MMUT SNV genotype. MTRDHH1/+ and +/- refer to heterozygous and non-carriers of MTRDHH1, respectively. The grey column represents the localization of the MMUT SNV (g.23,776,347G>A) within the MTRDHH1 haplotype. For each animal, only the phase supposed to host the MMUT SNV variant allele is represented. The blue color indicates the portion of local haplotype matching with the MTRDHH1 haplotype.



S4 Fig. Distribution of genotyped animals over time. The bar charts represent the number of genotyped animals according to sex and year of birth. The genomic selection in MTR dairy sheep was implemented in 2017.

S3 Table. EMBL-EBI accession numbers of the 100 whole-genome sequences used in the analysis.

Breed	Number of animals	ENA run accession	EBI project accession	Associated references
Belclare	2	SRR14934360; SRR14935129	PRJNA698548	
Berrichon du Cher	3	ERS1205899; ERS1205900; ERS1205901	PRJEB14418	[1]
Cambridge	7	ERR1419201; ERR1419202; ERR1419203; ERR1419204; ERR1419205; ERR1419206; ERR1419207	PRJEB14098	
Charollais	1	SRR14934359	PRJNA698548	
Lacaune (Dairy)	31	ERR3276357; ERR3276358; ERR3276359; ERR3276360; ERR3276361; ERR3276362; ERR3276363; ERR3276364; ERR3276365; ERR3276366; ERR3276367; ERR3276368; ERR3276369; ERR3276370; ERR3276371; ERR3276372; ERR3276373; ERR3276374; ERR3276375; ERR3276376; ERR3276377; ERR3276378; ERR3276379; ERR7891349; ERR7891350; ERR7891351	PRJEB32110	
Lacaune (Meat)	3	ERR968423; ERR968424; ERR968425 SRR501850; SRR501851	PRJEB9911 PRJNA160933	[2,3]
Manech Tête Rousse	22	ERR3276380; ERR3276381; ERR3276382 ERR3712282; ERR3712283; ERR3712284; ERR3712285*; ERR3712286; ERR3712287; ERR3712288; ERR3712289; ERR3712290; ERR3712291; ERR3712292; ERR3712293; ERR3712294; ERR3712295; ERR3712296; ERR3712297; ERR3712298; ERR3712299; ERR3712300; ERR3712301*; ERR7889920; ERR7889921	PRJEB31930 PRJEB35553	[4]
Martinik Blackbelly	1	ERR3255914	PRJEB14418	[1]
Noire du velay	2	ERR3828659; ERR3828660	PRJEB31930	
Romane	4	ERS1205902; ERS1205903; ERR2818429; ERR2818430	PRJEB35553	[4]
Romane x Martinik Blackbelly	13	ERR3255915; ERR3255916; ERR3255917; ERR3255918; ERR3255919; ERR3988554; ERR3988555; ERR3988556; ERR3988557; ERR3988558; ERR3988559; ERR3988560; ERR3988561	PRJEB31930	
Romanov	2	ERS1205904; ERS1205905	PRJEB14418	[1]
Suffolk	2	SRR14934357; SRR14934358	PRJNA698548	
Texel	2	SRR14934355; SRR14934356	PRJNA698548	
Vendéen	5	ERR4236129; ERR4236130; ERR4236131 SRR14934353; SRR14934354	PRJEB37460 PRJNA698548	[5]
Total	100			

* MTRDHH1 heterozygous carrier

1. Demars J, Cano M, Drouilhet L, Plisson-Petit F, Bardou P, Fabre S, et al. Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep. *Mol Biol Evol.* 2017;34: 1722–1729. doi:10.1093/molbev/msx114
2. Rupp R, Senin P, Sarry J, Allain C, Tasca C, Ligat L, et al. A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model. *PLoS Genet.* 2015;11: e1005629. doi:10.1371/journal.pgen.1005629
3. Ben Braiek M, Moreno-Romieux C, Allain C, Bardou P, Bordes A, Debat F, et al. A Nonsense Variant in CCDC65 Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep. *Genes.* 2022;13: 45. doi:10.3390/genes13010045
4. Chantepie L, Bodin L, Sarry J, Woloszyn F, Plisson-Petit F, Ruesche J, et al. Genome-Wide Identification of a Regulatory Mutation in BMP15 Controlling Prolificacy in Sheep. *Front Genet.* 2020;11. doi:10.3389/fgene.2020.00585
5. Fabre S, Chantepie L, Plisson-Petit F, Sarry J, Woloszyn F, Genet C, et al. A novel homozygous nonsense mutation in ITGB4 gene causes epidermolysis bullosa in Mouton Vendéen sheep. *Anim Genet.* 2020 [cited 27 Nov 2020]. doi:https://doi.org/10.1111/age.13026

S4 Table. Markers on SheepLD chip for genotyping at the *RXFP2* (1.8 kb InDel) and *ASIP* (5pb InDel) loci.

Mutation	Markers (LD chip)	Possible genotype at the mutation
OAR13:g.66,475,132_66,475,136del	oar3_OAR13_63047423 (D/I)	Ins/Ins Ins/Del Del/Del
OAR10: 1.8kb insertion in the 3'-UTR of <i>RXFP2</i>	RXFP2_insert_L1 (A/G) RXFP2_insert_R1 (A/T)	Ins/Ins (RXFP2_insert_L1 [G/G] & RXFP2_insert_R1 [A/A]) Ins/Del (RXFP2_insert_L1 [A/G] & RXFP2_insert_R1 [A/T]) Del/Del (RXFP2_insert_L1 [A/A] & RXFP2_insert_R1 [T/T])

The interpretation of the 1.8kb insertion in *RXFP2* is possible by the combination of two markers RXFP2_insert_L1 and RXFP2_insert_R1.

S5 Table. List of PCR primer sequences.

Application	Gene symbol	Sequence (5'→3')	fragment size (pb)	Efficiency
Terra PCR	<i>MMUT</i> NC_040271.1	F: GCAGATTTTGGCAAAAGAAAG R: ACCTTCAAGGCAGCATCATA	402	-
		F1: GAAGGTGACCAAGTTCATGCTAATCCAGATTCTTCTTGAATAATGATTTG F2: GAAGGTCCGAGTCAACGGATTGAATCCCAGATTCTTCTTGAATAATGATTTA R: GTGAAAAGTGCTCGAATTGCCAGGAA	81/82	-
PACE PCR	<i>ASIP (SNP T>A)</i> NC_040264.1	F1: GAAGGTGACCAAGTTCATGCTTCCGCAGCGCCTGCTCCT F2: GAAGGTCCGAGTCAACGGATTCCGCAGCGCCTGCTCCA R: CGCGCTCAGCAGGTGGGTT	70	-
		F: GCTCATCAAGGAACTCAA R: TTCAAATATCATCAAGCACTTG	169	1.80
		F: CGACTTCAACAGCGACACTC R: TGCTGTAGCCGAATTCATTG	113	1.90
Quantitative PCR	<i>YWHAZ</i> NM_001267887	F: ATTAAGTGAAGAGTCAATACAA R: GTATCCGATGTCCACAAT	81	2.00
		F: AATGCCAATGCCAACTC R: CCCTTTCGCTACCTATAAC	149	2.00
	<i>SDHA</i> XM_027980212	F: GAATGGTCTGGAACACTG R: AGTAAATCGTACTCGTCAAC	156	2.00

F: Forward primer, **R:** Reverse primer.

A single base pair duplication in *SLC33A1* gene causes fetal losses and neonatal lethality in Manech Tête Rouse dairy sheep

Maxime Ben Braiek¹, Soline Szymczak¹, Céline André², Philippe Bardou³, Francis Fidelle², Florence Plisson-Petit¹, Julien Sarry¹, Florent Woloszyn¹, Carole Moreno-Romieux^{1S} and Stéphane Fabre^{1*}

¹GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet-Tolosan, France

²CDEO, Quartier Ahetzia, 64130, Ordiarp, France

³Sigenae, INRAE, 31326 Castanet-Tolosan, France

\$ *in memoriam*

*Corresponding author

Email: stephane.fabre@inrae.fr (SF)

Abstract

Recently, we evidenced that the Manech Tête Rouse (MTR) deficient homozygous haplotype 2 (MTRDHH2) was likely to harbor a recessive lethal variant in ovine. In the present study, we fine mapped this region by analyzing the whole genome sequence of five MTRDHH2 heterozygous carriers compared to 95 sequences of non-carrier animals from MTR and others ovine breeds. We successfully identified a single base pair duplication in the *SLC33A1* gene, resulting in a frameshift leading to a premature stop codon (p.Arg246Alafs*3). *SLC33A1* acts as a transmembrane transporter of acetyl-Coenzyme A, essential for cellular metabolism. In order to assess for the lethal phenotype in homozygous MTR sheep, we generated at-risk matings by artificial insemination (AI) between rams and ewes heterozygous for the *SLC33A1* variant named *SLC33A1_dupG*. Gestation status was checked 15 days post-AI by a molecular test from blood expression of the interferon Tau-stimulated *MX1* gene, and by ultrasonography performed between 45 days and 60 days post-AI. Based on ultrasonography, the AI success was reduced by 12% compared to safe matings suggesting embryonic/fetal losses further confirmed by the molecular test based on *MX1* differential expression. Forty-nine lambs were born from at-risk matings with a mortality rate of 34.7% observed before weaning. Homozygous *SLC33A1_dupG* lambs contributed to 47% of this mortality occurring mainly in the first five days after lambing with no obvious clinical signs. Thus, an appropriate management of *SLC33A1_dupG* (allele frequency of 0.04) in the MTR selection scheme should increase the overall fertility and lamb survival.

Keywords

Ovis aries, homozygous haplotype deficiency, MTRDHH2, lethal variant, abortion, neonatal mortality, *SLC33A1*, acetyl-CoA transporter

Introduction

Most individuals are expected to carry between 1 and 5 highly deleterious variants in their genome known as loss-of-function alleles (Georges et al., 2019; MacArthur et al., 2012). When homozygous, these variants often cause severe defects leading to lethality (from embryonic stage to adulthood) or morphological disorders. The management of these defects must go through the identification of the causal variants. In order to tackle this issue, recent advances in the use of genomic tools have enhanced the discovery of the causative variants for many genetic disorders. In human, the genome was for the first time sequenced in 2001 opening up many perspectives, especially to fine-map causal variants (Lander et al., 2001). Nowadays in human, 7,342 entries are reported in the OMIM database (<https://www.omim.org>, consulted on 07/03/2023) for which the molecular basis is known. In livestock also, the emergence of next-generation sequencing has allowed to get access to the whole genome of many individuals and an exhaustive number of genetic markers as single nucleotide polymorphisms (SNP), representative of the individual genome variability (Eggen, 2012; Rupp et al., 2016).

In order to fine-map causal variants responsible for genetic defects, two main strategies were developed. The classic “top-down” strategy is based on available phenotypes and associated biological samples genotyped using SNP arrays. Based on genome-wide association study with a case-control approach (Uffelmann et al., 2021), and/or homozygosity mapping (Charlier et al., 2008; Lander and Botstein, 1987) coupled with whole genome sequencing data, geneticists can successfully fine-map causal variants. In contrast, the “bottom-up” strategy utilizes a reverse genetic screen method initially developed to identify lethal variant without any available phenotype (Fritz et al., 2013; VanRaden et al., 2011). This method uses high throughput SNP genotyping data to detect deficit in homozygous animals leading to a significant deviation from Hardy-Weinberg equilibrium is based on within trios transmission probability (VanRaden et al., 2011). The establishment of routine SNP genotyping within the framework of genomic selection in livestock has made possible the use this method in many species as cattle (Fritz et al., 2013; VanRaden et al., 2011), beef (Jenko, 2019), pigs (Derks et al., 2017), chicken (Derks et al., 2018), turkey (Abdalla et al., 2020), horses (Todd et al., 2020) and sheep (Ben Braiek et al., 2023, 2021). Especially in sheep, we recently identified numerous independent deficient homozygous haplotypes (DHH) in Lacaune and Manech Tête Rousse (MTR) dairy sheep (Ben Braiek et al., 2023, 2021). Thanks to WGS data and generation of at-risk mating between DHH heterozygous carriers, we were able to identify two nonsense variants, one in the *CCDC65* gene causative of respiratory disorders leading to juvenile lethality in Lacaune (LDHH6, OMIA 002342-9940) (Ben Braiek et al., 2022), and another one located in the *MMUT* gene affecting the propionic acid metabolism responsible for neonatal lethality in MTR (MTRDHH1) (Ben Braiek et al., 2023). These two DHH were firstly associated with significant increased stillbirth rate in at-risk mating based on population records. Additionally, some of the other evidenced DHH (LDHH1-2-8-9 in Lacaune, and MTRDHH2 in MTR) also affected the artificial insemination (AI) success by a 3% decrease in at-risk mating compared to safe mating, assuming the action of lethal embryonic variants (Ben Braiek et al., 2023, 2021).

Embryonic and early fetal losses are difficult to observe and have strong economic impact for sheep breeders (Diskin and Morris, 2008; Dixon et al., 2007). Although the causes of these losses are mainly environmental, lethal genetic variants could also explain a part of this problem (VanRaden et al., 2011). In sheep, fertilization failure and embryonic losses could be associated with a negative gestation diagnosis realized by ultrasonography at day 45-60 post-fertilization. However, most of embryonic losses are expected during the period from the fertilization (day 0) to the implantation of the conceptus (days 12-16) (Bindon, 1971; Spencer et al., 2008; Wilmut et al., 1986). During implantation, the embryo binds to the endometrium and leads to the secretion of interferon tau (IFNT) as a pregnancy recognition signal (Bazer, 2013). As a response, interferon stimulated genes (ISG) expression is greatly enhanced in the endometrium

but also in circulating immune cells in the blood (Ott and Gifford, 2010). Among these ISG, *MXI* (Myxovirus-influenza virus resistance 1), *STAT1* (Signal transducer and activator of transcription 1) and *CXCL10* (Chemokine C-X-C motif ligand 10) blood expression have been used to predict the gestation status at day 14-15 in sheep (Mauffré et al., 2016).

As presented above, the MTRDHH2 deficient haplotype in MTR dairy sheep reduced significantly the AI success by 3.0%, but it also increased the stillbirth rate by 4.3%, indicating a possible action of a lethal variant all along the gestation (Ben Braiek et al., 2023). MTRDHH2 was located on the ovine chromosome 1 (NC_040252.1, OAR1:251.9-256.4 Mb on sheep genome Rambouillet_v1.0) with an estimated frequency of heterozygous carriers of 8.7%, and we have reported only one functional candidate gene (*SLC33A1*) within the MTRDHH2 genomic region (Ben Braiek et al., 2023). The objective of this study was to evidence the causal variant associated with MTRDHH2 and to validate the lethal embryonic effect by generating at-risk mating and monitoring the early gestation time at day 15 by a molecular diagnosis and day 45-60 by an ultrasonography diagnosis.

Materials and Methods

Sequencing data, WGS variant calling and annotation

The list of the 100 publicly available ovine whole genome sequences (WGS, short read Illumina HiSeq/Nova Seq) coming from 14 different breeds generated in various INRAE and Teagasc research projects is detailed in Table S1 (project and sample accession numbers). Among them, 22 WGS were from Manech Tête Rousse dairy sheep also genotyped with the Illumina OvineSNP50 Beadchip in the framework of dairy sheep genomic selection program (Astruc et al., 2016). They all had a known status at the MTRDHH2 haplotype encoded 0 for non-carrier (n=17), 1 for heterozygous carrier (n=5) and 2 for homozygous carrier (n=0) (Ben Braiek et al., 2023). Read mapping, WGS variant calling and annotation were performed using Nextflow v20.10.0 and Sarek pipeline v2.6.1 as already described (Ben Braiek et al., 2023).

Identification of candidate causal variants

All SNPs and InDels variants located within the MTRDHH2 haplotype region extended by 1 Mb from each side were extracted from OAR1 (Oar_rambouillet_v1.0; NC_040252.1:250,858,291-257,412,373pb) using SnpSift Filter (Cingolani et al., 2012). Correlation analysis was performed between MTRDHH2 haplotype status (0, 1 and 2) and allele dosage for bi-allelic variants (also encoded 0, 1 or 2) using `geno--r2` command of VCFtools (Danecek et al., 2011). The putative causal variant was checked manually using the Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013) using the BAM files of the 22 MTR dairy sheep.

Biological samples

The experimental design is described in Fig. 1. Jugular blood samples from 419 MTR dairy ewes (all daughters of MTRDHH2 carrier sires) located in 6 private farms were collected by a habilitated veterinarian (Venoject system containing EDTA, Terumo, Tokyo, Japan) and stored at -20°C either for further genotyping (n=419) or RNA extraction (n=137). Ear biopsies (1 mm³) from the 49 newborn lambs were obtained with a tissue sampling unit (TSU, Allflex Europe, Vitré, France) and directly placed in the TSU storage buffer at 4°C. Ear biopsies were treated with consecutive action of NaOH and Tris-HCl for subsequent genotyping as previously described (Ben Braiek et al., 2023).

DNA sets from MTR genomic lambs (n=714) with a known status at MTRDHH2 (Ben Braiek et al., 2023), Manech-related Laxta Spanish sheep (n=100) and a diversity panel from 25 French sheep breeds (n=749) (Rochus et al., 2018) were also used for single marker genotyping.

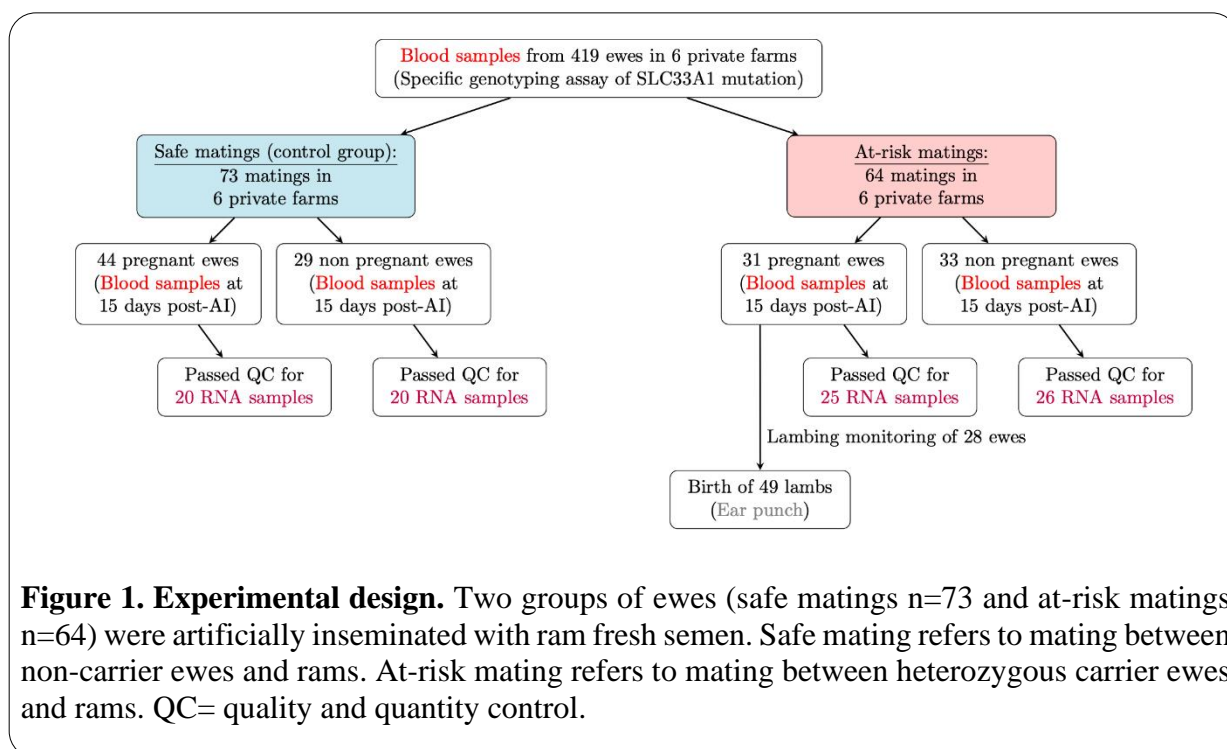


Figure 1. Experimental design. Two groups of ewes (safe matings n=73 and at-risk matings n=64) were artificially inseminated with ram fresh semen. Safe mating refers to mating between non-carrier ewes and rams. At-risk mating refers to mating between heterozygous carrier ewes and rams. QC= quality and quantity control.

***SLC33A1* specific genotyping assay**

Among the candidate variants located in MTRDHH2, a PCR allele competitive extension (PACE) genotyping assay was developed for the variant NC_040252.1:g.252,649,023dupG located in the *SLC33A1* gene. Fluorescent PACE analysis was done with 15ng of purified DNA (from DNA panels) using the PACE-IR 2x Genotyping Master mix (3CR Bioscience) in the presence of 12 μ M of a mix of extended allele specific forward primers (Table S2) in a final volume of 10 μ L. The touch-down PCR amplification condition was 15 min at 94°C for the hot-start activation, 10 cycles of 20s at 94°C, 54–62°C for 60s (dropping 0.8°C per cycle), then 36 cycles of 20s at 94°C and 60s at 54°C performed on an ABI9700 thermocycler followed by a final point read of the fluorescence on an ABI QuantStudio 6 real-time PCR system and using the QuantStudio software 1.3 (Applied Biosystems). For the genotyping of crude biological samples (whole blood or neutralized NaOH treatment solution of ear biopsies), a preliminary Terra PCR Direct Polymerase mix amplification (Takara Bio, Kusatsu, Japan) using 1 μ L of crude sample was used for direct genotyping without DNA purification. The following amplification primers (Table S2) were designed using Primer3Plus software (Untergasser et al., 2007). This preliminary PCR was performed on an ABI2720 thermocycler (Applied Biosystems, Waltham, MA, USA) with the following conditions: 5 min at 94°C, 35 cycles of 30s at 94°C, 30s at 58°C and 30s at 72°C, followed by 5 min final extension at 72°C. Then, 1 μ L of the PCR product were used for subsequent PACE genotyping.

Programmed mating

Among the 419 daughters of MTRDHH2 carrier sires and according to *SLC33A1* genotyping results, 137 ewes were retained for programmed mating. Two groups of mating were

constituted: safe matings (n=73 ewes) between non-carrier ewes and rams, and at-risk matings (n=64 ewes) between heterozygous-carriers. All ewes were mated by artificial insemination (AI) with fresh semen. A jugular blood sample was collected from each mated ewe 15 days after AI for further gestation molecular diagnosis test. An ultrasound diagnosis of gestation was realized between 45 and 60 days after AI. Gestations were followed and each lamb was monitored from birth to weaning age.

Molecular diagnosis of gestation

RNA extraction, reverse transcription and real-time PCR

Total RNAs were extracted from blood of 137 ewes with the Nucleospin® RNA Blood Kit (Macherey-Nagel, #Ref 740200.50) according to the manufacturer's protocol starting with 800µL of whole blood with a DNaseI digestion treatment to eliminate contaminating genomic DNA. RNAs was quantified by spectrophotometry (NanoDrop® ND-8000 spectrophotometer, ThermoFischer) and stored at -80°C. After quality and quantity control, 91 RNA samples (n=40 ewes in safe mating control group and n=51 ewes in at-risk mating group) were kept for reverse transcription. Reverse transcription was carried out from 500 ng of total RNA in solution with anchored oligo(dT) T22V (1µL at 100µM), random oligo-dN9 (1µL at 100µM) and dNTPs (2µL at 10 mM) in a reaction volume of 54µL. This mixture was incubated at 65°C for 5 min in an ABI2700 thermocycler (Applied Biosystems) then ramped down to 4°C. A second reaction mixture (18.5µL/reaction) containing the reaction buffer (14µL of First strand Buffer 5X, Invitrogen, France), DTT (Dithiothreitol, 3µL at 0.1M), Rnasine (0.5µL, 40 units/µL, Promega, France) and Superscript II reverse transcriptase (1µL, 200 units/µL, Invitrogen, France) was added to the denatured RNA solution (final volume reaction of 72.5µL) then incubated for 50 minutes at 42°C and placed for 15 minutes at 70°C. The complementary DNA (cDNA) solution obtained was directly diluted at 1:2 ratio and stored at -20°C. Quantitative PCR (qPCR) was performed using 3µL of cDNA, 5µL of SYBR Green real-time PCR Master Mix 2X (Applied Biosystems) and 2µL of primers at 3µM in a total reaction volume of 10µL. Each sample was tested in quadruplicate. qPCR was realized on a QuantStudio 6 Flex Real-Time PCR system (ThermoFisher). For each pair of primers, amplification efficiency was evaluated by $E = e^{-1/\alpha}$ which α is the slope of a linear curve obtained from cDNA serial dilution and corresponding Ct (cycle threshold) values. RNA transcript abundance was quantified using the delta Ct (ΔCt) method corrected by two reference genes (*GAPDH*, *YWHAZ*). Primers were design using Beacon Designer 8 (Premier Biosoft). qPCR primer sequences, amplification lengths and amplification efficiencies are available in Table S2.

Statistical analyses

The ewe gestation status (pregnant/non-pregnant) was firstly determined by the ultrasound diagnosis and thereafter corrected at the time of lambing. *MXI* and *STAT1* (Interferon Stimulated Genes, ISG) relative expressions between pregnant and non-pregnant ewes were compared using Wilcoxon non-parametric test.

The assessment of the gestation diagnosis molecular test (GDMT) at day 15 was based on the ISG expression. The prediction was performed by ROC (Receiver Operating Characteristic) curve analysis using easyROC (Goksuluk et al., 2016). The model was first fitted using the data expression of the safe mating control group (training data) and thereafter transposed on at-risk mating data expression supposed under the influence of the *SLC33AI* lethal embryonic variant (testing data). For training data, the area under the curve (AUC) was evaluated and compared to the expected value of 0.5 under the null hypothesis, the cut-off method ROC01 (which minimizes distance between ROC curve and point (100-Sp=0, Se=100)) was used to maximize Sensitivity (Se; i.e., ability of the GDMT to correctly detect pregnant ewes by ultrasonography). This cut-off value was used to classified the ISG relative expression in four categories: true positive (TP i.e., GDMT+ and pregnant), false positive (FP i.e., GDMT+ and non-pregnant),

true negative (TN i.e., GDMT- and pregnant) and false negative (FN i.e., GDMT- and non-pregnant). The test estimators were generalized using the prevalence (Pr) (e.g, the probability to be pregnant at day 45-60, corresponding to the population mean of 60.3% in the Manech Tête Rousse overall population (Ben Braiek et al., 2023)). The positive predictive value (PPV) i.e., the number of ewes with a positive GDMT and pregnant at day 45-60 among the number of ewes with positive GDMT, and negative predictive value (NPV), i.e., the number of ewes with a negative GDMT and non-pregnant at day 45-60 among the number of ewes with negative GDMT, was determined by $PPV = \frac{Se \times Pr}{Se \times Pr + (1-Sp) \times (1-Pr)}$ and $NPV = \frac{Sp \times (1-Pr)}{Sp \times (1-Pr) + (1-Se) \times Pr}$.

For testing data, the cut-off value of the training model was used to classified the ewes of the at-risk mating group in the TP, FP, TN and FN categories (observed numbers). The model parameters (Se, Sp, Pr, PPV and NPV) were used to estimate the expected numbers in TP, FP, TN and FN categories. Chi-squared tests were performed between the observed and expected number of ewes with positive and negative GDMT, respectively.

Results

Identification of a single base pair duplication in *SLC33A1* gene

In order to identify the causative variant located in MTRDHH2 haplotype, we have scanned the genome of 100 ovine WGS focusing on the polymorphisms located in MTRDHH2 region extended by 1 Mb from each side (OAR1:250,858,291-257,412,373pb). In this 6.5 Mb region, we detected 111,984 polymorphisms (variant call rate $\geq 95\%$, quality score > 30). Among the WGS animals, 22 were from the Manech Tête Rousse breed and 5 of them were heterozygous carriers of MTRDHH2. After Pearson correlation analysis between biallelic variant status (SNPs and InDels) and MTRDHH2 haplotype status (all encoded by 0, 1 and 2), 189 polymorphisms had a perfect correlation with MTRDHH2 status (Fig. 2a). Nevertheless, only one variant (NC_040252.1, OAR1:g.252,649,023dupG; Table S3) was located in a coding sequence and was predicted to highly alter the *Solute carrier family 33 member 1 (SLC33A1)* gene (XM_012100950.3, c.735dupG). This variant, thereafter called *SLC33A1_dupG*, is a single base pair duplication of a guanine in the first exon of the gene (Fig. 2b, c). It is predicted to create a premature stop codon three amino acids after the duplication (XP_011956340.1, p.Arg246Alafs*3) resulting in a truncated protein of 248 amino acids compared to the 550 amino acids full length protein (Fig. 2d). The *SLC33A1* protein is composed of 9 transmembrane domains (UniProt: A0A6P3TI15_SHEEP) while in the mutant form, only 3 transmembrane domains can be translated.

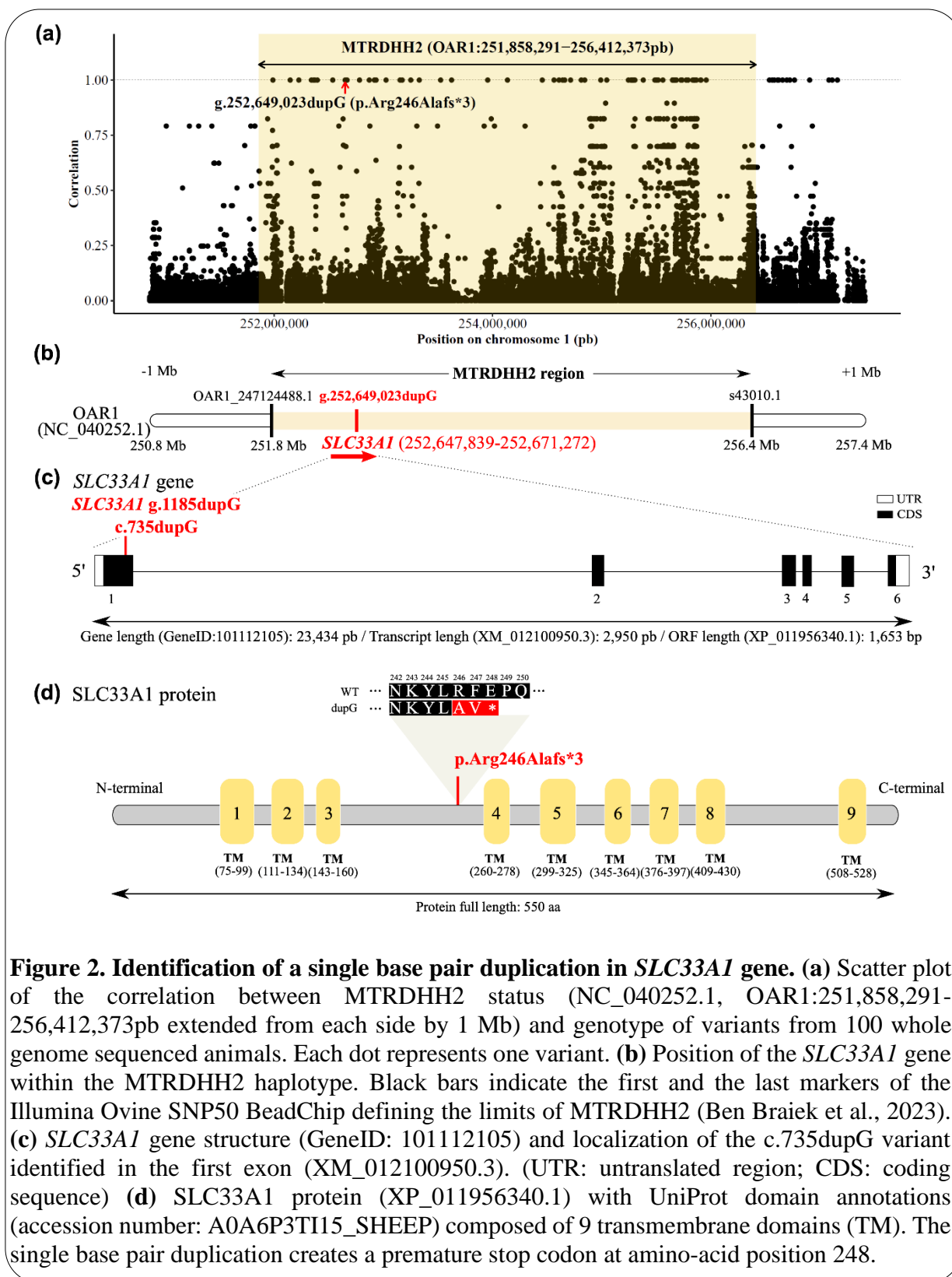


Figure 2. Identification of a single base pair duplication in *SLC33A1* gene. (a) Scatter plot of the correlation between MTRDHH2 status (NC_040252.1, OAR1:251,858,291–256,412,373pb extended from each side by 1 Mb) and genotype of variants from 100 whole genome sequenced animals. Each dot represents one variant. (b) Position of the *SLC33A1* gene within the MTRDHH2 haplotype. Black bars indicate the first and the last markers of the Illumina Ovine SNP50 BeadChip defining the limits of MTRDHH2 (Ben Braiek et al., 2023). (c) *SLC33A1* gene structure (GeneID: 101112105) and localization of the c.735dupG variant identified in the first exon (XM_012100950.3). (UTR: untranslated region; CDS: coding sequence) (d) *SLC33A1* protein (XP_011956340.1) with UniProt domain annotations (accession number: A0A6P3TI15_SHEEP) composed of 9 transmembrane domains (TM). The single base pair duplication creates a premature stop codon at amino-acid position 248.

Variant association with MTRDHH2 and diversity analysis

In order to validate the association of the putative causal variant located in *SLC33A1* gene with MTRDHH2, we genotyped a DNA set from 714 Manech Tête Rousse lambs of the 2021 genomic cohort with a known status at MTRDHH2 (Ben Braiek et al., 2023), and we observed

a variant allele frequency of 4%. The contingency table showed a strong association between the OAR1: g.252,649,023dupG genotype and MTRDHH2 haplotype status (Table 1, Fischer's exact test, $P < 0.0001$, without the homozygous carrier individuals). All the MTRDHH2 heterozygous carriers were also heterozygous for the *SLC33A1_dupG* variant. However, among the 56 heterozygous animals for the *SLC33A1_dupG*, 14 were not MTRDHH2 heterozygous carriers. A specific focus on the 66 SNP markers composing the MTRDHH2 haplotype showed shorter recombinant versions of MTRDHH2 (from 5 to 65 SNPs) possibly explaining this discrepancy (Fig. S1).

Table 1. Contingency table between MTRDHH2 status and NC_040252.1:g.252,649,023dupG genotype.

Genotype	+/+	MTRDHH2/+	MTRDHH2/ MTRDHH2	Total
WT/WT	658	0	0	658
WT/dupG	14	42	0	56
dupG/dupG	0	0	0	0
Total	672	42	0	714

+/+= non-carriers; MTRDHH2/+ heterozygous carriers and MTRDHH2/MTRDHH2= homozygous carriers (Fisher's exact test, $p < 0.001$, without the homozygous MTRDHH2 carriers).

In order to search for the segregation of the *SLC33A1_dupG* variant in other sheep breeds, we genotyped a DNA diversity panel of French (FR) and Spanish (ES) ovine breeds composed of 28 different breeds in total. The variant was found in the heterozygous state in 3 French MTR animals and one Spanish Latxa Cara Rubia animal (Table 2).

Table 2. NC_040252.1:g.252,649,023dupG genotype distribution from a DNA diversity panel of French (FR) and Spanish (ES) ovine breeds.

Breed	Total	Genotype		Breed	Total	Genotype	
		WT/WT	WT/dupG			WT/WT	WT/dupG
Berrichon du Cher (FR)	30	30		Martinik (FR)	23	23	
Blanche du Massif Central (FR)	31	31		Merinos d'Arles (FR)	27	27	
Causse du Lot (FR)	32	32		Mourerous (FR)	26	26	
Charmoise (FR)	31	31		Mouton Vendéen (FR)	30	30	
Charollais (FR)	30	30		Noir du Velay (FR)	28	28	
Corse (FR)	30	30		Préalpes du sud (FR)	27	27	
Ile de France (FR)	28	28		Rava (FR)	29	29	
Lacaune (Meat) (FR)	45	45		Romane (FR)	30	30	
Lacaune (Milk) (FR)	40	40		Romanov (FR)	26	26	
Latxa Cara Negra Euskadi (ES)	30	30		Rouge de l'Ouest (FR)	30	30	
Latxa Cara Negra Navarra (ES)	40	40		Roussin (FR)	30	30	
Latxa Cara Rubia (ES)	30	29	1	Suffolk (FR)	29	29	
Limousine (FR)	29	29		Tarasconnaise (FR)	33	33	
Manech Tête Rousse (FR)	28	25	3	Texel (FR)	27	27	
				Total	849	845	4

***SLC33AI_dupG* variant associated with decreased AI success and lamb mortality**

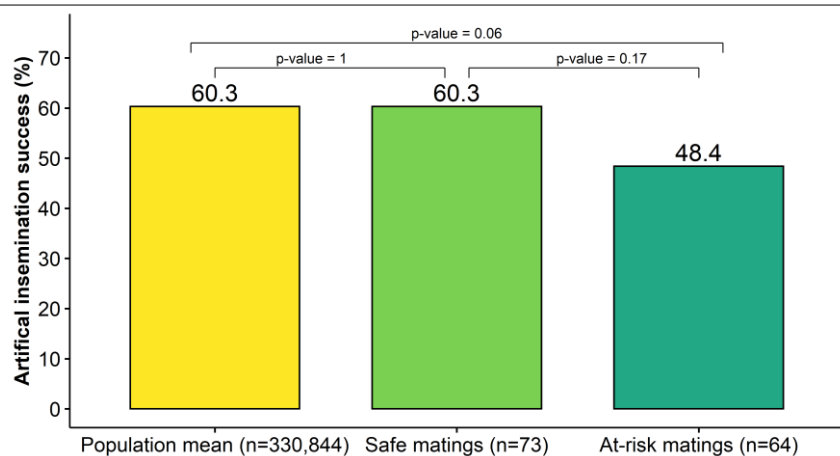


Figure 3. Artificial insemination success in MTR. MTR population artificial insemination (AI) success is based on lambing date according to the gestation length starting from the day of AI (151 ± 7 days) and was established using populational matings records ($n=330,844$; Ben Braiek et al., 2023). Safe and at-risk matings between *SLC33AI_dupG* carriers were realized in 6 private farms ($n=137$ ewes) with a gestation diagnosis realized by ultrasonography between days 45 and 60. In these experimental groups, AI success could have been corrected by lambing record. P-values were obtained by Fisher's exact test.

MTR ewes ($n=419$) from 6 private farms where genotyped for the *SLC33AI_dupG* variant. Among them, 137 ewes were selected to constitute two mating groups, a safe mating control group with 73 non-carrier ewes mated through AI with non-carrier rams, and an at-risk mating group with 64 heterozygous ewes for *SLC33AI_dupG* mated through AI with heterozygous rams in order to generate *SLC33AI_dupG* homozygous lambs. Gestation was monitored by a blood sampling at day 15, ultrasonography between day 45 and day 60, and lambing records at 151 ± 7 days after AI. As shown in Fig. 3, the AI success at ultrasonography (possibly corrected by lambing results) was 60.3% in the safe mating group in line with the overall AI success in the MTR population (Ben Braiek et al., 2023). However, the AI success was reduced to 48.4% in the at-risk mating group of *SLC33AI_dupG* heterozygous carriers. This reduction of 12% is not statistically significant ($p=0.17$) but is relevant in ovine breeding.

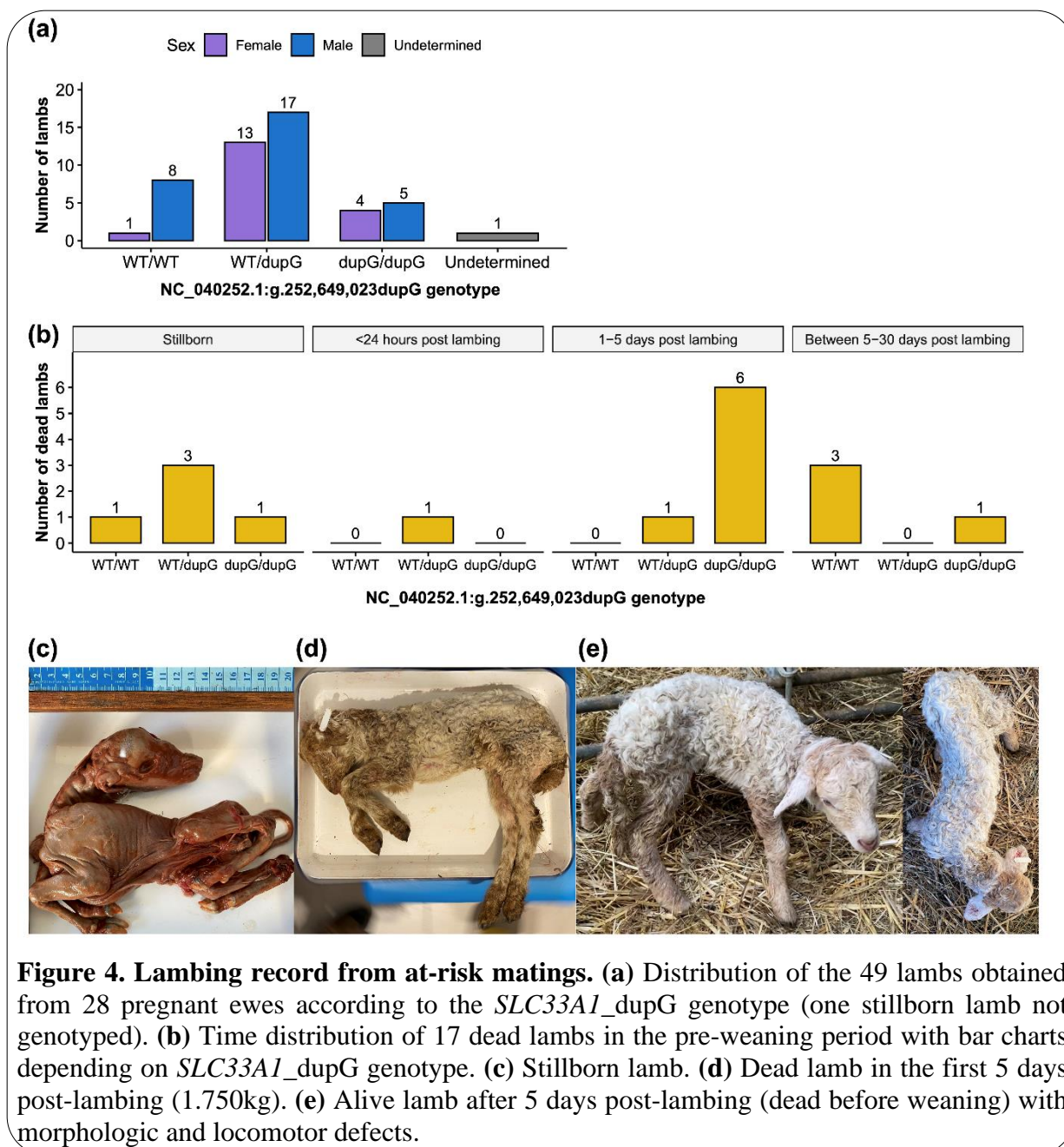
Among the 31 pregnant ewes in the at-risk mating group, 28 ewes had complete observation in farm for gestation, lambing and lamb monitoring between birth and weaning at 1 month of age. No abortion was observed between the time of ultrasonography (day 45-60 post-AI) and the parturition. The length of the gestation ranged between 147 and 154 days in line with the MTR population mean (151 ± 7 days) and 49 lambs were born (Fig. 4a). The contingency table (Table 3) between lamb genotypes and viability (alive or dead) indicated a significant lower viability rate for homozygous dupG/dupG lambs (Fisher's exact test, $p\text{-value} < 0.001$). The *SLC33AI_dupG* homozygous lambs contributed to 47% of the mortality in at-risk matings and the mortality occurred mainly in the first five days after lambing (Fig. 4b). A stillborn lamb (Fig. 4c) has been expelled with the placenta at full term of gestation, and it showed developmental arrest characteristic of a mid-gestation stage. Another stillborn lamb (same phenotype as Fig. 4c) was obtained but no ear punch was done to genotype the animal (undermined genotype). Most of homozygous animals died between 1 and 5 days with no apparent morphological defects (Fig. 4d). Only one homozygous lamb died in the 5-30 days

period and showed leg weakness and stiffness, and also spine deformity leading to locomotor problem (Fig. 4e).

Table 3. Contingency table between lamb NC_040252.1:g.252,649,023dupG genotypes and viability.

Viability	WT/WT	WT/dupG	dupG/dupG	-/-	Total
Alive	5	25	1	0	31
Dead	4	5	8	1	18
Total	9	30	9	1	49

-/-: undermined genotype due to absence of biological samples for a stillborn lamb. This animal is not taking into account for mortality rate calculation.



Molecular diagnosis of gestation at day 15 to assess for fetal losses

The decreased AI success in at-risk mating between *SLC33A1_dupG* heterozygous carriers suggested possible losses of homozygous embryos during the gestation period between AI and the ultrasound diagnosis at day 45-60. We then applied a gestation diagnosis molecular test (GDMT) based on blood mRNA expression levels of two interferon stimulated genes, *MX1* and *STAT1*, at day 15 as described by Mauffré et al. (Mauffré et al., 2016). RNA blood samples from 91 ewes (n=40 in safe mating group, n=51 in at-risk mating group) were analyzed by RT-qPCR (Fig. 1). The assessment of the GDMT at day 15 to predict the gestation status (pregnant /non-pregnant) at day 45-60 was performed on the safe mating control group. As shown in Fig. 5a and as expected, the *MX1* relative expression in non-pregnant ewes (67.7%) was significantly increased in pregnant ewes (84.3%, p-value=0.043, Wilcoxon test). The same suggestive observation could be made for *STAT1* expression, even if the difference was not significant (p-value=0.076, Wilcoxon test). Thus, only *MX1* relative expression was retained for GDMT and reliability of the diagnostic was tested using ROC curve on data from the safe mating group as training data. ROC plot showed an AUC of 0.687, differing significantly from 0.5 (p-value=0.033), a cut-off value (i.e., decision threshold) set at 63% and characterized by a sensitivity of 75% and a specificity of 60% (Table 4, Fig. 5c and d). The prevalence (Pr) of AI success was set at 60.3% (MTR population mean) and helped us to calculate the positive predictive value (PPV = 74%) and the negative predictive value (NPV = 61%). The cut-off value was used to classify ewes in four classes (TP, FP, TN and FN) according to GDMT (+ or -) and ultrasound diagnostic at day 45-60 (Fig. 5e). Positive GDMT enabled to detect 65% (TP/(TP+FP)) of pregnant ewes in the safe mating control group. In contrast, no significant difference was observed for *MX1* expression between pregnant and non-pregnant ewes of the at-risk mating group (p-value=0.27, Wilcoxon test, Fig. 5f). Thus, the molecular diagnostic parameters (cut-off=63%, Pr=60.3%, PPV= 74% and NPV= 61%) were transposed into the *MX1* expression data from at-risk mating group. In this group, the comparison of the observed and expected number of ewes with a positive GDMT (*MX1* relative expression >63%) has evidenced a trend to reduce the number of pregnant ewes (Chi-squared test, p=0.055) i.e. 16 pregnant ewes were expected with a positive GDMT while only 12 pregnant ewes were observed. No significant difference was pointed out between observed and expected number of pregnant ewes with a negative GDMT (Chi-squared test, p=0.44). Thus, the molecular diagnosis likely evidenced fetal losses in the at-risk mating group between 15 days and 60 days of gestation, with four ewes supposed to host homozygous fetuses for the *SLC33A1_dupG* variant.

Table 4. ROC curve analysis parameter for the gestation molecular diagnosis test based on *MX1* mRNA level expression.

Gene	AUC ± se	p-value	Se (%)	Sp (%)	Cut-Off (%) ¹	Pr (%)	PPV (%) ²	NPV (%) ²
<i>MX1</i>	0.687 ± 0.088	0.033	75	60	63	60.3	74	61

¹ROC01 method (minimizes distance between ROC curve and point (100-Sp=0, Se=100)) was used for optimal cut-off.

²PPV and NPV were calculated as described in Material and Methods section taking account Se, Sp, Pr. *MX1* = Myxovirus (influenza virus) resistance 1; AUC = Area under the curve; se = standard error; Se = Sensitivity; Sp = Specificity; Pr = Prevalence e.g., artificial insemination success in Manech Tête Rousse dairy sheep; PPV = Positive predictive value; NPV= Negative predictive value.

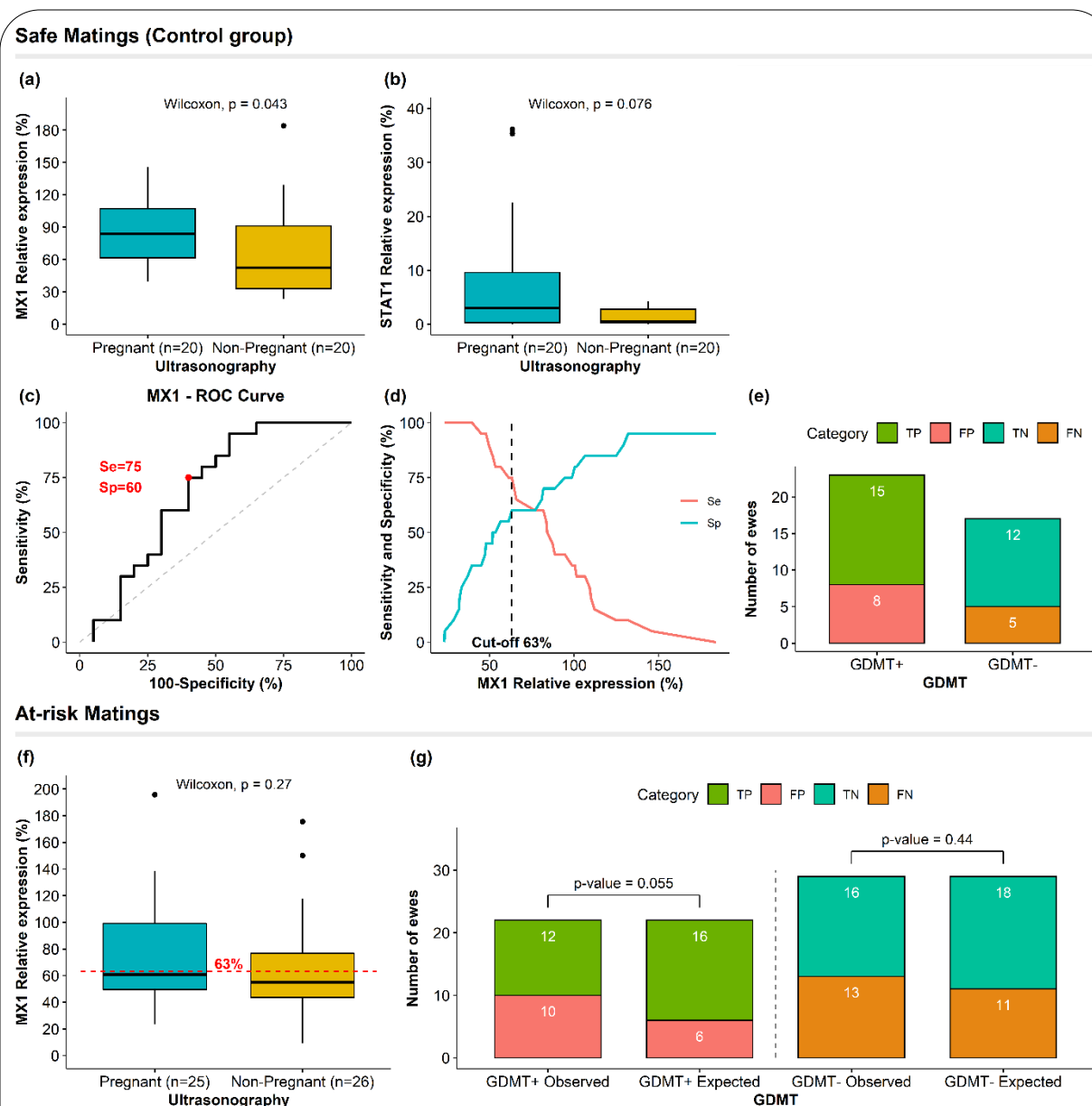


Figure 5. Gestation diagnosis molecular test (GDMT). Relative expression at 15 days post-IA of *MX1* (a) and *STAT1* (b) mRNA in pregnant and non-pregnant ewes assessed by ultrasonography in safe matings as control group. (c) ROC curve based on sensitivity (Se) and specificity (Sp) of GDMT using *MX1* relative expression in the control group. (d) Determination of *MX1* relative expression cut-off value (63%) associated with $Se=75\%$ and $Sp=60\%$ in the control group using the ROC01 method. (e) Distribution of pregnant and non-pregnant ewes according to GDMT based on *MX1* relative expression in the control group (TP, true positive; FP, false positive; TN, true negative; FN, false negative). (f) *MX1* relative expression between pregnant and non-pregnant ewes in at-risk matings. (g) Distribution of observed and expected pregnant and non-pregnant ewes according to GDMT based on *MX1* relative expression with a cut-off value of 63% in at-risk matings. The expected numbers were calculated based on prevalence ($Pr = 60.3\%$), positive predictive value ($PPV = 74\%$) and negative predictive value ($NPV = 61\%$). Differences between observed and expected numbers from GDMT (+ and -) were assessed by a Chi-squared test.

Discussion

We recently reported the segregation of five MTRDHH haplotypes with deficit in homozygous animals in MTR dairy sheep followed by the identification of a recessive lethal variant in the *MMUT* gene carried by the MTRDHH1 haplotype increasing stillbirth rate (Ben Braiek et al., 2023). In the present study, we focused on MTRDHH2, the second most significant haplotype, having a negative impact on both AI success (-3.0%) and stillbirth rate (+4.3%), suggesting that MTRDHH2 could also host a recessive lethal variant. Using WGS data, we fine-mapped a single base pair duplication (NC_040252.1:g.252,649,023dupG) in the *SLC33A1* gene previously highlighted as the only positional and functional candidate gene for MTRDHH2 (Ben Braiek et al., 2023).

SLC33A1 (also known as AT-1, acetyl-CoA transporter 1) is an essential protein involved in metabolism for transporting acetyl-CoA through the endoplasmic reticulum (ER) membrane (Jonas et al., 2010). It plays a role in N-lysine acetylation of ER proteins and regulates the degradation of protein aggregates by autophagy. When SLC33A1 is not functional, protein aggregates accumulate in the ER lumen resulting in autophagic cell death (Peng et al., 2014). Based on the sheep gene atlas (<http://biogps.org/sheepatlas/>; accessed 18 July 2022), *SLC33A1* appeared to be ubiquitously expressed (Clark et al., 2017). In the present study, the *SLC33A1_dupG* variant could result in a premature stop codon leading to a truncated protein of 248 amino acids (XP_011956340.1:p.Arg246Alafs*3) compared to the full-length protein encompassing 550 amino acids. Consequently, only three out of nine transmembrane domains would be translated into the mutant form of ovine SLC33A1, resembling a natural knock-out in sheep. Interestingly, *SLC33A1* is known to be embryonic lethal and decreases survival rate in knock-out mouse (MGI:1332247, <http://www.informatics.jax.org>). Knockdown of *SLC33A1* in zebrafish was also reported to cause defective axon outgrowth affecting BMP signaling (Liu et al., 2017). Additionally, point variants in this gene were associated with congenital cataracts, hearing loss, and neurodegeneration ((Huppke et al., 2012), OMIM 614482, <https://omim.org>), and also autosomal dominant spastic paraplegia (Lin et al., 2008; Liu et al., 2017; Mao et al., 2015; OMIM 612539) in human. These observations fit well with the decreased AI success in crosses between *SLC33A1_dupG* heterozygous carriers, the increased mortality of homozygous *dupG/dupG* newborn lambs, and the observation of a *dupG/dupG* lamb with locomotor problem resembling spastic paraplegia. Several variants in *SLC33A1* were also associated with low serum copper and ceruloplasmin in human (Huppke et al., 2012), however biochemical analyses were not performed in our study.

The generation and birth of *SLC33A1_dupG/dupG* homozygous lambs allowed us to confirm the increased stillbirth rate firstly observed for MTRDHH2 and thus the recessive lethal characteristic of this variant, heterozygous being not significantly affected. By oriented mating we also observed a 12% reduction of AI success that could be associated with embryonic losses in at-risk mating compared to safe mating. In order to explore this hypothesis, we tested a molecular diagnosis of gestation 15 days after IA (the time of embryo implantation) to be compared to the classic ultrasound test done between 45 days and 60 days post-IA. This molecular test, based on expression in blood cells of the *MX1* gene, a well-known interferon-tau stimulated gene (Mauffré et al., 2016), predicted 16 pregnant ewes while only 12 were really pregnant. This led us to think that for 4 heterozygous ewes, fetal losses had occurred between 15 days and 60 days of gestation due to lethality of *dupG/dupG* homozygous fetuses. The role of SLC33A1 in regulating acetyl-CoA metabolism could largely explain embryogenesis defect. Indeed, acetyl-CoA may play a major role in the regulation of cell growth, proliferation and apoptosis, suggesting that metabolic deficiency in acetyl-CoA is responsible for embryo failure (Tsuchiya et al., 2014).

Other methods of early pregnancy detection are available such as early transrectal ultrasonography (Rickard et al., 2017) or circulating biomarker dosages as progesterone,

protein B (PSP-B) or pregnancy associated glycoproteins (PAGs) (Karen et al., 2003). However, these approaches are only really effective from the 28th day of gestation after the expected time of return to estrous (Mauffré et al., 2016). Even if the GDMT helped us to highlight early fetal losses and offer good opportunities to predict the gestation status at day 15, its application in farm is still limited today. Indeed, only 66% of the RNA samples of this study were exploitable. This issue was mainly due to RNA degradation with limited storage control of the blood sample when collected in farms. The samples should be frozen directly within adequate buffer inactivating RNase. Focusing on the diagnosis method, the AUC of a ROC curve is a good estimator for assessing the power of a diagnosis test (Janssens and Martens, 2020). In the present case, we get an AUC of 0.687 which is considered as acceptable (Swets, 1988). The cut-off value of 63% for the *MXI* relative expression was determined by the ROC01 method instead of the Youden index generally used in ROC analysis (Perkins and Schisterman, 2006). Indeed, the Youden index gives an equal weight to sensitivity and specificity, while the optimal cut-off point of the ROC01 method favors the sensitivity in order to identify ewes with a positive GDMT and pregnant at days 45-60. Initially, the GDMT was based on the expression of two ISG, *MXI* and *STAT1*, but only *MXI* mRNA relative expression was different between pregnant and non-pregnant ewes in the control group. As indicated earlier, IFNT is mainly produced by the conceptus in a period between 14 to 17 days after AI and blood samples were collected at day 15 as proposed in (Mauffré et al., 2016). However, *STAT1* showed a higher mRNA expression at days 16-17 (Spencer et al., 2008) and the overexpression of *MXI* mRNA could be maintained until day 21 in peripheral blood mononuclear cells (Yankey et al., 2001). Thus, day 17 should be preferable to collect blood samples for further analysis.

Various variants are known in the superfamily of *SLC* genes accounting more than 400 members. About a hundred of *SLC* genes are associated with mendelian disorders in human mainly neurometabolic diseases under an autosomal recessive mode of inheritance (Schaller and Lauschke, 2019). Also in domestic animals, numerous causative variants in genes of the *SLC* family were evidenced, the vast majority being in dog and cattle (see omia.org). For example, and by using the reverse genetic screen approach in cattle, a missense variant in *SLC35A3* gene was evidenced as responsible for CVM, the complex vertebral malformation syndrome (rs438228855; (Thomsen et al., 2006; VanRaden et al., 2011)). CVM was previously identified with alive homozygous calves but a huge proportion of homozygous animals died during gestation. The homozygote deficient haplotype MH2 in Montbéliarde cattle was associated with decreased calving rate likely caused by a nonsense variant in *SLC37A2* (p.R12*, (Fritz et al., 2013)). The same *SLC37A2* gene is likely to harbor causal variants for craniomandibular osteopathy in dog (Hytönen et al., 2016; Letko et al., 2020). In a rather original way, alterations of the *SLC45A2* gene affected coat color in species as varied as horse (Mariat et al., 2003), chicken and quail (Gunnarsson et al., 2007), medaka (Fukamachi et al., 2008), tiger (Xu et al., 2013), gorilla (Prado-Martinez et al., 2013), dog (Winkler et al., 2014) and cattle (Rothhammer et al., 2017). In sheep, only one variant was reported as a 1bp deletion in *SLC13A1* to cause chondrodysplasia (Zhao et al., 2012). As in human, genetic screen based on WGS data should allow to identify loss-of-function variants located in *SLC* family genes in livestock (Charlier et al., 2016). The exploration of Ensembl variation database for sheep (a dataset composed of 453 animals from 38 breeds all over the world, <https://www.sheepmap.org/>) revealed three different deleterious variants located in the coding sequence of *SLC33A1*, one nonsense SNV segregating in Composite and D'Man breeds (rs1093927723: p.Trp546*), and two missense SNV with a null SIFT score segregating in Norduz breed (rs1089904504: p.Asp172Tyr; rs1093173882: p.Tyr380Cys). These variants would likely result in loss-of-function of *SLC33A1*, and further studies would be of interest to assess for an impact on stillbirth rate in these breeds as observed for the *SLC33A1_dupG* we evidenced in MTR dairy sheep.

Conclusion

Reverse genetic strategy is really effective in identifying embryonic lethal variant but also hidden genetic diseases such as those affecting metabolism which can be easily confused with normal lambing death due to environmental causes. The present study identified a single base pair duplication in *SLC33A1* gene linked to the homozygous deficient haplotype MTRDHH2 previously identified in MTR dairy sheep. *SLC33A1* encodes for a transporter of acetyl-CoA, essential for cellular metabolism. This could explain the impact of the mutation at different moments of lamb development from fetal to perinatal stages. Anyway, with a proven role on AI success and lamb mortality, an appropriate management of *SLC33A1_dupG* variant in the MTR dairy sheep selection scheme should improve the overall fertility and lamb survival.

Author contributions

MBB performed the analyses, interpreted the results, and drafted the manuscript. SS performed and analyzed the molecular diagnostic test of gestation. CA and FF managed mating in farms. PB performed bioinformatics. FPP, JS and FW managed the biological samples and their genotyping. MBB, CMR and SF conceived and designed the research. SF supervised the analyses, helped interpret the results, wrote, reviewed and edited the final manuscript. All authors read and approved the final manuscript.

Acknowledgments

We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) for providing help and/or computing and/or storage resources. The authors thank the Manech Tête Rousse breeding company who provided access to private genomic data and biological samples. We thank also the breeders involved in the project. We are grateful to NEIKER for providing samples for Spanish ovine diversity panel.

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

The WGS data used in this study are publicly available, EMBL-EBI accession numbers are described in Table S1.

Ethic approval

The experimental procedures on animals (blood sampling, ear biopsies) were approved 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 (approval numbers APAFIS#30615-2021032318054889 v5).

ORCID

Maxime BEN BRAIEK, ORCID iD <https://orcid.org/0000-0003-4770-0867>
Stéphane FABRE, ORCID iD <https://orcid.org/0000-0001-7350-9500>

References

- Abdalla, E.A., Id-Lahoucine, S., Cánovas, A., Casellas, J., Schenkel, F.S., Wood, B.J., Baes, C.F., 2020. Discovering lethal alleles across the turkey genome using a transmission ratio distortion approach. *Anim. Genet.* 51, 876–889. <https://doi.org/10.1111/age.13003>
- Astruc, J.-M., Baloché, G., Buisson, D., Labatut, J., Lagriffoul, G., Larroque, H., Robert-Granie, C., Legarra, A., Barillet, F., 2016. Genomic selection in French dairy sheep. *INRAE Prod. Anim.* 29, 41–55.
- Bazer, F.W., 2013. Pregnancy recognition signaling mechanisms in ruminants and pigs. *J. Anim. Sci. Biotechnol.* 4, 23. <https://doi.org/10.1186/2049-1891-4-23>
- Ben Braiek, M., Fabre, S., Hozé, C., Astruc, J.-M., Moreno-Romieux, C., 2021. Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep. *Genet. Sel. Evol.* 53, 41. <https://doi.org/10.1186/s12711-021-00634-1>
- Ben Braiek, M., Moreno-Romieux, C., Allain, C., Bardou, P., Bordes, A., Debat, F., Drögemüller, C., Plisson-Petit, F., Portes, D., Sarry, J., Tadi, N., Woloszyn, F., Fabre, S., 2022. A Nonsense Variant in CCDC65 Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep. *Genes* 13, 45. <https://doi.org/10.3390/genes13010045>
- Ben Braiek, M., Moreno-Romieux, C., André, C., Astruc, J.-M., Bardou, P., Bordes, A., Debat, F., Fidelle, F., Hozé, C., Plisson-Petit, F., Rivemale, F., Sarry, J., Tadi, N., Woloszyn, F., Fabre, S., 2023. Homozygous haplotype deficiency in Manech Tête Rousse dairy sheep revealed a nonsense variant in MMUT gene affecting newborn lamb viability. <https://doi.org/10.1101/2023.03.10.531894>
- Bindon, B.M., 1971. Systematic Study of Preimplantation Stages of Pregnancy in the Sheep. *Aust. J. Biol. Sci.* 24, 131–148. <https://doi.org/10.1071/bi9710131>
- Chantepie, L., Bodin, L., Sarry, J., Woloszyn, F., Plisson-Petit, F., Ruesche, J., Drouilhet, L., Fabre, S., 2020. Genome-Wide Identification of a Regulatory Mutation in BMP15 Controlling Prolificacy in Sheep. *Front. Genet.* 11. <https://doi.org/10.3389/fgene.2020.00585>
- Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J.S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., Frennet, J.-C., Hanset, R., Hubin, X., Jorgensen, C., Karim, L., Kent, M., Harvey, K., Pearce, B.R., Simon, P., Tama, N., Nie, H., Vandeputte, S., Lien, S., Longeri, M., Fredholm, M., Harvey, R.J., Georges, M., 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet.* 40, 449–454. <https://doi.org/10.1038/ng.96>
- Charlier, C., Li, W., Harland, C., Littlejohn, M., Coppieters, W., Creagh, F., Davis, S., Druet, T., Faux, P., Guillaume, F., Karim, L., Keehan, M., Kadri, N.K., Tamma, N., Spelman, R., Georges, M., 2016. NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Res.* 26, 1333–1341. <https://doi.org/10.1101/gr.207076.116>
- Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., Lu, X., 2012. Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet.* 3, 35. <https://doi.org/10.3389/fgene.2012.00035>
- Clark, E.L., Bush, S.J., McCulloch, M.E.B., Farquhar, I.L., Young, R., Lefevre, L., Pridans, C., Tsang, H.G., Wu, C., Afrasiabi, C., Watson, M., Whitelaw, C.B., Freeman, T.C., Summers, K.M., Archibald, A.L., Hume, D.A., 2017. A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLOS Genet.* 13, e1006997. <https://doi.org/10.1371/journal.pgen.1006997>
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes

- Project Analysis Group, 2011. The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Demars, J., Cano, M., Drouilhet, L., Plisson-Petit, F., Bardou, P., Fabre, S., Servin, B., Sarry, J., Woloszyn, F., Mulsant, P., Foulquier, D., Carrière, F., Aletru, M., Rodde, N., Cauet, S., Bouchez, O., Pirson, M., Tosser-Klopp, G., Allain, D., 2017. Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep. *Mol. Biol. Evol.* 34, 1722–1729. <https://doi.org/10.1093/molbev/msx114>
- Derks, M.F.L., Megens, H.-J., Bosse, M., Lopes, M.S., Harlizius, B., Groenen, M.A.M., 2017. A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics* 18, 858. <https://doi.org/10.1186/s12864-017-4278-1>
- Derks, M.F.L., Megens, H.-J., Bosse, M., Visscher, J., Peeters, K., Bink, M.C.A.M., Vereijken, A., Gross, C., de Ridder, D., Reinders, M.J.T., Groenen, M.A.M., 2018. A survey of functional genomic variation in domesticated chickens. *Genet. Sel. Evol.* 50, 17. <https://doi.org/10.1186/s12711-018-0390-1>
- Diskin, M., Morris, D., 2008. Embryonic and Early Foetal Losses in Cattle and Other Ruminants. *Reprod. Domest. Anim.* 43, 260–267. <https://doi.org/10.1111/j.1439-0531.2008.01171.x>
- Dixon, A.B., Knights, M., Winkler, J.L., Marsh, D.J., Pate, J.L., Wilson, M.E., Dailey, R.A., Seidel, G., Inskeep, E.K., 2007. Patterns of late embryonic and fetal mortality and association with several factors in sheep1. *J. Anim. Sci.* 85, 1274–1284. <https://doi.org/10.2527/jas.2006-129>
- Eggen, A., 2012. The development and application of genomic selection as a new breeding paradigm. *Anim. Front.* 2, 10–15. <https://doi.org/10.2527/af.2011-0027>
- Fabre, S., Chantepie, L., Plisson-Petit, F., Sarry, J., Woloszyn, F., Genet, C., Drouilhet, L., Tosser-Klopp, G., 2020. A novel homozygous nonsense mutation in ITGB4 gene causes epidermolysis bullosa in Mouton Vendéen sheep. *Anim. Genet.* <https://doi.org/10.1111/age.13026>
- Fritz, S., Capitan, A., Djari, A., Rodriguez, S.C., Barbat, A., Baur, A., Grohs, C., Weiss, B., Boussaha, M., Esquerré, D., Klopp, C., Rocha, D., Boichard, D., 2013. Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2. *PloS One* 8, e65550. <https://doi.org/10.1371/journal.pone.0065550>
- Fukamachi, S., Kinoshita, M., Tsujimura, T., Shimada, A., Oda, S., Shima, A., Meyer, A., Kawamura, S., Mitani, H., 2008. Rescue From Oculocutaneous Albinism Type 4 Using Medaka slc45a2 cDNA Driven by Its Own Promoter. *Genetics* 178, 761–769. <https://doi.org/10.1534/genetics.107.073387>
- Georges, M., Charlier, C., Hayes, B., 2019. Harnessing genomic information for livestock improvement. *Nat. Rev. Genet.* 20, 135–156. <https://doi.org/10.1038/s41576-018-0082-2>
- Goksuluk, D., Korkmaz, S., Zararsiz, G., Karaagaoglu, A., Ergun, 2016. easyROC: An Interactive Web-tool for ROC Curve Analysis Using R Language Environment. *R J.* 8, 213. <https://doi.org/10.32614/RJ-2016-042>
- Gunnarsson, U., Hellström, A.R., Tixier-Boichard, M., Minvielle, F., Bed'hom, B., Ito, S., Jensen, P., Rattink, A., Vereijken, A., Andersson, L., 2007. Mutations in SLC45A2 Cause Plumage Color Variation in Chicken and Japanese Quail. *Genetics* 175, 867–877. <https://doi.org/10.1534/genetics.106.063107>
- Huppke, P., Brendel, C., Kalscheuer, V., Korenke, G.C., Marquardt, I., Freisinger, P., Christodoulou, J., Hillebrand, M., Pitelet, G., Wilson, C., Gruber-Sedlmayr, U., Ullmann, R., Haas, S., Elpeleg, O., Nürnberg, G., Nürnberg, P., Dad, S., Møller, L.B.,

- Kaler, S.G., Gärtner, J., 2012. Mutations in SLC33A1 Cause a Lethal Autosomal-Recessive Disorder with Congenital Cataracts, Hearing Loss, and Low Serum Copper and Ceruloplasmin. *Am. J. Hum. Genet.* 90, 61–68. <https://doi.org/10.1016/j.ajhg.2011.11.030>
- Hytönen, M.K., Arumilli, M., Lappalainen, A.K., Owczarek-Lipska, M., Jagannathan, V., Hundi, S., Salmela, E., Venta, P., Sarkiala, E., Jokinen, T., Gorgas, D., Kere, J., Nieminen, P., Drögemüller, C., Lohi, H., 2016. Molecular Characterization of Three Canine Models of Human Rare Bone Diseases: Caffey, van den Ende-Gupta, and Raine Syndromes. *PLOS Genet.* 12, e1006037. <https://doi.org/10.1371/journal.pgen.1006037>
- Janssens, A.C.J.W., Martens, F.K., 2020. Reflection on modern methods: Revisiting the area under the ROC Curve. *Int. J. Epidemiol.* 49, 1397–1403. <https://doi.org/10.1093/ije/dyz274>
- Jenko, J., 2019. Analysis of a large dataset reveals haplotypes carrying putatively recessive lethal and semi-lethal alleles with pleiotropic effects on economically important traits in beef cattle 14.
- Jonas, M.C., Pehar, M., Puglielli, L., 2010. AT-1 is the ER membrane acetyl-CoA transporter and is essential for cell viability. *J. Cell Sci.* 123, 3378–3388. <https://doi.org/10.1242/jcs.068841>
- Karen, A., Beckers, J.-F., Sulon, J., de Sousa, N.M., Szabados, K., Reczigel, J., Szenci, O., 2003. Early pregnancy diagnosis in sheep by progesterone and pregnancy-associated glycoprotein tests. *Theriogenology* 59, 1941–1948. [https://doi.org/10.1016/S0093-691X\(02\)01289-X](https://doi.org/10.1016/S0093-691X(02)01289-X)
- Lander, E.S., Botstein, D., 1987. Homozygosity Mapping: A Way to Map Human Recessive Traits with the DNA of Inbred Children. *Science* 236, 1567–1570. <https://doi.org/10.1126/science.2884728>
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, Christina, Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H.,

- Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglu, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, Christopher, Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M.J., International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, C. for G.R., The Sanger Centre, Washington University Genome Sequencing Center, US DOE Joint Genome Institute, Baylor College of Medicine Human Genome Sequencing Center, RIKEN Genomic Sciences Center, Genoscope and CNRS UMR-8030, Department of Genome Analysis, I. of M.B., GTC Sequencing Center, Beijing Genomics Institute/Human Genome Center, Multimegabase Sequencing Center, T.I. for S.B., Stanford Genome Technology Center, University of Oklahoma's Advanced Center for Genome Technology, Max Planck Institute for Molecular Genetics, Cold Spring Harbor Laboratory, L.A.H.G.C., GBF—German Research Centre for Biotechnology, *Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings), Scientific management: National Human Genome Research Institute, U.N.I. of H., Stanford Human Genome Center, University of Washington Genome Center, Department of Molecular Biology, K.U.S. of M., University of Texas Southwestern Medical Center at Dallas, Office of Science, U.D. of E., The Wellcome Trust, 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. <https://doi.org/10.1038/35057062>
- Letko, A., Leuthard, F., Jagannathan, V., Corlazzoli, D., Matiasek, K., Schweizer, D., Hytönen, M.K., Lohi, H., Leeb, T., Drögemüller, C., 2020. Whole Genome Sequencing Indicates Heterogeneity of Hyperostotic Disorders in Dogs. *Genes* 11, 163. <https://doi.org/10.3390/genes11020163>
- Lin, P., Li, Jianwei, Liu, Q., Mao, F., Li, Jisheng, Qiu, R., Hu, H., Song, Y., Yang, Y., Gao, G., Yan, C., Yang, W., Shao, C., Gong, Y., 2008. A Missense Mutation in SLC33A1, which Encodes the Acetyl-CoA Transporter, Causes Autosomal-Dominant Spastic Paraplegia (SPG42). *Am. J. Hum. Genet.* 83, 752–759. <https://doi.org/10.1016/j.ajhg.2008.11.003>
- Liu, P., Jiang, B., Ma, J., Lin, P., Zhang, Y., Shao, C., Sun, W., Gong, Y., 2017. S113R mutation in SLC33A1 leads to neurodegeneration and augmented BMP signaling in a mouse model. *Dis. Model. Mech.* 10, 53–62. <https://doi.org/10.1242/dmm.026880>
- MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., Albers, C.A., Zhang, Z.D., Conrad, D.F., Lunter, G., Zheng, H., Ayub, Q., DePristo, M.A., Banks, E., Hu, M., Handsaker, R.E., Rosenfeld, J.A., Fromer, M., Jin, M., Mu, X.J., Khurana, E., Ye, K., Kay, M., Saunders, G.I., Suner, M.-M., Hunt, T., Barnes, I.H.A., Amid, C., Carvalho-Silva, D.R., Bignell, A.H., Snow, C., Yngvadottir, B., Bumpstead, S., Cooper, D.N., Xue, Y., Romero, I.G., Consortium, 1000 Genomes Project, Wang, J., Li, Y., Gibbs, R.A., McCarroll, S.A., Dermitzakis, E.T., Pritchard, J.K., Barrett, J.C., Harrow, J., Hurles, M.E., Gerstein, M.B., Tyler-Smith, C., 2012. A Systematic Survey of Loss-of-Function

- Variants in Human Protein-Coding Genes. *Science* 335, 823–828. <https://doi.org/10.1126/science.1215040>
- Mao, F., Li, Z., Zhao, B., Lin, P., Liu, P., Zhai, M., Liu, Q., Shao, C., Sun, W., Gong, Y., 2015. Identification and Functional Analysis of a SLC33A1: c.339T>G (p.Ser113Arg) Variant in the Original SPG42 Family. *Hum. Mutat.* 36, 240–249. <https://doi.org/10.1002/humu.22732>
- Mariat, D., Taourit, S., Guérin, G., 2003. A mutation in the MATP gene causes the cream coat colour in the horse. *Genet. Sel. Evol.* 35, 119. <https://doi.org/10.1186/1297-9686-35-1-119>
- Mauffré, V., Grimard, B., Eozenou, C., Inghels, S., Silva, L., Giraud-Delville, C., Capo, D., Sandra, O., Constant, F., 2016. Interferon stimulated genes as peripheral diagnostic markers of early pregnancy in sheep: a critical assessment. *Animal* 10, 1856–1863. <https://doi.org/10.1017/S175173111600077X>
- Ott, T.L., Gifford, C.A., 2010. Effects of Early Conceptus Signals on Circulating Immune Cells: Lessons from Domestic Ruminants. *Am. J. Reprod. Immunol.* 64, 245–254. <https://doi.org/10.1111/j.1600-0897.2010.00912.x>
- Peng, Y., Li, M., Clarkson, B.D., Pehar, M., Lao, P.J., Hillmer, A.T., Barnhart, T.E., Christian, B.T., Mitchell, H.A., Bendlin, B.B., Sandor, M., Puglielli, L., 2014. Deficient Import of Acetyl-CoA into the ER Lumen Causes Neurodegeneration and Propensity to Infections, Inflammation, and Cancer. *J. Neurosci.* 34, 6772–6789. <https://doi.org/10.1523/JNEUROSCI.0077-14.2014>
- Perkins, N.J., Schisterman, E.F., 2006. The Inconsistency of “Optimal” Cut-points Using Two ROC Based Criteria. 11.
- Prado-Martinez, J., Hernando-Herraez, I., Lorente-Galdos, B., Dabad, M., Ramirez, O., Baeza-Delgado, C., Morcillo-Suarez, C., Alkan, C., Hormozdiari, F., Raineri, E., Estellé, J., Fernandez-Callejo, M., Valles, M., Ritscher, L., Schöneberg, T., de la Calle-Mustienes, E., Casillas, S., Rubio-Acero, R., Melé, M., Engelken, J., Caceres, M., Gomez-Skarmeta, J.L., Gut, M., Bertranpetit, J., Gut, I.G., Abello, T., Eichler, E.E., Mingarro, I., Lalueza-Fox, C., Navarro, A., Marques-Bonet, T., 2013. The genome sequencing of an albino Western lowland gorilla reveals inbreeding in the wild. *BMC Genomics* 14, 363. <https://doi.org/10.1186/1471-2164-14-363>
- Rickard, J.P., Ryan, G., Hall, E., de Graaf, S.P., Hermes, R., 2017. Using transrectal ultrasound to examine the effect of exogenous progesterone on early embryonic loss in sheep. *PLOS ONE* 12, e0183659. <https://doi.org/10.1371/journal.pone.0183659>
- Rochus, C.M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tosser-Klopp, G., Servin, B., 2018. Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. *BMC Genomics* 19, 71. <https://doi.org/10.1186/s12864-018-4447-x>
- Rothhammer, S., Kunz, E., Seichter, D., Krebs, S., Wassertheurer, M., Fries, R., Brem, G., Medugorac, I., 2017. Detection of two non-synonymous SNPs in SLC45A2 on BTA20 as candidate causal mutations for oculocutaneous albinism in Braunvieh cattle. *Genet. Sel. Evol.* 49, 73. <https://doi.org/10.1186/s12711-017-0349-7>
- Rupp, R., Mucha, S., Larroque, H., McEwan, J., Conington, J., 2016. Genomic application in sheep and goat breeding. *Anim. Front.* 6, 39–44. <https://doi.org/10.2527/af.2016-0006>
- Rupp, R., Senin, P., Sarry, J., Allain, C., Tasca, C., Ligat, L., Portes, D., Woloszyn, F., Bouchez, O., Tabouret, G., Lebastard, M., Caubet, C., Foucras, G., Tosser-Klopp, G., 2015. A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model. *PLoS Genet.* 11, e1005629. <https://doi.org/10.1371/journal.pgen.1005629>

- Schaller, L., Lauschke, V.M., 2019. The genetic landscape of the human solute carrier (SLC) transporter superfamily. *Hum. Genet.* 138, 1359–1377. <https://doi.org/10.1007/s00439-019-02081-x>
- Spencer, T.E., Sandra, O., Wolf, E., 2008. Genes involved in conceptus–endometrial interactions in ruminants: insights from reductionism and thoughts on holistic approaches. *Reproduction* 135, 165–179. <https://doi.org/10.1530/REP-07-0327>
- Swets, J.A., 1988. Measuring the Accuracy of Diagnostic Systems. *Science* 240, 1285–1293. <https://doi.org/10.1126/science.3287615>
- Thomsen, B., Horn, P., Panitz, F., Bendixen, E., Petersen, A.H., Holm, L.-E., Nielsen, V.H., Agerholm, J.S., Arnbjerg, J., Bendixen, C., 2006. A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Res.* 16, 97–105. <https://doi.org/10.1101/gr.3690506>
- Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. <https://doi.org/10.1093/bib/bbs017>
- Todd, E.T., Thomson, P.C., Hamilton, N.A., Ang, R.A., Lindgren, G., Viklund, Å., Eriksson, S., Mikko, S., Strand, E., Velie, B.D., 2020. A genome-wide scan for candidate lethal variants in Thoroughbred horses. *Sci. Rep.* 10, 13153. <https://doi.org/10.1038/s41598-020-68946-8>
- Tsuchiya, Y., Pham, U., Hu, W., Ohnuma, S., Gout, I., 2014. Changes in Acetyl CoA Levels during the Early Embryonic Development of *Xenopus laevis*. *PLOS ONE* 9, e97693. <https://doi.org/10.1371/journal.pone.0097693>
- Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., Posthuma, D., 2021. Genome-wide association studies. *Nat. Rev. Methods Primer* 1, 1–21. <https://doi.org/10.1038/s43586-021-00056-9>
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M., 2007. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35, W71–W74. <https://doi.org/10.1093/nar/gkm306>
- VanRaden, P.M., Olson, K.M., Null, D.J., Hutchison, J.L., 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* 94, 6153–6161. <https://doi.org/10.3168/jds.2011-4624>
- Wilmut, I., Sales, D.I., Ashworth, C.J., 1986. Maternal and embryonic factors associated with prenatal loss in mammals. *Reproduction* 76, 851–864. <https://doi.org/10.1530/jrf.0.0760851>
- Winkler, P.A., Gornik, K.R., Ramsey, D.T., Dubielzig, R.R., Venta, P.J., Petersen-Jones, S.M., Bartoe, J.T., 2014. A Partial Gene Deletion of SLC45A2 Causes Oculocutaneous Albinism in Doberman Pinscher Dogs. *PLOS ONE* 9, e92127. <https://doi.org/10.1371/journal.pone.0092127>
- Xu, X., Dong, G.-X., Hu, X.-S., Miao, L., Zhang, X.-L., Zhang, D.-L., Yang, H.-D., Zhang, T.-Y., Zou, Z.-T., Zhang, T.-T., Zhuang, Y., Bhak, J., Cho, Y.S., Dai, W.-T., Jiang, T.-J., Xie, C., Li, R., Luo, S.-J., 2013. The Genetic Basis of White Tigers. *Curr. Biol.* 23, 1031–1035. <https://doi.org/10.1016/j.cub.2013.04.054>
- Yankey, S.J., Hicks, B.A., Carnahan, K.G., Assiri, A.M., Sinor, S.J., Kodali, K., Stellflug, J.N., Ott, T.L., 2001. Expression of the antiviral protein Mx in peripheral blood mononuclear cells of pregnant and bred, non-pregnant ewes. *J. Endocrinol.* 170, R7-11. <https://doi.org/10.1677/joe.0.170r007>
- Zhao, X., Onteru, S.K., Piripi, S., Thompson, K.G., Blair, H.T., Garrick, D.J., Rothschild, M.F., 2012. In a shake of a lamb's tail: using genomics to unravel a cause of chondrodysplasia in Texel sheep. *Anim. Genet.* 43, 9–18. <https://doi.org/10.1111/j.1365-2052.2011.02304.x>

Supporting information

Table S1. EMBL-EBI accession numbers of the 100 whole-genome sequences used in the analysis.

Breed	Number of animals	ENA run accession	EBI project accession	Associated references
Belclare	2	SRR14934360; SRR14935129	PRJNA698548	
Berrichon du Cher	3	ERS1205899; ERS1205900; ERS1205901	PRJEB14418	(Demars et al., 2017)
Cambridge	7	ERR1419201; ERR1419202; ERR1419203; ERR1419204; ERR1419205; ERR1419206; ERR1419207	PRJEB14098	
Charollais	1	SRR14934359	PRJNA698548	
Lacaune (Dairy)	31	ERR3276357; ERR3276358; ERR3276359; ERR3276360; ERR3276361; ERR3276362; ERR3276363; ERR3276364; ERR3276365; ERR3276366; ERR3276367; ERR3276368; ERR3276369; ERR3276370; ERR3276371; ERR3276372; ERR3276373; ERR3276374; ERR3276375; ERR3276376; ERR3276377; ERR3276378; ERR3276379; ERR7891349; ERR7891350; ERR7891351 ERR968423; ERR968424; ERR968425 SRR501850; SRR501851	PRJEB32110 PRJEB9911 PRJNA160933	(Ben Braiek et al., 2022; Rupp et al., 2015)
Lacaune (Meat)	3	ERR3276380; ERR3276381; ERR3276382	PRJEB32110	
Manech Tête Rousse	22	ERR3712282; ERR3712283; ERR3712284; ERR3712285; ERR3712286; ERR3712287*; ERR3712288*; ERR3712289; ERR3712290*; ERR3712291; ERR3712292; ERR3712293; ERR3712294; ERR3712295*; ERR3712296; ERR3712297*; ERR3712298; ERR3712299; ERR3712300; ERR3712301; ERR7889920; ERR7889921	PRJEB35682	(Ben Braiek et al., 2023)
Martinik Blackbelly	1	ERR3255914	PRJEB31930	
Noire du velay	2	ERR3828659; ERR3828660	PRJEB35553	(Chantepie et al., 2020)
Romane	4	ERS1205902; ERS1205903; ERR2818429; ERR2818430	PRJEB14418	(Demars et al., 2017)
Romane x Martinik Blackbelly	13	ERR3255915; ERR3255916; ERR3255917; ERR3255918; ERR3255919; ERR3988554; ERR3988555; ERR3988556; ERR3988557; ERR3988558; ERR3988559; ERR3988560; ERR3988561	PRJEB31930	
Romanov	2	ERS1205904; ERS1205905	PRJEB14418	(Demars et al., 2017)
Suffolk	2	SRR14934357; SRR14934358	PRJNA698548	
Texel	2	SRR14934355; SRR14934356	PRJNA698548	
Vendéen	5	ERR4236129; ERR4236130; ERR4236131 SRR14934353; SRR14934354	PRJEB37460 PRJNA698548	(Fabre et al., 2020)
Total	100			

* MTRDHH2 heterozygous carrier

- Ben Braiek, M., Moreno-Romieux, C., Allain, C., Bardou, P., Bordes, A., Debat, F., Drögemüller, C., Plisson-Petit, F., Portes, D., Sarry, J., Tadi, N., Woloszyn, F., Fabre, S., 2022. A Nonsense Variant in CCDC65 Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep. *Genes* 13, 45. <https://doi.org/10.3390/genes13010045>
- Ben Braiek, M., Moreno-Romieux, C., André, C., Astruc, J.-M., Bardou, P., Bordes, A., Debat, F., Fidelle, F., Hozé, C., Plisson-Petit, F., Rivemale, F., Sarry, J., Tadi, N., Woloszyn, F., Fabre, S., 2023. Homozygous haplotype deficiency in Manech Tête Rousse dairy sheep revealed a nonsense variant in MMUT gene affecting newborn lamb viability. <https://doi.org/10.1101/2023.03.10.531894>
- Chantepie, L., Bodin, L., Sarry, J., Woloszyn, F., Plisson-Petit, F., Ruesche, J., Drouilhet, L., Fabre, S., 2020. Genome-Wide Identification of a Regulatory Mutation in BMP15 Controlling Prolificacy in Sheep. *Front. Genet.* 11. <https://doi.org/10.3389/fgene.2020.00585>
- Demars, J., Cano, M., Drouilhet, L., Plisson-Petit, F., Bardou, P., Fabre, S., Servin, B., Sarry, J., Woloszyn, F., Mulsant, P., Foulquier, D., Carrière, F., Aletru, M., Rodde, N., Cauet, S., Bouchez, O., Pirson, M., Tosser-Klopp, G., Allain, D., 2017. Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep. *Mol. Biol. Evol.* 34, 1722–1729. <https://doi.org/10.1093/molbev/msx114>
- Fabre, S., Chantepie, L., Plisson-Petit, F., Sarry, J., Woloszyn, F., Genet, C., Drouilhet, L., Tosser-Klopp, G., 2020. A novel homozygous nonsense mutation in ITGB4 gene causes epidermolysis bullosa in Mouton Vendéen sheep. *Anim. Genet.* <https://doi.org/10.1111/age.13026>
- Rupp, R., Senin, P., Sarry, J., Allain, C., Tasca, C., Ligat, L., Portes, D., Woloszyn, F., Bouchez, O., Tabouret, G., Lebastard, M., Caubet, C., Foucras, G., Tosser-Klopp, G., 2015. A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model. *PLoS Genet.* 11, e1005629. <https://doi.org/10.1371/journal.pgen.1005629>

Table S2. List of PCR primer sequences.

Application	Gene symbol (Transcript accession number)	Gene description	Sequence (5'→3')	Size (pb)	Efficacy
Terra PCR	<i>SLC33A1</i> NC_040252.1	Solute carrier family 33 member 1	F: TGGTTGGGCGTTAACTATGT R: CCTTTCCTTGCTCACITTCAG	402	-
PACE PCR	<i>SLC33A1</i> NC_040252.1	Solute carrier family 33 member 1	F1: GAAGGTGACCAAGTTCATGCTGGGTTGAGGCTCAAACCGCA F2: GAAGGTCGGAGTCAACGGATTGGGTTGAGGCTCAAACCGCC R: CTCGAGTCTGCCGACTTTTGTAAACA	72/73	-
Quantitative PCR	<i>MX1</i> * NM_001009753	Myxovirus (influenza virus) resistance 1	F: CCACCCACCGACAGCTCCCCCT R: GCAGGTGTGGGCGTGAAGCA	167	2.35
	<i>STAT1</i> NM_001166203	Signal transducer and activator of transcription 1	F: CTGTCCCTTCTTCTGTAAC R: TTCCCTACAGAACCTTGTC	186	1.95
	<i>GAPDH</i> NM_001190390	Glyceraldehyde-3-phosphate dehydrogenase	F: CGACTTCAAACAGCGACACTC R: TGCTGTAGCCGAATTCATTG	113	1.90
	<i>YWHAZ</i> NM_001267887	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta	F: ATTAAGTGAAGAGTCATACAA R: GTATCCGATGTCCACAAT	81	2.05

F= Forward primer; R= Reverse primer.

*Same primer used in (Mauffré et al., 2016)

Table S3. List of variants located in MTRDHH2 extended by 1 Mb from each side with perfect correlation with MTRDHH2 status.

Position	Ref/Alt	quality score	Location Annotation	Impact prediction, Functional Consequence
251986773	AG/A	844.5	intronic, <i>LOC114109706</i> , <i>KCNABI</i>	
252145590	C/T	937.55	intronic, <i>KCNABI</i>	
252224500	T/A	512.26	intronic, <i>KCNABI</i>	
252339893	G/A	895.52	Intergenic	
252359862	AG/A	1018.83	Intergenic	
252371225	TG/T	903.55	Intergenic	
252394185	C/T	841.11	Intergenic	
252535183	G/A	1020.15	Intergenic	
252649023	dupG	492.62	Exonic, <i>SLC33A1</i> (c.735dupG)	High, single base pair duplication (frameshift) (p.Arg246fs)
252666223	A/G	741.98	Intronic, <i>SLC33A1</i>	
252780758	C/T	873.31	Intronic, <i>PLCHI</i>	
252870096	C/T	986.41	Intronic, <i>PLCHI</i>	
252916791	T/G	1269.81	Intronic, <i>PLCHI</i>	
252932261	G/A	889.79	Intronic, <i>PLCHI</i>	
253023490	A/G	706.08	Intergenic	
253152130	G/C	914.67	Intergenic	
253171663	G/GC	663.97	Intergenic	
253249851	C/T	928.86	Intergenic	
253328216	G/A	1031.66	Intronic, <i>MME</i>	
253523817	A/AT	737.7	Intergenic	
253623685	C/A	568.47	Intergenic	
253954730	C/T	1005.62	Intergenic	
254136111	G/C	793.98	Intergenic	
254454799	C/G	694.24	Intronic, <i>ARHGEF26</i>	
254562583	C/CG	625.31	Intronic, <i>ARHGEF26</i>	
254608121	G/T	1207.14	Intergenic	
254641567	G/A	885.8	Intergenic	
254660722	G/T	788.47	Intergenic	
254714255	G/T	849.41	Intergenic	
254811992	C/T	1019.52	Intergenic	
254843691	C/A	1368.63	Intergenic	
254990600	T/A	316.16	Intergenic	
255045557	A/C	805.8	Intergenic	
255244108	G/A	1036.98	Intergenic	
255267011	G/A	563.02	Intergenic	
255273842	G/A	767.08	Intergenic	
255274025	C/A	826.64	Intergenic	
255274893	T/G	977.58	Intergenic	
255276345	G/A	1146.74	Intergenic	

255277890	C/T	709.91	Intergenic
255278315	A/G	971.62	Intergenic
255279387	TG/T	796.08	Intergenic
255279578	A/G	1010.91	Intergenic
255281448	G/A	1117.71	Intergenic
255289447	G/A	860.72	Intergenic
255290723	T/C	873.14	Intergenic
255291778	A/G	1191.43	Intergenic
255292025	A/G	864.76	Intergenic
255303312	T/C	1004.49	Intergenic
255380418	A/G	502.3	Intergenic
255486316	G/A	355.22	Intergenic
255501900	A/G	770.81	Intergenic
255501928	C/G	758.16	Intergenic
255517085	G/A	971.39	Intergenic
255559980	C/T	889.74	Intergenic
255648413	C/T	914.6	Intergenic
255649190	A/G	1108.66	Intergenic
255649208	C/A	1026.97	Intergenic
255649895	C/T	1097.46	Intergenic
255650084	C/T	994.22	Intergenic
255650637	C/A	947.1	Intergenic
255650736	G/A	1042.9	Intergenic
255651058	C/A	1109.74	Intergenic
255651820	G/T	1000.01	Intergenic
255651983	T/A	757.47	Intergenic
255652382	T/G	732.98	Intergenic
255652785	C/T	1284.14	Intergenic
255654050	T/C	1091.05	Intergenic
255654283	G/T	1004.03	Intergenic
255655935	C/T	981.43	Intergenic
255655960	G/C	746.61	Intergenic
255656034	G/A	773.74	Intergenic
255656913	A/C	1010.63	Intergenic
255657037	C/A	1093.98	Intergenic
255657602	CTA/C	1236.48	Intergenic
255657673	T/C	872.88	Intergenic
255658018	G/A	944.03	Intergenic
255658626	A/G	1148.34	Intergenic
255659501	T/G	603.63	Intergenic
255660246	C/T	1213.23	Intergenic
255660247	A/G	1213.26	Intergenic
255660523	A/G	970.32	Intergenic
255660735	C/T	1032.79	Intergenic
255660825	A/G	1275.1	Intergenic
255661059	T/C	939.97	Intergenic
255661119	T/G	1058.82	Intergenic
255661183	A/G	1098.68	Intergenic

255661215	G/C	1149.61	Intergenic
255661350	A/AG	872.08	Intergenic
255661387	A/G	776.74	Intergenic
255661508	T/C	791.75	Intergenic
255661663	G/A	745.48	Intergenic
255661669	A/G	742.5	Intergenic
255661700	G/A	741.92	Intergenic
255661789	G/A	1105.62	Intergenic
255661830	A/G	1315.49	Intergenic
255661837	T/C	1201.79	Intergenic
255661866	T/C	959.61	Intergenic
255661916	C/T	746.19	Intergenic
255662102	GGATTGC TACAATC/ G	1006.86	Intergenic
255662260	G/A	1108.66	Intergenic
255662325	T/C	906.57	Intergenic
255662513	TAC/T	809.16	Intergenic
255662987	G/A	977.53	Intergenic
255663287	A/C	1021.77	Intergenic
255663295	C/T	1031.66	Intergenic
255663376	C/T	1080.46	Intergenic
255663498	C/A	941.52	Intergenic
255663553	C/T	1089.18	Intergenic
255663618	G/T	1153.79	Intergenic
255663709	C/G	1374.97	Intergenic
255663768	T/C	871.75	Intergenic
255664021	C/G	788.96	Intergenic
255664041	A/G	924.15	Intergenic
255664045	A/G	923.8	Intergenic
255664158	T/TG	880.8	Intergenic
255664645	C/T	505.5	Intergenic
255664677	T/C	620.41	Intergenic
255664733	C/G	640.1	Intergenic
255664849	G/A	771.16	Intergenic
255664877	C/T	889.07	Intergenic
255664998	T/C	908.78	Intergenic
255665014	C/T	996.79	Intergenic
255665199	A/G	1045.93	Intergenic
255665537	C/T	1022.44	Intergenic
255665753	G/T	1570.61	Intergenic
255665771	C/T	1580.55	Intergenic
255665797	A/T	1047.72	Intergenic
255665846	G/T	1061.42	Intergenic
255666039	C/G	1188.79	Intergenic
255666209	T/A	937.87	Intergenic
255666388	A/G	774.94	Intergenic
255666417	C/A	873.59	Intergenic

255666483	T/G	553.45	Intergenic
255666579	T/G	996.96	Intergenic
255666773	C/T	892.27	Intergenic
255666828	A/G	627.67	Intergenic
255666895	C/A	1147.62	Intergenic
255666919	C/T	1026.75	Intergenic
255666962	C/G	776.72	Intergenic
255667506	A/C	1349.29	Intergenic
255667516	G/A	1241	Intergenic
255667609	T/C	984.62	Intergenic
255667671	A/G	1489.2	Intergenic
255667683	C/T	1394.96	Intergenic
255667728	TC/T	675.22	Intergenic
255668500	C/A	732.49	Intergenic
255669488	G/T	1186.35	Intergenic
255671765	A/G	581.78	Intergenic
255672974	C/T	990.66	Intergenic
255674723	G/T	569.21	Intergenic
255676215	C/A	707.01	Intergenic
255677716	G/A	699.73	Intergenic
255682308	C/T	1380.7	Intergenic
255682715	C/T	1069.48	Intergenic
255724888	A/AG	764.22	Intergenic
255735791	C/A	956.95	Intergenic
255773831	A/G	607.09	Intergenic
255790802	A/T	622.32	Intergenic
255833231	A/T	526.77	Intergenic
255838634	C/T	992.87	Intergenic
255842991	A/G	588.13	Intergenic
255843113	G/T	934.77	Intergenic
255849460	ATCTT/A	1059.1	Intergenic
255849517	A/C	1182.63	Intergenic
255849526	G/A	1257.7	Intergenic
255849927	A/G	996.45	Intergenic
255849944	A/G	883.22	Intergenic
255849967	G/A	1041.6	Intergenic
255850242	A/G	690.14	Intergenic
255850863	G/T	992.34	Intergenic
255861296	C/T	999.92	Intergenic
255867068	A/AT	698.86	Intergenic
255905551	C/G	949.33	Intergenic
255908866	G/A	957.2	Intergenic
255956889	A/T	1037.84	intergenic
256536066	A/AC	1025.7	intronic, <i>MBNLI</i>
256561382	GTTC/G	1088.43	intronic, <i>MBNLI</i>
256581070	G/GA	653.79	intronic, <i>MBNLI</i>
256611589	C/G	790.78	intronic, <i>MBNLI</i>
256620750	G/A	681.68	intronic, <i>MBNLI</i>

256656528	A/AGT	817.74	Intergenic
256694804	T/C	630.05	Intergenic
256723100	C/T	613.51	Intergenic
256765697	C/T	777.35	Intergenic
256904420	T/TTA	1011.39	Intergenic
257080648	T/C	802.51	Intergenic
257111381	G/A	1022.06	Intergenic, upstream of <i>SUCNR1</i>
257158836	G/A	445.97	intronic, <i>LOC114112796</i>

Animal	MTRDHH2 status	SLC33A1 genotype	MTRDHH2 haplotype (66 markers)	
			+	-
Animal 1	+/-	WT/dupG	A	SNP1
Animal 2	+/-	WT/dupG	A	SNP2
Animal 3	+/-	WT/dupG	G	SNP3
Animal 4	+/-	WT/dupG	G	SNP4
Animal 5	+/-	WT/dupG	G	SNP5
Animal 6	+/-	WT/dupG	C	SNP6
Animal 7	+/-	WT/dupG	G	SNP7
Animal 8	+/-	WT/dupG	A	SNP8
Animal 9	+/-	WT/dupG	A	SNP9
Animal 10	+/-	WT/dupG	A	SNP10
Animal 11	+/-	WT/dupG	G	SNP11
Animal 12	+/-	WT/dupG	A	SNP12
Animal 13	+/-	WT/dupG	A	SNP13
Animal 14	+/-	WT/dupG	A	SNP14
			A	SNP15
			A	SNP16
			A	SNP17
			A	SNP18
			A	SNP19
			A	SNP20
			C	SNP21
			G	SNP22
			C	SNP23
			G	SNP24
			G	SNP25
			G	SNP26
			A	SNP27
			A	SNP28
			A	SNP29
			A	SNP30
			G	SNP31
			G	SNP32
			A	SNP33
			A	SNP34
			A	SNP35
			C	SNP36
			A	SNP37
			G	SNP38
			G	SNP39
			G	SNP40
			G	SNP41
			A	SNP42
			A	SNP43
			A	SNP44
			A	SNP45
			A	SNP46
			A	SNP47
			G	SNP48
			G	SNP49
			T	SNP50
			A	SNP51
			A	SNP52
			A	SNP53
			C	SNP54
			G	SNP55
			G	SNP56
			A	SNP57
			C	SNP58
			G	SNP59
			G	SNP60
			G	SNP61
			C	SNP62
			G	SNP63
			G	SNP64
			A	SNP65
			A	SNP66

Figure S1. MTRDHH2 recombinant haplotypes from 14 animals showing mismatch between MTRDHH2 status and SLC33A1_dupG genotype. MTRDHH2+/+ and +/- refer to heterozygous and non-carriers of MTRDHH2, respectively. The grey column represents the localization of SLC33A1_dupG (g.252,649,023dupG) within the MTRDHH2 haplotype. For each animal, only the phase supposed to host the SLC33A1_dupG is represented. The blue color indicates the portion of local haplotype matching with the MTRDHH2 haplotype.

21. Communication lors de congrès

Résumé et communication orale : [Maxime Ben Braiek](#), Stéphane Fabre, Chris Hozé, Jean-Michel Astruc, and Carole Moreno-Romieux. Identification of homozygous haplotype compromising fertility traits in dairy sheep. In *Annual Meeting of the European Federation of Animal Science (EAAP)*, Virtual meeting, 1-4 December, 2019.

Résumé et communication orale : [Maxime Ben Braiek](#), Carole Moreno-Romieux, Charlotte Allain, Philippe Bardou, Arnaud Bordes, Frederic Debat, Florence Plisson-Petit, David Portes, Julien Sarry, Némuel Tadi, Florent Woloszyn, and Stéphane Fabre. A recessive stop-gain mutation in *CCDC65* is associated with lamb mortality in French Lacaune dairy sheep. In *International Sheep Genomics and the International Goat Genome Consortiums combined virtual meeting*, Virtual meeting, 8-10 June, 2021.

Article et communication orale : [Maxime Ben Braiek](#), Carole Moreno-Romieux, Chris Hozé, Jean-Michel Astruc, and Stéphane Fabre. Characterization of homozygous haplotype deficiency compromising fertility traits in Manech Tête Rousse dairy sheep. In *World Congress on Genetics Applied to Livestock Production (WCGALP)*, Rotterdam, The Netherlands, 3-8 July, 2022. doi : 10.3920/978-90-8686-940-4_745.



Identification of homozygous haplotype compromising fertility traits in dairy sheep

Maxime Ben Braiek, Stéphane Fabre, Chris Hozé, Jean-Michel Astruc, Carole Moreno

► To cite this version:

Maxime Ben Braiek, Stéphane Fabre, Chris Hozé, Jean-Michel Astruc, Carole Moreno. Identification of homozygous haplotype compromising fertility traits in dairy sheep. Annual Meeting of the European Federation of Animal Science (EAAP), Dec 2020, On line, France. hal-03627607

HAL Id: hal-03627607

<https://hal.inrae.fr/hal-03627607>

Submitted on 1 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Annual Meeting of the European Federation of Animal Science



Abstract n° 34261 - Theatre presentation

Session 61. SMARTER: small ruminants breeding for efficiency and resilience



Date: Friday 4 December 2020; 9.00 – 11.45
Chair: Conington J. / Moreno C.

Identification of homozygous haplotype compromising fertility traits in dairy sheep

M. Ben Braiek¹, S. Fabre¹, C. Hozé², J.M. Astruc³ and C. Moreno-Romieux¹

¹INRAE, 24 chemin de Borde-Rouge, 31326 Castanet-Tolosan, France, ²Allice, 149 rue de Bercy, 75595 Paris, France, ³Institut de l'Élevage, 24 chemin de Borde-Rouge, 31326 Castanet-Tolosan, France; maxime.ben-braiek@inrae.fr

In the frame of H2020 project SMARTER no. 772787 and HOMLET (APIS-GENE, Région Occitanie) project

In livestock population under selection, the small effective size and inbreeding can cause the emergence of recessive deleterious mutations. When homozygous, these mutations may be responsible for embryo or neonate lethality, or genetic defects, affecting female fertility and animal welfare. To detect such mutations, a reverse genetic screen was applied on phased 50k SNP genotypes and pedigree data to identify Homozygous Haplotype Deficiency (HHD) in Lacaune (LAC, n=19,102) and Manech Tête Rousse (MTR, n=4,900) dairy populations. We detected 10 significant HHD, 5 presenting a complete deficit of homozygous animals (3 in LAC, 2 in MTR), and 5, a partial deficit of 79 to 96% compared to the expected (4 in LAC, 1 in MTR). These haplotypes spanned regions from 1.2 to 3.0 Mb with a frequency of heterozygous carriers between 4.4 to 17.4%. Then, we defined risky mating as mating between heterozygous rams at a HHD and females coming from heterozygous rams at the same HHD. Using logistic binary models, we tested the effect of risky mating for each HHD on two fertility traits: conception at AI and stillbirth rates. HHD in complete deficit were mainly associated with a decrease in conception rate in LAC breed and an increase of stillbirth rate in MTR breed. Only 2 HHD in partial deficit in the LAC breed were associated with an increase of stillbirth rate. Finally, we tested a putative selective advantage of heterozygous rams at the 10 HHD. The daughter yield deviation of 4 main traits (milk, fat and protein yields, somatic cell score) selected in dairy sheep in France were tested by variance analyses comparing carrier and non-carrier rams. Among the 10 HHD, 4 detected in LAC breed had significant positive impact on at least one of the selected traits. Thereafter, using available and newly generated sequence data of HHD carriers and non-carriers, we will try to identify putative causal mutations. The further management of these mutations in the LAC and MTR selection schemes will allow to improve the overall fertility and lamb viability.



A recessive stop-gain mutation in CCDC65 is associated with lamb mortality in French Lacaune dairy sheep

Maxime Ben Braiek, Carole Moreno-Romieu, Charlotte Allain, Philippe Bardou, Arnaud Bordes, Frederic Debat, Florence Plisson-Petit, David Portes, Julien Sarry, Némuel Tadi, et al.

► **To cite this version:**

Maxime Ben Braiek, Carole Moreno-Romieu, Charlotte Allain, Philippe Bardou, Arnaud Bordes, et al.. A recessive stop-gain mutation in CCDC65 is associated with lamb mortality in French Lacaune dairy sheep. International Sheep Genomics and the International Goat Genome Consortiums combined virtual meeting, Jun 2021, On line, France. hal-03360133

HAL Id: hal-03360133

<https://hal.inrae.fr/hal-03360133>

Submitted on 30 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

International Sheep Genomics and the International Goat Genome Consortiums combined virtual meeting

June 8th/9th to June 10th/11th 2021



Presenter abstracts

13. A recessive stop-gain mutation in *CCDC65* is associated with lamb mortality in French Lacaune dairy sheep

Maxime Ben Braiek¹, Carole Moreno-Romieux¹, Charlotte Allain², Philippe Bardou³, Arnaud Bordes¹, Frédéric Debat¹, Florence Plisson-Petit¹, David Portes², Julien Sarry¹, Némuel Tadi¹, Florent Woloszyn¹ and Stéphane Fabre¹

¹GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, France

²UE Domaine de La Fage, INRAE, 12250 Saint Jean et Saint Paul, France

³Sigenae, INRAE, 31326 Castanet-Tolosan, France

maxime.ben-braiek@inrae.fr

Funded by SMARTER (EU H2020 n°772787), PRESAGE (CASDAR n°20ART1532777) and HOMLET (APIS-GENE, Région Occitanie) projects

In a previous work using 50k SNP data from Lacaune dairy sheep genotypes, we have identified 11 haplotypes with homozygous deficiency (LDHH) supposed to host recessive lethal mutations. Among them, LDHH6 located on OAR3 was the most frequent haplotype (12.1% of heterozygous carriers) and it was associated with an increase in the stillbirth rate.

In the present study, we have analyzed the whole-genome sequences of two Lacaune rams heterozygous carriers of LDHH6, and we have compared them to the sequences of 23 non-carriers Lacaune rams and 68 other animals from 14 different sheep breeds. After variant filtering, we have identified a SNP in the two LDHH6 carriers whose variant allele induced a premature stop-gain mutation in the Coiled-Coil Domain Containing 65 (*CCDC65*) gene. *CCDC65* is involved in the assembly of the nexin-dynein complex for the formation of microtubules in ciliated cells and mutations in *CCDC65* are associated with “*Ciliary dyskinesia, primary 27*” causing respiratory distress in human.

In order to identify the affected phenotype in sheep, we generated 17 at-risk matings between carrier rams and carrier ewes genotyped for the candidate mutation in *CCDC65*. A total of 16 lambs were born alive, and among them, five were identified as homozygous carriers and died between the age of one week and one month, all suffering from respiratory problems as tachypnea. At the autopsy, we particularly observed a broad hepatization of lungs involving infectious pneumonia.

Management of this causal mutation in the Lacaune dairy sheep selection scheme through reasoned mating of carrier rams and ewes could improve overall lamb viability by 2%.

Characterization of homozygous haplotype deficiency compromising fertility traits in Manech Tête Rousse dairy sheep

M. Ben Braiek^{1*}, C. Moreno-Romieux¹, C. Hozé², J-M. Astruc³ and S. Fabre¹

¹GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet-Tolosan, France; ²Allice, 75595 Paris, 149 rue de Bercy, France; ³Institut de l'Élevage, 24 chemin de Borde-Rouge, F-31321, Castanet-Tolosan, France; *maxime.ben-braiek@inrae.fr

Abstract

With the objective to identify deleterious recessive mutations, we have scanned the genome of 5,271 French Manech Tête Rousse (MTR) dairy sheep using 50k SNP phased genotypes and pedigree data. Five deficient homozygous haplotypes, named MTRDHH1 to MTRDHH5, were identified. These haplotypes showed a highly significant deficit of homozygous animals ranging from 84 to 100%, and a carrier frequency ranging from 7.8 to 16.6%. By comparing at-risk matings (between carriers) and safe matings, some of these haplotypes were associated with a reduced success of artificial insemination (MTRDHH2), and/or an increased stillbirth rate (MTRDHH1 and 2) suggesting the segregation of recessive variants leading to embryo and/or perinatal lethality affecting fertility. Other haplotypes are likely to be associated with lamb morphological disorders or counter-selection based on breed standard. For each haplotype, the most probable candidate genes were highlighted based on their roles in lethal phenotypes, genetic disorders, hornless or coat color.

Introduction

In livestock under selection, the small effective population size and a non-negligible inbreeding rate could increase homozygosity and thereafter lead to the emergence of genetic disorders under the control of recessive deleterious variants. When homozygous, these variants very often lead to early embryo death, or to developmental disorders that affect fetuses and subsequently young individuals altering health and welfare (Georges et al. 2019). Detection of homozygous haplotype deficiency using reverse genetic screen is a useful method for mapping specific haplotypes associated with deleterious recessive mutations (VanRaden et al. 2011). The analysis is based on high throughput genotyping data usually available in the framework of genomic selection. For the first time in sheep, we recently validated this approach in Lacaune dairy sheep by the identification of eight independent deficient homozygous haplotypes (Ben Braiek et al. 2021). Complementary whole genome sequence studies have led to identify the causal variant in the *CCDC65* gene within one of these haplotypes (LDHH6, OMIA 002342-9940). Homozygous lambs for the *CCDC65* mutation suffered from a lethal respiratory syndrome (Ben Braiek et al. 2022). The objective of the present study is to search for homozygous haplotype deficiency in the Manech Tête Rousse (MTR) dairy sheep also having a large set of genotyping data from genomic selection. This opens the opportunity to discover new deleterious recessive mutations in sheep.

Materials & Methods

Animal and genotyping data. The total dataset consists of 6,845 genotyped Manech Tête Rousse animals (82% male and 18% female) born between 1993 and 2021. Genotypes were obtained in the framework of the French dairy sheep genomic selection from Ovine SNP50 BeadChip (MD; n=3,889) and SheepLD (LD; n=2,956) (Illumina Inc., San Diego, USA). Quality control for each SNP was based on (i) a call frequency >97%, (ii) a minor allele frequency >1% and (iii) the respect of the Hardy-Weinberg equilibrium ($P > 10^{-5}$). Phasing and

imputation (from LD to MD SNP chip) of genotypes were done using *FImpute* v2.2. After quality control, the remaining 38,523 autosomal SNPs were mapped onto the *Ovis aries* genome assembly Oar v3.1.

Detection of homozygous haplotype deficiency. The genome of 5,271 genotyped animals belonging to trios (77 offspring with both parents genotyped and 4,799 offspring with sire and maternal grandsire genotyped) was scanned by a 20 SNP sliding window as described (Ben Braiek et al. 2021). Consecutive 20 SNP windows with the same parameters (haplotype frequency, significant homozygous animal deficit) were clustered to define larger region called “Manech Tête Rousse Deficient Homozygous Haplotype” (MTRDHH). For each MTRDHH, the genotypical status (homozygous non-carriers, heterozygous and homozygous carriers) was determined for all genotypes available (n=6,845).

Analysis of fertility traits. Trait records of MTR matings between 2006 and 2019 were obtained from the national database and based on ewes with a genotyped sire and genotyped rams (maternal sire and ram both having a known status at each MTRDHH; n=330,844 matings). We analyzed two fertility-associated traits, artificial insemination success (AIS) and stillbirth rate (SBR). We have considered “at-risk” a mating between a carrier ram and a ewe from a carrier sire and the other combinations were “safe” mating. AIS and SBR were considered to differ significantly when the fixed effect “mating type” had a P-value lower than 1.0×10^{-3} after correcting for multiple testing with a level of significance at 1% following the SAS GLIMMIX procedure previously described (Ben Braiek et al. 2021).

Identification of candidate genes in MTRDHH regions. We extracted all annotated genes located in the MTRDHH region extended by 1 Mb from each side from the ovine genome Oar_v3.1 using CLC export annotation function (QIAGEN CLC Main Workbench 7.9). All genes related to mortality/aging phenotypes when knocked-out in mice and/or associated to mammalian genetic disorders were extracted from Mouse Genome Informatics (MGI, <http://www.informatics.jax.org>), Online Mendelian Inheritance in Man (OMIM, <https://omim.org>) and Online Mendelian Inheritance in Animal (OMIA, <https://omia.org>) databases.

Table 1. List of Manech Tête Rousse deficient homozygous haplotypes.

Haplotype	OAR	Number of markers	aPosition (Mb)	bCarrier frequency (%)	Number of homozygotes			
					cExp	dObs	Deficit	Poisson P-value
MTRDHH1	20	32	21.4-23.2	9.7%	13	0	100%	2.9×10^{-6}
MTRDHH2	1	66	229.2-233.4	8.7%	10	0	100%	3.8×10^{-5}
MTRDHH3	1	39	97.2-99.8	7.8%	9	0	100%	9.6×10^{-5}
MTRDHH4	10	26	28.2-29.9	8.7%	11	1	91%	1.5×10^{-4}
MTRDHH5	13	29	61.3-64.0	16.6%	49	8	84%	5.3×10^{-13}

aPosition on ovine genome assembly Oar_v3.1; bwhole genotyped population (n=6,845); cExpected; dObserved.

Results

Identification of deficient homozygous haplotypes in MTR dairy sheep. Using a reverse genetic screen in MTR genotyped population, we identified 5 independent regions showing a significant deficit of homozygous animals and named MTRDHH (Table 1). Among those regions, three haplotypes (MTRDHH1, 2 and 3) showed a complete deficit of observed homozygous animal while 9 to 13 were expected. The two other haplotypes (MTRDHH4 and 5) presented a partial deficit between 84 and 91% (some homozygous animals were genotyped). The length of these haplotypes ranged from 1.8 to 4.2 Mb and carrier frequencies ranged from 7.8% to 16.6% in the whole genotyped population. MTRDHH2 and 3 are located on the same

chromosome (OAR1) but are not in linkage disequilibrium. The other haplotypes are located on different chromosomes, MTRHH1 on OAR20, MTRDHH4 on OAR10 and MTRDHH5 on OAR13. Consequently, the five MTRDHH detected are likely to host five recessive deleterious variants.

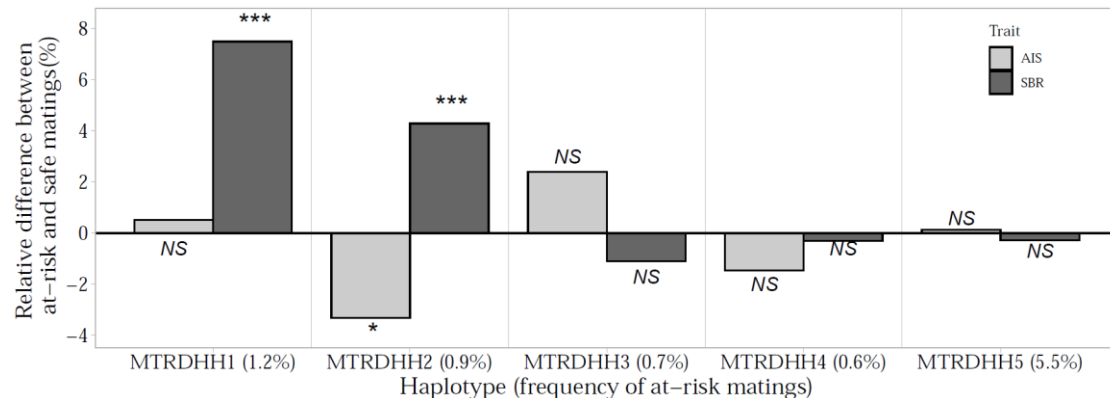


Figure 1. Effects of MTRDHH on the success rate of artificial insemination and on the stillbirth rate in at-risk matings compared to safe matings. AIS: artificial insemination success, SBR: stillbirth rate. Significant effects are indicated by the corrected P-value for multiple tests with a threshold set at $\alpha=1\%$: *: $P < 1.0 \times 10^{-3}$; **: $P < 1.0 \times 10^{-4}$ and ***: $P < 1.0 \times 10^{-5}$. NS: not significant.

Impact of MTRDHH on the success rate of AI and on stillbirth rate. In order to identify the putative lethal effect of the MTRDHH, we studied two fertility binary traits (AIS and SBR) by comparing at-risk matings to safe matings. AIS (population mean = 60.9%) is a good proxy for embryonic loss, while SBR (population mean = 7.5%) is a good proxy for perinatal lethality. As shown in Figure 1, MTRDHH1 has a significant negative impact on SBR (+7.5% in at-risk matings) and MTRDHH2 have negative effects on both AIS (-3.3%) and SBR (+4.3%). The other haplotypes showed no significant effect on the two tested traits.

Table 2. Candidate genes located in MTRDHH regions

Haplotype (phenotype ^a)	Gene name/Phenotype disorder in mouse and human
MTRDHH1 (Neonatal lethality)	MMUT : neonatal, postnatal lethality [MGI:97239]/ methylmalonic aciduria [OMIM# 609058]; TFAP2B : neonatal, postnatal, preweaning lethality [MGI:104672]/ patent ductus arteriosus, Char syndrome [OMIM# 601601]; PKHD1 : perinatal, postnatal, preweaning lethality [MGI:2155808]/ polycystic kidney/liver disease [OMIM# 606702]
MTRDHH2 (Embryo/fetus lethality)	DHX36 : embryonic lethality [MGI:1919412]; SSR3 : neonatal lethality [MGI:1914687]
MTRDHH3 (Juvenile lethality)	SV2A : premature death [MGI:1927139]; ECM1 : preweaning lethality, premature death [MGI:103060]; skin/mucosa/viscera lipid proteinosis [OMIM# 602201]
MTRDHH4 (Juvenile lethality, morphological disorder)	KL : premature death [MGI:1101771]/ tumoral calcinosis [OMIM# 604824]; BRCA2 : postnatal, prenatal lethality, premature death [MGI:109337]/ Wilms tumor, Fanconi anemia D1 [OMIM# 600185]; RXFP2 : Polled/Horns phenotype in sheep [OMIA 000483-9940]
MTRDHH5 (Morphological disorder)	ASIP : Coat color in sheep [OMIA 000201-9940]; skin/hair/eye pigmentation [OMIM#600201]

^a Supposed deleterious phenotype in sheep

Candidate genes for MTRHH regions. Among the 459 genes annotated in the five MTRDHH regions, 61 are implicated in lethal phenotype when knocked-out in mice. Among them, several are associated also with mammalian genetic disorders (Table 2). The remaining candidate genes are involved in important functions such as metabolism (*MMUT* and *SSR3*), DNA/RNA processing or transcription (*TFAP2B*, *DHX36* and *BRCA2*), or cell signaling (*PKHD1*, *SV2A*, *ECM1*, *KL* and *RXFP2*).

Discussion

In this study, we have confirmed the usefulness of reverse genetic screen to detect homozygous haplotype deficiency in sheep, even with a more reduced dataset of genotyped trios (MTR ~ 5,000) compared to our previous work (Lacaune ~15,000, (Ben Braiek et al. 2021)). However, using the same parameters we identified less significant haplotypes (5 in MTR vs. 11 in Lacaune) with the lower haplotype carrier frequency at 7.8% compared to 3.7% in Lacaune.

Among the five MTRDHH, three of them are totally deficient in homozygous animals leading to the hypothesis that these haplotypes host recessive lethal mutations. Accordingly, MTRDHH1 and 2 are associated with defect in fertility traits, AIS and/or SBR. This lets us think that when homozygous these haplotypes induce embryonic or fetal/neonatal lethality. To support this hypothesis, we found obvious candidate genes in these regions (*MMUT*, *TFAP2B*, *PKHB1*, *DHX36* and *SSR3*) all associated with embryonic, perinatal or neonatal lethality when altered in mouse or in human (Table 2). In contrast, MTRDHH3 showing also a complete deficit in homozygous animals was not associated with alteration of AIS or SBR. Since lambs are genotyped around three months of age, this haplotype could affect lambs after birth in the preweaning period as supposed by premature death caused by the alteration of the two candidate genes in this region (*SV2A* and *ECM1*, Table 2).

MTRDHH4 and 5 showed a partial deficit in homozygous animals and no significant effect on fertility traits. We hypothesize that these deficits are likely due to either a defect after birth in the juvenile period (alteration of *KL* or *BRCA2* genes for MTRDHH4) or counter-selection based on breed standard. Indeed, horned female and black animals are unwanted in MTR selection scheme. Interestingly, MTRDHH4 hosts the *RXFP2* gene known to be implicated in horn/polled phenotype (Wiedemar and Drögemüller 2015), and *ASIP*, known to be involved in coat color determinism, locates in MTRDHH5 (Norris and Whan 2008).

Acknowledgments: This research has received funding from the European Union's Horizon 2020 research and innovation program under the Grant Agreement No. 772787 (SMARTER), PRESAGE project (CASDAR No. 20ART1532777) and APIS-GENE and Région Occitanie.

References

- Ben Braiek M, Fabre S, Hozé C, Astruc J-M and Moreno-Romieux C. (2021) *Genet Sel Evol* 53(1):41. <https://doi.org/10.1186/s12711-021-00634-1>
- Ben Braiek M, Moreno-Romieux C, Allain C, Bardou P, Bordes A, et al. (2022) *Genes* 13(1):45. <https://doi.org/10.3390/genes13010045>
- Georges M, Charlier C and Hayes B. (2019) *Nat Rev Genet* 20(3):135-56. <https://doi.org/10.1038/s41576-018-0082-2>
- Norris BJ and Whan VA. (2008) *Genome Research* 18(8):1282-93. <https://doi.org/10.1101/gr.072090.107>
- VanRaden PM, Olson KM, Null DJ and Hutchison JL. (2011) *Journal of Dairy Science* 94(12):6153-61. <https://doi.org/10.3168/jds.2011-4624>
- Wiedemar N and Drögemüller C. (2015) *Animal Genetics* 46(4):457-61. <https://doi.org/10.1111/age.12309>

VIII

Annexes

Breed	Number of animals	ENA run accession	EBI project accession
Belclare	2	SRR14934360; SRR14935129	PRJNA698548
Berrichon du Cher	3	ERS1205899; ERS1205900; ERS1205901	PRJEB14418
Cambridge	7	ERR1419201; ERR1419202; ERR1419203; ERR1419204; ERR1419205; ERR1419206; ERR1419207	PRJEB14098
Charollais	1	SRR14934359	PRJNA698548
Lacaune (Dairy)	31	ERR3276357; ERR3276358; ERR3276359; ERR3276360; ERR3276361; ERR3276362; ERR3276363; ERR3276364; ERR3276365; ERR3276366; ERR3276367; ERR3276368; ERR3276369; ERR3276370; ERR3276371; ERR3276372; ERR3276373; ERR3276374; ERR3276375; ERR3276376; ERR3276377; ERR3276378; ERR3276379; ERR7891349; ERR7891350; ERR7891351 ERR968423; ERR968424; ERR968425 SRR501850; SRR501851	PRJEB32110 PRJEB9911 PRJNA160933 PRJEB32110
Lacaune (Meat)	3	ERR3276380; ERR3276381; ERR3276382	PRJEB32110
Manech Tête Rousse	22	ERR3712282; ERR3712283; ERR3712284; ERR3712285; ERR3712286; ERR3712287; ERR3712288; ERR3712289; ERR3712290; ERR3712291; ERR3712292; ERR3712293; ERR3712294; ERR3712295; ERR3712296; ERR3712297; ERR3712298; ERR3712299; ERR3712300; ERR3712301; ERR7889920; ERR7889921	PRJEB35682
Martinik Blackbelly	1	ERR3255914	PRJEB31930
Noire du velay	2	ERR3828659; ERR3828660	PRJEB35553
Romane	4	ERS1205902; ERS1205903; ERR2818429; ERR2818430	PRJEB14418
Romane x Martinik Blackbelly	13	ERR3255915; ERR3255916; ERR3255917; ERR3255918; ERR3255919; ERR3988554; ERR3988555; ERR3988556; ERR3988557; ERR3988558; ERR3988559; ERR3988560; ERR3988561	PRJEB31930
Romanov	2	ERS1205904; ERS1205905	PRJEB14418
Suffolk	2	SRR14934357; SRR14934358	PRJNA698548
Texel	2	SRR14934355; SRR14934356	PRJNA698548
Vendéen	5	ERR4236129; ERR4236130; ERR4236131 SRR14934353; SRR14934354	PRJEB37460 PRJNA698548
Total	100		

FIGURE 21.1 – Séquences ovines disponibles WGS

TABLE 21.1 – Liste des amorces utilisées pour géotyper les différentes mutations. F=Forward; R= Reverse; PACE= PCR Allele Compe-
titive Extension. Dans le cas des génotypes PACE, F1 correspond à l'amorce spécifique de l'allèle sauvage (ajout amorce universelle fluorescente :
GAAGGTGACCAAGTTCATGCT) et F2 correspond à l'amorce de l'allèle muté (ajout amorce universelle fluorescente : GAAGGTGGGAGTCAACGGATT).

Région/ Ha- plotype	Gène	Mutation	Méthode génoty- page	Amorces (Séquence 5' → 3')	Taille (pb)	Enzyme (RFLP)	Interprétation
LDHH1	<i>ORC5</i> (NC_040255.1)	OAR4 g.51,214,530dupT (c.839dupA)	PCR (amplifica- tion)	F :CACTGGGTGATTTGTAGCA R :TGCTCTTTAGTGTGAACTTTGG	310	-	-
			PACE	F1 :GAAGGTGACCAAGTTCATGCTATCTGTGTCATCTTTCTGTAGCTTTTC F2 :GAAGGTGGAGTCAACGGAGTGAATGATCTGTGTCATCTTTCTGTAGCTTTT R :GATTTTTTTTTTCCAGTTCGCAATGGAA		-	-
LDHH2	<i>ID11/LOC101108851</i> (NC_040264.1)	OAR13 g.47,939,558_ 47,939,560delAAG (c.287_289delAAG)	PCR+ RFLP	F :AGATGTGCATCCTCATCGAC R :ACTGAAAGCTCGGTGCAA	250	MboII	+ / +=65 / 179 ; m / +=65 / 179 / 241 ; m / m =241

PACE
 F1 :GAAGGTGACCAAGTTCATGCTCGTTCAGGTGGCAGTTCCTTC
 F2 :GAAGGTGGAGTCAACGGATTCTCGTTCAGGTGGCAGTTCCTTC
 R :TCGACGAGAACGACCCGGAGGAT

+/+ = 53/178 ;
 m/+ = 53/178/231 ;
 m/m = 231

231 BpmI

F :TGAATCGGGCACTTTTCTG
 R :CCAGTTTGGAAAGATTCCTT

PCR+
 RFLP

OAR3
 g.35,745,589G>A
 (c.997C>T)

PREB
 (NC_040254.1)

LDHH3

+/+ = 116/288 ;
 m/+ = 116/288/404 ;
 m/m = 404

BsaJI

404

F :GAGCTGCGTGTGTAAGATGA
 R :CCTCCAGCTCCATGTTGTAA

PCR+
 RFLP

OAR3
 g.147,207,999C>A
 (c.331G>T)

GDC65
 (NC_040254.1)

LDHH6

PACE
 F1 :GAAGGTGACCAAGTTCATGCTGGACCTGTCAGAAAGCCGAGG
 F2 :GAAGGTGGAGTCAACGGATTAGGACCTGTCAGAAAGCCGAGT
 R :AGGGGTGGGGTGTGCTGCT

PACE

LDHH9	<i>EDC3</i> (NC_040269.1)	OAR18 g.31,922,931T>G (c.291T>G)	PCR+ RFLP	F : CTCACGGAAATCTCACATC R : GCAAGAGAATATCGGCAATG	272	AvaII	+ / + = 272 ; m / + = 55 / 217 / 272 ; m / m = 55 / 217
MTRDHH1	<i>MMUT</i> (NC_040271.1)	OAR20 g.23,776,347G>A (c.1225C>T)	PCR (amplification)	F : GCAGATTTGGCAAAAGAAAG R : ACCTTCAAGGCAGCATCATA	358	-	-
MTRDHH2	<i>SLC33A1</i> (NC_040252.1)	OAR1 g.252,649,023dupG (c.735dupG)	PCR (amplification)	F1 : CAAGGTGACCAAGTTCATGCTAATCCCAGATTCCTTCTTGAATAATGATTTG F2 : GAAGGTGGAGTCAACGGATGAATCCCAGATTCCTTCTTGAATAATGATTTA R : GTGAAAAGTGCTCGAATTCGAGGAA	402	-	-
MTRDHH2	<i>SLC33A1</i> (NC_040252.1)	OAR1 g.252,649,023dupG (c.735dupG)	PCR (amplification)	F1 : GAAGGTGACCAAGTTCATGCTGGTTGAGGCTCAAACCGCA F2 : GAAGGTGGAGTCAACGGATGGTTGAGGCTCAAACCGCC R : CTCGAGTCTGGCGACTTTTGTAAACA	402	-	-

MTRDHH3	<i>FCGR1A</i> (NC_040252.1)	OAR1 g.105,662,733T>A (c.973T>A)	PCR+ RFLP	F :GTCTCCCTCAGGCACAGATA R :GTGTCCTTTCTCGCAGTTGT	368	MboII	+/-=272 ; m/+ =55/217 /272 ; m/m=272
				F1 :GAAGGTGACCAAGTTCATGCTAGGAACTACAAAAGAAAATGTGGAATT F2 :GAAGGTCCGAGTCAACGGATTAGGAAGTACAAAAGAAAATGTGGAATA R :AGAGTCCAGAGAGTCCAAAATAGATTCTA			
Mammites	<i>SOCS2</i> (NC_040254.1)	OAR3 g.139,302,270C>T (c.286C>T)	Puce_LD	15k_OAR3_129722200 (C/T)	-	-	-
Cornage	<i>RXFP2</i> (NC_040261.1)	OAR10 1.8kb insertion dans 3'UTR de <i>RXFP2</i>	Puce_LD	RXFP2_insert_L1 (A/G) ; RXFP2_insert_R1 (A/T) Interprétation : As/As (RXFP2_insert_L1 [C/G] & RXFP2_insert_R1 [A/A]) ; + /As (RXFP2_insert_L1 [A/G] & RXFP2_insert_R1 [A/T]) ; + /+ (RXFP2_insert_L1 [A/A] & RXFP2_insert_R1 [T/T])	-	-	-

					PCR	F :TTGAAAAAGCAAGGAAAGTG R1 :GCACAAAATGCTACTGTGTCTG R2 :GCACCTGGGAGCATAGTAGGA	-	-	+ / +=328 ; + / As=328 / 733 ; As / As=733
Cryptorchidie	<i>RXFP2</i> (NC_040261.1)	OAR10 g.30,964,387delC (c.2050delC)	PCR+ RFLP	F :TGAATCCGTAGCTCCTATGC R :CCACACCGATGGAAGTAGAC	344/345	BsaJI	+ / +=178/167 ; + / DelG=178 /167 /345 ; DelG/DelG=344		
Toison noire	<i>ASIP</i> (NC_040264.1)	OAR13 g.66,475,132_ 66,475,136del (Del5p)	Puce_LD	oar3_ oar13_ 63047423 (D/I)	-	-	-		
Toison noire	<i>ASIP</i> (NC_040264.1)	OAR13 g.66,474,980T>A (c.376T>A)	PCR (amplification)	F :AGAGAAAAGGCTCCGATGAAG R :CCCTAGCTGAGACTTCCTGA	322	-	-		
			PAGE	F1 :GAAGGTGACCAAGTTCATGCTTCCGACGGCCTGCTCCT F2 :GAAGGTGGAGTCAACGGATTCCGACGGCCTGCTCCA R :CGCGCTCAGCAGGTGGGTT	-	-	-		

TABLE 21.2 – Liste des amorces de qPCR utilisées pour le test moléculaire de gestation.

Gène	Séquence	Taille (pb)	Efficacité
<i>MX1*</i> NM_001009753	F : CCACCACCGACAGCTCCCCT R : GCAGGTGTGGGCGTGAAGCA	167	2.35
<i>STAT1</i> NM_001166203	F : CTGTCCTTCTTCCTGAAC R : TTCCTTACAGAACCTTGTC	186	1.95
<i>GAPDH</i> NM_001190390	F : CGACTTCAACAGCGACACTC R : TGCTGTAGCCGAATTCATTG	113	1.90
<i>YWHAZ</i> NM_001267887	F : ATTAAGTGAAGAGTCATACAA R : GTATCCGATGTCCACAAT	81	2.05

IX Références bibliographiques

Références

- Abdalla, E. A., Id-Lahoucine, S., Cánovas, A., Casellas, J., Schenkel, F. S., Wood, B. J., & Baes, C. F. (2020). Discovering lethal alleles across the turkey genome using a transmission ratio distortion approach. *Animal Genetics*, *51*(6), 876–889. doi: 10.1111/age.13003
- Agerholm, J. S., Basse, A., & Christensen, K. (1993, septembre). Investigations on the Occurrence of Hereditary Diseases in the Danish Cattle Population 1989-1991. *Acta Veterinaria Scandinavica*, *34*(3), 245–253. doi: 10.1186/BF03548188
- Agreste. (2020). *Filière caprins* (Rapport technique). Consulté le 2022-07-30, sur https://draaf.nouvelle-aquitaine.agriculture.gouv.fr/IMG/pdf/AgresteNA_Etudes_14_oct2020_FFcaprins_cle05471d-1.pdf
- Agreste. (2021). *Graph'agri 2021. L'agriculture, la forêt, la pêche et les industries agroalimentaires* (Rapport technique). Consulté le 2022-07-27, sur https://agreste.agriculture.gouv.fr/agreste-web/download/publication/publie/GraFra2021Integral/GraFra2021_integral.pdf
- Agreste. (2022). *Ovins : une production en léger repli et des prix record en 2021 et des prix record en 2021* (Rapport technique N° 387). Consulté le 2022-07-30, sur <https://agreste.agriculture.gouv.fr/agreste-web/download/publication/publie/SynAbo22387/consyn3872022030vins.pdf>
- Ahmed, I., Buchert, R., Zhou, M., Jiao, X., Mittal, K., Sheikh, T. I., Scheller, U., Vasli, N., Rafiq, M. A., Brohi, M. Q., Mikhailov, A., Ayaz, M., Bhatti, A., Sticht, H., Nasr, T., Carter, M. T., Uebe, S., Reis, A., Ayub, M., John, P., Kiledjian, M., Vincent, J. B., & Jamra, R. A. (2015, juin). Mutations in DCPS and EDC3 in autosomal recessive intellectual disability indicate a crucial role for mRNA decapping in neurodevelopment. *Human Molecular Genetics*, *24*(11), 3172–3180. doi: 10.1093/hmg/ddv069
- Albalat, R., & Cañestro, C. (2016, juillet). Evolution by gene loss. *Nature Reviews Genetics*, *17*(7), 379–391. doi: 10.1038/nrg.2016.39
- Albarella, S., D'Anza, E., Galdiero, G., Esposito, L., De Biase, D., Paciello, O., Ciotola, F., & Peretti, V. (2019, octobre). Cytogenetic Analyses in Ewes with Congenital Abnormalities of the Genital Apparatus. *Animals*, *9*(10), 776. doi: 10.3390/ani9100776
- Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., de Villemereuil, P., Benjelloun, B., Librado, P., Biscarini, F., Colli, L., Barbato, M., Zamani, W., Alberti, A., Engelen, S., Stella, A., Joost, S., Ajmone-Marsan, P., Negrini, R., Orlando, L., Rezaei, H. R., Naderi, S., Clarke, L., Flicek, P., Wincker, P., Coissac, E., Kijas, J., Tosser-Klopp, G., Chikhi, A., Bruford, M. W., Taberlet, P., & Pompanon, F. (2018, mars). Convergent genomic signatures of domestication in sheep and goats. *Nature Communications*, *9*(1), 813. doi: 10.1038/s41467-018-03206-y
- Alkan, C., Coe, B. P., & Eichler, E. E. (2011, mai). Genome structural variation discovery and genotyping. *Nature Reviews Genetics*, *12*(5), 363–376. doi: 10.1038/nrg2958
- Allais-Bonnet, A., Hintermann, A., Deloche, M.-C., Cornette, R., Bardou, P., Naval-Sanchez, M., Pinton, A., Haruda, A., Grohs, C., Zakany, J., Bigi, D., Medugorac, I., Putelat, O., Greyvenstein, O., Hadfield, T., Jemaa, S. B., Bunevski, G., Menzi, F., Hirter, N., Paris,

- J. M., Hedges, J., Palhiere, I., Rupp, R., Lenstra, J. A., Gidney, L., Lesur, J., Schafberg, R., Stache, M., Wandhammer, M.-D., Arbogast, R.-M., Guintard, C., Blin, A., Boukadiri, A., Rivière, J., Esquerré, D., Donnadiou, C., Danchin-Burge, C., Reich, C. M., Riley, D. G., Marle-Koster, E. v., Cockett, N., Hayes, B. J., Drögemüller, C., Kijas, J., Pailhoux, E., Tosser-Klopp, G., Duboule, D., & Capitan, A. (2021, février). Analysis of Polycerate Mutants Reveals the Evolutionary Co-option of HOXD1 for Horn Patterning in Bovidae. Molecular Biology and Evolution, msab021. doi: 10.1093/molbev/msab021
- Anamthawat-Jonsson, K., Long, S. E., Basrur, P. K., & Adalsteinsson, S. (1992, mai). Reciprocal translocation (13;20)(q12;q22) in an Icelandic sheep. Research in Veterinary Science, 52(3), 367–370. doi: 10.1016/0034-5288(92)90039-5
- Andersen, J. S., Wilkinson, C. J., Mayor, T., Mortensen, P., Nigg, E. A., & Mann, M. (2003, décembre). Proteomic characterization of the human centrosome by protein correlation profiling. Nature, 426(6966), 570–574. doi: 10.1038/nature02166
- Astruc, J.-M., Baloche, G., Buisson, D., Labatut, J., Lagriffoul, G., Larroque, H., Robert-Granie, C., Legarra, A., & Barillet, F. (2016a). Genomic selection in French dairy sheep. INRAE Productions Animales, 29(1), 41–55.
- Astruc, J.-M., Baloche, G., Buisson, D., Labatut, J., Lagriffoul, G., Larroque, H., Robert-Granie, C., Legarra, A., & Barillet, F. (2016b). La sélection génomique des ovins laitiers en France. , 16.
- Astruc, J.-M., & Buisson, D. (2018a, décembre). La pression de sélection génomique. Dossier : Les ovins laitiers résolument tournés vers la sélection génomique, 3.
- Astruc, J.-M., & Buisson, D. (2018b, décembre). Le fonctionnement contractuel de la sélection génomique. Dossier : Les ovins laitiers résolument tournés vers la sélection génomique, 3.
- Astruc, J.-M., & Buisson, D. (2018c, décembre). Le schéma de sélection génomique. Dossier : Les ovins laitiers résolument tournés vers la sélection génomique, 4.
- Astruc, J.-M., & Buisson, D. (2018d, décembre). Les nouveaux caractères. Dossier : Les ovins laitiers résolument tournés vers la sélection génomique, 5.
- Astruc, J.-M., & Buisson, D. (2018e, décembre). Les populations de référence en ovins laitiers. Dossier : Les ovins laitiers résolument tournés vers la sélection génomique, 3.
- Astruc, J.-M., & Buisson, D. (2018f, décembre). L'évaluation génomique en ovins laitiers. Dossier : Les ovins laitiers résolument tournés vers la sélection génomique, 3.
- Austin-Tse, C., Halbritter, J., Zariwala, M., Gilberti, R., Gee, H., Hellman, N., Pathak, N., Liu, Y., Panizzi, J., Patel-King, R., Tritschler, D., Bower, R., O'Toole, E., Porath, J. D., Hurd, T., Chaki, M., Diaz, K., Kohl, S., Lovric, S., Hwang, D.-Y., Braun, D., Schueler, M., Airik, R., Otto, E., Leigh, M., Noone, P., Carson, J., Davis, S., Pittman, J., Ferkol, T., Atkinson, J., Olivier, K., Sagel, S., Dell, S., Rosenfeld, M., Milla, C., Loges, N., Omran, H., Porter, M., King, S., Knowles, M., Drummond, I., & Hildebrandt, F. (2013, octobre). Zebrafish Ciliopathy Screen Plus Human Mutational Analysis Identifies C21orf59 and CCDC65 Defects as Causing Primary Ciliary Dyskinesia. The American Journal of Human Genetics, 93(4), 672–686. doi: 10.1016/j.ajhg.2013.08.015
- Ayers, K., Kumar, R., Robevska, G., Bruell, S., Bell, K., Malik, M. A., Bathgate, R. A., & Sinclair,

- A. (2019, novembre). Familial bilateral cryptorchidism is caused by recessive variants in RXFP2. *Journal of Medical Genetics*, *56*(11), 727–733. doi: 10.1136/jmedgenet-2019-106203
- Banstola, A., & Reynolds, J. N. J. (2022, septembre). The Sheep as a Large Animal Model for the Investigation and Treatment of Human Disorders. *Biology*, *11*(9), 1251. doi: 10.3390/biology11091251
- Barillet, F., Lagriffoul, G., Marnet, P.-G., Larroque, H., Rupp, R., Portes, D., Bocquier, F., & Astruc, J.-M. (2016). Breeding objectives and reasoned strategy of implementation at the population level for French dairy sheep breeds. *INRAE Productions Animales*, *29*(1), 19–40.
- Barillet, F., Mariat, D., Amigues, Y., Faugeras, R., Caillat, H., Moazami-Goudarzi, K., Rupp, R., Babilliot, J. M., Lacroux, C., Lugan, S., Schelcher, F., Chartier, C., Corbière, F., Andréoletti, O., & Perrin-Chauvineau, C. . (2009). Identification of seven haplotypes of the caprine PrP gene at codons 127, 142, 154, 211, 222 and 240 in French Alpine and Saanen breeds and their association with classical scrapie. *Journal of General Virology*, *90*(3), 769–776. doi: 10.1099/vir.0.006114-0
- Barré-Sinoussi, F., & Montagnutelli, X. (2015, novembre). Animal models are essential to biological research : issues and perspectives. *Future Science OA*, *1*(4), FSO63. doi: 10.4155/fso.15.63
- Baumgartner, M. R., Hörster, F., Dionisi-Vici, C., Haliloglu, G., Karall, D., Chapman, K. A., Huemer, M., Hochuli, M., Assoun, M., Ballhausen, D., Burlina, A., Fowler, B., Grünert, S. C., Grünewald, S., Honzik, T., Merinero, B., Pérez-Cerdá, C., Scholl-Bürgi, S., Skovby, F., Wijburg, F., MacDonald, A., Martinelli, D., Sass, J. O., Valayannopoulos, V., & Chakrapani, A. (2014, décembre). Proposed guidelines for the diagnosis and management of methylmalonic and propionic acidemia. *Orphanet Journal of Rare Diseases*, *9*(1), 130. doi: 10.1186/s13023-014-0130-8
- Baylis, M. S., Wayte, D. M., & Owen, J. B. (1984, janvier). An XO/XX mosaic sheep with associated gonadal dysgenesis. *Research in Veterinary Science*, *36*(1), 125–126. doi: 10.1016/S0034-5288(18)32013-7
- Bazer, F. W. (2013, juin). Pregnancy recognition signaling mechanisms in ruminants and pigs. *Journal of Animal Science and Biotechnology*, *4*(1), 23. doi: 10.1186/2049-1891-4-23
- Bazzi, H., & Anderson, K. V. (2014, avril). Acentriolar mitosis activates a p53-dependent apoptosis pathway in the mouse embryo. *Proceedings of the National Academy of Sciences*, *111*(15), E1491–E1500. doi: 10.1073/pnas.1400568111
- Beaujean, N., Boutinaud, M., Devinoy, , Jammes, H., Le Guillou, S., Le Provost, F., Leroux, C., Mobuchon, L., Pannetier, M., Sellem, E., & Kiefer, H. (2020, septembre). L'épigénétique et la construction du phénotype chez le bovin. doi: 10.20870/productions-animales.2020.33.2.4477
- Beck, C. L., Fahlke, C., & George, A. L. (1996, octobre). Molecular basis for decreased muscle chloride conductance in the myotonic goat. *Proceedings of the National Academy of Sciences*, *93*(20), 11248–11252. doi: 10.1073/pnas.93.20.11248
- Becker, D., Otto, M., Ammann, P., Keller, I., Drögemüller, C., & Leeb, T. (2014). The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the

- TYRP1 locus on chromosome 8. *Animal Genetics*, *46*(1), 50–54. doi: 10.1111/age.12240
- Becker, D., Tetens, J., Brunner, A., Bürstel, D., Ganter, M., Kijas, J., Consortium, f. t. I. S. G., & Drögemüller, C. (2010, janvier). Microphthalmia in Texel Sheep Is Associated with a Missense Mutation in the Paired-Like Homeodomain 3 (PITX3) Gene. *PLOS ONE*, *5*(1), e8689. doi: 10.1371/journal.pone.0008689
- Beever, J. E., Smit, M. A., Meyers, S. N., Hadfield, T. S., Bottema, C., Albretsen, J., & Cockett, N. E. (2006). A single-base change in the tyrosine kinase II domain of ovine FGFR3 causes hereditary chondrodysplasia in sheep. *Animal Genetics*, *37*(1), 66–71. doi: 10.1111/j.1365-2052.2005.01398.x
- Ben Braiek, M., Fabre, S., Hozé, C., Astruc, J.-M., & Moreno-Romieux, C. (2021, mai). Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep. *Genetics Selection Evolution*, *53*(1), 41. doi: 10.1186/s12711-021-00634-1
- Ben Braiek, M., Moreno-Romieux, C., Allain, C., Bardou, P., Bordes, A., Debat, F., Drögemüller, C., Plisson-Petit, F., Portes, D., Sarry, J., Tadi, N., Woloszyn, F., & Fabre, S. (2022, janvier). A Nonsense Variant in CCDC65 Gene Causes Respiratory Failure Associated with Increased Lamb Mortality in French Lacaune Dairy Sheep. *Genes*, *13*(1), 45. doi: 10.3390/genes13010045
- Benestad, S. L., Austbø, L., Tranulis, M. A., Espenes, A., & Olsaker, I. (2012, décembre). Healthy goats naturally devoid of prion protein. *Veterinary Research*, *43*(1), 87. doi: 10.1186/1297-9716-43-87
- Beraldi, D., McRae, A. F., Gratten, J., Slate, J., Visscher, P. M., & Pemberton, J. M. (2006, juillet). Development of a Linkage Map and Mapping of Phenotypic Polymorphisms in a Free-Living Population of Soay Sheep (*Ovis aries*). *Genetics*, *173*(3), 1521–1537. doi: 10.1534/genetics.106.057141
- Berger, S. L., Kouzarides, T., Shiekhattar, R., & Shilatifard, A. (2009, avril). An operational definition of epigenetics. *Genes & Development*, *23*(7), 781–783. doi: 10.1101/gad.1787609
- Berry, D. P., O'Brien, A., O'Donovan, J., McHugh, N., Wall, E., Randles, S., McDermott, K., O'Connor, R. E., Patil, M. A., Ho, J., Kennedy, A., Byrne, N., & Purfield, D. C. (2018, janvier). Aneuploidy in dizygotic twin sheep detected using genome-wide single nucleotide polymorphism data from two commonly used commercial vendors. *Animal*, *12*(12), 2462–2469. doi: 10.1017/S1751731118000204
- Bhatia, S., & Shanker, V. (1990, mars). A case report on XO/XX/XXX mixoploidy in a goat. *The Veterinary Record*, *126*(13), 312–313.
- Bhatia, S., & Shanker, V. (1992, mars). First report of a XX/XXY fertile goat buck. doi: 10.1136/vr.130.13.271
- Billinis, C., Panagiotidis, C. H., Psychas, V., Argyroudis, S., Nicolaou, A., Leontides, S., Papadopoulos, O., & Sklaviadis, T. . (2002). Prion protein gene polymorphisms in natural goat scrapie. *Journal of General Virology*, *83*(3), 713–721. doi: 10.1099/0022-1317-83-3-713
- Bindon, B. M. (1971). Systematic Study of Preimplantation Stages of Pregnancy in the Sheep. *Australian Journal of Biological Sciences*, *24*(1), 131–148. doi: 10.1071/bi9710131

- Bodin, L., Di Pasquale, E., Fabre, S., Bontoux, M., Monget, P., Persani, L., & Mulsant, P. (2007, janvier). 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*, *148*(1), 393–400. doi: 10.1210/en.2006-0764
- Boichard, D., Grohs, C., Danchin-Burge, C., & Capitan, A. (2016, décembre). Genetic defects : definition, origin, transmission and evolution, and mode of action. *INRAE Productions Animales*, *29*(5), 297–306. doi: 10.20870/productions-animales.2016.29.5.2997
- Boichard, D., Grohs, C., Michot, P., Danchin-Burge, C., Capitan, A., Genestout, L., Barbier, S., & Fritz, S. (2016, décembre). Taking into account the genetic defects in selection : the case of cattle. *INRAE Productions Animales*, *29*(5), 351–358. doi: 10.20870/productions-animales.2016.29.5.3003
- Bond, M., Holthaus, S.-M. k., Tammen, I., Tear, G., & Russell, C. (2013, novembre). Use of model organisms for the study of neuronal ceroid lipofuscinosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1832*(11), 1842–1865. doi: 10.1016/j.bbadis.2013.01.009
- Boross, P., Lent, P. L. v., Martin-Ramirez, J., Kaa, J. v. d., Mulder, M. H. C. M., Claassens, J. W. C., Berg, W. B. v. d., Arandhara, V. L., & Verbeek, J. S. (2008, avril). Destructive Arthritis in the Absence of Both FcRI and FcRIII. *The Journal of Immunology*, *180*(7), 5083–5091. doi: 10.4049/jimmunol.180.7.5083
- Bourneuf, E., Otz, P., Pausch, H., Jagannathan, V., Michot, P., Grohs, C., Piton, G., Ammermüller, S., Deloche, M.-C., Fritz, S., Leclerc, H., Péchoux, C., Boukadiri, A., Hozé, C., Saintilan, R., Créchet, F., Mosca, M., Segelke, D., Guillaume, F., Bouet, S., Baur, A., Vasilescu, A., Genestout, L., Thomas, A., Allais-Bonnet, A., Rocha, D., Colle, M.-A., Klopp, C., Esquerré, D., Wurmser, C., Flisikowski, K., Schwarzenbacher, H., Burgstaller, J., Brüggmann, M., Dietschi, E., Rudolph, N., Freick, M., Barbey, S., Fayolle, G., Danchin-Burge, C., Schibler, L., Bed'Hom, B., Hayes, B. J., Daetwyler, H. D., Fries, R., Boichard, D., Pin, D., Drögemüller, C., & Capitan, A. (2017, septembre). Rapid Discovery of De Novo Deleterious Mutations in Cattle Enhances the Value of Livestock as Model Species. *Scientific Reports*, *7*(1), 11466. doi: 10.1038/s41598-017-11523-3
- Bower, R., Tritschler, D., Mills, K. V., Heuser, T., Nicastro, D., & Porter, M. E. (2018, janvier). DRC2/CCDC65 is a central hub for assembly of the nexin–dynein regulatory complex and other regulators of ciliary and flagellar motility. *Molecular Biology of the Cell*, *29*(2), 137–153. doi: 10.1091/mbc.E17-08-0510
- Bower, R., Tritschler, D., VanderWaal, K., Perrone, C. A., Mueller, J., Fox, L., Sale, W. S., & Porter, M. E. (2013, février). The N-DRC forms a conserved biochemical complex that maintains outer doublet alignment and limits microtubule sliding in motile axonemes. *Molecular Biology of the Cell*, *24*(8), 1134–1152. doi: 10.1091/mbc.e12-11-0801
- Breitling, R., Laubner, D., Clizbe, D., Adamski, J., & Krisans, S. K. (2003, septembre). Isopentenyl-Diphosphate Isomerases in Human and Mouse : Evolutionary Analysis of a Mammalian Gene Duplication. *Journal of Molecular Evolution*, *57*(3), 282–291. doi: 10.1007/s00239-003-2476-8
- Bresciani, C., Parma, P., De Lorenzi, L., Di Ianni, F., Bertocchi, M., Bertani, V., Cantoni, A. M.,

- & Parmigiani, E. (2015). A Clinical Case of an SRY-Positive Intersex/Hermaphrodite Holstein Cattle. doi: 10.1159/000440691
- Brochard, M., Boichard, D., Capitan, A., Fayolle, G., Fritz, S., Nicod, L., & Vacelet, J. (2018). pANO, le risque d'anomalie létale pour les produits d'accouplements : principe et utilisation en race Montbéliarde sur le zone GEN'IAtest. In In Proceedings of the 24ème Rencontres Autour des Recherches Sur les Ruminants : 5-6 December.
- Brochard, M., Palhière, I. P., Verrier, E., Moazami-Goudarzi, K., Amigues, Y., Barillet, F. F., Hom, B. B., Bibé, B. B., Bouix, J. J., Crihiu, E. E., François, D. D., & Leymarie, C. (2006, décembre). Impact d'une sélection forte pour un gène majeur sur la variabilité et les niveaux génétiques : cas de la sélection pour la résistance à la tremblante ovine. In (Vol. 13). Institut de l'Élevage. Consulté le 2022-08-17, sur <https://hal.inrae.fr/hal-02758378>
- Bruere, A. N., Marshall, R. B., & Ward, D. P. J. (1969, juin). Testicular Hypoplasia And XXY Sex Chromosome Complement In Two Rams : The Ovine Counterpart Of Klinefelter's Syndrome In Man. Reproduction, *19*(1), 103–108. doi: 10.1530/jrf.0.0190103
- Bruère, A. N., & Mills, R. A. (1971). Observations on the incidence of Robertsonian translocations and associated testicular changes in a flock of New Zealand Romney sheep. doi: 10.1159/000130146
- Buisson, D. (2021). Préparer la création d'un observatoire des anomalies génétiques en petits ruminants (PRESAGE). Consulté sur https://idele.fr/presage/?eID=cmis_download&oID=workspace%3A%2F%2FSpacesStore%2F9c264865-b1d4-48fa-8172-181f79945f1d&cHash=e9d214668d366fc45ae16a241557362b
- Buisson, D., Astruc, J.-M., & Barillet, F. (2018, mai). Overview and perspectives of the management of genetic diversity in French dairy sheep. INRAE Productions Animales, *31*(1), 1–12. doi: 10.20870/productions-animales.2018.31.1.2202
- Burguete, I., Di Berardino, D., Lioi, M., Taibi, L., & Matassino, D. (1987). Cytogenetic observations on a Robertsonian translocation in Saanen goats. doi: 10.1186/1297-9686-19-4-391
- Cabianca, G., Rota, A., Cozzi, B., & Ballarin, C. (2007). Expression of AMH in Female Fetal Intersex Gonads in the Bovine. Anatomia, Histologia, Embryologia, *36*(1), 24–26. doi: 10.1111/j.1439-0264.2006.00713.x
- Calvo, J. H., Chantepie, L., Serrano, M., Sarto, M. P., Iguacel, L. P., Jiménez, M., Alabart, J. L., Folch, J., Fabre, S., & Lahoz, B. (2020, mars). A new allele in the BMP15 gene (FecXRA) that affects prolificacy co-segregates with FecXR and FecXGR in Rasa aragonesa sheep. Theriogenology, *144*, 107–111. doi: 10.1016/j.theriogenology.2020.01.010
- Capgènes. (2018a). La race Alpine française (Rapport technique). Consulté le 2022-07-30, sur https://www.capgenes.com/wp-content/uploads/2018/04/la_race_alpine_francaise_FR.pdf
- Capgènes. (2018b). La race Saanen française (Rapport technique). Consulté le 2022-07-30, sur https://www.capgenes.com/wp-content/uploads/2018/04/la_race_saanen_francaise_FR.pdf
- Capitan, A., Michot, P., Guillaume, F., Grohs, C., Djari, A., Fritz, S., Barbey, S., Otz, P.,

- Bourneuf, E., Esquerre, D., Gallard, Y., Klopp, C. C., & Boichard, D. (2014, août). Rapid discovery of mutations responsible for sporadic dominant genetic defects in livestock using genome sequence data : enhancing the value of farm animals as model species. In 10. World Congress of Genetics Applied to Livestock Production (p. np). Vancouver, Canada. Consulté le 2022-08-22, sur <https://hal.archives-ouvertes.fr/hal-01193912>
- Cavanagh, K. T., Leipprandt, J. R., Jones, M. Z., & Friderici, K. (1995). Molecular defect of caprine N-acetylglucosamine-6-sulphatase deficiency. A single base substitution creates a stop codon in the 5'-region of the coding sequence. Journal of Inherited Metabolic Disease, *18*(1), 96–96. doi: 10.1007/BF00711390
- Chaban, C., Broussou, D., Brass, S., Pailhoux, E., Lesueur, E., Pinton, A., Burger, M., Lucas, M.-N., & Maillard, S. C. (2015). Particularité du freemartinisme chez les ovins : à propos d'un cas d'intersexualité. Point Vétérinaire, *46*(359 (Part 1)), 72.
- Chantepie, L., Bodin, L., Sarry, J., Woloszyn, F., Plisson-Petit, F., Ruesche, J., Drouilhet, L., & Fabre, S. (2020). Genome-Wide Identification of a Regulatory Mutation in BMP15 Controlling Prolificacy in Sheep. Frontiers in Genetics, *11*. doi: 10.3389/fgene.2020.00585
- Charlier, C., Agerholm, J. S., Coppieters, W., Karlskov-Mortensen, P., Li, W., Jong, G. d., Fasquelle, C., Karim, L., Cirera, S., Cambisano, N., Ahariz, N., Mullaart, E., Georges, M., & Fredholm, M. (2012, août). A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina. PLOS ONE, *7*(8), e43085. doi: 10.1371/journal.pone.0043085
- Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J. S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., Frennet, J.-C., Hanset, R., Hubin, X., Jorgensen, C., Karim, L., Kent, M., Harvey, K., Pearce, B. R., Simon, P., Tama, N., Nie, H., Vandeputte, S., Lien, S., Longeri, M., Fredholm, M., Harvey, R. J., & Georges, M. (2008, avril). Highly effective SNP-based association mapping and management of recessive defects in livestock. Nature Genetics, *40*(4), 449–454. doi: 10.1038/ng.96
- Charlier, C., Li, W., Harland, C., Littlejohn, M., Coppieters, W., Creagh, F., Davis, S., Druet, T., Faux, P., Guillaume, F., Karim, L., Keehan, M., Kadri, N. K., Tamma, N., Spelman, R., & Georges, M. (2016, octobre). NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. Genome Research, *26*(10), 1333–1341. doi: 10.1101/gr.207076.116
- Chaves, R., Adegá, F., Wienberg, J., Guedes-Pinto, H., & Heslop-Harrison, J. S. (2003, octobre). Molecular cytogenetic analysis and centromeric satellite organization of a novel 8;11 translocation in sheep : a possible intermediate in banded chromosome evolution. doi: 10.1007/s00335-003-3004-2
- Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F., Mainland, I., Kao, R. R., Pemberton, J. M., Beraldi, D., Stear, M. J., Alberti, A., Pittau, M., Iannuzzi, L., Banabazi, M. H., Kazwala, R. R., Zhang, Y.-p., Arranz, J. J., Ali, B. A., Wang, Z., Uzun, M., Dione, M. M., Olsaker, I., Holm, L.-E., Saarma, U., Ahmad, S., Marzanov, N., Eythorsdottir, E., Holland, M. J., Ajmone-Marsan, P., Bruford, M. W., Kantanen, J., Spencer, T. E., & Palmarini, M. (2009, avril). Revealing the History of Sheep Domestication Using Retrovirus Integrations.

- Science, 324(5926), 532–536. doi: 10.1126/science.1170587
- Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., & Lu, X. (2012). Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. Frontiers in Genetics, 3, 35. doi: 10.3389/fgene.2012.00035
- Clark, E. L., Archibald, A. L., Daetwyler, H. D., Groenen, M. A. M., Harrison, P. W., Houston, R. D., Kühn, C., Lien, S., Macqueen, D. J., Reecy, J. M., Robledo, D., Watson, M., Tuggle, C. K., & Giuffra, E. (2020, novembre). From FAANG to fork : application of highly annotated genomes to improve farmed animal production. Genome Biology, 21(1), 285. doi: 10.1186/s13059-020-02197-8
- Clark, E. L., Bush, S. J., McCulloch, M. E. B., Farquhar, I. L., Young, R., Lefevre, L., Pridans, C., Tsang, H. G., Wu, C., Afrasiabi, C., Watson, M., Whitelaw, C. B., Freeman, T. C., Summers, K. M., Archibald, A. L., & Hume, D. A. (2017, septembre). A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). PLOS Genetics, 13(9), e1006997. doi: 10.1371/journal.pgen.1006997
- Clayton, J. S., McNamara, E. L., Goulee, H., Conijn, S., Muthsam, K., Musk, G. C., Coote, D., Kijas, J., Testa, A. C., Taylor, R. L., O'Hara, A. J., Groth, D., Ottenheijm, C., Ravenscroft, G., Laing, N. G., & Nowak, K. J. (2020, août). Ovine congenital progressive muscular dystrophy (OCPMD) is a model of TNNT1 congenital myopathy. Acta Neuropathologica Communications, 8(1), 142. doi: 10.1186/s40478-020-01017-1
- Clement, V., Carillier-Jacquin, C., Tessier, M., Rupp, R., Tosser-Klopp, G., Martin, P., Robert-Granié, C. C., Palhiere, I., & Larroque, H. (2020, janvier). Genomic selection in French Dairy Goats.. Consulté le 2022-08-17, sur <https://hal.inrae.fr/hal-03155964>
- Clizbe, D. B., Owens, M. L., Masuda, K. R., Shackelford, J. E., & Krisans, S. K. (2007, mars). IDI2, a Second Isopentenyl Diphosphate Isomerase in Mammals*. Journal of Biological Chemistry, 282(9), 6668–6676. doi: 10.1074/jbc.M610922200
- Clop, A., Marcq, F., Takeda, H., Pirottin, D., Tordoir, X., Bibé, B., Bouix, J., Caiment, F., Elsen, J.-M., Eychenne, F., Larzul, C., Laville, E., Meish, F., Milenkovic, D., Tobin, J., Charlier, C., & Georges, M. (2006, juillet). A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nature Genetics, 38(7), 813–818. doi: 10.1038/ng1810
- Cockett, N. E., Jackson, S. P., Shay, T. L., Farnir, F., Berghmans, S., Snowden, G. D., Nielsen, D. M., & Georges, M. (1996, juillet). Polar Overdominance at the Ovine callipyge Locus. Science, 273(5272), 236–238. doi: 10.1126/science.273.5272.236
- Cole, J., Null, D., & VanRaden, P. (2016, septembre). Phenotypic and genetic effects of recessive haplotypes on yield, longevity, and fertility. Journal of Dairy Science, 99(9), 7274–7288. doi: 10.3168/jds.2015-10777
- Cole, J. B. (2015, novembre). A simple strategy for managing many recessive disorders in a dairy cattle breeding program. Genetics Selection Evolution, 47(1), 94. doi: 10.1186/s12711-015-0174-9
- Cooper, T., Wiggans, G., VanRaden, P., Hutchison, J., Cole, J., & Null, D. (2013). Genomic evaluation of Ayrshire dairy cattle and new haplotypes affecting fertility and stillbirth in

- Holstein, Brown Swiss and Ayrshire breeds. ADSA-ASAS Joint Annual Meeting :poster T206.
- Cooper, T. A., Wiggans, G. R., Null, D. J., Hutchison, J. L., & Cole, J. B. (2014, juin). Genomic evaluation, breed identification, and discovery of a haplotype affecting fertility for Ayrshire dairy cattle. *Journal of Dairy Science*, *97*(6), 3878–3882. doi: 10.3168/jds.2013-7427
- Cosenza, G., Pauciuolo, A., Colimoro, L., Mancusi, A., Rando, A., Di Berardino, D., & Ramunno, L. (2007). An SNP in the goat CSN2 promoter region is associated with the absence of -casein in milk. *Animal Genetics*, *38*(6), 655–658. doi: 10.1111/j.1365-2052.2007.01649.x
- Daetwyler, H. D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R. F., Liao, X., Djari, A., Rodriguez, S. C., Grohs, C., Esquerré, D., Bouchez, O., Rossignol, M.-N., Klopp, C., Rocha, D., Fritz, S., Eggen, A., Bowman, P. J., Coote, D., Chamberlain, A. J., Anderson, C., VanTassell, C. P., Hulsegege, I., Goddard, M. E., Guldbbrandtsen, B., Lund, M. S., Veerkamp, R. F., Boichard, D. A., Fries, R., & Hayes, B. J. (2014, août). Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics*, *46*(8), 858–865. doi: 10.1038/ng.3034
- Danchin-Burge, C., Danvy, S., Laloë, D., & Verrier, E. (2017). Création d'un observatoire de la VARIabilité génétique des RUMinants et des Equidés (VARUME). *Innovations Agronomiques*, *55*, 235–245. doi: 10.15454/1.5137783633347627E12
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011, août). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi: 10.1093/bioinformatics/btr330
- Darré, G., Quéinnec, G., & Berland, H. (1972). La translocation 1-29 des bovins. Étude générale et importance du phénomène dans le Sud-Ouest. *Rev. Méd. Vet.*, *123*, 477–494.
- Demars, J., Cano, M., Drouilhet, L., Plisson-Petit, F., Bardou, P., Fabre, S., Servin, B., Sarry, J., Woloszyn, F., Mulsant, P., Foulquier, D., Carrière, F., Aletru, M., Rodde, N., Cauet, S., Bouchez, O., Pirson, M., Tosser-Klopp, G., & Allain, D. (2017, juillet). Genome-Wide Identification of the Mutation Underlying Fleece Variation and Discriminating Ancestral Hairy Species from Modern Woolly Sheep. *Molecular Biology and Evolution*, *34*(7), 1722–1729. doi: 10.1093/molbev/msx114
- Demars, J., Fabre, S., Sarry, J., Rossetti, R., Gilbert, H., Persani, L., Tosser-Klopp, G., Mulsant, P., Nowak, Z., Drobik, W., Martyniuk, E., & Bodin, L. (2013, avril). Genome-Wide Association Studies Identify Two Novel BMP15 Mutations Responsible for an Atypical Hyperprolificacy Phenotype in Sheep. *PLOS Genetics*, *9*(4), e1003482. doi: 10.1371/journal.pgen.1003482
- Deng, J., Xie, X.-L., Wang, D.-F., Zhao, C., Lv, F.-H., Li, X., Yang, J., Yu, J.-L., Shen, M., Gao, L., Yang, J.-Q., Liu, M.-J., Li, W.-R., Wang, Y.-T., Wang, F., Li, J.-Q., Hehua, E., Liu, Y.-G., Shen, Z.-Q., Ren, Y.-L., Liu, G.-J., Chen, Z.-H., Gorkhali, N. A., Rushdi, H. E., Salehian-Dehkordi, H., Esmailzadeh, A., Nosrati, M., Paiva, S. R., Caetano, A. R., Štěpánek, O., Olsaker, I., Weimann, C., Erhardt, G., Curik, I., Kantanen, J., Mwacharo, J. M., Hanotte, O., Bruford, M. W., Ciani, E., Periasamy, K., Amills, M., Lenstra, J. A., Han, J.-L., Zhang, H.-P., Li, L., & Li, M.-H. (2020, octobre). Paternal Origins and Migratory Episodes of

- Domestic Sheep. *Current Biology*, 30(20), 4085–4095.e6. doi: 10.1016/j.cub.2020.07.077
- Denoyelle, L., Talouarn, E., Bardou, P., Colli, L., Alberti, A., Danchin, C., Del Corvo, M., Engelen, S., Orvain, C., Palhière, I., Rupp, R., Sarry, J., Salavati, M., Amills, M., Clark, E., Crepaldi, P., Faraut, T., Masiga, C. W., Pompanon, F., Rosen, B. D., Stella, A., Van Tassell, C. P., Tosser-Klopp, G., Kijas, J., Guldbbrandtsen, B., Kantanen, J., Duby, D., Martin, P., Danchin, C., Duclos, D., Allain, D., Arquet, R., Mandonnet, N., Naves, M., Palhière, I., Rupp, R., breeders, C., Pompanon, F., Rezaei, H. R., Carolan, S., Foran, M., Stella, A., Ajmone-Marsan, P., Colli, L., Crisà, A., Marletta, D., Crepaldi, P., Ottino, M., Randi, E., Benjelloun, B., Lenstra, H., Moaen-ud Din, M., Reecy, J., Goyache, F., Alvarez, I., Amills, M., Sánchez, A., Capote, J., Jordana, J., Pons, A., Martínez, A., Molina, A., Rosen, B., Visser, C., Drögemüller, C., Luikart, G., Masiga, C. W., Mujibi, D. F., Mruttu, H. A., Gondwe, T., Sikosana, J., Da Gloria, M. T., Nash, O., & The VarGoats Consortium. (2021, novembre). VarGoats project : a dataset of 1159 whole-genome sequences to dissect *Capra hircus* global diversity. *Genetics Selection Evolution*, 53(1), 86. doi: 10.1186/s12711-021-00659-6
- Deodato, F., Boenzi, S., Santorelli, F. M., & Dionisi-Vici, C. (2006). Methylmalonic and propionic aciduria. *American Journal of Medical Genetics Part C : Seminars in Medical Genetics*, 142C(2), 104–112. doi: 10.1002/ajmg.c.30090
- Derks, M. F. L., Lopes, M. S., Bosse, M., Madsen, O., Dibbits, B., Harlizius, B., Groenen, M. A. M., & Megens, H.-J. (2018, septembre). Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS Genetics*, 14(9), e1007661. doi: 10.1371/journal.pgen.1007661
- Derks, M. F. L., Megens, H.-J., Bosse, M., Lopes, M. S., Harlizius, B., & Groenen, M. A. M. (2017, novembre). A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics*, 18(1), 858. doi: 10.1186/s12864-017-4278-1
- Derks, M. F. L., Megens, H.-J., Bosse, M., Visscher, J., Peeters, K., Bink, M. C. A. M., Vereijken, A., Gross, C., de Ridder, D., Reinders, M. J. T., & Groenen, M. A. M. (2018, décembre). A survey of functional genomic variation in domesticated chickens. *Genetics Selection Evolution*, 50(1), 17. doi: 10.1186/s12711-018-0390-1
- Derks, M. F. L., & Steensma, M. (2021). Review : Balancing Selection for Deleterious Alleles in Livestock. *Frontiers in Genetics*, 12.
- Dillingh, M. R., van Poelgeest, E. P., Malone, K. E., Kemper, E. M., Stroes, E. S. G., Moerland, M., & Burggraaf, J. (2014, septembre). Characterization of inflammation and immune cell modulation induced by low-dose LPS administration to healthy volunteers. *Journal of Inflammation*, 11(1), 28. doi: 10.1186/s12950-014-0028-1
- Dittmer, K. E., Jolly, R. D., Mayhew, I. G., Ridler, A. L., Chernyavtseva, A., Garrick, D. J., & Blair, H. T. (2017, décembre). Familial episodic ataxia in lambs is potentially associated with a mutation in the fibroblast growth factor 14 (FGF14) gene. *PLOS ONE*, 12(12), e0190030. doi: 10.1371/journal.pone.0190030
- Drouilhet, L., Lecerf, F., Bodin, L., Fabre, S., & Mulsant, P. (2009). Fine mapping of the FeCL locus influencing prolificacy in Lacaune sheep. *Animal Genetics*, 40(6), 804–812. doi: 10.1111/j.1365-2052.2009.01919.x

- Drouilhet, L., Mansanet, C., Sarry, J., Tabet, K., Bardou, P., Woloszyn, F., Lluch, J., Harichaux, G., Viguié, C., Monniaux, D., Bodin, L., Mulsant, P., & Fabre, S. (2013, septembre). The Highly Prolific Phenotype of Lacaune Sheep Is Associated with an Ectopic Expression of the B4GALNT2 Gene within the Ovary. *PLOS Genetics*, *9*(9), e1003809. doi: 10.1371/journal.pgen.1003809
- Duchesne, A., Grohs, C., Michot, P., Bertaud, M., Boichard, D., Floriot, S., & Capitan, A. (2016, décembre). From phenotype to the causal mutation : recessive defects. *INRAE Productions Animales*, *29*(5), 319–328. doi: 10.20870/productions-animales.2016.29.5.3000
- Ducos, A., Eggen, A. A., Darré, R., & Boichard, D. (2002, décembre). Les anomalies génétiques dans l'espèce bovine. In *9. Rencontres autour des recherches sur les ruminants*. Paris, France : Institut de l'Élevage - INRA. Consulté le 2022-09-01, sur <https://hal.inrae.fr/hal-02764333>
- Ducos, A., Revay, T., Kovacs, A., Hidas, A., Pinton, A., Bonnet-Garnier, A., Molteni, L., Slota, E., Switonski, M., Arruga, M. V., Haeringen, W. A. v., Nicolae, I., Chaves, R., Guedes-Pinto, H., Andersson, M., & Iannuzzi, L. (2008). Cytogenetic screening of livestock populations in Europe : an overview. *Cytogenetic and Genome Research*, *120*(1-2), 26–41. doi: 10.1159/000118738
- Embree, C. M., Alhasan, R. A., & Singh, G. (2022, octobre). Features and factors that dictate if terminating ribosomes cause or counteract nonsense-mediated mRNA decay. *Journal of Biological Chemistry*, 102592. doi: 10.1016/j.jbc.2022.102592
- Esteban-Lopez, M., & Agoulnik, A. I. (2020, octobre). Diverse functions of insulin-like 3 peptide. *Journal of Endocrinology*, *247*(1), R1–R12. doi: 10.1530/JOE-20-0168
- Evans, H. J., Buckland, R. A., & Sumner, A. T. (1973, juillet). Chromosome homology and heterochromatin in goat, sheep and ox studied by banding techniques. doi: 10.1007/BF00399407
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016, octobre). MultiQC : summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. doi: 10.1093/bioinformatics/btw354
- Fabre, S., Chantepie, L., Plisson-Petit, F., Sarry, J., Woloszyn, F., Genet, C., Drouilhet, L., & Tosser-Klopp, G. (2020). A novel homozygous nonsense mutation in ITGB4 gene causes epidermolysis bullosa in Mouton Vendéen sheep. *Animal Genetics*. doi: <https://doi.org/10.1111/age.13026>
- Ferrer, L. M., Monteagudo, L. V., García de Jalón, J. A., Tejedor, M. T., Ramos, J. J., & Lacasta, D. (2009). A case of ovine female XY sex reversal syndrome not related to anomalies in the sex-determining region Y (SRY). doi: 10.1159/000253308
- Finocchiaro, R., Portolano, B., Damiani, G., Caroli, A., Budelli, E., Bolla, P., & Pagnacco, G. (2003, juin). The hairless (hr) gene is involved in the congenital hypotrichosis of Valle del Belice sheep. *Genetics Selection Evolution*, *35*(1), S147. doi: 10.1186/1297-9686-35-S1-S147
- Fliss, M. S., Hinkle, P. M., & Bancroft, C. (1999). Expression Cloning and Characterization of PREB (Prolactin Regulatory Element Binding), a Novel WD Motif DNA-Binding Protein with a Capacity to Regulate Prolactin Promoter Activity. , *13*(4), 14.

- Fontanesi, L., Beretti, F., Riggio, V., Dall'Olio, S., González, E. G., Finocchiaro, R., Davoli, R., Russo, V., & Portolano, B. (2009, août). Missense and nonsense mutations in melanocortin 1 receptor (MC1R) gene of different goat breeds : association with red and black coat colour phenotypes but with unexpected evidences. *BMC Genetics*, *10*(1), 47. doi: 10.1186/1471-2156-10-47
- Fontanesi, L., Dall'Olio, S., Beretti, F., Portolano, B., & Russo, V. (2011, janvier). Coat colours in the Massese sheep breed are associated with mutations in the agouti signalling protein (ASIP) and melanocortin 1 receptor (MC1R) genes. *Animal*, *5*(1), 8–17. doi: 10.1017/S1751731110001382
- Freking, B. A., Murphy, S. K., Wylie, A. A., Rhodes, S. J., Keele, J. W., Leymaster, K. A., Jirtle, R. L., & Smith, T. P. L. (2002, janvier). Identification of the Single Base Change Causing the Callipyge Muscle Hypertrophy Phenotype, the Only Known Example of Polar Overdominance in Mammals. *Genome Research*, *12*(10), 1496–1506. doi: 10.1101/gr.571002
- Fritz, S., Capitan, A., Djari, A., Rodriguez, S. C., Barbat, A., Baur, A., Grohs, C., Weiss, B., Boussaha, M., Esquerré, D., Klopp, C., Rocha, D., & Boichard, D. (2013, juin). Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2. *PloS One*, *8*(6), e65550. doi: 10.1371/journal.pone.0065550
- Fritz, S., Hozé, C., Rebours, E., Barbat, A., Bizard, M., Chamberlain, A., Escoufflaire, C., Vander Jagt, C., Boussaha, M., Grohs, C., Allais-Bonnet, A., Philippe, M., Vallée, A., Amigues, Y., Hayes, B. J., Boichard, D., & Capitan, A. (2018, juillet). An initiator codon mutation in SDE2 causes recessive embryonic lethality in Holstein cattle. *Journal of Dairy Science*, *101*(7), 6220–6231. doi: 10.3168/jds.2017-14119
- Fritz, S., Michot, P., Hozé, C., Grohs, C., Boussaha, M., Boichard, D., & Capitan, A. (2016, décembre). Anticipate the emergence of genetic defects from genomics data. *INRAE Productions Animales*, *29*(5), 339–350. doi: 10.20870/productions-animales.2016.29.5.3002
- Frugier, T., Mitchell, N. L., Tammen, I., Houweling, P. J., Arthur, D. G., Kay, G. W., van Diggelen, O. P., Jolly, R. D., & Palmer, D. N. (2008, février). A new large animal model of CLN5 neuronal ceroid lipofuscinosis in Borderdale sheep is caused by a nucleotide substitution at a consensus splice site (c.571+1G>>>A) leading to excision of exon 3. *Neurobiology of Disease*, *29*(2), 306–315. doi: 10.1016/j.nbd.2007.09.006
- Galloway, S. M., McNatty, K. P., Cambridge, L. M., Laitinen, M. P. E., Juengel, J. L., Jokiranta, T. S., McLaren, R. J., Luiro, K., Dodds, K. G., Montgomery, G. W., Beattie, A. E., Davis, G. H., & Ritvos, O. (2000, juillet). Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and infertility in a dosage-sensitive manner. *Nature Genetics*, *25*(3), 279–283. doi: 10.1038/77033
- Garcia, M., Juhos, S., Larsson, M., Olason, P. I., Martin, M., Einfeldt, J., DiLorenzo, S., Sandgren, J., Ståhl, T. D. D., Ewels, P., Wirta, V., Nistér, M., Käller, M., & Nystedt, B. (2020, septembre). Sarek : A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Research*. doi: 10.12688/f1000research.16665.2
- Georges, M., Charlier, C., & Cockett, N. (2003, mai). The callipyge locus : evidence for the

- trans interaction of reciprocally imprinted genes. Trends in Genetics, 19(5), 248–252. doi: 10.1016/S0168-9525(03)00082-9
- Georges, M., Charlier, C., & Hayes, B. (2019, mars). Harnessing genomic information for livestock improvement. Nature Reviews Genetics, 20(3), 135–156. doi: 10.1038/s41576-018-0082-2
- Goksuluk, D., Korkmaz, S., Zararsiz, G., & Karaagaoglu, A., Ergun. (2016). easyROC : An Interactive Web-tool for ROC Curve Analysis Using R Language Environment. The R Journal, 8(2), 213. doi: 10.32614/RJ-2016-042
- Goldmann, W., Hunter, N., Foster, J. D., Salbaum, J. M., Beyreuther, K., & Hope, J. (1990, avril). Two alleles of a neural protein gene linked to scrapie in sheep. Proceedings of the National Academy of Sciences, 87(7), 2476–2480. doi: 10.1073/pnas.87.7.2476
- Goldmann, W., Ryan, K., Stewart, P., Parnham, D., Xicohtencatl, R., Fernandez, N., Saunders, G., Windl, O., González, L., Bossers, A., & Foster, J. (2011, octobre). Caprine prion gene polymorphisms are associated with decreased incidence of classical scrapie in goat herds in the United Kingdom. Veterinary Research, 42(1), 110. doi: 10.1186/1297-9716-42-110
- González-Montaña, J.-R., Escalera-Valente, F., Alonso, A. J., Lomillos, J. M., Robles, R., & Alonso, M. E. (2020, octobre). Relationship between Vitamin B12 and Cobalt Metabolism in Domestic Ruminant : An Update. Animals, 10(10), 1855. doi: 10.3390/ani10101855
- Gonçalves, H. C., Jorge, W., & Cury, P. R. (1992, septembre). Distribution of a Robertsonian translocation in goats. Small Ruminant Research, 8(4), 345–352. doi: 10.1016/0921-4488(92)90215-P
- Gootwine, E., Abu-Siam, M., Obolensky, A., Rosov, A., Honig, H., Nitzan, T., Shirak, A., Ezra-Elia, R., Yamin, E., Banin, E., Averbukh, E., Hauswirth, W. W., Ofri, R., & Seroussi, E. (2017, mars). Gene Augmentation Therapy for a Missense Substitution in the cGMP-Binding Domain of Ovine CNGA3 Gene Restores Vision in Day-Blind Sheep. Investigative Ophthalmology & Visual Science, 58(3), 1577–1584. doi: 10.1167/iovs.16-20986
- Gorlov, I. P., Kamat, A., Bogatcheva, N. V., Jones, E., Lamb, D. J., Truong, A., Bishop, C. E., McElreavey, K., & Agoulnik, A. I. (2002, septembre). Mutations of the GREAT gene cause cryptorchidism. Human Molecular Genetics, 11(19), 2309–2318. doi: 10.1093/hmg/11.19.2309
- Gratten, J., Beraldi, D., Lowder, B., McRae, A., Visscher, P., Pemberton, J., & Slate, J. (2007, mars). Compelling evidence that a single nucleotide substitution in *TYRP1* is responsible for coat-colour polymorphism in a free-living population of Soay sheep. Proceedings of the Royal Society B : Biological Sciences, 274(1610), 619–626. doi: 10.1098/rspb.2006.3762
- Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.-J., Li, S., Larkin, D. M., Kim, H., Frantz, L. A. F., Caccamo, M., Ahn, H., Aken, B. L., Anselmo, A., Anthon, C., Auvil, L., Badaoui, B., Beattie, C. W., Bendixen, C., Berman, D., Blecha, F., Blomberg, J., Bolund, L., Bosse, M., Botti, S., Bujie, Z., Bystrom, M., Capitanu, B., Carvalho-Silva, D., Chardon, P., Chen, C., Cheng, R., Choi, S.-H., Chow, W., Clark, R. C., Clee, C., Crooijmans, R. P. M. A., Dawson, H. D., Dehais, P., De Sapio, F., Dibbits, B., Drou, N., Du, Z.-Q., Eversole, K., Fadista, J., Fairley, S., Faraut, T., Faulkner, G. J., Fowler, K. E., Fredholm, M., Fritz,

- E., Gilbert, J. G. R., Giuffra, E., Gorodkin, J., Griffin, D. K., Harrow, J. L., Hayward, A., Howe, K., Hu, Z.-L., Humphray, S. J., Hunt, T., Hornshøj, H., Jeon, J.-T., Jern, P., Jones, M., Jurka, J., Kanamori, H., Kapetanovic, R., Kim, J., Kim, J.-H., Kim, K.-W., Kim, T.-H., Larson, G., Lee, K., Lee, K.-T., Leggett, R., Lewin, H. A., Li, Y., Liu, W., Loveland, J. E., Lu, Y., Lunney, J. K., Ma, J., Madsen, O., Mann, K., Matthews, L., McLaren, S., Morozumi, T., Murtaugh, M. P., Narayan, J., Truong Nguyen, D., Ni, P., Oh, S.-J., Onteru, S., Panitz, F., Park, E.-W., Park, H.-S., Pascal, G., Paudel, Y., Perez-Enciso, M., Ramirez-Gonzalez, R., Reecy, J. M., Rodriguez-Zas, S., Rohrer, G. A., Rund, L., Sang, Y., Schachtschneider, K., Schraiber, J. G., Schwartz, J., Scobie, L., Scott, C., Searle, S., Servin, B., Southey, B. R., Sperber, G., Stadler, P., Sweedler, J. V., Tafer, H., Thomsen, B., Wali, R., Wang, J., Wang, J., White, S., Xu, X., Yerle, M., Zhang, G., Zhang, J., Zhang, J., Zhao, S., Rogers, J., Churcher, C., & Schook, L. B. (2012, novembre). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, *491*(7424), 393–398. doi: 10.1038/nature11622
- Grohs, C., Duchesne, A., Floriot, S., Deloche, M.-C., Boichard, D., Ducos, A., & Danchin-Burge, C. (2016, décembre). The National Observatory of Bovine Abnormalities : actions and results for effective assistance to the management of genetic defects. *INRAE Productions Animales*, *29*(5), 307–318. doi: 10.20870/productions-animales.2016.29.5.2998
- Gui, L., Song, K., Tritschler, D., Bower, R., Yan, S., Dai, A., Augspurger, K., Sakizadeh, J., Grzemska, M., Ni, T., Porter, M. E., & Nicastro, D. (2019, novembre). Scaffold subunits support associated subunit assembly in the Chlamydomonas ciliary nexin–dynein regulatory complex. *Proceedings of the National Academy of Sciences*, *116*(46), 23152–23162. doi: 10.1073/pnas.1910960116
- Gunderson, K. L., Steemers, F. J., Lee, G., Mendoza, L. G., & Chee, M. S. (2005, mai). A genome-wide scalable SNP genotyping assay using microarray technology. *Nature Genetics*, *37*(5), 549–554. doi: 10.1038/ng1547
- Hanrahan, J. P., Gregan, S. M., Mulsant, P., Mullen, M., Davis, G. H., Powell, R., & Galloway, S. M. (2004, avril). 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. *Biology of Reproduction*, *70*(4), 900–909. doi: 10.1095/biolreprod.103.023093
- Has, C., Bauer, J., Bodemer, C., Bolling, M., Bruckner-Tuderman, L., Diem, A., Fine, J.-D., Heagerty, A., Hovnanian, A., Marinkovich, M., Martinez, A., McGrath, J., Moss, C., Murrell, D., Palisson, F., Schwieger-Briel, A., Sprecher, E., Tamai, K., Uitto, J., Woodley, D., Zambruno, G., & Mellerio, J. (2020). Consensus reclassification of inherited epidermolysis bullosa and other disorders with skin fragility. *British Journal of Dermatology*, *183*(4), 614–627. doi: 10.1111/bjd.18921
- He, S., Zhang, Z., Sun, Y., Ren, T., Li, W., Zhou, X., Michal, J. J., Jiang, Z., & Liu, M. (2020). Genome-wide association study shows that microtia in Altay sheep is caused by a 76 bp duplication of HMX1. *Animal Genetics*, *51*(1), 132–136. doi: 10.1111/age.12876
- Heaton, M. P., Clawson, M. L., Chitko-Mckown, C. G., Leymaster, K. A., Smith, T. P. L.,

- Harhay, G. P., White, S. N., Herrmann-Hoesing, L. M., Mousel, M. R., Lewis, G. S., Kalbfleisch, T. S., Keen, J. E., & Laegreid, W. W. (2012, janvier). Reduced Lentivirus Susceptibility in Sheep with TMEM154 Mutations. *PLOS Genetics*, *8*(1), e1002467. doi: 10.1371/journal.pgen.1002467
- Hedrick, P. W. (2015). Heterozygote Advantage : The Effect of Artificial Selection in Livestock and Pets. *Journal of Heredity*, *106*(2), 141–154. doi: 10.1093/jhered/esu070
- Henkel, J., Saif, R., Jagannathan, V., Schmocker, C., Zeindler, F., Bangerter, E., Herren, U., Posantzis, D., Bulut, Z., Ammann, P., Drögemüller, C., Flury, C., & Leeb, T. (2019, décembre). Selection signatures in goats reveal copy number variants underlying breed-defining coat color phenotypes. *PLOS Genetics*, *15*(12), e1008536. doi: 10.1371/journal.pgen.1008536
- Henry, C. J., Huang, Y., Wynne, A. M., & Godbout, J. P. (2009, mars). Peripheral lipopolysaccharide (LPS) challenge promotes microglial hyperactivity in aged mice that is associated with exaggerated induction of both pro-inflammatory IL-1 and anti-inflammatory IL-10 cytokines. *Brain, Behavior, and Immunity*, *23*(3), 309–317. doi: 10.1016/j.bbi.2008.09.002
- Heuser, T., Raytchev, M., Krell, J., Porter, M. E., & Nicastro, D. (2009, décembre). The dynein regulatory complex is the nexin link and a major regulatory node in cilia and flagella. *Journal of Cell Biology*, *187*(6), 921–933. doi: 10.1083/jcb.200908067
- Hill, W. G., & Robertson, A. (1968, juin). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, *38*(6), 226–231. doi: 10.1007/BF01245622
- Hollox, E. J., Zuccherato, L. W., & Tucci, S. (2022, janvier). Genome structural variation in human evolution. *Trends in Genetics*, *38*(1), 45–58. doi: 10.1016/j.tig.2021.06.015
- Holm, L.-E., Bendixen, C., Eythorsdottir, E., & Hallsson, J. H. (2022). A frameshift deletion in the GDF9 gene in Icelandic Loa prolific sheep. *Animal Genetics*, *53*(2), 220–223. doi: 10.1111/age.13168
- Horani, A., Brody, S. L., Ferkol, T. W., Shoseyov, D., Wasserman, M. G., Ta-shma, A., Wilson, K. S., Bayly, P. V., Amirav, I., Cohen-Cymberek, M., Dutcher, S. K., Elpeleg, O., & Kerem, E. (2013, août). CCDC65 Mutation Causes Primary Ciliary Dyskinesia with Normal Ultrastructure and Hyperkinetic Cilia. *PLOS ONE*, *8*(8), e72299. doi: 10.1371/journal.pone.0072299
- Hozé, C., Escoufflaire, C., Mesbah-Uddin, M., Barbat, A., Boussaha, M., Deloche, M., Boichard, D., Fritz, S., & Capitan, A. (2020, janvier). Short communication : A splice site mutation in CENPU is associated with recessive embryonic lethality in Holstein cattle. *Journal of Dairy Science*, *103*(1), 607–612. doi: 10.3168/jds.2019-17056
- Hozé, C., Fritz, S., Baur, A., Grohs, C., Danchin-Burge, C., & Boichard, D. (2018, décembre). Accounting for genetic conditions in breeding objectives in dairy cattle. In *In Proceedings of the 24ème Rencontres Autour des Recherches Sur les Ruminants : 5-6 December*. Paris.
- Huang, Z., Zang, K., & Reichardt, L. F. (2005, août). The origin recognition core complex regulates dendrite and spine development in postmitotic neurons. *Journal of Cell Biology*, *170*(4), 527–535. doi: 10.1083/jcb.200505075
- Huntzinger, E., & Izaurralde, E. (2011, février). Gene silencing by microRNAs : contributions of

- translational repression and mRNA decay. *Nature Reviews Genetics*, 12(2), 99–110. doi: 10.1038/nrg2936
- Huppke, P., Brendel, C., Kalscheuer, V., Korenke, G., Marquardt, I., Freisinger, P., Christodoulou, J., Hillebrand, M., Pitelet, G., Wilson, C., Gruber-Sedlmayr, U., Ullmann, R., Haas, S., Elpeleg, O., Nürnberg, G., Nürnberg, P., Dad, S., Møller, L., Kaler, S., & Gärtner, J. (2012, janvier). Mutations in SLC33A1 Cause a Lethal Autosomal-Recessive Disorder with Congenital Cataracts, Hearing Loss, and Low Serum Copper and Ceruloplasmin. *The American Journal of Human Genetics*, 90(1), 61–68. doi: 10.1016/j.ajhg.2011.11.030
- Häfliger, I. M., Seefried, F. R., & Drögemüller, C. (2021, décembre). Reverse Genetic Screen for Deleterious Recessive Variants in the Local Simmental Cattle Population of Switzerland. *Animals*, 11(12), 3535. doi: 10.3390/ani11123535
- Häfliger, I. M., Seefried, F. R., Spengeler, M., & Drögemüller, C. (2021, décembre). Mining massive genomic data of two Swiss Braunvieh cattle populations reveals six novel candidate variants that impair reproductive success. *Genetics Selection Evolution*, 53(1), 95. doi: 10.1186/s12711-021-00686-3
- Häfliger, I. M., Spengeler, M., Seefried, F. R., & Drögemüller, C. (2022, décembre). Four novel candidate causal variants for deficient homozygous haplotypes in Holstein cattle. *Scientific Reports*, 12(1), 5435. doi: 10.1038/s41598-022-09403-6
- Iannuzzi, A., Parma, P., & Iannuzzi, L. (2021, mars). Chromosome Abnormalities and Fertility in Domestic Bovids : A Review. *Animals*, 11(3), 802. doi: 10.3390/ani11030802
- Iannuzzi, A., Perucatti, A., Genuardo, V., De Lorenzi, L., Di Bernardino, D., Parma, P., & Iannuzzi, L. (2013). Cytogenetic elaboration of a novel reciprocal translocation in sheep. doi: 10.1159/000346193
- Iannuzzi, A., Perucatti, A., Genuardo, V., Pauciullo, A., Incarnato, D., Musilova, P., Rubes, J., & Iannuzzi, C. (2014). The Utility of Chromosome Microdissection in Clinical Cytogenetics : A New Reciprocal Translocation in Sheep. *Cytogenetic and Genome Research*, 142(3), 174–178. doi: 10.1159/000361000
- IDELE. (2016). Comment sont évalués génétiquement les ovins laitiers en France (Rapport technique). Consulté le 2020-04-02, sur http://idele.fr/no_cache/recherche/publication/idelesolr/recommends/comment-evaluer-genetiquement-les-ovins-laitiers-en-france.html
- IDELE. (2022a). Résultats de contrôle laitier 2021 - Espèce caprine (Rapport technique). Consulté le 2022-07-30, sur https://idele.fr/?eID=cmis_download&oID=workspace%3A%2F%2FSpacesStore%2Ffe302ebd2-1b8f-4646-a9f1-4124d825297b&cHash=05f50afceeadc6bc0c8df6dd05fdd9c3
- IDELE. (2022b). Résultats de contrôle laitier 2021 - Espèce ovine (Rapport technique). Consulté le 2022-07-30, sur https://idele.fr/?eID=cmis_download&oID=workspace%3A%2F%2FSpacesStore%2F289791dc-515e-4803-8a21-d1314cf53753&cHash=ca6196e07cac7aea74d067e064538292
- IDELE, & CNE. (2021a, septembre). caprins 2021 - Productions lait et viande (Rapport technique).

- IDELE, & CNE. (2021b, septembre). ovins 2021 - Productions lait et viande (Rapport technique).
- Illumina. (2009). Infinium HD Assay Ultra Protocol Guide (Rapport technique). Consulté sur http://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/infinium_assays/infinium-hd-ultra/11328087_RevB_Infinium_HD_Ultra_Assay_Guide_press.pdf
- Illumina. (2010). Interpreting Infinium Assay Data for Whole-Genome Structural Variation (Rapport technique). Consulté sur https://emea.illumina.com/Documents/products/technotes/technote_cytoanalysis.pdf
- Illumina. (2015). OvineSNP50 Genotyping BeadChip (Rapport technique). Consulté sur https://www.illumina.com/documents/products/datasheets/datasheet_ovinesnp50.pdf
- Illumina. (2022a). Evaluation of Infinium Staining Controls in the GenomeStudio Controls Dashboard. Consulté le 2022-08-27, sur <https://support.illumina.com/bulletins/2022/04/evaluation-of-infinium-staining-controls-in-the-genomestudio-controls.html>
- Illumina. (2022b). Illumina Microarray Technology. Consulté le 2022-08-27, sur <https://emea.illumina.com/science/technology/microarray.html>
- Illumina. (2022c). Sequencing Technology | Sequencing by synthesis. Consulté le 2022-09-06, sur <https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology.html>
- Ishiai, M., Dean, F. B., Okumura, K., Abe, M., Moon, K.-Y., Amin, A. A., Kagotani, K., Taguchi, H., Murakami, Y., Hanaoka, F., O'donnell, M., Hurwitz, J., & Eki, T. (1997, décembre). Isolation of Human and Fission Yeast Homologues of the Budding Yeast Origin Recognition Complex Subunit ORC5 : Human Homologue (ORC5L) Maps to 7q22. Genomics, *46*(2), 294–298. doi: 10.1006/geno.1997.5003
- Ivell, R., & Anand-Ivell, R. (2018, novembre). Insulin-like peptide 3 (INSL3) is a major regulator of female reproductive physiology. Human Reproduction Update, *24*(6), 639–651. doi: 10.1093/humupd/dmy029
- Jammes, H., & Renard, J. P. (2010, mars). Epigénétique et construction du phénotype, un enjeu pour les productions animales? INRAE Productions Animales, *23*(1), 23–42. doi: 10.20870/productions-animales.2010.23.1.3283
- Janssens, A. C. J. W., & Martens, F. K. (2020, août). Reflection on modern methods : Revisiting the area under the ROC Curve. International Journal of Epidemiology, *49*(4), 1397–1403. doi: 10.1093/ije/dyz274
- Jenko, J. (2019). Analysis of a large dataset reveals haplotypes carrying putatively recessive lethal and semi-lethal alleles with pleiotropic effects on economically important traits in beef cattle. , 14.
- Jiang, Y., Xie, M., Chen, W., Talbot, R., Maddox, J. F., Faraut, T., Wu, C., Muzny, D. M., Li, Y., Zhang, W., Stanton, J.-A., Brauning, R., Barris, W. C., Hourlier, T., Aken, B. L., Searle, S. M. J., Adelson, D. L., Bian, C., Cam, G. R., Chen, Y., Cheng, S., DeSilva, U., Dixen, K., Dong, Y., Fan, G., Franklin, I. R., Fu, S., Fuentes-Utrilla, P., Guan, R., Highland, M. A., Holder, M. E., Huang, G., Ingham, A. B., Jhangiani, S. N., Kalra, D., Kovar, C. L.,

- Lee, S. L., Liu, W., Liu, X., Lu, C., Lv, T., Mathew, T., McWilliam, S., Menzies, M., Pan, S., Robelin, D., Servin, B., Townley, D., Wang, W., Wei, B., White, S. N., Yang, X., Ye, C., Yue, Y., Zeng, P., Zhou, Q., Hansen, J. B., Kristiansen, K., Gibbs, R. A., Flicek, P., Warkup, C. C., Jones, H. E., Oddy, V. H., Nicholas, F. W., McEwan, J. C., Kijas, J. W., Wang, J., Worley, K. C., Archibald, A. L., Cockett, N., Xu, X., Wang, W., & Dalrymple, B. P. (2014, juin). The sheep genome illuminates biology of the rumen and lipid metabolism. *Science*, *344*(6188), 1168–1173. doi: 10.1126/science.1252806
- Johnston, S. E., Gratten, J., Berenos, C., Pilkington, J. G., Clutton-Brock, T. H., Pemberton, J. M., & Slate, J. (2013, octobre). Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature*, *502*(7469), 93–95. doi: 10.1038/nature12489
- Kadri, N. K., Sahana, G., Charlier, C., Iso-Touru, T., Guldbandsen, B., Karim, L., Nielsen, U. S., Panitz, F., Aamand, G. P., Schulman, N., Georges, M., Vilkki, J., Lund, M. S., & Druet, T. (2014, janvier). A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle : Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. *PLoS Genetics*, *10*(1), e1004049. doi: 10.1371/journal.pgen.1004049
- Karen, A., Beckers, J.-F., Sulon, J., de Sousa, N. M., Szabados, K., Reczigel, J., & Szenci, O. (2003, mai). Early pregnancy diagnosis in sheep by progesterone and pregnancy-associated glycoprotein tests. *Theriogenology*, *59*(9), 1941–1948. doi: 10.1016/S0093-691X(02)01289-X
- Ka-Shu Wong, G., Liu, B., Wang, J., Zhang, Y., Yang, X., Zhang, Z., Meng, Q., Zhou, J., Li, D., Zhang, J., Ni, P., Li, S., Ran, L., Li, H., Zhang, J., Li, R., Li, S., Zheng, H., Lin, W., Li, G., Wang, X., Zhao, W., Li, J., Ye, C., Dai, M., Ruan, J., Zhou, Y., Li, Y., He, X., Zhang, Y., Wang, J., Huang, X., Tong, W., Chen, J., Ye, J., Chen, C., Wei, N., Li, G., Dong, L., Lan, F., Sun, Y., Zhang, Z., Yang, Z., Yu, Y., Huang, Y., He, D., Xi, Y., Wei, D., Qi, Q., Li, W., Shi, J., Wang, M., Xie, F., Wang, J., Zhang, X., Wang, P., Zhao, Y., Li, N., Yang, N., Dong, W., Hu, S., Zeng, C., Zheng, W., Hao, B., Hillier, L. W., Yang, S.-P., Warren, W. C., Wilson, R. K., Brandström, M., Ellegren, H., Crooijmans, R. P. M. A., van der Poel, J. J., Bovenhuis, H., Groenen, M. A. M., Ovcharenko, I., Gordon, L., Stubbs, L., Lucas, S., Glavina, T., Aerts, A., Kaiser, P., Rothwell, L., Young, J. R., Rogers, S., Walker, B. A., van Hateren, A., Kaufman, J., Bumstead, N., Lamont, S. J., Zhou, H., Hocking, P. M., Morrice, D., de Koning, D.-J., Law, A., Bartley, N., Burt, D. W., Hunt, H., Cheng, H. H., Gunnarsson, U., Wahlberg, P., Andersson, L., Kindlund, E., Tammi, M. T., Andersson, B., Webber, C., Ponting, C. P., Overton, I. M., Boardman, P. E., Tang, H., Hubbard, S. J., Wilson, S. A., Yu, J., Wang, J., Yang, H., International Chicken Polymorphism Map Consortium, Polymorphism discovery and analysis : Beijing Institute of Genomics of Chinese Academy of Sciences and James D. Watson Institute of Genome Sciences of Zhejiang University, Genome sequence of red jungle fowl : Washington University School of Medicine, Molecular evolution : Uppsala University, Population genotyping, B. s. a. h. W. U., Lawrence Livermore National Laboratory, DOE Joint Genome Institute, Examples of application to complex traits : Institute for Animal Health, Iowa State University, Roslin Institute, USDA-ARS Avian Disease and Oncology Laboratory, Domestication and selection : Uppsala University,

- Karolinska Institutet, Human disease genes : University of Oxford, EST-based SNP data : University of Manchester Institute of Science and Technology, University of Sheffield, & Scientific management : Beijing Institute of Genomics of Chinese Academy of Sciences. (2004, décembre). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature*, *432*(7018), 717–722. doi: 10.1038/nature03156
- Kent, M., Moser, M., Boman, I. A., Lindtveit, K., Árnýasi, M., Sundsaasen, K. K., & Våge, D. I. (2021, juin). Insertion of an endogenous Jaagsiekte sheep retrovirus element into the BCO2 - gene abolishes its function and leads to yellow discoloration of adipose tissue in Norwegian Spælsau (*Ovis aries*). *BMC Genomics*, *22*(1), 492. doi: 10.1186/s12864-021-07826-5
- Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., Wilson, P., Ingersoll, R. G., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V. H., Nicholas, F. W., Raadsma, H., & Consortium, f. t. I. S. G. (2009, mars). A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *PLOS ONE*, *4*(3), e4668. doi: 10.1371/journal.pone.0004668
- Kouzarides, T. (2007, février). Chromatin Modifications and Their Function. *Cell*, *128*(4), 693–705. doi: 10.1016/j.cell.2007.02.005
- Krisans, S. K., Ericsson, J., Edwards, P. A., & Keller, G. A. (1994, mai). Farnesyl-diphosphate synthase is localized in peroxisomes. *Journal of Biological Chemistry*, *269*(19), 14165–14169. doi: 10.1016/S0021-9258(17)36769-8
- Kurkowiak, M., Ziętkiewicz, E., & Witt, M. (2015, janvier). Recent advances in primary ciliary dyskinesia genetics. *Journal of Medical Genetics*, *52*(1), 1–9. doi: 10.1136/jmedgenet-2014-102755
- Lander, E. S., & Botstein, D. (1987, juin). Homozygosity Mapping : A Way to Map Human Recessive Traits with the DNA of Inbred Children. *Science*, *236*(4808), 1567–1570. doi: 10.1126/science.2884728
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczy, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissole, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M.,

Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglu, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G. R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F. A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M. J., International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, C. f. G. R., The Sanger Centre :, Washington University Genome Sequencing Center, US DOE Joint Genome Institute :, Baylor College of Medicine Human Genome Sequencing Center :, RIKEN Genomic Sciences Center :, Genoscope and CNRS UMR-8030 :, Department of Genome Analysis, I. o. M. B., GTC Sequencing Center :, Beijing Genomics Institute/Human Genome Center :, Multimegabase Sequencing Center, T. I. f. S. B., Stanford Genome Technology Center :, University of Oklahoma's Advanced Center for Genome Technology :, Max Planck Institute for Molecular Genetics :, Cold Spring Harbor Laboratory, L. A. H. G. C., GBF—German Research Centre for Biotechnology :, *Genome Analysis Group (listed in alphabetical order, a. i. i. l. u. o. h., Scientific management : National Human Genome Research Institute, U. N. I. o. H., Stanford Human Genome Center :, University of Washington Genome Center :, Department of Molecular Biology, K. U. S. o. M., University of Texas Southwestern Medical Center at Dallas :, Office of Science, U. D. o. E., & The Wellcome Trust :. (2001, février). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. doi: 10.1038/35057062

Laplanche, J. L., Chatelain, J., Westaway, D., Thomas, S., Dussaucy, M., Brugere-Picoux, J., & Launay, J. M. (1993, janvier). PrP Polymorphisms Associated with Natural Scrapie Discovered by Denaturing Gradient Gel Electrophoresis. *Genomics*, *15*(1), 30–37. doi: 10.1006/geno.1993.1006

- Larroque, H., & Astruc, J.-M. (2017). Integration of a low density chip in dairy sheep genomic selection (Rapport technique).
- Larroque, H., Chassier, M., Saintilan, R., & Astruc, J.-M. (2017, septembre). Imputation accuracy from a low density SNP panel in 5 dairy sheep breeds in France. In In Proceedings of the 68th Annual Meeting of the European Association for Animal Production : 28 August-1 September. Tallinn.
- Lassoued, N., Benkhilil, Z., Woloszyn, F., Rejeb, A., Aouina, M., Rekik, M., Fabre, S., & Bedhiaf-Romdhani, S. (2017, mai). FecXBara Novel BMP15 mutation responsible for prolificacy and female sterility in Tunisian Barbarine Sheep. BMC Genetics, *18*(1), 43. doi: 10.1186/s12863-017-0510-x
- Lau, Y.-F. C. (2020, août). Y chromosome in health and diseases. Cell & Bioscience, *10*(1), 97. doi: 10.1186/s13578-020-00452-w
- Lawson, H. A., Cheverud, J. M., & Wolf, J. B. (2013, septembre). Genomic imprinting and parent-of-origin effects on complex traits. Nature Reviews Genetics, *14*(9), 609–617. doi: 10.1038/nrg3543
- Legarra, A., Baloche, G., Barillet, F., Astruc, J. M., Soulas, C., Aguerre, X., Arrese, F., Mintegi, L., Lasarte, M., Maeztu, F., Beltrán de Heredia, I., & Ugarte, E. (2014, mai). Within- and across-breed genomic predictions and genomic relationships for Western Pyrenees dairy sheep breeds Latxa, Manech, and Basco-Béarnaise. Journal of Dairy Science, *97*(5), 3200–3212. doi: 10.3168/jds.2013-7745
- Leipprandt, J. R., Kraemer, S. A., Haithcock, B. E., Chen, H., Dyme, J. L., Cavanagh, K. T., Friderici, K. H., & Jones, M. Z. (1996, octobre). Caprine α -Mannosidase : Sequencing and Characterization of the cDNA and Identification of the Molecular Defect of Caprine α -Mannosidosis. Genomics, *37*(1), 51–56. doi: 10.1006/geno.1996.0519
- Letko, A., Dijkman, R., Strugnell, B., Häfliger, I. M., Paris, J. M., Henderson, K., Geraghty, T., Orr, H., Scholes, S., & Drögemüller, C. (2020, octobre). Deleterious AGXT Missense Variant Associated with Type 1 Primary Hyperoxaluria (PH1) in Zwartbles Sheep. Genes, *11*(10), 1147. doi: 10.3390/genes11101147
- Letko, A., Strugnell, B., Häfliger, I. M., Paris, J. M., Waine, K., Drögemüller, C., & Scholes, S. (2021, janvier). Compound heterozygous PLA2G6 loss-of-function variants in Swaledale sheep with neuroaxonal dystrophy. Molecular Genetics and Genomics, *296*(1), 235–242. doi: 10.1007/s00438-020-01742-1
- Lewis, A., & Redrup, L. (2005, avril). Genetic Imprinting : Conflict at the Callipyge Locus. Current Biology, *15*(8), R291–R294. doi: 10.1016/j.cub.2005.04.003
- Li, H. (2013, mai). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv :1303.3997 [q-bio].
- Li, H., & Durbin, R. (2010, mars). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics, *26*(5), 589–595. doi: 10.1093/bioinformatics/btp698
- Lin, J., Tritschler, D., Song, K., Barber, C. F., Cobb, J. S., Porter, M. E., & Nicastro, D. (2011, août). Building Blocks of the Nexin-Dynein Regulatory Complex in Chlamydomonas Flagella*. Journal of Biological Chemistry, *286*(33), 29175–29191. doi: 10.1074/jbc.M111

- .241760
- Lin, P., Li, J., Liu, Q., Mao, F., Li, J., Qiu, R., Hu, H., Song, Y., Yang, Y., Gao, G., Yan, C., Yang, W., Shao, C., & Gong, Y. (2008, décembre). A Missense Mutation in SLC33A1, which Encodes the Acetyl-CoA Transporter, Causes Autosomal-Dominant Spastic Paraplegia (SPG42). *The American Journal of Human Genetics*, *83*(6), 752–759. doi: 10.1016/j.ajhg.2008.11.003
- Liu, W., Chen, L., Zhang, S., Hu, F., Wang, Z., Lyu, J., Wang, B., Xiang, H., Zhao, R., Tian, Z., Ge, S., & Wang, W. (2019, janvier). Decrease of gene expression diversity during domestication of animals and plants. *BMC Evolutionary Biology*, *19*(1), 19. doi: 10.1186/s12862-018-1340-9
- Lucienne, M., Aguilar-Pimentel, J. A., Amarie, O. V., Becker, L., Calzada-Wack, J., da Silva-Buttkus, P., Garrett, L., Hölter, S. M., Mayer-Kuckuk, P., Rathkolb, B., Rozman, J., Spielmann, N., Treise, I., Busch, D. H., Klopstock, T., Schmidt-Weber, C., Wolf, E., Wurst, W., Forny, M., Mathis, D., Fingerhut, R., Froese, D. S., Gailus-Durner, V., Fuchs, H., de Angelis, M. H., & Baumgartner, M. R. (2020, mars). In-depth phenotyping reveals common and novel disease symptoms in a hemizygous knock-in mouse model (Mut-ko/ki) of mut-type methylmalonic aciduria. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1866*(3), 165622. doi: 10.1016/j.bbadis.2019.165622
- Lühken, G., Fleck, K., Pauciullo, A., Huisinga, M., & Erhardt, G. (2012, décembre). Familial Hypopigmentation Syndrome in Sheep Associated with Homozygous Deletion of the Entire Endothelin Type-B Receptor Gene. *PLOS ONE*, *7*(12), e53020. doi: 10.1371/journal.pone.0053020
- MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J. K., Montgomery, S. B., Albers, C. A., Zhang, Z. D., Conrad, D. F., Lunter, G., Zheng, H., Ayub, Q., DePristo, M. A., Banks, E., Hu, M., Handsaker, R. E., Rosenfeld, J. A., Fromer, M., Jin, M., Mu, X. J., Khurana, E., Ye, K., Kay, M., Saunders, G. I., Suner, M.-M., Hunt, T., Barnes, I. H. A., Amid, C., Carvalho-Silva, D. R., Bignell, A. H., Snow, C., Yngvadottir, B., Bumpstead, S., Cooper, D. N., Xue, Y., Romero, I. G., Consortium, . G. P., Wang, J., Li, Y., Gibbs, R. A., McCarroll, S. A., Dermitzakis, E. T., Pritchard, J. K., Barrett, J. C., Harrow, J., Hurles, M. E., Gerstein, M. B., & Tyler-Smith, C. (2012, février). A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. *Science*, *335*(6070), 823–828. doi: 10.1126/science.1215040
- Marchini, J., & Howie, B. (2010, juillet). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, *11*(7), 499–511. doi: 10.1038/nrg2796
- Martinez-Royo, A., Jurado, J. J., Smulders, J. P., Martí, J. I., Alabart, J. L., Roche, A., Fantova, E., Bodin, L., Mulsant, P., Serrano, M., Folch, J., & Calvo, J. H. (2008). A deletion in the bone morphogenetic protein 15 gene causes sterility and increased prolificacy in Rasa Aragonesa sheep. *Animal Genetics*, *39*(3), 294–297. doi: 10.1111/j.1365-2052.2008.01707.x
- Mauffré, V., Grimard, B., Eozenou, C., Inghels, S., Silva, L., Giraud-Delville, C., Capo, D., Sandra, O., & Constant, F. (2016, janvier). Interferon stimulated genes as peripheral diagnostic markers of early pregnancy in sheep : a critical assessment. *Animal*, *10*(11), 1856–1863. doi: 10.1017/S175173111600077X

- McClure, M. C., Bickhart, D., Null, D., VanRaden, P., Xu, L., Wiggans, G., Liu, G., Schroeder, S., Glasscock, J., Armstrong, J., Cole, J. B., Tassell, C. P. V., & Sonstegard, T. S. (2014, mars). Bovine Exome Sequence Analysis and Targeted SNP Genotyping of Recessive Fertility Defects BH1, HH2, and HH3 Reveal a Putative Causative Mutation in SMC2 for HH3. *PloS One*, *9*(3), e92769. doi: 10.1371/journal.pone.0092769
- McHugo, G. P., Dover, M. J., & MacHugh, D. E. (2019, décembre). Unlocking the origins and biology of domestic animals using ancient DNA and paleogenomics. *BMC Biology*, *17*(1), 98. doi: 10.1186/s12915-019-0724-7
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010, janvier). The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. doi: 10.1101/gr.107524.110
- McRae, K. M., Baird, H. J., Dodds, K. G., Bixley, M. J., & Clarke, S. M. (2016, décembre). Incidence and heritability of ovine pneumonia, and the relationship with production traits in New Zealand sheep. *Small Ruminant Research*, *145*, 136–141. doi: 10.1016/j.smallrumres.2016.11.003
- Menzi, F., Keller, I., Reber, I., Beck, J., Brenig, B., Schütz, E., Leeb, T., & Drögemüller, C. (2016, juin). Genomic amplification of the caprine EDNRA locus might lead to a dose dependent loss of pigmentation. *Scientific Reports*, *6*(1), 28438. doi: 10.1038/srep28438
- Mesbah-Uddin, M., Hozé, C., Michot, P., Barbat, A., Lefebvre, R., Boussaha, M., Sahana, G., Fritz, S., Boichard, D., & Capitan, A. (2019, juillet). A missense mutation (p.Tyr452Cys) in the CAD gene compromises reproductive success in French Normande cattle. *Journal of Dairy Science*, *102*(7), 6340–6356. doi: 10.3168/jds.2018-16100
- Michot, P. (2017). *Utilisation des séquences de génome complet pour l'identification de mutations délétères responsables d'anomalies génétiques récessives chez le bovin.* (Theses, Université Paris Saclay (COMUE)). Consulté le 2020-12-09, sur <https://tel.archives-ouvertes.fr/tel-02068853> (Issue : 2017SACLA011)
- Ministère de l'Agriculture et de l'Alimentation. (2018a). *Référentiel de contrôle des performances dans le cadre du service public d'enregistrement et de contrôle des performances des ruminants pour la filière "production de lait de chèvre"* (Rapport technique). Consulté sur <https://info.agriculture.gouv.fr>
- Ministère de l'Agriculture et de l'Alimentation. (2018b). *Référentiel de contrôle des performances dans le cadre du service public d'enregistrement et de contrôle des performances des ruminants pour la filière « production de lait de brebis »* (Rapport technique). Consulté sur <https://info.agriculture.gouv.fr>
- Minvielle, F. (2010). *Principes d'amélioration génétique des animaux domestiques.* QUAE.
- Monniaux, D., Caraty, A., clement, F., Dalbies-Tran, R., Dupont, J., Fabre, S., Gerard, N., Mermillod, P., Monget, P., & Uzbekova, S. (2009, avril). Développement folliculaire ovarien et ovulation chez les mammifères. *INRAE Productions Animales*, *22*(2), 59–76. doi: 10.20870/productions-animales.2009.22.2.3335
- Monniaux, D., Clément, F., Dalbiès-Tran, R., Estienne, A., Fabre, S., Mansanet, C., & Monget,

- P. (2014, avril). The Ovarian Reserve of Primordial Follicles and the Dynamic Reserve of Antral Growing Follicles : What Is the Link?1. Biology of Reproduction, 90(4), 85, 1–11. doi: 10.1095/biolreprod.113.117077
- Monteagudo, L. V., Ferrer, L. M., Catalan-Insa, E., Savva, D., McGuffin, L. J., & Tejedor, M. T. (2015). In silico identification and three-dimensional modelling of the missense mutation in ADAMTS2 in a sheep flock with dermatosparaxis. Veterinary Dermatology, 26(1), 49–e16. doi: 10.1111/vde.12178
- Monteagudo, L. V., Tejedor, M. T., Ramos, J. J., Lacasta, D., & Ferrer, L. M. (2015, avril). Ovine congenital myotonia associated with a mutation in the muscle chloride channel gene. The Veterinary Journal, 204(1), 128–129. doi: 10.1016/j.tvjl.2015.01.014
- Moraes, J. C., Mattevi, M. S., & Ferreira, J. M. (1980, novembre). Chromosome studies in Brazilian rams. doi: 10.1136/vr.107.21.489
- Morohoshi, A., Miyata, H., Shimada, K., Nozawa, K., Matsumura, T., Yanase, R., Shiba, K., Inaba, K., & Ikawa, M. (2020, janvier). Nexin-Dynein regulatory complex component DRC7 but not FBXL13 is required for sperm flagellum formation and male fertility in mice. PLOS Genetics, 16(1), e1008585. doi: 10.1371/journal.pgen.1008585
- Moum, T., Olsaker, I., Hopp, P., Moldal, T., Valheim, M., Moum, T., & Benestad, S. L. . (2005). Polymorphisms at codons 141 and 154 in the ovine prion protein gene are associated with scrapie Nor98 cases. Journal of General Virology, 86(1), 231–235. doi: 10.1099/vir.0.80437-0
- Mömke, S., Kerkmann, A., Wöhlke, A., Ostmeier, M., Hewicker-Trautwein, M., Ganter, M., Kijas, J., Consortium, f. t. I. S., & Distl, O. (2011, mai). A Frameshift Mutation within LAMC2 Is Responsible for Herlitz Type Junctional Epidermolysis Bullosa (HJEB) in Black Headed Mutton Sheep. PLOS ONE, 6(5), e18943. doi: 10.1371/journal.pone.0018943
- Nezamzadeh, R., Seubert, A., Pohlenz, J., & Brenig, B. (2005). Identification of a mutation in the ovine uroporphyrinogen decarboxylase (UROD) gene associated with a type of porphyria. Animal Genetics, 36(4), 297–302. doi: 10.1111/j.1365-2052.2005.01301.x
- Ni, Z., Olsen, J. B., Guo, X., Zhong, G., Ruan, E. D., Marcon, E., Young, P., Guo, H., Li, J., Moffat, J., Emili, A., & Greenblatt, J. F. (2011, octobre). Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B. doi: 10.4161/trns.2.5.17803
- Nicholas, F. W., & Hobbs, M. (2014). Mutation discovery for Mendelian traits in non-laboratory animals : a review of achievements up to 2012. Animal Genetics, 45(2), 157–170. doi: 10.1111/age.12103
- Nicol, L., Bishop, S. C., Pong-Wong, R., Bendixen, C., Holm, L.-E., Rhind, S. M., & McNeilly, A. S. (2009, décembre). Homozygosity for a single base-pair mutation in the oocyte-specific GDF9 gene results in sterility in Thoka sheep. Reproduction, 138(6), 921–933. doi: 10.1530/REP-09-0193
- Nimmerjahn, F., & Ravetch, J. V. (2008, janvier). Fc receptors as regulators of immune responses. Nature Reviews Immunology, 8(1), 34–47. doi: 10.1038/nri2206
- Norris, B. J., & Whan, V. A. (2008, janvier). A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Research, 18(8), 1282–1293.

- doi: 10.1101/gr.072090.107
- Null, D., Hutchison, J., Bickhart, D., VanRaden, P., & Cole, J. (2017, juin). Discovery of a haplotype affecting fertility in Ayrshire dairy dattle and identification of a putative causal variant. Journal of Dairy Science, 100(Suppl. 2) : 199(abstr. 205).
- Olbrich, H., Cremers, C., Loges, N., Werner, C., Nielsen, K., Marthin, J., Philipsen, M., Wallmeier, J., Pennekamp, P., Menchen, T., Edelbusch, C., Dougherty, G., Schwartz, O., Thiele, H., Altmüller, J., Rommelmann, F., & Omran, H. (2015, octobre). Loss-of-Function GAS8 Mutations Cause Primary Ciliary Dyskinesia and Disrupt the Nexin-Dynein Regulatory Complex. The American Journal of Human Genetics, 97(4), 546–554. doi: 10.1016/j.ajhg.2015.08.012
- Old, R., Sewell, H., Norris, R., Steel, M., Joyce, C. R. B., & Old, R. (1993, janvier). HETEROZYGOTE ADVANTAGE : Why are some deleterious genes so common ? The Lancet, 341(8839), 214. doi: 10.1016/0140-6736(93)90076-S
- ONAB. (2021, décembre). Observatoire National des Anomalies Bovines. doi: 10.15454/BRKOV3
- Ott, T. L., & Gifford, C. A. (2010). Effects of Early Conceptus Signals on Circulating Immune Cells : Lessons from Domestic Ruminants. American Journal of Reproductive Immunology, 64(4), 245–254. doi: 10.1111/j.1600-0897.2010.00912.x
- O’Doherty, A. M., MacHugh, D. E., Spillane, C., & Magee, D. A. (2015). Genomic imprinting effects on complex traits in domesticated animal species. Frontiers in Genetics, 6.
- Pailhoux, E., Cribiu, E. P., Chaffaux, S., Darre, R., Fellous, M., & Cotinot, C. (1994, mars). Molecular analysis of 60,XX pseudohermaphrodite polled goats for the presence of SRY and ZFY genes. doi: 10.1530/jrf.0.1000491
- Pailhoux, E., Vigier, B., Chaffaux, S., Serval, N., Taourit, S., Furet, J.-P., Fellous, M., Grosclaude, F., Cribiu, E. P., Cotinot, C., & Vaiman, D. (2001, décembre). A 11.7-kb deletion triggers intersexuality and polledness in goats. Nature Genetics, 29(4), 453–458. doi: 10.1038/ng769
- Pal, A., Pal, A., Banerjee, S., Batabyal, S., & Chatterjee, P. N. (2019, mai). Mutation in Cytochrome B gene causes debility and adverse effects on health of sheep. Mitochondrion, 46, 393–404. doi: 10.1016/j.mito.2018.10.003
- Palhière, I., Gousseau, V., & Colleau, J.-J. (2022, décembre). La Sélection à Parenté Minimum Génomique : principes et résultats pour les deux races caprines laitières principales françaises. INRAE Productions Animales, 35(3), 171–186. doi: 10.20870/productions-animales.2022.35.3.7152
- Papasavva-Stylianou, P., Kleanthous, M., Toumazos, P., Mavrikiou, P., & Loucaides, P. (2007, mars). Novel polymorphisms at codons 146 and 151 in the prion protein gene of Cyprus goats, and their association with natural scrapie. The Veterinary Journal, 173(2), 459–462. doi: 10.1016/j.tvjl.2005.09.013
- Paris, J. M., Letko, A., Häfliger, I. M., Ammann, P., Flury, C., & Drögemüller, C. (2019). Identification of two TYRP1 loss-of-function alleles in Valais Red sheep. Animal Genetics, 50(6), 778–782. doi: 10.1111/age.12863
- Paris, J. M., Letko, A., Häfliger, I. M., Švara, T., Gombač, M., Klinc, P., Škibin, A., Pogorevc, E., & Drögemüller, C. (2020, janvier). A de novo variant in OTX2 in a lamb with otocephaly.

- Acta Veterinaria Scandinavica*, 62(1), 5. doi: 10.1186/s13028-020-0503-z
- Pausch, H., Schwarzenbacher, H., Burgstaller, J., Flisikowski, K., Wurmser, C., Jansen, S., Jung, S., Schnieke, A., Wittek, T., & Fries, R. (2015, avril). Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. *BMC Genomics*, 16(1), 312. doi: 10.1186/s12864-015-1483-7
- Pearce, P. D., Ansari, H. A., Maher, D. W., Malcolm, A. A., Stewart-Scott, I. A., & Broad, T. E. (1994). New Robertsonian translocation chromosomes in domestic sheep (*Ovis aries*). *Cytogenetic and Genome Research*, 67(2), 137–140. doi: 10.1159/000133814
- Perkins, N. J., & Schisterman, E. F. (2006). The Inconsistency of “Optimal” Cut-points Using Two ROC Based Criteria. , 11.
- Persuy, M.-A., Printz, C., Medrano, J. F., & Mercier, J.-C. (1999). A single nucleotide deletion resulting in a premature stop codon is associated with marked reduction of transcripts from a goat -casein null allele. *Animal Genetics*, 30(6), 444–451. doi: 10.1046/j.1365-2052.1999.00547.x
- Peters, H., Nefedov, M., Sarsero, J., Pitt, J., Fowler, K. J., Gazeas, S., Kahler, S. G., & Ioannou, P. A. (2003, décembre). A Knock-out Mouse Model for Methylmalonic Aciduria Resulting in Neonatal Lethality*. *Journal of Biological Chemistry*, 278(52), 52909–52913. doi: 10.1074/jbc.M310533200
- Petit, M., Fritz, S., Astruc, J.-M., Saintilan, R., & Moreno-Romieux, C. (2015, janvier). Detection of lethal mutations in sheep : research for homozygous chromosome segments with high density molecular information (54 000 SNPs). In 23. Plant and Animal Genome (PAG). San Diego, Californie, United States. Consulté le 2022-09-07, sur <https://hal.archives-ouvertes.fr/hal-01535402>
- Pincetic, A., Bournazos, S., DiLillo, D. J., Maamary, J., Wang, T. T., Dahan, R., Fiebiger, B.-M., & Ravetch, J. V. (2014, août). Type I and type II Fc receptors regulate innate and adaptive immunity. *Nature Immunology*, 15(8), 707–716. doi: 10.1038/ni.2939
- Popescu, C., & Boscher, J. (1972). [Not Available]. doi: 10.1186/1297-9686-4-3-355
- Popescu, C. P. (1990). Conséquences des anomalies de la structure chromosomique chez les animaux domestiques. *Reproduction Nutrition Development*, 30(Suppl. 1), 105s–116s. doi: 10.1051/rnd:19900711
- Popescu, C. P., & Tixier, M. (1984). The frequency of chromosome abnormalities in farm animals and their economic consequences. *Annales De Genetique*, 27(2), 69–72.
- Popova, V. V., Brechalov, A. V., Georgieva, S. G., & Kopytova, D. V. (2018, décembre). Nonreplicative functions of the origin recognition complex. *Nucleus*, 9(1), 460–473. doi: 10.1080/19491034.2018.1516484
- Poprzeczko, M., Bicka, M., Farahat, H., Bazan, R., Osinka, A., Fabczak, H., Joachimiak, E., & Wloga, D. (2019, décembre). Rare Human Diseases : Model Organisms in Deciphering the Molecular Basis of Primary Ciliary Dyskinesia. *Cells*, 8(12), 1614. doi: 10.3390/cells8121614
- Porada, C. D., Sanada, C., Long, C. R., Wood, J. A., Desai, J., Frederick, N., Millsap, L., Bormann, C., Menges, S. L., Hanna, C., Flores-Foxworth, G., Shin, T., Westhusin, M. E., Liu, W., Glimp, H., Zanjani, E. D., Lozier, J. N., Pliska, V., Stranzinger, G., Joerg, H.,

- Kraemer, D. C., & Almeida-Porada, G. (2010). Clinical and molecular characterization of a re-established line of sheep exhibiting hemophilia A. *Journal of Thrombosis and Haemostasis*, *8*(2), 276–285. doi: 10.1111/j.1538-7836.2009.03697.x
- Posbergh, C. J., Staiger, E. A., & Huson, H. J. (2020, juin). A Stop-Gain Mutation within MLPH Is Responsible for the Lilac Dilution Observed in Jacob Sheep. *Genes*, *11*(6), 618. doi: 10.3390/genes11060618
- Prache, S., Caillat, H., & Lagriffoul, G. (2018). Diversité dans la filière petits ruminants : une source de résilience ? *Innovations Agronomiques* 68, 171-191. doi: 10.15454/LB7VKP
- Pryce, J. E., Royal, M. D., Garnsworthy, P. C., & Mao, I. L. (2004, mars). Fertility in the high-producing dairy cow. *Livestock Production Science*, *86*(1), 125–135. doi: 10.1016/S0301-6226(03)00145-3
- Pérez, M. J., Leroux, C., Bonastre, A. S., & Martin, P. (1994, septembre). Occurrence of a LINE sequence in the 3' UTR of the goat s1-casein E-encoding allele associated with reduced protein synthesis level. *Gene*, *147*(2), 179–187. doi: 10.1016/0378-1119(94)90063-9
- Quintana, D. G., Thome, K. C., Hou, Z.-h., Ligon, A. H., Morton, C. C., & Dutta, A. (1998, octobre). ORC5L, a New Member of the Human Origin Recognition Complex, Is Deleted in Uterine Leiomyomas and Malignant Myeloid Diseases*. *Journal of Biological Chemistry*, *273*(42), 27137–27145. doi: 10.1074/jbc.273.42.27137
- Raadsma, H. W., Jonas, E., Fleet, M. R., Fullard, K., Gongora, J., Cavanagh, C. R., Tammen, I., & Thomson, P. C. (2013). QTL and association analysis for skin and fibre pigmentation in sheep provides evidence of a major causative mutation and epistatic effects. *Animal Genetics*, *44*(5), 547–559. doi: 10.1111/age.12033
- Ramunno, L., Longobardi, E., Pappalardo, M., Rando, A., Di Gregorio, P., Cosenza, G., Mariani, P., Pastore, N., & Masina, P. (2001). An allele associated with a non-detectable amount of s2 casein in goat milk. *Animal Genetics*, *32*(1), 19–26. doi: 10.1046/j.1365-2052.2001.00710.x
- Raoul, J., Elsen, J., Rodríguez-Ramilo, S., & Palhière, I. (2022, décembre). Management of undesired alleles at multiple loci for a goat breeding program. In *Proceeding of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)* (pp. 1833–1836). Rotterdam, the Netherlands : Wageningen Academic Publishers. doi: 10.3920/978-90-8686-940-4_440
- Reicher, S., Seroussi, E., & Gootwine, E. (2010, février). A mutation in gene CNGA3 is associated with day blindness in sheep. *Genomics*, *95*(2), 101–104. doi: 10.1016/j.ygeno.2009.10.003
- Revell, L. J. (2019). learnPopGen : An R package for population genetic simulation and numerical analysis. *Ecology and Evolution*, *9*(14), 7896–7902. doi: https://doi.org/10.1002/ece3.5412
- Reyes-Pardo, H., Barbosa-Camacho, A. A., Pérez-Mejía, A. E., Lara-Chacón, B., Salas-Estrada, L. A., Robledo-Rivera, A. Y., Montero-Morán, G. M., Lara-González, S., Calera, M. R., & Sánchez-Olea, R. (2012, octobre). A nuclear export sequence in GPN-loop GTPase 1, an essential protein for nuclear targeting of RNA polymerase II, is necessary and sufficient for nuclear export. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1823*(10), 1756–1766. doi: 10.1016/j.bbamcr.2012.07.001
- Rickard, J. P., Ryan, G., Hall, E., de Graaf, S. P., & Hermes, R. (2017, août). Using transrectal

- ultrasound to examine the effect of exogenous progesterone on early embryonic loss in sheep. *PLOS ONE*, *12*(8), e0183659. doi: 10.1371/journal.pone.0183659
- Riquet, J., Rousseau, S., Mercat, M. J., & Pailhoux, E. (2016, décembre). Hereditary congenital diseases in pigs. *INRAE Productions Animales*, *29*(5), 329–338. doi: 10.20870/productions-animales.2016.29.5.3001
- Rochus, C. M., Tortereau, F., Plisson-Petit, F., Restoux, G., Moreno-Romieux, C., Tosser-Klopp, G., & Servin, B. (2018, janvier). Revealing the selection history of adaptive loci using genome-wide scans for selection : an example from domestic sheep. *BMC Genomics*, *19*(1), 71. doi: 10.1186/s12864-018-4447-x
- Rochus, C. M., Westberg Sunesson, K., Jonas, E., Mikko, S., & Johansson, A. M. (2019). Mutations in ASIP and MC1R : dominant black and recessive black alleles segregate in native Swedish sheep populations. *Animal Genetics*, *50*(6), 712–717. doi: 10.1111/age.12837
- Rupp, R., Lagriffoul, G., Astruc, J., & Barillet, F. (2003, avril). Genetic Parameters for Milk Somatic Cell Scores and Relationships with Production Traits in French Lacaune Dairy Sheep. *Journal of Dairy Science*, *86*(4), 1476–1481. doi: 10.3168/jds.S0022-0302(03)73732-1
- Rupp, R., Mucha, S., Larroque, H., McEwan, J., & Conington, J. (2016, janvier). Genomic application in sheep and goat breeding. *Animal Frontiers*, *6*(1), 39–44. doi: 10.2527/af.2016-0006
- Rupp, R., Senin, P., Sarry, J., Allain, C., Tasca, C., Ligat, L., Portes, D., Woloszyn, F., Bouchez, O., Tabouret, G., Lebastard, M., Caubet, C., Foucras, G., & Tosser-Klopp, G. (2015, décembre). A Point Mutation in Suppressor of Cytokine Signalling 2 (Socs2) Increases the Susceptibility to Inflammation of the Mammary Gland while Associated with Higher Body Weight and Size and Higher Milk Production in a Sheep Model. *PLoS Genetics*, *11*(12), e1005629. doi: 10.1371/journal.pgen.1005629
- Sahana, G., Nielsen, U. S., Aamand, G. P., Lund, M. S., & Guldbbrandtsen, B. (2013, décembre). Novel Harmful Recessive Haplotypes Identified for Fertility Traits in Nordic Holstein Cattle. *PLoS ONE*, *8*(12), e82909. doi: 10.1371/journal.pone.0082909
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977, décembre). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463–5467. doi: 10.1073/pnas.74.12.5463
- Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014, juin). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, *15*(1), 478. doi: 10.1186/1471-2164-15-478
- Schwarzenbacher, H., Burgstaller, J., Seefried, F. R., Wurmser, C., Hilbe, M., Jung, S., Fuerst, C., Dinholpl, N., Weissenböck, H., Fuerst-Waltl, B., Dolezal, M., Winkler, R., Grueter, O., Bleul, U., Wittek, T., Fries, R., & Pausch, H. (2016, décembre). A missense mutation in TUBD1 is associated with high juvenile mortality in Braunvieh and Fleckvieh cattle. *BMC Genomics*, *17*(1), 400. doi: 10.1186/s12864-016-2742-y
- Schübeler, D. (2015, janvier). Function and information content of DNA methylation. *Nature*, *517*(7534), 321–326. doi: 10.1038/nature14192
- Shuster, D. E., Kehrl, M. E., Ackermann, M. R., & Gilbert, R. O. (1992, octobre). Identification

- and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. Proceedings of the National Academy of Sciences, 89(19), 9225–9229. doi: 10.1073/pnas.89.19.9225
- Silva, B. D. M., Castro, E. A., Souza, C. J. H., Paiva, S. R., Sartori, R., Franco, M. M., Azevedo, H. C., Silva, T. a. S. N., Vieira, A. M. C., Neves, J. P., & Melo, E. O. (2011). A new polymorphism in the Growth and Differentiation Factor 9 (GDF9) gene is associated with increased ovulation rate and prolificacy in homozygous sheep. Animal Genetics, 42(1), 89–92. doi: 10.1111/j.1365-2052.2010.02078.x
- Simon, R., Lischer, H. E. L., Pieńkowska-Schelling, A., Keller, I., Häfliger, I. M., Letko, A., Schelling, C., Lühken, G., & Drögemüller, C. (2020). New genomic features of the polled intersex syndrome variant in goats unraveled by long-read whole-genome sequencing. Animal Genetics, 51(3), 439–448. doi: 10.1111/age.12918
- Smit, M. A., Shay, T. L., Beever, J. E., Notter, D. R., & Cockett, N. E. (2002). Identification of an agouti-like locus in sheep. Animal Genetics, 33(5), 383–385. doi: 10.1046/j.1365-2052.2002.00896_5.x
- Souza, C. J. H., McNeilly, A. S., Benavides, M. V., Melo, E. O., & Moraes, J. C. F. (2014). Mutation in the protease cleavage site of GDF9 increases ovulation rate and litter size in heterozygous ewes and causes infertility in homozygous ewes. Animal Genetics, 45(5), 732–739. doi: 10.1111/age.12190
- Spencer, T. E., Johnson, G. A., Bazer, F. W., Burghardt, R. C., Palmarini, M., Spencer, T. E., Johnson, G. A., Bazer, F. W., Burghardt, R. C., & Palmarini, M. (2006, décembre). Pregnancy recognition and conceptus implantation in domestic ruminants : roles of progesterone, interferons and endogenous retroviruses. Reproduction, Fertility and Development, 19(1), 65–78. doi: 10.1071/RD06102
- Spencer, T. E., Sandra, O., & Wolf, E. (2008, février). Genes involved in conceptus–endometrial interactions in ruminants : insights from reductionism and thoughts on holistic approaches. Reproduction, 135(2), 165–179. doi: 10.1530/REP-07-0327
- Spinner, N. B., & Conlin, L. K. (2014). Mosaicism and clinical genetics. American Journal of Medical Genetics Part C : Seminars in Medical Genetics, 166(4), 397–405. doi: 10.1002/ajmg.c.31421
- Statello, L., Guo, C.-J., Chen, L.-L., & Huarte, M. (2021, février). Gene regulation by long non-coding RNAs and its biological functions. Nature Reviews Molecular Cell Biology, 22(2), 96–118. doi: 10.1038/s41580-020-00315-9
- Stayner, C., Poole, C. A., McGlashan, S. R., Pilanthanonond, M., Brauning, R., Markie, D., Lett, B., Slobbe, L., Chae, A., Johnstone, A. C., Jensen, C. G., McEwan, J. C., Dittmer, K., Parker, K., Wiles, A., Blackburne, W., Leichter, A., Leask, M., Pinnapureddy, A., Jennings, M., Horsfield, J. A., Walker, R. J., & Eccles, M. R. (2017, mai). An ovine hepatorenal fibrocystic model of a Meckel-like syndrome associated with dysmorphic primary cilia and TMEM67 mutations. Scientific Reports, 7(1), 1601. doi: 10.1038/s41598-017-01519-4
- Suárez-Vega, A., Gutiérrez-Gil, B., Benavides, J., Perez, V., Tosser-Klopp, G., Klopp, C., Kennel, S. J., & Arranz, J. J. (2015, mai). Combining GWAS and RNA-Seq Approaches for

- Detection of the Causal Mutation for Hereditary Junctional Epidermolysis Bullosa in Sheep. *PLOS ONE*, 10(5), e0126416. doi: 10.1371/journal.pone.0126416
- Suárez-Vega, A., Gutiérrez-Gil, B., Cuchillo-Ibáñez, I., Sáez-Valero, J., Pérez, V., García-Gámez, E., Benavides, J., & Arranz, J. J. (2013, novembre). Identification of a 31-bp Deletion in the RELN Gene Causing Lissencephaly with Cerebellar Hypoplasia in Sheep. *PLOS ONE*, 8(11), e81072. doi: 10.1371/journal.pone.0081072
- Swets, J. A. (1988, juin). Measuring the Accuracy of Diagnostic Systems. *Science*, 240(4857), 1285–1293. doi: 10.1126/science.3287615
- Switoński, M., & Stranzinger, G. (1998, décembre). Studies of synaptonemal complexes in farm mammals—a review. doi: 10.1093/jhered/89.6.473
- Tagu, D., Jaubert-Possamai, S., & Méreau, A. (2018). *Principes des techniques de biologie moléculaire et génomique : 3e édition revue et augmentée*. Consulté le 2022-09-06, sur <https://hal.inrae.fr/hal-03318953> (Pages : 312 p.)
- Takebayashi, S. S. G., & Jorge, W. (1986). Testicular hypoplasia in a horned goat with 61, XXY/60, XY karyotype. doi: 10.1266/jjg.61.177
- Tammen, I., Houweling, P. J., Frugier, T., Mitchell, N. L., Kay, G. W., Cavanagh, J. A. L., Cook, R. W., Raadsma, H. W., & Palmer, D. N. (2006, octobre). A missense mutation (c.184C>T) in ovine CLN6 causes neuronal ceroid lipofuscinosis in Merino sheep whereas affected South Hampshire sheep have reduced levels of CLN6 mRNA. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1762(10), 898–905. doi: 10.1016/j.bbadis.2006.09.004
- Tan, P., Allen, J. G., Wilton, S. D., Akkari, P. A., Huxtable, C. R., & Laing, N. G. (1997, juillet). A splice-site mutation causing ovine McArdle's disease. *Neuromuscular Disorders*, 7(5), 336–342. doi: 10.1016/S0960-8966(97)00062-X
- Taylor Clelland, C. L., Levy, B., McKie, J. M., Duncan, A. M., Hirschhorn, K., & Bancroft, C. (2000, août). Cloning and characterization of human PREB; a gene that maps to a genomic region associated with trisomy 2p syndrome. *Mammalian Genome*, 11(8), 675–681. doi: 10.1007/s003350010142
- Teissier, M., Larroque, H., & Robert-Granie, C. (2019, avril). Accuracy of genomic evaluation with weighted single-step genomic best linear unbiased prediction for milk production traits, udder type traits, and somatic cell scores in French dairy goats. *Journal of Dairy Science*, 102(4), 3142–3154. doi: 10.3168/jds.2018-15650
- Thomsen, B., Horn, P., Panitz, F., Bendixen, E., Petersen, A. H., Holm, L.-E., Nielsen, V. H., Agerholm, J. S., Arnbjerg, J., & Bendixen, C. (2006, janvier). A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Research*, 16(1), 97–105. doi: 10.1101/gr.3690506
- Thorvaldssdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013, mars). Integrative Genomics Viewer (IGV) : high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. doi: 10.1093/bib/bbs017
- Todd, E. T., Thomson, P. C., Hamilton, N. A., Ang, R. A., Lindgren, G., Viklund, , Eriksson, S., Mikko, S., Strand, E., & Velie, B. D. (2020, août). A genome-wide scan for candidate lethal variants in Thoroughbred horses. *Scientific Reports*, 10(1), 13153. doi: 10.1038/

- s41598-020-68946-8
- Torres, P. A., Zeng, B. J., Porter, B. F., Alroy, J., Horak, F., Horak, J., & Kolodny, E. H. (2010, décembre). Tay-Sachs disease in Jacob sheep. *Molecular Genetics and Metabolism*, *101*(4), 357–363. doi: 10.1016/j.ymgme.2010.08.006
- Tortereau, F. (2016, mai). Conséquences de la sélection pour la résistance à la tremblante sur la diversité génétique du génome ovin. In *Séminaire Ressources Génétiques Animales* (p. 20 p.). Paris, France. Consulté le 2022-08-17, sur <https://hal.inrae.fr/hal-02797279>
- Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., Donnadiu-Tonon, C., Eggen, A., Heuven, H. C. M., Jamli, S., Jiken, A. J., Klopp, C., Lawley, C. T., McEwan, J., Martin, P., Moreno, C. R., Mulsant, P., Nabihoudine, I., Pailhoux, E., Palhière, I., Rupp, R., Sarry, J., Sayre, B. L., Tircazes, A., Wang, J., Wang, W., Zhang, W., & Consortium, a. t. I. G. G. (2014, janvier). Design and Characterization of a 52K SNP Chip for Goats. *PLOS ONE*, *9*(1), e86227. doi: 10.1371/journal.pone.0086227
- Tritschler, F., Eulalio, A., Truffault, V., Hartmann, M. D., Helms, S., Schmidt, S., Coles, M., Izaurralde, E., & Weichenrieder, O. (2007, décembre). A Divergent Sm Fold in EDC3 Proteins Mediates DCP1 Binding and P-Body Targeting. *Molecular and Cellular Biology*, *27*(24), 8600–8611. doi: 10.1128/MCB.01506-07
- Tsuchiya, Y., Pham, U., Hu, W., Ohnuma, S.-i., & Gout, I. (2014, mai). Changes in Acetyl CoA Levels during the Early Embryonic Development of *Xenopus laevis*. *PLOS ONE*, *9*(5), e97693. doi: 10.1371/journal.pone.0097693
- Tyynela, J. (2000, juin). A mutation in the ovine cathepsin D gene causes a congenital lysosomal storage disease with profound neurodegeneration. *The EMBO Journal*, *19*(12), 2786–2792. doi: 10.1093/emboj/19.12.2786
- Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021, août). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), 1–21. doi: 10.1038/s43586-021-00056-9
- Vaccari, G., Bari, M. A. D., Morelli, L., Nonno, R., Chiappini, B., Antonucci, G., Marcon, S., Esposito, E., Fazzi, P., Palazzini, N., Troiano, P., Petrella, A., Guardo, G. D., & Agrimi, U. (2006). Identification of an allelic variant of the goat PrP gene associated with resistance to scrapie. *Journal of General Virology*, *87*(5), 1395–1402. doi: 10.1099/vir.0.81485-0
- Vaiman, D., Koutita, O., Oustry, A., Elsen, J. M., Manfredi, E., Fellous, M., & Cribiu, E. P. (1996, février). Genetic mapping of the autosomal region involved in XX sex-reversal and horn development in goats. doi: 10.1007/s003359900033
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018, septembre). The Third Revolution in Sequencing Technology. *Trends in Genetics*, *34*(9), 666–681. doi: 10.1016/j.tig.2018.05.008
- VanRaden, P. M. (2020, juin). Symposium review : How to implement genomic selection. *Journal of Dairy Science*, *103*(6), 5291–5301. doi: 10.3168/jds.2019-17684
- VanRaden, P. M., Olson, K. M., Null, D. J., & Hutchison, J. L. (2011, décembre). Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy Science*, *94*(12), 6153–6161. doi: 10.3168/jds.2011-4624

- van Steenbeek, F. G., Hytönen, M. K., Leegwater, P. a. J., & Lohi, H. (2016). The canine era : the rise of a biomedical model. *Animal Genetics*, *47*(5), 519–527. doi: 10.1111/age.12460
- Veenboer, G. J., & de Vijlder, J. J. (1993, janvier). Molecular basis of the thyroglobulin synthesis defect in Dutch goats. *Endocrinology*, *132*(1), 377–381. doi: 10.1210/endo.132.1.8380383
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V. D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R.-R., Ke, Z., Ketchum, K. A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z. Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S. C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigó, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen,

- M., Wu, D., Wu, M., Xia, A., Zandieh, A., & Zhu, X. (2001, février). The Sequence of the Human Genome. *Science*, *291*(5507), 1304–1351. doi: 10.1126/science.1058040
- Vignal, A. (2011, septembre). Etat actuel du séquençage et de la connaissance du génome des espèces animales. *INRAE Productions Animales*, *24*(4), 387–404. doi: 10.20870/productions-animales.2011.24.4.3272
- Våge, D. I., & Boman, I. A. (2010, février). A nonsense mutation in the beta-carotene oxygenase 2 (BCO2) gene is tightly associated with accumulation of carotenoids in adipose tissue in sheep (*Ovis aries*). *BMC Genetics*, *11*(1), 10. doi: 10.1186/1471-2156-11-10
- Våge, D. I., Husdal, M., Kent, M. P., Klemetsdal, G., & Boman, I. A. (2013, janvier). A missense mutation in growth differentiation factor 9 (GDF9) is strongly associated with litter size in sheep. *BMC Genetics*, *14*(1), 1. doi: 10.1186/1471-2156-14-1
- Våge, D. I., Klungland, H., Lu, D., & Cone, R. D. (1999, janvier). Molecular and pharmacological characterization of dominant black coat color in sheep. *Mammalian Genome*, *10*(1), 39–43. doi: 10.1007/s003359900939
- Walker, K., Holler, L., & Beever, J. E. (2012). Ovine GM1-gangliosidosis is caused by a mutation in GLB1. *Plant and Animal Genome*.
- Weil, D., Piton, A., Lessel, D., & Standart, N. (2020, mai). Mutations in genes encoding regulators of mRNA decapping and translation initiation : links to intellectual disability. *Biochemical Society Transactions*, *48*(3), 1199–1211. doi: 10.1042/BST20200109
- White, S. N., Mousel, M. R., Reynolds, J. O., Herrmann-Hoesing, L. M., & Knowles, D. P. (2014). Deletion variant near ZNF389 is associated with control of ovine lentivirus in multiple sheep flocks. *Animal Genetics*, *45*(2), 297–300. doi: 10.1111/age.12107
- Wiedemar, N., & Drögemüller, C. (2015). A 1.8-kb insertion in the 3-UTR of RXFP2 is associated with polledness in sheep. *Animal Genetics*, *46*(4), 457–461. doi: https://doi.org/10.1111/age.12309
- Wilmot, I., Sales, D. I., & Ashworth, C. J. (1986, mars). Maternal and embryonic factors associated with prenatal loss in mammals. *Reproduction*, *76*(2), 851–864. doi: 10.1530/jrf.0.0760851
- Wilson, T., Wu, X.-Y., Juengel, J. L., Ross, I. K., Lumsden, J. M., Lord, E. A., Dodds, K. G., Walling, G. A., McEwan, J. C., O’Connell, A. R., McNatty, K. P., & Montgomery, G. W. (2001, avril). 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 Cells1. *Biology of Reproduction*, *64*(4), 1225–1235. doi: 10.1095/biolreprod64.4.1225
- Wirschell, M., Olbrich, H., Werner, C., Tritschler, D., Bower, R., Sale, W. S., Loges, N. T., Pennekamp, P., Lindberg, S., Stenram, U., Carlén, B., Horak, E., Köhler, G., Nürnberg, P., Nürnberg, G., Porter, M. E., & Omran, H. (2013, mars). The nexin-dynein regulatory complex subunit DRC1 is essential for motile cilia function in algae and humans. *Nature Genetics*, *45*(3), 262–268. doi: 10.1038/ng.2533
- Woolley, S. A., Hayes, S. E., Shariflou, M. R., Nicholas, F. W., Willet, C. E., O’Rourke, B. A., & Tammen, I. (2020, septembre). Molecular basis of a new ovine model for human 3M syndrome-2. *BMC Genetics*, *21*(1), 106. doi: 10.1186/s12863-020-00913-8

- Wu, J., Li, Y., Rendahl, A., & Bhargava, M. (2022). Novel Human FCGR1A Variants Affect CD64 Functions and Are Risk Factors for Sarcoidosis. *Frontiers in Immunology*, *13*.
- Xu, X., Ectors, F., Davis, E. E., Pirottin, D., Cheng, H., Farnir, F., Hadfield, T., Cockett, N., Charlier, C., Georges, M., & Takeda, H. (2015, octobre). Ectopic Expression of Retrotransposon-Derived PEG11/RTL1 Contributes to the Callipyge Muscular Hypertrophy. *PLOS ONE*, *10*(10), e0140594. doi: 10.1371/journal.pone.0140594
- Yankey, S. J., Hicks, B. A., Carnahan, K. G., Assiri, A. M., Sinor, S. J., Kodali, K., Stellflug, J. N., & Ott, T. L. (2001, août). Expression of the antiviral protein Mx in peripheral blood mononuclear cells of pregnant and bred, non-pregnant ewes. *Journal of Endocrinology*, *170*(2), R7–11. doi: 10.1677/joe.0.170r007
- Ytournal, F., Gilbert, H., & Boichard, D. (2008, juin). Comment affiner la localisation d'un QTL ? *INRAE Productions Animales*, *21*(2), 147–158. doi: 10.20870/productions-animales.2008.21.2.3388
- Zartman, D. L., Hinesley, L. L., & Gnatkowski, M. W. (1981). A 53, X female sheep (*Ovis aries*). *Cytogenetic and Genome Research*, *30*(1), 54–58. doi: 10.1159/000131589
- Zhang, Y., Wang, K., Yang, K., Shi, Y., & Hong, J. (2021, octobre). Insight into the interaction between the RNA helicase CGH-1 and EDC-3 and its implications. *Scientific Reports*, *11*(1), 20359. doi: 10.1038/s41598-021-99919-0
- Zhao, X., Dittmer, K. E., Blair, H. T., Thompson, K. G., Rothschild, M. F., & Garrick, D. J. (2011, juillet). A Novel Nonsense Mutation in the DMP1 Gene Identified by a Genome-Wide Association Study Is Responsible for Inherited Rickets in Corriedale Sheep. *PLOS ONE*, *6*(7), e21739. doi: 10.1371/journal.pone.0021739
- Zhao, X., Onteru, S. K., Dittmer, K. E., Parton, K., Blair, H. T., Rothschild, M. F., & Garrick, D. J. (2012, septembre). A missense mutation in AGTPBP1 was identified in sheep with a lower motor neuron disease. *Heredity*, *109*(3), 156–162. doi: 10.1038/hdy.2012.23
- Zhao, X., Onteru, S. K., Piripi, S., Thompson, K. G., Blair, H. T., Garrick, D. J., & Rothschild, M. F. (2012). In a shake of a lamb's tail : using genomics to unravel a cause of chondrodysplasia in Texel sheep. *Animal Genetics*, *43*(s1), 9–18. doi: 10.1111/j.1365-2052.2011.02304.x
- Zheng, Z., Wang, X., Li, M., Li, Y., Yang, Z., Wang, X., Pan, X., Gong, M., Zhang, Y., Guo, Y., Wang, Y., Liu, J., Cai, Y., Chen, Q., Okpeku, M., Colli, L., Cai, D., Wang, K., Huang, S., Sonstegard, T. S., Esmailzadeh, A., Zhang, W., Zhang, T., Xu, Y., Xu, N., Yang, Y., Han, J., Chen, L., Lesur, J., Daly, K. G., Bradley, D. G., Heller, R., Zhang, G., Wang, W., Chen, Y., & Jiang, Y. (2020, mai). The origin of domestication genes in goats. *Science Advances*, *6*(21), eaaz5216. doi: 10.1126/sciadv.aaz5216
- Zhou, H., Hickford, J. G. H., & Fang, Q. (2012). A premature stop codon in the ADAMTS2 gene is likely to be responsible for dermatosparaxis in Dorper sheep. *Animal Genetics*, *43*(4), 471–473. doi: 10.1111/j.1365-2052.2011.02275.x
- Zhou, H., Zhang, Y., Suter, R., Gong, H., Fang, Q., Zhou, P., & Hickford, J. G. H. (2017). A nucleotide substitution in exon 8 of the glucosylceramidase beta gene is associated with Gaucher disease in sheep. *Animal Genetics*, *48*(6), 733–734. doi: 10.1111/age.12613
- Zhou, X., Carbonetto, P., & Stephens, M. (2013, février). Polygenic Modeling with Bayesian Sparse

- Linear Mixed Models. PLOS Genetics, 9(2), e1003264. doi: 10.1371/journal.pgen.1003264
- Zuo, Y., & Deng, G.-M. (2021). Fc Gamma Receptors as Regulators of Bone Destruction in Inflammatory Arthritis. Frontiers in Immunology, 12.

Détection de mutations homozygotes létales chez les petits ruminants

Résumé

Les anomalies génétiques résultent de mutations naturelles uniques de l'ADN. Elles sont inévitables, rares, héréditaires, affectant la morphologie, la santé et la viabilité. Ce travail de thèse, mené chez le mouton, s'intéresse à une catégorie de ces mutations appelées "embryonnaire létale à transmission récessive". Ces mutations induisent la mortalité de l'individu homozygote, qui porte deux copies de ces mutations, au cours de la gestation ou autour de la naissance. Pour les découvrir, cette thèse s'est appuyée sur des milliers de données de génotypages issues des programmes de sélection génomique et sur la capacité de séquencer des génomes complets. Ainsi, 9 mutations délétères dans 9 gènes différents causant des anomalies génétiques ont été identifiées, dont 3 sont létales pour les agneaux. La gestion génétique raisonnée de ces anomalies découvertes permettra de limiter la mortalité des agneaux, et plus généralement d'améliorer la santé et le bien-être des animaux.

Mots-clés : Petits ruminants, génomique, anomalies génétiques, mutations délétères, génétique inverse.

Detection of homozygous lethal mutations in small ruminants

Abstract

Genetic defects result from unique natural DNA mutations. They are inevitable, rare, hereditary, affecting morphology, health and viability. This thesis, conducted in sheep, focuses on a category of mutations called "recessive lethal embryonic mutations". These mutations induce the mortality of the homozygous individual, carrying two copies of these mutations, during gestation or around birth. To discover them, this thesis relied on thousands of genotyping data from genomic breeding programs and on the ability to sequence whole genomes. Thus, 9 deleterious mutations in 9 different genes causing genetic defects were identified, 3 of which are lethal to lambs. The rational genetic management of these discovered defects will limit the mortality of lambs, and more generally improve animal health and welfare.

Keywords : Small ruminants, genomics, genetic defects, deleterious mutations, reverse genetics.