

Exploration génomique de la mortalité juvénile des bovins

Genomic exploration of juvenile mortality in dairy cattle farming

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n° 581 : Agriculture, alimentation, biologie,
environnement, santé (ABIES)

Spécialité de doctorat : Génétique Animale

Graduate School : Biosphera. Référent : AgroParisTech

Thèse préparée dans l'UMR **GABI** (Université Paris-Saclay, INRAE, AgroParisTech),
sous la direction de **Didier BOICHARD**, Directeur de recherche,
la co-direction de **Aurélien CAPITAN**, Directeur de recherche,
et le co-encadrement de **Sophie MATTALIA**, Cheffe de service IDELE.

Thèse soutenue à Paris-Saclay, le 17 juin 2024, par

Florian BESNARD

Composition du Jury

Membres du jury avec voix délibérative

Xavier ROGNON Professeur, AgroParisTech (Université Paris-Saclay)	Président
Hélène GILBERT Directrice de recherche, INRAE (INP Toulouse)	Rapporteur & Examinatrice
Marie ABITBOL Professeure, VetAgro Sup	Rapporteur & Examinatrice
Xavier MONTAGUTELLI Directeur de recherche, Institut Pasteur (Université Paris-Saclay)	Examineur
Daniel VAIMAN Directeur de recherche, INSERM (Université Paris-Cité)	Examineur

Titre : Exploration génomique de la mortalité juvénile des bovins

Mots clés : Mortalité -Génomique - Génétique - Veaux - Bovins - Anomalies

Résumé : Dans le secteur de l'élevage bovin laitier, la mortalité des jeunes animaux représente un problème important, avec environ 10% des veaux mourant de manière naturelle au cours de leur première année. Ce phénomène est complexe, influencé par divers facteurs environnementaux et génétiques, ce qui le rend difficile à mesurer à grande échelle.

L'Observatoire National des Anomalies Bovines (ONAB) en France a permis d'identifier de nombreuses mutations génétiques impliquées dans des anomalies, bien que ces découvertes ne représentent qu'une petite partie des morts observées.

Depuis la fin des années 2000, l'accumulation de vastes bases de données sur le pedigree, le phénotype et le génotype bovins offre de nouvelles opportunités. Ces données, bien que parfois peu précises dans leur description des historiques de vie des animaux, offrent un volume considérable pour mener des analyses approfondies.

Dans le cadre de ma thèse, j'ai développé plusieurs approches exploitant ces vastes ensembles de données pour identifier des *loci* impliqués dans la mortalité avec des déterminismes génétiques dominants, récessifs et parfois avec une pénétrance incomplète. Ces *loci*, non remontés à l'ONAB en raison de l'absence de signes cliniques caractéristiques, comprennent notamment les gènes *ITGB7*, *NOA1* et *RFC5*, associés respectivement aux anomalies génétiques BLIRD, MTCP et à l'alopécie des extrémités. Ces découvertes ont été confirmées par des observations phénotypiques fines. De plus, j'ai développé des évaluations génomiques pilotes sur la mortalité.

L'intégration de ces travaux dans les programmes de sélection promet de réduire la mortalité en élevage. Par ailleurs, la découverte de mutations délétères affectant des gènes peu connus ouvre des perspectives prometteuses pour la recherche fondamentale.

Title : Genomic exploration of juvenile mortality in dairy cattle farming

Keywords : Mortality -Genomics - Genetics - Calves - Cattle - Abnormalities

Abstract : Juvenile animal mortality is a major problem in the dairy industry, with approximately 10% of calves dying naturally in their first year. This is a complex phenomenon influenced by a variety of environmental and genetic factors, making it difficult to measure on a large scale.

The French "Observatoire National des Anomalies Bovines" (ONAB) has made it possible to identify numerous genetic mutations associated with abnormalities, although these discoveries represent only a small proportion of observed deaths.

Since the late 2000s, the accumulation of vast databases of bovine pedigrees, phenotypes and genotypes has opened up new possibilities. These data, although sometimes not very precise in their description of the animal's life history, offer a considerable volume for in-depth analysis.

As part of my Ph.D. thesis, I developed several approaches using these large datasets to identify *loci* involved in mortality with dominant, recessive and sometimes incomplete penetrance genetic determinism. These *loci*, which cannot be tracked with ONAB due to the absence of characteristic clinical signs, include mutations in the *ITGB7*, *NOA1*, and *RFC5* genes, which are associated with BLIRD, MTCP, and limb alopecia genetic abnormalities, respectively. These findings were supported by detailed phenotypic observations. In addition, I developed pilot genomic evaluations of mortality.

The integration of this work into selection programs promises to reduce mortality in breeding. In addition, the discovery of deleterious mutations in little known genes opens up promising prospects for basic research.

Remerciements

Je tiens à remercier mes encadrants de thèse qui m'ont fait découvrir le monde passionnant de la recherche, Aurélien, Didier et Sophie. Vous êtes tous les trois une source d'inspiration pour moi tant par vos connaissances sans fond que par votre rigueur et votre passion pour la recherche. J'apprécie tout le temps que vous avez passé à être pédagogue avec moi et j'espère pouvoir mettre en œuvre ces connaissances dans les prochaines années. Merci à Pierre Larraufie et l'Ecole Doctorale pour l'implication dans ce projet et les différentes formations proposées.

J'aimerais remercier mes co-bureaux Sébastien Taussat et Jeanlin Jourdain pour nos discussions et pour m'avoir fait découvrir une filière bovine dont je découvre encore les particularités.

Tous les doctorants avec qui j'ai pu discuter, que ce soit pour se plaindre ensemble de « reviews » trop longues ou partiales, ou pour échanger sur différents loisirs et sujets de recherche.

Merci Mekki, Thierry, Bea, Clémentine, Stéphanie, Romain, Marie-Pierre, Hélène, Sophie et tous les autres d'avoir pris du temps pour m'aider et pour partager avec moi vos connaissances.

Merci à toute ma famille, ils ont toujours été là pour m'accompagner et me supporter, notamment pour les conseils de Maïlys et Morgane, les discussions avec Alban sur la génétique et la recherche avec François. Merci Sabine de m'avoir aidé durant ces trois années.

Julie, merci de m'accompagner depuis tant d'années et de m'avoir suivi pour ce projet de thèse.

Cette thèse a été financée par l'ANRT (convention°2020/1664) dans le cadre d'une bourse CIFRE et par APIS GENE (programme de recherche SelViG²). Le financement de l'encadrement par IDELE a bénéficié du soutien du Ministère de l'Agriculture dans le cadre du CASDAR. Merci donc à l'ANRT, APIS GENE et au Ministère de l'Agriculture pour le soutien financier qu'ils ont accordé à cette thèse.



Valorisations

Article en premier auteur :

Detailed analysis of mortality rates in the female progeny of 1,001 Holstein bulls allows the discovery of new dominant genetic defects

Journal of Dairy Science

2023-01 | Journal article

DOI: 10.3168/jds.2022-22365

AUTEURS: F. Besnard; H. Leclerc; M. Boussaha; C. Grohs; N. Jewell; A. Pinton; H. Barasc; J. Jourdain; M. Femenia; L. Dorso et al.

Massive detection of cryptic recessive genetic defects in dairy cattle mining millions of life histories

2023-09-23 | Preprint

DOI: 10.1101/2023.09.22.558782

AUTEURS: Florian Besnard; Ana Guintard; Cécile Grohs; Laurence Guzylack-Piriou; Margarita Cano; Clémentine Escouflaire; Chris Hozé; Hélène Leclerc; Thierry Buronfosse; Lucie Dutheil et al.

Article avec participation en tant que co-auteur :

Large-scale detection and characterization of interchromosomal rearrangements in normozoospermic bulls using massive genotype and phenotype data sets

Genome Research

2023-06 | Journal article

DOI: 10.1101/gr.277787.123

AUTEURS: Jeanlin Jourdain; Harmonie Barasc; Thomas Faraut; Anne Calgaro; Nathalie Bonnet; Camille Marcuzzo; Amandine Suin; Anne Barbat; Chris Hozé; Florian Besnard et al.

Investigating the impact of paternal age, paternal heat stress, and estimation of non-genetic paternal variance on dairy cow phenotype

Genetics Selection Evolution

2024-06-18 | Journal article

DOI: 10.1186/s12711-024-00918-2

AUTEURS: Corentin Fouéré; Chris Hozé; Florian Besnard; Mekki Boussaha; Didier Boichard; Marie-Pierre Sanchez

A recurrent de novo missense mutation in COL1A1 causes osteogenesis imperfecta type II and preterm delivery in Normande cattle

Genetics Selection Evolution

2024-05-21 | Journal article

DOI: 10.1186/s12711-024-00909-3

CONTRIBUTORS: Julien Corbeau; Cecile Grohs; Jeanlin Jourdain; Mekki Boussaha; Florian Besnard; Anne Barbat; Vincent Plassard; Julie Rivière; Christophe Hamelin; Jeremy Mortier et al.

Communications en congrès:

Detailed analysis of mortality rates in the female progeny of Holstein bulls allows the discovery of new dominant genetic defects

2022-07-03 | Conference paper WCGALP

HAL: hal-03731355

AUTEURS: F Besnard; H Leclerc; N Cesbron; M Boussaha; C Grohs; N Jewell; A Pinton; H Barasc; M Femenia; L Dorso et al.

Large scale screening for genetic defects in Holstein cattle using transmission disequilibrium test

2023-08-26 | Conference paper EAAP

HAL: hal-04195821

AUTEURS: F Besnard; M Boussaha; H Leclerc; J Jourdain; D Boichard; A Capitan

Table des matières

Remerciements.....	i
Table des matières	v
Listes des tableaux.....	viii
Liste des figures	ix
Liste des abréviations.....	xii
Introduction générale	1
Chapitre I. Synthèse bibliographique	3
1. Introduction bibliographique	3
2. Données générales sur le cheptel bovin français	3
3. L'amélioration génétique en France	5
4. Mortalité juvénile en élevage laitier	14
5. Structure des populations bovines et émergence régulière d'anomalies génétiques ..	17
6. Présentation de l'ONAB.....	22
7. Méthodes de cartographie et identification de <i>loci</i> responsables d'anomalies génétiques.....	24
Chapitre II. Exploration phénotypique de la mortalité	36
1. Introduction	36
2. Matériel et méthodes	36
3. Résultats et discussion	38
4. Conclusion	44
Chapitre III. Identifications d'anomalies génétiques dominantes.....	45
1. Introduction	45
2. Analyse de deux anomalies dominantes pour des taureaux avec des taux de mortalité extrêmes.....	47
3. Discussion.....	61

Chapitre IV. Analyse d'association (GWAS) des taux de mortalité dans les descendance des taureaux	63
1. Introduction	63
2. Matériel et méthodes	63
3. Résultats	66
4. Discussion	74
5. Conclusion	74
Chapitre V. Détection des anomalies récessives chez les bovins laitiers : une approche intégrant l'analyse des trajectoires de vie et le déficit et l'enrichissement en homozygotes ..	76
1. Introduction	76
2. Article soumis : “Massive detection of cryptic recessive genetic defects in cattle mining millions of life histories “	78
3. Discussion	115
Chapitre VI. Détection de <i>loci</i> influençant la létalité embryonnaire et la survie juvénile avec un déterminisme autosomique dominant à pénétrance incomplète.	117
1. Introduction	117
2. Matériel et Méthodes.....	118
3. Résultats et discussion	123
4. Discussion	131
Chapitre VII. Développement d'une évaluation génomique sur la mortalité juvénile	134
1. Introduction	134
2. Matériel, Méthodes et Résultats intermédiaires.....	136
3. Résultats de l'évaluation pilote par Single step.....	143
4. Exploration de pistes d'amélioration du modèle d'évaluation génomique.....	150
Chapitre VIII. Discussion générale et perspectives	160
1. Résultats principaux des études sur la mortalité, avantages et limites	160
2. Gestion des anomalies génétiques et sélection pour une amélioration de la survie des jeunes bovins.....	167
3. Anticipation des futures émergences	173

4. Travaux à approfondir dans un futur proche	175
Conclusion générale	180
Références bibliographiques	181
Annexe.....	190

Listes des tableaux

Tableau I-1 : Agents pathogènes des diarrhées néonatales du veau. (adapté de Gourreau 2011) :	16
Tableau I-2 : Statistiques illustrant les différences de variabilité génétique entre les 3 principales races laitières et allaitantes françaises.	19
Tableau I-3 : Liste des anomalies récessives détectées dans les populations bovines Holstein, Montbéliarde et Normande avec un variant causal connu.	32
Tableau II-1 : Effectifs des veaux femelles étudiés par année de naissance pour les trois principales races laitières françaises.	37
Tableau IV-1 : Liste des QTL significatifs et de leurs variants « cime » pour les 3 races et 5 caractères de mortalité juvénile étudiés.	71
Tableau IV-2 : Liste des variants candidats affectant une région codante du génome.	73
Tableau VI-1 : Résumé des jeux de données étudiés à chaque étape de l'étude HSD et de leurs effectifs pour les trois races.	122
Tableau VI-2 : Résultats significatifs des taux de mortalité liés aux haplotypes « cime » dans la population de validation.	125
Tableau VI-3 : Résultats significatifs des taux de conception liés aux haplotypes « cime » dans la population de validation.	126
Tableau VI-4 : Informations sur les variants candidats et leurs conséquences phénotypiques présumées.	130
Tableau VII-1 : Estimation des paramètres génétiques dans les races Montbéliarde, Normande et Holstein.	142
Tableau VII-2 : Estimations des CD à partir d'information purement génomique dans une population de validation.	148
Tableau VII-3 : Tableau récapitulatif le nombre de veaux morts et vivants disponibles selon l'information génétique disponible pour ces animaux.	153
Tableau VII-4 : Liste des haplotypes les plus significatifs détectés dans les 3 races. La référence OMIA est indiquée si l'anomalie est déjà connue.	156

Liste des figures

Figure I-1 : Illustration de la distribution géographique des races bovines françaises (source : Races De France ; www.racesdefrance.fr).....	4
Figure I-2 : Illustration d'un schéma de sélection.	5
Figure I-3 : Frise chronologique présentant les avancées technologiques et méthodologiques en sélection bovine.	13
Figure I-4 : Schéma récapitulant les principales caractéristiques des déterminisme génétiques récessif, dominant ou co-dominant.	19
Figure I-5 : Représentation schématique de la génétique en avant "forward genetic" et de la génétique inverse "reverse genetic".	25
Figure I-6 : Représentation de la cartographie par homozygotie d'un locus récessif (Source ONAB 2023).	26
Figure II-1 : Evolution annuelle du taux de mortalité juvénile totale des veaux femelles de races Holstein, Montbéliarde et Normande de l'année 2000 à 2022.....	38
Figure II-2 : Evolution des taux de mortalité en pourcentage de 2000 à 2022 pour chaque période (0_2 périnatale, 3_14 post-natale, 15_55 pré-sevrage, 56_365 post-sevrage) pour les races Holstein, Montbéliarde et Normande.	40
Figure II-3 : Histogramme montrant les différences de taux de mortalité sur la période de 0 à 365 jours pour les veaux femelles de races Holstein, Montbéliarde et Normande selon la saison de naissance.	42
Figure II-4 : Histogramme des taux de mortalité pour les taureaux avec plus de 100 filles pour 4 périodes (périnatale, post-natale, pré-sevrage et post-sevrage).	43
Figure II-5 : Histogramme des taux de mortalité juvénile au cours de la première année (0 à 365 jours) pour l'ensemble des taureaux des races Holstein, Montbéliarde et Normande.....	44
Figure IV-1 : Manhattan Plots de l'analyse GWAS des 5 caractères de mortalité pour la race Holstein (66).	68
Figure IV-2 : Manhattan Plots de l'analyse GWAS des 5 caractères de mortalité pour la race Montbéliarde (46).	69
Figure IV-3 : Manhattan Plots de l'analyse GWAS des 5 caractères de mortalité pour la race Normande (46).	70
Figure 1. Overview of the analysis plan developed in this study to detect and characterize cryptic recessive genetic defects in cattle.	82
Figure 2. Results of HHED mapping.....	83
Figure 3. Effects of the 34 validated haplotypes on female survival.	84

Figure 4. Evolution of haplotype frequencies, pedigree analyses, and estimation of haplotype effects on recorded traits.....	86
Figure 5. Characterization of the ITGB7 substitution.	88
Figure 6. Pathophysiological characterization of RFC5 and NOA1 mutants in Normande and Montbeliarde cattle, respectively.	90
Figure VI-1 : Schéma résumant les différentes causes possibles de déséquilibres de transmissions de segments chromosomiques observés entre un taureau et ses descendants vivants à un temps « t ».	118
Figure VI-2 : Description de l’approche utilisée pour la détection et la validation d’haplotypes responsables de mortalité juvénile ou de mort embryonnaire.	122
Figure VI-3 : Weblogo montrant la conservation de la séquence protéique de MRPS18B autour de la substitution p.96P>L +/- 10 acides aminés.....	128
Figure VI-4 : Weblogo montrant la conservation de la séquence protéique de FNDC3B autour de la substitution p.125P>L +/-10 acides aminés.	129
Figure VII-1 : Histogramme de la distribution des taux de mortalité entre les jours 1 et 56 en race Holstein.	137
Figure VII-2 : Histogramme montrant la différence des taux de mortalité entre les primipares et les vaches pour les veaux femelles nés entre 2015 et 2021 de la race Holstein.	139
Figure VII-3 : Histogramme montrant les taux de mortalité journaliers pour les veaux femelles nés entre 2015 et 2021 de la race Holstein en fonction de leurs conditions de naissance.	140
Figure VII-4 : Histogramme montrant les taux de mortalité journaliers pour les veaux femelles nés entre 2015 et 2021 de la race Holstein en fonction de la taille des élevages.	141
Figure VII-5 : Courbes de tendance des taux de mortalité et des BLUP sur la période 3 à 365 jours.	145
Figure VII-6 : Comparaison des taux de mortalité par descendance de taureaux candidats durant la période de validation (2021-2022) avec les prédictions génomiques issues de la population d'apprentissage (2015-2020) en races Montbéliarde (a), Normande (b) et Holstein (c).	147
Figure VII-7 : Comparaison des taux de mortalité des veaux nés durant la période de validation (2021-2022) avec les classes d’index génomiques prédites sans les performances propres (découpés en déciles) en races Montbéliarde (a.), Normande (b.) et Holstein (c.).	149
Figure VII-8 : Schéma du calcul de ROH pour un haplotype donné selon les typages disponibles.	152
Figure VII-9 : Déficit de veaux pour chaque haplotype sur le chromosome 11 en race Holstein.	154

Figure VII-10 : Déficit de veaux pour chaque haplotype du chromosome 6 en race Montbéliarde.	155
Figure VII-11 : Relation entre les résultats des tests de Chi2 en log 10 de p-value en abscisse et les solutions en % du RoH dans l'évaluation en ordonné pour la race Montbéliarde.	158
Figure VIII-1 : Graphique présentant la contribution brute au patrimoine génétique de leur race, l'année de naissance et la disponibilité d'une séquence de génome complet ou non pour les principaux taureaux Holstein, Montbéliards et Normands.....	174

Liste des abréviations

ANRT : Agence Nationale de la Recherche et de la Technologie

ADN : Acide Désoxyribonucléique Nucléique

BDNI : Base de Données Nationale d'Identification

BLAD : Bovine Leukocyte Adhesion Deficiency

BLIRD : Bovine Lymphocyte Intestinal Retention Defect

BLUP : Best Linear Unbiased Predictor

CASDAR : Compte d'Affectation Spéciale Développement Agricole et Rural

CD : Coefficient de Détermination

CDH : Cholesterol Deficiency Haplotype

CTIG : Centre de Traitement des Informations Génétiques

CIFRE : Conventions Industrielles de Formation par la REcherche

DGF : Dispositif Génétique Français

DL : Déséquilibre de liaison

ECEL : Entreprises de Conseil en Elevage

EDE : Etablissements De l'Elevage

EMP : Entreprise de Mise en Place

ES : Entreprise de Sélection

FGE : France Génétique Elevage

GDS : Groupement de Défense Sanitaire

GWAS : Genome-Wide Association Study

HSD : Haplotype Segregation Distorsion

HHED : Homozygous Haplotype Enrichment / Depletion

HWE : Hardy Weinberg Equilibrium

h² : héritabilité

IA : Insémination Animale

ICAR : International Committee for Animal Recording

IDELE : l'Institut de l'Elevage

INDEL : insertion/déletion

INRAE (auparavant INRA) : Institut National de la Recherche pour l'Agriculture, l'Alimentation et l' Environnement

IPG : Identification Pérenne Généralisée

ISU : Index de Synthèse Unique

Lof : Loss of function

MAF : Minor Allele Frequency

Mb : Megabase

MGI : Mouse Genome Informatics

MTCP : Mitochondropathie

NCBI : National Center for Biotechnology Information

OMIA : Online Mendelian Inheritance in Animals

OMIM : Online Mendelian Inheritance in Man

ONAB : Observatoire National des Anomalies Bovines

OS : Organisme de Sélection

QTL : Quantitative Trait Loci

ROH : Run of Homozygosity

SNIG : Système National d'Information Génétique

SNP : Single Nucleotide Polymorphism

SNGTV : Société Nationale des Groupements Techniques Vétérinaires

UMT : Unité Mixte Technologique

RZUE : Règlement Zootechnique Européen

Introduction générale

Aujourd'hui, réduire la mortalité des jeunes bovins est un enjeu majeur pour les filières bovines. Le taux de survie de la naissance jusqu'à la puberté est en effet faible en races laitières, estimé à 89,5% pour les femelles et 84,9% pour les mâles selon l'étude MORPHE (Leclerc et *al.* 2016). Cette question revêt une importance significative à plusieurs niveaux, que ce soit sur le plan économique, sociétal ou environnemental, surtout dans un contexte où la réduction des périodes improductives est primordiale pour la rentabilité et la durabilité des exploitations. Si les causes de mortalité sont multiples, de nombreux travaux ont démontré que la survie des jeunes animaux pouvait être fortement impactée par des prédispositions génétiques. La plupart des travaux se sont cependant essentiellement penchés sur la mortalité périnatale et les anomalies génétiques, et la connaissance des facteurs génétiques influençant la survie des veaux jusqu'à leur puberté est encore limitée.

Mon projet de thèse s'inscrit donc dans ce cadre. Il vise à approfondir les connaissances sur le déterminisme génétique de la mortalité juvénile des bovins, grâce à une exploitation large des bases de données nationales (identification, généalogies, géotypes) et des données de séquences génomiques disponibles pour les travaux de recherche. Il aborde à la fois la recherche de mutations causales de différents types (effets additifs ou récessifs), des évaluations pilotes à l'aide d'un modèle de sélection adapté à l'élevage bovin laitier et les perspectives d'optimisation de la gestion des troupeaux laitiers.

Ma thèse s'est déroulée dans l'UMT eBIS, structure partenariale associant l'Institut de l'Élevage (IDELE), INRAE et ELIANCE au sein de l'Unité Mixte de Recherche Génétique Animale et Biologie Intégrative d'INRAE située à Jouy-en-Josas (78). Elle a été encadrée par IDELE, mon employeur, et par INRAE (Didier Boichard et Aurélien Capitan) sur le plan scientifique. Elle a bénéficié du financement de l'ANRT sous la forme d'une bourse CIFRE, et d'APIS GENE (programme SelViG²).

Dans le cadre de cette thèse en partie sur articles, nous proposons une démarche méthodologique articulée autour d'une synthèse bibliographique, d'une exploration phénotypique, ainsi que d'une recherche des anomalies génétiques dominantes et récessives dans les populations bovines laitières. Enfin, nous proposons une évaluation génomique pilote visant à réduire la mortalité juvénile par sélection et proposons des perspectives d'optimisation de la gestion des troupeaux laitiers.

Ce mémoire de thèse s'articule autour de 8 chapitres, comprenant une synthèse bibliographique, une exploration des données phénotypiques, les résultats de plusieurs études portant sur l'identification d'anomalies génétiques dominantes et récessives dans les populations bovines laitières, et le développement d'évaluations génomiques. Le dernier chapitre permet de synthétiser et de discuter l'ensemble des résultats que j'ai obtenus au cours de mes travaux.

Chapitre I. Synthèse bibliographique

1. Introduction bibliographique

Ce chapitre présente une synthèse bibliographique explorant le contexte qui a encadré le déroulement de cette thèse. Nous commencerons par examiner les caractéristiques distinctives de la filière bovine et du rôle crucial du dispositif génétique français dans le recueil de données et l'amélioration génétique du cheptel national. Une attention particulière sera accordée à la mortalité en élevage bovin, en se concentrant spécifiquement sur les anomalies génétiques et leur détection.

2. Données générales sur le cheptel bovin français

La filière bovine constitue un pilier essentiel de l'agriculture et de l'entretien du paysage en France. Elle comprend environ 17 millions de bovins répartis dans une quarantaine de races, dont 7 millions de vaches reproductrices (Agrete 2023). Au-delà de cet effectif important, l'élevage bovin revêt une grande diversité en termes de territoires, systèmes d'élevage ou races. Ces races peuvent être catégorisées en fonction de leurs finalités, que ce soient les races bouchères (également désignées comme races « allaitantes »), les races laitières spécialisées, ou celles dotées d'aptitudes mixtes.

Parmi les races laitières, la Holstein se démarque par sa capacité de production exceptionnelle et sa diffusion mondiale. En France, elle contribue pour les deux tiers des effectifs de vaches laitières et 80% du lait récolté dans le pays. La Montbéliarde est la seconde race laitière avec environ 560 000 vaches majoritairement concentrées dans l'Est et le Centre, même si elle est en expansion géographique. Cette race est très présente dans des filières AOC, en particulier le Comté (Figure I-1). La Normande est la troisième race en effectif. Historiquement très développée, elle a perdu progressivement du terrain, du fait de son statut très mixte et de la concurrence avec la Holstein sur les mêmes territoires. Enfin, plusieurs races ont une importance régionale, associées à un bassin de production et souvent à une filière qualité. Notons les races Abondance et Tarentaise dans les Alpes du Nord, la Brune et la Simmental dans

l'Est, la Vosgienne dans les Vosges, ou encore la Jersiaise dans l'Ouest. Plusieurs races ont des effectifs encore plus faibles, de quelques centaines à quelques milliers d'individus, comme la Flamande ou la Bleue du Nord (Maigret 2024).

Parmi les races allaitantes, la Charolaise présente l'effectif le plus important, débordant largement de son noyau du Nord-Est du Massif Central. La Limousine et la Blonde d'Aquitaine prennent les seconde et troisième places. Plusieurs autres races ont des qualités de rusticité et se trouvent dans des régions plus difficiles dans le grand Sud-Ouest : Salers, Aubrac, Gasconne. Quelques races à effectifs modérés présentent des qualités de production de viande : Parthenaise, Rouge des Prés, Bazadaise.

Enfin, une quinzaine de races sont en effectifs très limités et font l'objet d'un programme de conservation sous l'égide de l'Institut de l'Élevage (IDELE).



Figure I-1 : Illustration de la distribution géographique des races bovines françaises (source : Races De France ; www.racesdefrance.fr).

Races laitières (de haut en bas et de gauche à droite) : Bleue du Nord - Rouge Flamande, Jersiaise - Normande, Simmental - Vosgienne, Holstein, Montbéliarde, Abondance, Tarentaise, Brune. Races allaitantes (de haut en bas et de gauche à droite) : Blanc Bleu Belge, Rouge des Prés, Charolaise, Parthenaise – Limousine – Salers – Hereford, Bazadaise – Aubrac, Blonde d'Aquitaine, Gasconne, Raço di Biou, Race de Combat.

Près de la moitié des races françaises bénéficient d'un programme de sélection (Figure I-2), dont l'objectif est d'identifier les meilleurs individus pour les faire reproduire en priorité et ainsi

améliorer génération après génération les performances moyennes de la population concernée. Sélection et reproduction sont intimement liées car le nombre de reproducteurs nécessaires dépend de leur pouvoir de diffusion. La vache a un pouvoir de diffusion faible, même en mobilisant des technologies comme le transfert embryonnaire. Chez le mâle, le potentiel est modéré et restreint à son troupeau en cas de monte naturelle mais il devient très important en insémination animale (IA). La sélection est donc bien plus efficace lorsqu'elle intègre largement cette technique (Boichard 2020).

Cette sélection a été rendue possible grâce à un système performant de collecte de données à l'échelle nationale, orchestré par une diversité d'acteurs impliqués dans la filière bovine (Frappreau 2023). Historiquement basée sur des tests sur descendance pour les taureaux d'insémination, cette sélection a évolué grâce aux avancées récentes en biotechnologie, donnant naissance aux évaluations génomiques.

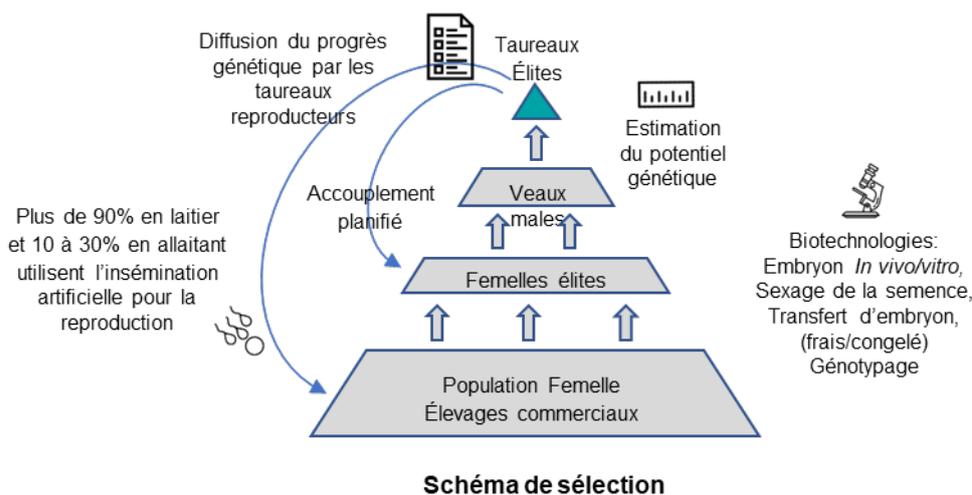


Figure I-2 : Illustration d'un schéma de sélection.

La figure montre une pyramide de sélection génétique en élevage bovin composée de quatre niveaux principaux et de deux flux génétiques représentés par des flèches bleues. La base de la pyramide représente la grande population de femelles dans les élevages commerciaux. Le deuxième niveau représente les femelles élites, sélectionnées pour leurs caractéristiques génétiques supérieures qui donneront naissance aux veaux mâles. Ces derniers, si leur niveau génétique est suffisant, pourront devenir de futurs taureaux reproducteurs. Enfin, par le biais de l'insémination artificielle, ces taureaux « élites » permettent l'amélioration globale du cheptel bovin.

3. L'amélioration génétique en France

3.1. L'organisation de la sélection

3.1.1. Principales informations enregistrées dans les bases de données

La sélection est réalisée intra race, les races étant des populations bovines relativement fermées constituées au cours des 200 dernières années à partir d'animaux phénotypiquement ressemblants et reproduits entre eux. En France, elle a été structurée par la Loi sur l'Elevage de 1966 qui a défini les missions nécessaires pour la sélection et les différents organismes en charge de ces missions. La structuration du Dispositif Génétique Français (DGF), restée stable durant 40 ans, a été revue dans la Loi d'Orientation Agricole de 2006 qui a, entre autres, abouti au transfert de l'essentiel du pilotage du dispositif à l'interprofession génétique France Génétique Elevage (FGE) créée à cette occasion. La dernière évolution majeure date de 2018 avec le Règlement Zootechnique Européen (RZUE), qui vise à harmoniser la structuration des dispositifs au niveau européen. Ce dispositif et les acteurs qui le constituent sont décrits dans la partie suivante.

Le DGF génère et utilise de très nombreuses informations (performances, généalogies, résultats d'évaluations génétiques ou index) qui sont apportées dans une base nationale unique (Système National d'Information Génétique ou SNIG) par plus de 200 organismes différents. Cette base est centralisée au Centre de Traitement des Informations Génétiques (CTIG) d'INRAE et des copies sont transférées aux centres régionaux pour les animaux qui les concernent. Les règles de collecte et d'enregistrement de ces informations sont définies dans le cadre d'un Système de Management de la Qualité supervisé par FGE et sont conformes aux normes établies par l'*International Committee for Animal Recording* (ICAR). Ces données sont utilisées par leurs apporteurs, mais aussi par les autres acteurs du DGF : par exemple la date d'une insémination permet de confirmer la parenté d'un animal, alors que les deux activités (insémination et certification des parentés) sont réalisées par des acteurs différents. Elles sont utilisées pour les évaluations génétiques et les travaux de recherche conduits par les équipes de recherche nationale (INRAE, Idele et Eliance au sein de l'Unité Mixte de Technologie ou UMT eBIS, voir plus loin). Une restructuration du système d'information est en cours et aboutira prochainement à la mise en œuvre d'une Base Professionnelle différente, mais le CTIG continuera à gérer toutes les données nécessaires aux évaluations génétiques et permettra un accès aux informations pour la recherche.

Les génotypages, de par leur volume et leur sensibilité en termes de droits d'accès, ne sont pas apportés dans le SNIG mais dans plusieurs bases de données différentes en fonction de l'organisme gérant cette activité. Au niveau national, les informations génomiques sont centralisées dans la base SIGENO du CTIG et sont utilisées dans les programmes de recherche et les évaluations génomiques.

Le DGF repose sur l'Identification Pérenne Généralisée (IPG) des animaux, même si celle-ci n'est pas une activité directement associée à la sélection. En France et ce depuis la crise de la vache folle (1998), tous les bovins sont identifiés dès la naissance avec un numéro unique enregistré dans la Base de Données Nationale d'Identification (BDNI). L'éleveur doit, dans les 48h suivant la naissance d'un veau, boucler l'animal et déclarer son numéro IPG ainsi que sa date de naissance, son sexe, son type racial, le père supposé et la mère. Ces informations, essentielles pour retracer tout le parcours d'un animal ou d'un groupe d'animaux en cas de crise sanitaire, sont aussi précieuses pour vérifier la cohérence de toutes les informations du SNIG, définir les groupes d'animaux contemporains d'étable pour les évaluations génétiques ou encore retracer la carrière de chaque animal pour divers travaux de recherche.

Notre étude repose intégralement sur des données extraites du SNIG et de SIGENO gérées par le CTIG.

3.1.2. Les acteurs de la sélection

Cette section vise à présenter les principaux acteurs impliqués dans le processus complexe de sélection génétique, en mettant l'accent sur leur rôle le plus significatif dans le panorama de l'élevage bovin et la gestion des données associées.

Historiquement, les Organismes de Sélection (OS) étaient en charge de l'orientation de leur race, ce qui se traduisait en particulier par la définition des caractéristiques de la population et des objectifs de sélection, la tenue des livres généalogiques, la gestion des gènes d'intérêt et l'ingénierie de la morphologie. Depuis 2018 et la mise en œuvre du RZUE 2016/1012 (Ministère de l'agriculture de l'agroalimentaire et de la forêt 2016), les OS sont responsables des programmes de sélection pour lesquels ils sont agréés. Le RZUE a ainsi étendu les anciennes missions à la quasi-totalité des activités puisque la certification de parenté, le contrôle de performances, l'évaluation génétique et la publication des résultats sont désormais de leur ressort, ces missions pouvant être réalisées en propre ou déléguées par contrat.

Les Entreprises de Sélection (ES), mettent en œuvre les schémas de sélection, c'est-à-dire les opérations visant à sélectionner les reproducteurs d'élite et en particulier les taureaux d'insémination. En général, l'activité d'une ES peut se résumer comme suit : les ES planifient des accouplements contractualisés avec des éleveurs, achètent et élèvent les jeunes taureaux nés de ces accouplements, les font évaluer (historiquement par testage sur descendance, à partir de

2010 par évaluation génomique), sélectionnent les meilleurs d'entre eux, et leur font produire de la semence qu'ils diffusent chez les éleveurs par l'intermédiaire des Entreprises de Mise en Place (EMP). Les ES disposent de centres d'élevage et de centres de collecte de taureaux.

Les EMP assurent l'activité d'insémination animale. Les doses sont fournies par les ES et mises en place par les techniciens des EMP, ou directement par l'éleveur. L'EMP assure aussi une activité de suivi de la fertilité et souvent de transfert embryonnaire. L'EMP est responsable de l'enregistrement des données d'inséminations dont la déclaration dans le SNIG est obligatoire.

Les Etablissements De l'Elevage (EDE) jouent un rôle central en agissant comme entités locales responsables de l'IPG. Elles réalisent les opérations associées à la certification des parentés, désormais par délégation des OS.

Les Entreprises de Conseil en Elevage (ECEL) assurent la production des données de performances individuelles des animaux pour différents caractères, entre autres de production laitière (quantité de lait, de matière protéique et de matière grasse, taux protéique et de matière grasse) et de santé de la mamelle (comptages cellulaires dans le lait, évènements de mammite clinique). Les analyses du lait pour les taux protéique et butyreux sont effectuées par spectrométrie Moyen Infra Rouge. Ces spectres sont actuellement exploités pour prédire de nouveaux phénotypes tels que les corps cétoniques ou les émissions de méthane (Barbat et al. 2016; Fresco et al. 2023). Dans leurs finalités de sélection, ces activités sont, depuis le RZUE, réalisées sous la responsabilité des OS.

Enfin, des organismes nationaux, tels que INRAE, IDELE et ELIANCE contribuent significativement à la recherche et au développement, apportant des avancées scientifiques pour soutenir l'amélioration génétique continue des races bovines. Historiquement, INRAE a été en charge de l'évaluation génétique des reproducteurs jusqu'en 2018 et, à ce titre, a développé des compétences approfondies dans ce domaine. INRAE, par le CTIG, héberge la base zootechnique nationale pour le compte du Ministère de l'Agriculture. Depuis le RZUE, l'activité d'évaluation génétique a été reprise par un nouvel organisme, GenEval, sous pilotage des deux fédérations représentant les OS et les ES (Races de France et Eliance). GenEval est constituée d'une équipe d'experts en génétique, en informatique et dans le traitement des données génétiques pour réaliser les évaluations génomiques des animaux d'élevage. INRAE fournit un support contractuel, incluant les données, l'infrastructure de calcul, les logiciels et l'appui méthodologique. IDELE est un institut technique qualifié par le Ministère de l'Agriculture pour conduire des actions de R&D dans le secteur des herbivores (bovins, caprins, ovins, équins). Dans le domaine de la génétique, IDELE participe à de multiples programmes de recherche. Il est

un acteur important pour le transfert des outils et des résultats, avec pour mission d'accélérer l'impact des travaux de issus de ces programmes en vue de la transition agro-écologique. De plus, l'organisme diffuse les index calculés par GenEval pour le compte des OS, et administre le SNIG. ELIANCE est la fédération des organismes de conseil et d'appui aux éleveurs (ES, EMP, ECEL). ELIANCE a fortement contribué à la mise en place de la sélection génomique en France avec INRAE.

INRAE, IDELE et ELIANCE sont partenaires de l'Unité Mixte Technologique eBis, dont l'objectif est de conduire des travaux de recherche en génétique et sélection bovine.

3.2. De la sélection phénotypique à la sélection génomique

3.2.1. La sélection polygénique

Durant des millénaires, la sélection s'est faite sur les performances propres des individus, de façon plus ou moins volontaire. Les travaux de Fisher ont initié un changement majeur grâce à la formalisation mathématique du modèle polygénique, stipulant que le phénotype dépend d'effets génétiques et de milieu, et que l'effet génétique résulte de l'effet cumulé d'un grand nombre de gènes. De ce modèle résulte la notion de coefficient d'apparentement proposé par Wright (1922) et qui mesure la covariance entre les valeurs génétiques de deux individus et donc leur degré de parenté. Dans le courant du XXe siècle, Hazel (1943) développe la théorie des index de sélection, où la valeur génétique est estimée par régression linéaire multiple des performances de l'individu et de ses apparentés, les coefficients de régression dépendant de l'apparentement. Puis en 1975, Henderson propose le modèle BLUP (*Best Linear Unbiased Prediction*) dans lequel les effets génétiques et du milieu sont estimés simultanément, permettant ainsi de s'affranchir de différents biais. Il propose également une méthode pour calculer très rapidement l'inverse de la matrice de parenté, ce qui rend possible la construction et la résolution des équations du BLUP à grande échelle. En 1988, Quaas formalise le modèle « animal » dans lequel les performances de chaque individu sont exprimées en fonction de la valeur génétique de cet individu (et pas seulement de son père) et où toutes les parentés sont prises en compte. Ce sont ces modèles successifs qui ont été appliqués en France pour estimer les index, avec notamment la mise en œuvre du modèle animal en 1989 (Bonaiti et al. 1990).

Même avec le meilleur modèle possible, le modèle polygénique sur parenté ne permet pas de prédire l'aléa de méiose (c'est-à-dire la différence entre la valeur d'un individu et la moyenne de celles de ses parents) à partir des données des parents ou des collatéraux. C'est pour cela que

le testage sur descendance est resté une constante pendant plus de 50 ans : les taureaux étaient utilisés pour produire 50 à 100 filles réparties dans différents élevages. Au cours de leur première lactation, ces filles étaient phénotypées pour tous les caractères correspondant aux objectifs de sélection de façon à évaluer précisément les pères sur des caractères exprimés par les femelles tout en prenant en compte différents facteurs environnementaux. A l'issue de ce processus, environ un taureau sur dix était ensuite sélectionné et largement diffusé grâce aux propriétés biologiques du sperme bovin, lequel se conserve très efficacement congelé dans l'azote liquide.

Cette méthode a fait ses preuves mais elle pose diverses contraintes : sa lenteur (5 ans), son coût (50 k€ pour évaluer un taureau) et l'intensité de sélection modérée des taureaux (1 sur 10), même si elle est précise pour la plupart des caractères.

Le coût du testage a induit d'autres problèmes, en particulier l'utilisation d'un nombre restreint de taureaux et donc une réduction notable de l'effectif génétique des races laitières, avec les conséquences associées. Nous en reparlerons dans la partie intitulée « Structure des populations bovines et émergence régulière d'anomalies génétiques ».

Le progrès génétique attendu peut être estimé par la formule suivante (Rendel et Robertson 1950) :

$$\Delta G = \frac{i R \sigma_g}{T}$$

ΔG : progrès génétique attendu, c'est-à-dire la différence génétique entre la génération actuelle et la génération suivante.

i : l'intensité de sélection, c'est-à-dire la supériorité des sélectionnés sur le critère de sélection, par rapport à l'ensemble des candidats

R : la précision, c'est-à-dire la corrélation entre valeur génétique vraie et la valeur génétique estimée (ou index)

σ_g : l'écart type génétique

T : l'intervalle de génération correspondant à l'âge moyen des parents à la naissance des descendants.

Cette formule montre l'intérêt des programmes de testage et de l'amélioration des méthodes de prédiction des valeurs génétiques, les deux permettant d'améliorer la précision des valeurs génétiques estimées à partir desquelles les reproducteurs étaient sélectionnés. Elle montre aussi l'intérêt d'une intensité de sélection forte, et donc de l'utilisation de taureaux

d'insémination artificielle. Cela dit, jusqu'à l'arrivée de la sélection génomique, le progrès génétique a été limité par la lenteur de l'obtention d'index précis, qui allongeait l'intervalle entre générations (T).

3.2.2. La révolution génomique

En 2001, Meuwissen et *al.* proposent un nouveau type de sélection qu'ils nomment « sélection génomique ». Le principe est d'estimer l'effet de marqueurs génétiques couvrant tout le génome à partir d'une population de référence phénotypée et génotypée, et ensuite d'appliquer les effets estimés de ces marqueurs aux candidats à la sélection, génotypés. Cette méthode n'a pu être utilisée efficacement qu'avec l'arrivée des premiers outils de génotypage haut débit, développés à partir de la séquence du génome bovin. En 2007 apparaissait la première puce réellement utilisable à grande échelle avec 54 001 SNP (*Single Nucleotide Polymorphism* ou substitution d'un nucléotide par un autre ; puce Illumina BovineSNP50 dite "50k" ; Matukumalli et *al.* 2009). Grâce à cet outil, les premières évaluations génomiques ont pu être mises en place en 2009 dans différents pays.

Au niveau méthodologique, deux modèles équivalents sont proposés : le GBLUP (VanRaden 2008) analogue au BLUP mais dans lequel la matrice de parenté pedigree est remplacée par la matrice de parenté génomique, et le SNPBLUP qui inclut les effets aléatoires de tous les SNP, chacun ayant un effet de variance a priori constante (Meuwissen et *al.* 2001; Legarra et Misztal 2008; VanRaden 2008). En France, les trois principales races laitières (Holstein, Montbéliarde, Normande) ont d'abord été concernées dès 2009, suivies par 7 autres races entre 2011 et 2019.

En 2009, les premières populations de référence étaient constituées des taureaux d'IA testés sur descendance, dont l'ADN était disponible sous forme de semence. Pour élargir ces populations de référence et donc gagner en précision, plusieurs voies ont été mises en œuvre, telles le partage d'informations entre pays (comme au sein des consortia Eurogenomics pour la race Holstein (Lund et al, 2011) ou Intergenomics pour la Brune) ou la prise en compte des génotypes de femelles en production (moins informatives individuellement mais beaucoup plus nombreuses, et source obligatoire des populations de référence futures).

Le « typage femelle » a pu se développer grâce à la forte baisse du prix du génotypage permis initialement par la conception de puces à densité plus faible en SNP (dites « LD » pour *Low Density* et comprenant au minimum 7 000 marqueurs, (Boichard et *al.* 2012). En parallèle, des

procédures d'imputation ont été mises en œuvre pour restaurer l'information 50k avec une précision élevée (Li et *al.* 2009; Boichard et *al.* 2012). L'imputation est un processus mathématique permettant de reconstituer les génotypes manquants et qui repose sur la connaissance des haplotypes existant au sein d'une population. Elle permet de disposer d'une information uniforme et complète même quand les typages sont obtenus avec des puces différentes. Le coût a également été réduit par les volumes croissants de génotypage et les achats groupés au sein de consortia tel Eurogenomics. Aujourd'hui, les prix sont attractifs même pour des puces de moyenne densité (50 à 100k) compte tenu des gros volumes pratiqués, de sorte que la puce LD n'est plus utilisée.

La stratégie dans la filière bovine est assez unique car les éleveurs, en génotypant et phénotypant leurs animaux, ont accès à des index précis et à un stade précoce. Ces animaux, lorsqu'ils sont plus âgés, alimentent ensuite la population de référence, assurant la durabilité du processus de sélection. Le génotypage femelle est aussi la seule solution pour disposer d'index sur de nouveaux caractères, le nombre de mâles indexés étant trop limité.

Les premières évaluations génomiques étaient réalisées en deux étapes : une évaluation polygénique de tous les animaux, suivie d'une évaluation génomique des animaux génotypés uniquement, en utilisant des phénotypes corrigés issus de la première étape. Ces deux étapes induisent des biais d'autant plus importants que la sélection génomique est ancienne et efficace. Pour intégrer toute l'information en une seule étape, une méthodologie dite « Single Step » a été proposée par Aguilar et *al.* (2010) et Christensen et Lund (2010). Elle a fait l'objet de développements progressifs permettant de la mettre en œuvre à grande échelle, d'assurer la compatibilité entre informations génomiques et pedigree, y compris en présence de groupes génétiques. Le Single Step existe aussi sous les deux formes GBLUP et SNPBlup. Le Single Step a été mis en œuvre en France chez les bovins en 2022.

L'ensemble des évolutions méthodologiques de la sélection phénotypique à l'évaluation génomique est résumé dans la Figure I-3.

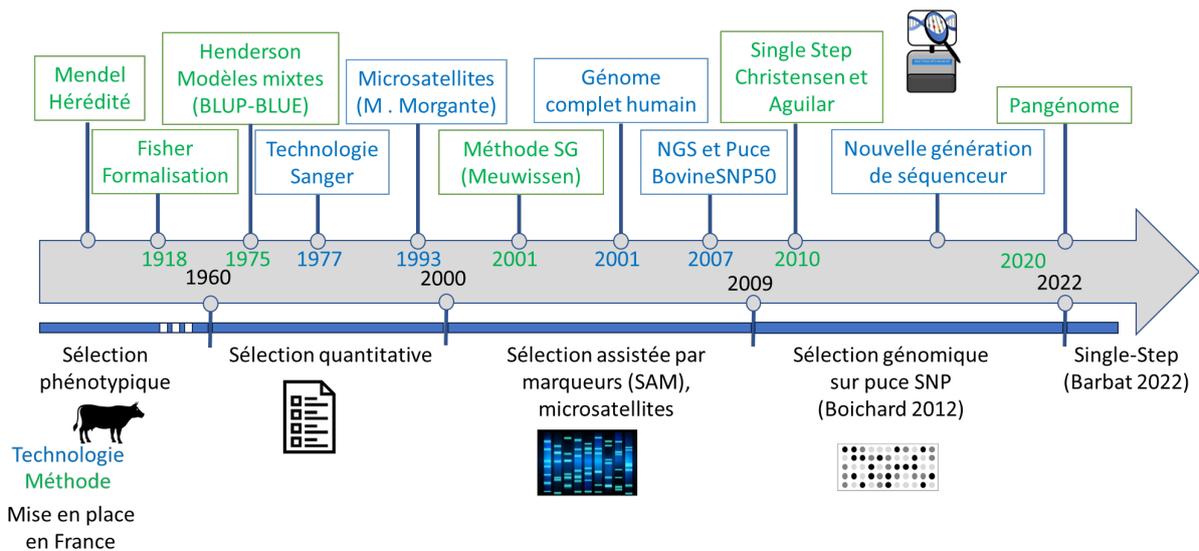


Figure I-3: Frise chronologique présentant les avancées technologiques et méthodologiques en sélection bovine.

Sont représentés en vert les avancées méthodologiques, en bleu les avancées technologiques et en noir les évolutions du système de sélection animal français.

3.2.3. Objectifs de sélection et nouveaux caractères d'intérêt

Les objectifs de sélection des races bovines ont beaucoup évolué au fil des décennies. Structurée dans les années 60, la sélection visait à augmenter la production moyenne des vaches laitières qui atteignait seulement 3000 litres de lait par lactation à l'époque contre 10 000 litres de nos jours (Stevenson et Britt 2017). Au cours des années 80, suite à l'instauration de quotas régulant la production laitière à l'échelle de l'UE, une transition significative s'est opérée, avec l'incorporation de caractères liés à la qualité du produit dans les critères de sélection. Cette orientation a été renforcée dans les années 90, avec l'intégration de caractères liés à la reproduction, la santé de la mamelle, la longévité et les conditions de naissance, comme souligné par Sellier et *al.* (2019).

Plus récemment, un effort important a été consacré aux caractères de santé, infectieux ou non. On peut citer par exemple les pathologies des pattes, la paratuberculose, ou les maladies métaboliques. Les anomalies génétiques font aussi l'objet d'un travail de détection, de caractérisation, et de gestion très actif grâce au génotypage massif (voir le paragraphe 7 « Méthodes de cartographie et identification de *loci* responsables d'anomalies génétiques » dédiée à ce sujet).

Aujourd'hui, la dynamique de la sélection génétique prend une nouvelle direction, intégrant des caractères comme l'efficacité alimentaire et les émissions de méthane associés à l'empreinte

carbone de l'élevage. Ces initiatives s'inscrivent dans le cadre des objectifs européens de réduction des émissions de gaz à effet de serre, visant une diminution drastique des émissions de l'élevage d'ici 2030, conformément à la législation européenne en la matière (Fresco et *al.* 2023).

Actuellement, au travers des projets tant européens que nationaux, des initiatives visant à diminuer les émissions de méthane ou à réduire les périodes improductives des animaux sont en cours de mise en place. L'introduction d'une sélection visant à minimiser la mortalité en élevage représente un enjeu capital pour l'avenir de l'élevage, compte tenu des pertes économiques, de l'impact environnemental et des atteintes au bien-être considérables que représente le décès prématuré d'un animal avant sa période de production.

4. Mortalité juvénile en élevage laitier

La mortalité des jeunes animaux est un enjeu majeur pour les filières bovines. Dans le programme MorPhé (FGE, 2013-2015), le taux de mortalité cumulé de la naissance à 150 jours a été estimé à 11,5% et 6,9% en moyenne pour les veaux femelles de races laitières et allaitantes, et 15,1% et 9,6% pour les mâles, respectivement (Leclerc et *al.* 2016). Cela a un impact important sur le revenu des éleveurs, notamment par les frais vétérinaires et le manque à gagner induits, en particulier pour l'élevage allaitant, dont la vente de veaux et jeunes bovins de boucherie est le principal revenu (contrairement à l'élevage laitier). La réduction du taux de mortalité est aussi un enjeu important en termes d'autonomie alimentaire sur le plan national : on peut estimer, sur la base des éléments statistiques de MorPhé et du cheptel bovin (chiffres clefs du cheptel bovin, Idele 2023 : 3,4 millions de vaches laitières et 3,6 d'allaitantes), que plus de 300 000 femelles des deux filières et 170 000 veaux allaitants meurent chaque année avant 6 mois. L'enjeu est aussi environnemental puisque ces animaux émettent du méthane sans déboucher sur une production quelconque. Enfin, une réduction de la mortalité des jeunes animaux participe aux objectifs d'amélioration du bien-être en élevage.

La mortalité juvénile est traditionnellement découpée en sous-périodes correspondant à des étiologies différentes (Gourreau 2011; Santman-Berends et *al.* 2019) détaillées ci-dessous.

4.1. Dystocies et morts périnatales/néonatales

La mortalité au cours des 48 premières heures affecte environ 4% des veaux de races Holstein, Montbéliarde et Normande. Elle est due à des problèmes de fragilité du veau dont l'origine peut être génétique, environnementale (infection *in-utero* ou absorption de substance toxique) ou la conséquence d'une dystocie (c'est-à-dire d'une difficulté de vêlage). Les principaux facteurs prédisposant à la dystocie en élevage bovin incluent la disparité entre la taille du veau et l'ouverture pelvienne de la mère (elle-même influencée par son âge et sa parité), les naissances gémeaux, le sexe du veau, et certaines anomalies génétiques (Cuttance et Laven 2019).

La mortalité à la naissance fait l'objet d'évaluations génétiques et génomiques dans les races bovines laitières depuis 2008 (Mattalia et Laloë 2009). Deux caractères sont évalués à partir du même phénotype (survie ou mortalité) : l'effet direct (donc du veau) appelé « vitalité à la naissance » (VIN) et l'effet maternel appelé « vitalité au vêlage » (VIV). L'information est disponible et diffusée, mais ce caractère n'est pas inclus dans l'objectif de sélection (ISU). Malgré ce suivi, on observe une tendance à la hausse de ces mortalités au cours du temps, qui justifieraient sans doute une prise en compte effective en sélection.

4.2. Les premières semaines de vie : maladies digestives, une histoire de virus, de bactéries et de parasites

La période de 3 à 30 jours peut être qualifiée de "période de transition de l'immunité". Pendant cette phase critique, le veau commence à développer ses propres défenses immunitaires, passant de l'immunité passive fournie par la mère et son colostrum à une immunité active. C'est une période où le veau présente un déficit immunitaire, le rendant susceptible aux infections.

La majorité des morts sont imputables à des entérites causées par divers pathogènes (voir Tableau I-1). Les virus, tels que le rotavirus (présent chez 30 à 60% des animaux diarrhéiques) et les parasites du genre *Cryptosporidium* (15 à 45%), présentant une faible spécificité d'hôte et infectant divers mammifères, sont les plus fréquemment identifiés. Cependant, le pouvoir pathogène de ces agents est généralement léger à modéré, et certains pathogènes moins fréquents peuvent avoir un impact plus significatif sur l'animal. Des bactéries du genre *Salmonella* et *Escherichia coli* entérotoxigènes sont présentes chez 0 à 15% des veaux diarrhéiques. Certains virus, comme le coronavirus bovin, peuvent aussi avoir un pouvoir pathogène marqué, même s'ils sont peu à moyennement présents dans un élevage (Gourreau 2011).

Tableau I-1 : Agents pathogènes des diarrhées néonatales du veau. (adapté de Gourreau 2011) :

Pouvoir pathogène: ?: mal connu; +/-: dans certaines circonstances; +: léger; ++: modéré; +++: marqué. Fréquence: ?: mal connue ; +: 0 à 15% ; ++ : 15 à 30%; +++: 30 à 45%; ++++: 45 à 60%.

Groupe	Nom	Pouvoir pathogène	Fréquence chez les veaux diarrhéiques
Bactéries	<i>E. coli</i> entérotoxigènes	+++	+
	<i>E. coli</i> entérohémorragiques	+/-	+/?
	<i>E. coli</i> autres pathotypes	?	?
	<i>Salmonella</i> sp.	+++	+
	<i>Campylobacter</i> sp	+/-	?
	<i>Clostridium perfringens</i>	+/-	?
Virus	Rotavirus	+ à ++	+++ à ++++
	Coronavirus	+++	+ à ++
	Virus BVD	+ à +++	+/?
	Torovirus	+/- à ?	+/?
	Calicivirus	+/- à ?	+/?
	Parvovirus	+/- à ?	+/?
Parasites	<i>Cryptosporidium</i>	+ à ++	++ à +++
	Giarda	+/-	?

4.3. Maladies respiratoires et immunitaires

Au-delà d'une trentaine de jours, le jeune bovin se révèle particulièrement vulnérable aux broncho-pneumonies car l'efficacité maximale des poumons n'est pas atteinte avant un an et ces derniers sont donc sous-dimensionnés par rapport à leur masse corporelle (environ 12 litres pour 500 kg, en comparaison aux 42 litres des chevaux (Veit et Farrell 1978)). Le virus respiratoire syncytial bovin (RS) et le virus parainfluenza III (PI3) sont les principaux agents pathogènes responsables de pneumonies virales en France identifiés par les Groupements de Défense Sanitaire (GDS) auxquels il faut ajouter les virus de la diarrhée virale bovine (BVD) et le coronavirus respiratoire bovin qui peuvent agir comme cofacteurs (Boubet 2021). Outre les infections purement virales, on recense également des pneumonies "mixtes" résultant de co-infections virales et pasteurelliques (dont la forme la plus fréquente associe le virus syncytial et

la bactérie *Mannheimia haemolytica*), la bronchite vermineuse due au nématode *Dictyocaulus viviparus*, et des complications mycosiques telles l'aspergillose. La diversité des pathogènes responsables de la mortalité après 30 jours souligne la complexité inhérente au défi sanitaire. Enfin, au-delà des causes environnementales, des facteurs génétiques peuvent entraîner des surmortalités sur cette période à la suite de déficits immunitaires ou de problèmes métaboliques, comme les anomalies génétiques récessives BLAD (*Bovine Leucocyte Adhesion Deficiency*) et CDH (*Cholesterol Deficiency Haplotype*) en race Holstein (Shuster et al. 1992; Menzi et al. 2016). Nous en reparlerons dans la section suivante.

Pour conclure sur la mortalité post-sevrage, elle représente environ un quart à un tiers des mortalités juvéniles totales (environ 3 à 6% dans les grandes races laitières). Rapporté au nombre de jours, ce taux de mortalité est relativement faible durant cette période mais son impact économique est élevé du fait des frais vétérinaires et des coûts d'élevage associés. En effet l'élevage de femelles pour le renouvellement du troupeau représente un investissement important qui ne commence à être amorti qu'à partir de leur premier vêlage et de leur entrée en production laitière vers l'âge de deux ou trois ans.

5. Structure des populations bovines et émergence régulière d'anomalies génétiques

Comme évoqué au début de ce chapitre bibliographique, les races bovines sont des populations fermées constituées au cours des 200 dernières années à partir d'animaux phénotypiquement proches et reproduits entre eux. L'utilisation intensive de taureaux « élite » dans le cadre de l'IA (dont certains ont dépassé le million de paillettes de semence produites) a créé des goulots d'étranglement engendrant une réduction importante de la variabilité génétique de chaque race. Cette perte de variabilité génétique est illustrée par différents indicateurs calculés à partir des données de généalogies. Prenons par exemple les résultats de la dernière étude menée dans le cadre du projet VARUME (Danchin-Burge et al. 2017) et dont un extrait est présenté dans le Tableau I-2. Le premier indicateur est la notion d'ancêtre efficace. Le nombre d'ancêtres efficaces F_e est dérivé de la contribution c_i de chaque ancêtre à la population actuelle : $F_e = 1 / \sum c_i^2$ (Boichard et al. 1997). F_e mesure donc le nombre théorique d'ancêtres différents s'ils contribueraient tous de façon identique. Ainsi les six principales races bovines françaises ont un F_e très faible (compris entre 17 et 150) par rapport aux centaines de milliers à millions de femelles nées sur la période 2019-2022. Les deuxième et troisième indicateurs sont la contribution

marginale de l'ancêtre principal au patrimoine génétique de la race et le nombre d'ancêtres expliquant 50% de ce patrimoine génétique. Comme le montre le Tableau I-2, ces valeurs sont particulièrement faibles dans les races laitières du fait de leur recours historique plus élevé à l'IA et l'impact majeur des lignées mâles de taureaux. Leur ancêtre principal a une contribution proche de 1/8, c'est donc comme si tous les individus de la race avaient un arrière grand-parent en commun. Les autres ancêtres influents ont aussi des contributions élevées, de sorte que seulement 6 à 8 ancêtres expliquent à eux seuls la moitié du patrimoine génétique de la race. Cela affecte en conséquence le taux moyen de consanguinité des animaux (estimé sur pedigree) qui est d'environ 6,0 % en race laitière et qui augmente d'environ 1% par génération. Un faible effectif génétique a un double impact : il augmente la dérive (ou fluctuation aléatoire des fréquences alléliques d'une génération à l'autre ; ainsi une anomalie peut voir sa fréquence devenir non négligeable en une ou deux générations) et l'augmentation de consanguinité entraîne une hausse de l'homozygotie sur le génome. La structure des races bovines augmente donc la probabilité que des individus soient homozygotes pour de grands segments chromosomiques hérités d'un ancêtre commun et donc la probabilité d'émergences d'anomalies génétiques récessives si ces segments portent une mutation délétère. C'est la notion de fardeau génétique, tout individu est porteur de quelques variants récessifs délétères à l'état hétérozygote. Chez l'Homme un article du consortium 1000 genomes estime que leur nombre varie entre 76 et 190 pour les variants rares délétères non-synonymes, et s'élève jusqu'à 20 pour les variants entraînant une perte de fonction et liés à des maladies connues (McVean et al. 2012). Ainsi, chez le bovin on connaît environ une anomalie génétique récessive pour chaque ancêtre influent et d'autres sont encore à découvrir.

Enfin, bien que la sélection génomique ait apporté un réel progrès pour l'amélioration génétique des espèces d'élevage, le raccourcissement des intervalles de génération a entraîné une augmentation de l'accroissement annuel du taux de consanguinité, au moins en race Holstein. Différents travaux ont porté sur la préservation de la variabilité génétique tout en augmentant le gain génétique (Colleau et al. 2015). Par ailleurs, la fin du testage sur descendance, qui agissait comme un filtre permettant d'écarter les taureaux avec un taux de mortalité sur descendants anormalement élevé, expose les populations à des anomalies génétiques dues à des mutations *de novo* dominantes avec mosaïcisme germinale chez le taureau (voir plus bas), ou encore à des remaniements interchromosomiques dans la descendance de taureaux au caryotype transloqué équilibré. La réalisation de plusieurs milliers d'IA en quelques mois avant que les premiers descendants ne naissent peut ainsi avoir des conséquences importantes pour la filière.

Tableau I-2 : Statistiques illustrant les différences de variabilité génétique entre les 3 principales races laitières et allaitantes françaises.

Source : projet VARUME, Minéry, 2022 (idele.fr/detail-dossier/varume-resultats-2022).

Race	Aptitude	Femelles 2019-2022	Nombre d'ancêtres efficaces	Contribution de l'ancêtre principal (%)	Nombre d'ancêtres expliquant 50% des gènes	Consanguinité moyenne 2019-2022(%)
Holstein	Laitière	3 058 373	19	12,4	7	5,8
Montbéliarde	Laitière	737 866	17	15,1	6	6,0
Normande	Laitière	333 778	21	12,2	8	5,9
Charolaise	Allaitante	726 827	150	2,9	58	1,1
Limousine	Allaitante	574 441	146	3,1	71	1,1
Blonde d'Aquitaine	Allaitante	300 704	58	7,5	29	2,0

Les déterminismes d'anomalies génétiques les plus fréquemment rencontrés chez le bovin sont présentés ci-dessous dans la Figure I-4.

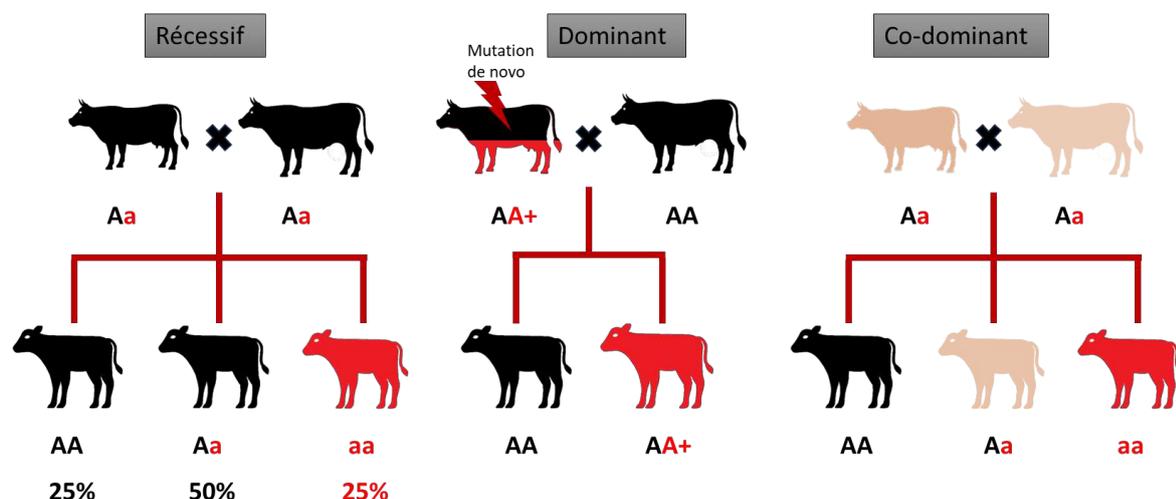


Figure I-4 : Schéma récapitulant les principales caractéristiques des déterminismes génétiques récessif, dominant ou co-dominant.

Sont représentés en noir les individus sains, en rouge les individus atteints et en gris-rouge les individus présentant un phénotype intermédiaire. En dessous de chaque animal figure son génotype pour un locus bi-allélique. Dans le cas d'une anomalie récessive, deux parents hétérozygotes sains auront 25% de leurs veaux atteints, 50% de porteurs hétérozygotes sains et 25% d'homozygotes pour l'allèle sauvage. Dans le cas d'une anomalie dominante, un parent porteur hétérozygote accouplé avec des partenaires non porteurs aura 50% de ses veaux atteints. Dans le cas particulier d'un mosaïcisme germlinal, le parent impliqué aura une proportion N de cellules germinales porteuses d'une mutation dominante et produira une proportion N/2 de gamètes porteuses de cette mutations dominantes et de veaux atteints. Dans le cas d'un déterminisme génétique co-dominant, chaque allèle s'exprime et les porteurs hétérozygotes

ont un phénotype intermédiaire entre ceux des autres génotypes. Deux hétérozygotes accouplés ensemble auront donc 25% de veaux normaux, 50% de veaux partiellement atteints et 25% d'entièrement atteints.

5.1. Les anomalies récessives

Les anomalies autosomiques récessives se caractérisent par le fait que le phénotype délétère ne se manifeste que chez les individus porteurs homozygotes de la mutation, les hétérozygotes étant porteurs sains. Comme évoqué précédemment l'augmentation de leur fréquence est due à une dérive génétique (variation brusque des fréquences alléliques) induite par la sur-utilisation de la semence de certains taureaux d'insémination. Au sein de la filière bovine, ces anomalies sont considérées comme particulièrement coûteuses en raison de la structure de la population et du schéma de sélection en vigueur. En effet, au moment où elles sont détectées chez les premiers homozygotes, elles sont déjà présentes chez un grand nombre d'hétérozygotes, et l'éradication ou la diminution de de la fréquence allélique peut être longue.

5.2. Les anomalies co-dominantes et la sélection équilibrante

On peut également rencontrer des variants génétiques qui ont un effet favorable à l'état hétérozygote et défavorable à l'état homozygote. Ces allèles sont sélectionnés favorablement et voient leur fréquence augmenter pendant des générations avant la naissance des premiers homozygotes et la découverte de l'anomalie génétique associée. On parle de sélection équilibrante et la plupart de ces allèles ont un effet co-dominant. C'est le cas par exemple des mutants du gène *ACAN*, codant une protéine essentielle à la croissance en longueur des os, à l'origine à la fois du phénotype « pattes courtes » à l'état hétérozygote (très prisé en race Dexter du fait de leur faible hauteur au garrot) mais aussi du syndrome léthal des veaux bouledogues à l'état homozygote (Cavanagh et al. 2007). De même en race Charolaise, la mutation dans le gène *KIF1C* responsable de l'ataxie à l'état homozygote donne un avantage en terme de développement musculaire chez l'hétérozygote (Duchesne et al. 2018). Elle a donc été sélectionnée pendant des décennies pour atteindre une fréquence allélique de 12% avant sa contre-sélection. Autres exemples, la sélection sur des phénotypes extrêmes de musculature en race Blanc Bleu Belge a été directement responsable de l'émergence de 3 des 7 défauts génétiques recensés dans cette race à cause de ce phénomène de sélection équilibrante (en l'occurrence les dystonies musculaires congénitales 1 et 2, et le syndrome de la queue tordue,

dues à des mutations des gènes *ATP2A1*, *SLC6A5* et *MRC2* ; Charlier et al. 2008; Fasquelle et al. 2009).

5.3. Les anomalies dominantes associées à un mosaïcisme germinale

Contrairement aux précédentes, les anomalies dominantes sont relativement faciles à gérer en sélection car les porteurs de la mutation expriment systématiquement le phénotype anormal lorsque la pénétrance est complète. Sauf si le phénotype apparaît tardivement, les taureaux porteurs peuvent donc être éliminés efficacement des schémas de reproduction. Il n'empêche que les anomalies dominantes peuvent avoir des conséquences sévères dans les cas de mosaïcisme somatique ou germinale affectant des taureaux d'insémination. Lors de la formation de l'embryon, des mutations *de novo* peuvent survenir, engendrant un mosaïcisme somatique. Par exemple, si une mutation se produit dans un embryon au stade quatre cellules, il en résultera un animal constitué d'une mosaïque de cellules non mutées et de cellules porteuses de cette mutation dont la proportion sera variable au niveau de chaque organe en fonction du devenir des différentes lignées cellulaires et d'environ ¼ à l'échelle de l'organisme entier. Si les cellules germinales sont affectées on parlera de mosaïcisme germinale. Dans ce cas l'individu mosaïque produira une proportion de gamètes porteurs de cette mutation égale à la moitié du pourcentage de mosaïcisme germinale et une partie de ses veaux seront porteurs hétérozygotes de cette mutation (dans l'ensemble de leurs cellules cette fois) et atteints dans le cas d'un déterminisme dominant. La plupart des individus mosaïques sont asymptomatiques même s'il arrive parfois qu'ils présentent des signes cliniques atténués lorsque la proportion de cellules mutées est élevée. Ce fut le cas pour une délétion de 3,7 Mb en race Charolaise responsable d'une absence complète de cornes et d'autres anomalies du développement (*Polled and Multisystemic Syndrome*) dans la descendance d'un taureau Charolais qui lui présentait des croûtes cornées de quelques centimètres et donc un phénotype intermédiaire (Capitan et al. 2012).

5.4. Les remaniements chromosomiques et aneuploïdies

Les remaniements interchromosomiques sont un type d'anomalie génétique défini par le transfert de matériel génétique entre chromosomes non-homologues (c'est-à-dire de paires différentes). Il en existe différentes sortes dont les plus impactantes sont les translocations

réciroques impliquant les bras longs des chromosomes et qui résultent le plus souvent d'erreurs de méiose mâle (Burssted et al. 2022). Les individus porteurs de caryotypes dits transloqués équilibrés pour les chromosomes i et j ont deux copies de chaque gène mais sont porteurs d'un fragment du chromosome i sur un de ses chromosomes j et inversement, ce qui va conduire à la production d'une proportion élevée de gamètes déséquilibrés. Si l'on schématise, un tel animal accouplé avec des femelles au caryotype normal donnera environ ¼ de descendants au caryotype normal, ¼ de descendants transloqués équilibrés (comme lui), ¼ de monosomiques partiels pour le chromosome i et trisomiques partiels pour le chromosome j, et enfin ¼ de monosomiques partiels pour chromosome j et trisomiques partiels pour le chromosome i. Ces deux dernières catégories d'animaux aneuploïdes (c'est-à-dire n'étant pas diploïdes pour l'ensemble de leur génome) auront des degrés d'affection divers en fonction du dosage et de la fonction des gènes impactés, allant de la mort embryonnaire à des malformations impactant leur survie à la naissance et au-delà. Bien que ce type d'anomalie n'affecte en général qu'un taureau, les conséquences dans sa descendance sont majeures.

6. Présentation de l'ONAB

Comme nous venons de le voir, l'émergence régulière de nouvelles anomalies génétiques est un processus naturel que l'on ne peut empêcher. Par contre, le plus tôt on agit, le mieux on réduit leurs impacts sur la filière. C'est ce qui a motivé la création de l'Observatoire National des Anomalies Bovines (ONAB) en France, une plateforme d'hérédo-surveillance créée en 2002 à l'initiative de l'INRA pour la gestion des anomalies héréditaires (www.onab.fr, Grohs et al. 2016). L'ONAB regroupe les principaux acteurs de l'élevage bovin de notre pays tels IDELE, ELIANCE, Races de France (la fédération des OS), les quatre Ecoles Nationales Vétérinaires françaises, la SNGTV (Société Nationale des Groupements Techniques Vétérinaires) et GDS France (la fédération des GDS), sous l'égide du Ministère de l'Agriculture. Ses missions sont de fédérer la filière, repérer les émergences, susciter les recherches, informer les professionnels et les conseiller sur la gestion de ces anomalies. Son activité de surveillance repose sur des remontées d'observations venant du terrain. En pratique différents acteurs de la filière (éleveur, inséminateur, technicien d'OS ou de contrôle de performances, vétérinaires...) déclarent environ 500 cas par an à l'aide d'une fiche dédiée qui peut être remplie en ligne ou téléchargée depuis le site de l'ONAB (<https://www.onab.fr/declarer-une-anomalie>).

Cette fiche de déclaration a été conçue pour permettre aux différents acteurs, quels que soient leurs niveaux de connaissances et de compétences, de décrire de façon simple et normalisée le phénotype des animaux atteints en cochant s'il existe ou non un défaut visible pour une liste de catégories et sous-catégories anatomiques. Cette normalisation concourt également à ce que des anomalies identiques déclarées de façon indépendante soient assimilées à la même entité pathologique. Les déclarations sont souvent accompagnées de prélèvements de différents types (biopsies d'oreille, prélèvement sanguins, poils, ...), et si possible d'informations complémentaires comme des photographies ou des comptes rendus d'examens cliniques ou autopsies.

Cette stratégie s'est révélée fructueuse et depuis sa création l'ONAB a contribué à la découverte et à la caractérisation d'une quarantaine d'anomalies génétiques dont une vingtaine a déjà fait l'objet d'articles scientifiques.

Parmi les plus représentées on trouve les anomalies affectant la peau et ses annexes comme l'épidermolyse bulleuse jonctionnelle récessive (p.ex. délétion d'*ITGB4* et variant affectant l'épissage d'*ITGB6* en race Charolaise ; Michot et al. 2015 ; Bousaha et al. 2023a), la dysplasie ectodermique anhidrotique (variant récessif entraînant un décalage de lecture d'*EDAR* en race Charolaise et inversion affectant *EDA* et le centre de l'inactivation du chromosome X chez une velle Holstein ; Bourneuf et al. 2017; Escouflaire et al. 2019), des dermatosparaxies dominantes en Holstein (Grohs et al. ; in prep) ou encore des défauts de différenciation des bourgeons de cornes (mutations dominantes de *TWIST1* et *ZEB2* en race Charolaise ; Capitan et al. 2009, Capitan et al. 2012).

Viennent ensuite des défauts du développement squelettique tels la syndactylie (substitutions de deux nucléotides adjacents de *LRP4* à déterminisme autosomique récessif ; Duchesne et al. 2006), l'achondrogénèse de type II causée par des mutations *de novo* dominantes de *COL2A1* (Daetwyler et al. 2014; Bourneuf et al. 2017) en race Holstein, l'ostéogénèse imparfaite en races Simmental et Normande causées par des mutations *de novo* dominantes de *COL1A1* (Bourneuf et al. 2017; Corbeau et al. 2024), ou encore des chondrodysplasies rhizoméliques ponctuées en races Aubrac et Bazadaise (Boulling et al. en prep ; Grohs et al. en prep).

L'ONAB s'est également penché sur plusieurs pathologies neuromusculaires récessives comme les veaux tourneurs en race Rouge des Prés (variant faux-sens de *SLC25A46* ; Duchesne et al. 2017), l'ataxie en Charolaise (variant faux-sens de *KIF1C* ; Duchesne et al. 2018), le palais-fendu en Limousin (décalage du cadre de lecture de *MYH3* ; Vaiman et al. 2022).

Les anomalies traitées à l'ONAB sont répertoriées sur son site internet (<https://www.onab.fr>) et dans le catalogue *Online Mendelian Inheritance in Animals* (OMIA ; <https://omia.org/>), qui liste l'ensemble des anomalies génétiques connues chez les animaux domestiques, notamment le bovin.

L'ensemble des variants causaux identifiés par l'ONAB ou par des équipes étrangères font l'objet de tests génétiques disponibles sur la puce utilisée en routine pour les évaluations génomiques (actuellement la puce Illumina EuroGMD) afin de permettre aux professionnels de chaque race de gérer leurs anomalies. L'ONAB communique régulièrement sur ses nouvelles découvertes et sur les projets en cours sur son site internet ou d'autres média (cours aux étudiants vétérinaires, aux techniciens d'IA, articles de vulgarisation, journées d'information) pour favoriser la remontée de cas et animer ce réseau.

7. Méthodes de cartographie et identification de *loci* responsables d'anomalies génétiques

La recherche des mutations causales et la compréhension des mécanismes impliqués reposent sur l'analyse de diverses informations, allant des caractéristiques phénotypiques à l'information génomique des animaux. Trois éléments sont essentiels dans les différentes méthodes de cartographie : le pedigree des individus, les données génétiques plus ou moins exhaustives allant de marqueurs aux séquences de génome complet, et les descriptions phénotypiques plus ou moins fines selon les sujets étudiés. Les approches usuelles de cartographie vont du phénotype au génotype, en recherchant une liaison (familiale) ou une association (à l'échelle de la population) entre un phénotype décrit et des variants génétiques. Ces approches sont parfois appelées génétiques « en avant » (*forward genetics*) ou « clonage positionnel » (Figure I-5).

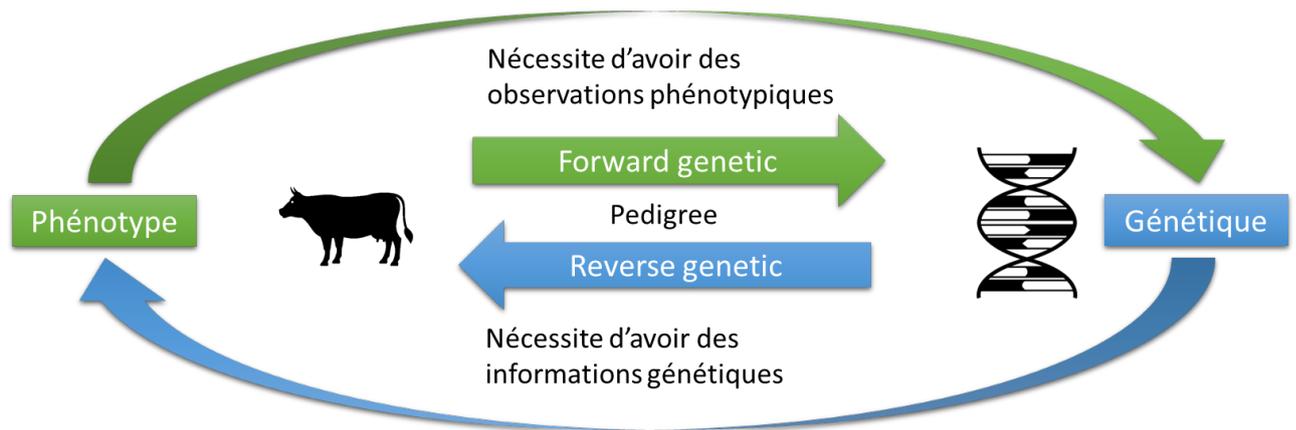


Figure I-5 : Représentation schématique de la génétique en avant "forward genetic" et de la génétique inverse "reverse genetic".

La flèche verte en haut de la figure représente les techniques de "forward genetics", dont l'objectif est de déterminer la cause génétique d'un phénotype spécifique observé. La flèche bleue représente les techniques de "reverse genetics", qui se basent sur l'étude des variations génétiques pour comprendre leurs effets sur le phénotype. Le pedigree est placé au centre de la figure car il est une information cruciale pour les deux méthodes. En effet, il permet de tracer la transmission des traits et des variations génétiques à travers les générations, facilitant ainsi l'identification des relations entre génotype et phénotype.

Dans ce contexte, la génétique « en avant » s'attache à découvrir les mutations causales du génome associées à des phénotypes spécifiques. Les méthodes telles que la cartographie par homozygotie, la recherche de déséquilibre de transmission haplotypique et les études d'association sur le génome entier (GWAS) seront abordées dans ce manuscrit (Chapitre V ; Chapitre IV).

Au contraire, la génétique inverse a comme objectif de découvrir des mutations qui seraient potentiellement responsables d'anomalies sur la base de leur annotation. Elle s'affranchit donc dans un premier temps des informations phénotypiques pour se focaliser sur les informations génomiques et de pedigree. Les phénotypes sont ensuite étudiés pour confirmer ou non ces variants candidats.

Dans ce manuscrit, nous explorerons les méthodes de génétique inverse à l'échelle de la séquence et dans son acception plus large à l'échelle d'haplotypes présentant un déficit en homozygotes parmi les animaux génotypés pour les évaluations génomiques.

7.1. Génétique « en avant »

7.1.1. Cartographie par homozygotie

La cartographie par homozygotie repose sur l'hypothèse d'un déterminisme récessif et d'une origine unique de l'allèle mutant (Figure I-6). Elle cherche à détecter une même région homozygote chez tous les animaux génotypés atteints, et rarement homozygote dans la population contrôle. Cette approche suppose l'identité par descendance du segment partagé, c'est-à-dire que le segment homozygote est hérité d'un ancêtre commun aux deux parents des animaux atteints. Chez le bovin elle a été employée pour la première fois avec des marqueurs microsatellites à la fin des années 1990. Elle a notamment contribué à l'identification de mutations causales dans le gène de la myostatine (*MSTN*) pour le phénotype culard ou hypermuscularité dans diverses races (Grobet et *al.* en 1997). Toutefois son essor a vraiment commencé avec l'arrivée sur le marché des puces à SNP qui ont permis de baliser le génome bovin avec des dizaines de milliers de marqueurs au lieu de centaines (voir la publication de Charlier et *al.* 2008 qui fait référence). Cette approche est particulièrement puissante et donne de bons résultats avec un nombre limité de cas ($n \geq 2$) et de contrôles ($n \geq 100$). Sa résolution, c'est-à-dire la taille du segment partagé entre les cas, se situe en général entre deux et une dizaine de mégabases. Elle dépend du nombre de recombinaisons, et donc du nombre de cas génotypés et du nombre de générations entre l'ancêtre et les cas.

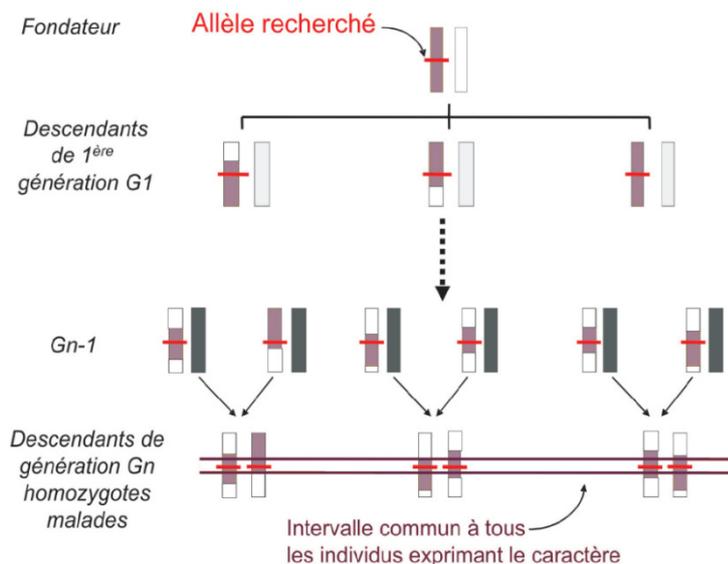


Figure I-6 : Représentation de la cartographie par homozygotie d'un locus récessif (Source ONAB 2023).

La mutation recherchée est apparue spontanément chez le fondateur, puis a été transmise à ses descendants de la génération G1, et ainsi de suite jusqu'à la génération Gn. Ces animaux (Gn) sont issus d'accouplements entre des individus hétérozygotes pour la mutation, et ils présentent tous le phénotype affecté. Ils sont homozygotes pour le

variant causal mais aussi pour un segment de longueur variable entourant ce variant. Cette approche de cartographie permet de réduire au maximum la zone où la mutation est susceptible de se situer, en identifiant l'intervalle partagé par tous les animaux affectés.

7.1.2. Déséquilibre de transmission

Une approche complémentaire au clonage positionnel repose sur le déséquilibre de transmission, qui se révèle un outil particulièrement efficace dans l'analyse de la descendance d'un taureau. On parle de déséquilibre lorsque le taux de transmission de chaque allèle dévie significativement de 0,5. Il peut être analysé par marqueur ou, plus efficacement, par haplotype. Cette dernière approche nécessite cependant un phasage préalable des génotypes, c'est-à-dire la détermination précise de l'origine parentale des allèles des marqueurs. On parle alors de déséquilibre de transmission haplotypique. Pour chaque descendant, on identifie l'haplotype parental reçu. Lorsqu'un phénotype délétère est observé chez certains veaux, cette technique permet d'identifier l'haplotype du taureau surreprésenté chez les individus atteints, tout en étant peu présent chez les individus témoins, et donc de cartographier une région / un *locus* d'intérêt.

Cette approche présente l'avantage de fournir une base génétique rigoureuse, en se concentrant spécifiquement sur les informations héritées du parent mâle. Sa précision est remarquable lorsque le nombre de cas et de contrôles est suffisant, offrant ainsi une méthode robuste pour identifier le *locus* responsable du phénotype observé sans a priori sur le déterminisme génétique attendu, même si elle est généralement mise en œuvre pour cartographier des *loci* dominants. Toutefois, elle requiert d'avoir des informations génétiques et phénotypiques détaillées sur les veaux atteints, les veaux témoins, et sur le père. Cette méthode s'avère particulièrement utile dans les filières bovines, caractérisées par une grande taille des descendance des mâles. Par ailleurs, dans le cadre de la sélection génomique, les typages subissent des traitements préalables et le phasage et l'imputation sont effectués en routine, facilitant ainsi la mise en place d'une telle méthode.

7.1.3. GWAS sur des phénotypes directs ou indirects (« proxy »)

Lorsque le phénotype est directement disponible en nombre suffisant, l'analyse d'association à l'échelle d'une population (et non de familles isolées) est la méthode de choix. Ainsi, si les animaux morts comme les vivants sont génotypés, l'approche la plus simple consiste à analyser

ce phénotype de mortalité, une information souvent disponible pour la plupart des animaux de la race. Les GWAS sur la mortalité seront abordées en détail dans le Chapitre IV. On peut alors utiliser un modèle additif comme c'est le cas pour cartographier des QTL de production par exemple ou bien opter pour un modèle prenant en compte un effet de dominance ou récessivité afin de détecter des *loci* mendéliens. Outre les taux de mortalité, on peut également étudier des « *proxies* » ou « phénotypes par procuration » en français comme la taille ou la note d'état corporel des animaux pour détecter des anomalies génétiques qui impactent la santé et la croissance des animaux mais ne sont pas repérées par les dispositifs du type de l'ONAB, faute de signes cliniques distinctifs. Cette approche combinant GWAS avec un modèle non-additif sur des phénotypes indirects a été utilisée pour la première fois chez le bovin par Reynolds *et al.* (2021). Après une étape préliminaire d'imputation des données de génotypage sur puce à SNP vers l'ensemble de la séquence du génome, elle leur a permis de mettre en évidence six *loci* récessifs responsables de maladies génétiques expliquant ainsi une partie de la dépression de consanguinité. Cette méthode exige un grand nombre d'informations phénotypiques, génétiques et généalogiques, et elle est généralement appliquée dans des populations sélectionnées depuis de nombreuses années.

Les approches de type GWAS qui nécessitent des individus avec phénotype et génotype ne sont donc pas adaptées à l'identification des mutations létales à l'état embryonnaire. Par ailleurs, lorsque la fréquence de l'allèle délétère est basse, la proportion d'homozygotes atteints est très faible et seul un effet très fort peut être mis en évidence. Enfin les allèles à faibles fréquences sont en général mal imputés, ce qui réduit encore la puissance de l'approche dans cette situation.

Lorsque les phénotypes sont disponibles mais pas les génotypages, ce qui est le cas le plus fréquent pour les mortalités précoces, une analyse d'association du phénotype moyen intra famille de père est une alternative efficace. L'effectif des familles est bien inférieur à l'effectif d'animaux phénotypés, mais le phénotype moyen est précis, ce qui compense au moins partiellement la perte de puissance de l'analyse. Cela constituera le fondement de notre analyse pour le Chapitre IV nommé « Analyse d'association (GWAS) des taux de mortalité dans les descendance des taureaux ».

7.2. Génétique inverse

7.2.1. Déficients en homozygotes

La génétique inverse au sens large constitue une approche novatrice dans l'identification de mutations létales. Une première approche, la recherche de déficits en homozygotes, a été initialement proposée par VanRaden et *al.* (2011) et a fait ensuite l'objet de nombreuses études.

Cette stratégie exploite la puissance des puces à SNP et le grand nombre de génotypages réalisé dans le cadre de la sélection génomique pour détecter des *loci* récessifs responsables de mort embryonnaire ou périnatale et donc d'une absence ou d'une sous-représentation des homozygotes. Dans une grande population génotypée, la méthode consiste donc à comparer le nombre d'homozygotes observés au nombre d'homozygotes attendus compte tenu des génotypes des parents pour des haplotypes coulissants de 10 à 20 marqueurs couvrant en général 1Mb. Les haplotypes en déséquilibre de liaison avec une mutation délétère présentent un nombre d'observés significativement inférieur au nombre d'homozygotes attendus. Cette stratégie offre l'avantage de détecter des variants létaux récessifs sans qu'il ne soit nécessaire de repérer des individus atteints, de les phénotyper et de les génotyper.

Cette méthode s'est avérée efficace pour détecter des anomalies qui ne sont pas repérées via l'ONAB faute de cas observables (mortalité embryonnaire ou fœtale) ou de signes cliniques spécifiques (mortalité périnatale ou juvénile sans malformation apparente). Son application initiale à la population Holstein américaine a conduit à la détection des haplotypes HH1, HH2 et HH3 (VanRaden et *al.* 2011). Son application aux populations françaises a également révélé des haplotypes liés à la mortalité embryonnaire, aux avortements et à la mortalité juvénile, dont plusieurs caractérisés jusqu'à la mutation causale (Tableau I-3).

7.2.2. Génétique inverse à partir d'annotations des séquences de génome complet (WGS)

La génétique inverse au sens strict exploite les séquences complètes des génomes (WGS). Elle utilise divers outils permettant d'annoter les variants et de prédire leurs conséquences biologiques selon leur nature (SNP, InDels ou variants structuraux), leur localisation génomique (dans des régions codantes, régulatrices ou sans rôle connu) et la fonction des gènes affectés. Elle exploite notamment les bases de données répertoriant les conséquences phénotypiques de

mutant naturels ou induits caractérisés chez l'homme ([omim](#)), la souris ([mgi](#)) et diverses espèces d'élevage ([omia.org](#)) La sélection de variants candidats est réalisée en se concentrant sur ceux provoquant une perte de fonction (c'est-à-dire les substitutions prédites comme délétères, gains ou pertes de codons stop, décalages du cadre de lecture, gains ou pertes du site initiateur de transcription ou encore les variants entraînant une modification de l'épissage) portés à l'état hétérozygote par des individus sains et avec un déterminisme génétique prédit comme récessif. Des sondes spécifiques sont ensuite dessinées permettant de génotyper ces variants candidats sur les puces utilisées en routines pour les évaluations génomiques et qui comportent une partie customisable. Il est alors possible d'obtenir gratuitement des informations génotypiques sur ces variants pour des centaines de milliers d'animaux par an et ensuite de cibler les homozygotes pour les phénotyper ou encore de rechercher un déficit en homozygotes sur variant (et non sur haplotype) afin de valider ou d'invalider la prédiction. C'est une approche très puissante qui permet de détecter des anomalies qui passent inaperçues comme la perte progressive de vision en race Normande (mutant de *RP1*, Michot et al. 2016) ou d'anticiper de futures émergences d'anomalies récessives en ciblant les mutations *de novo* comme démontré dans une étude « preuve de concept » avec la dysplasie ectodermique anhidrotique confirmée par l'étude des premiers accouplements consanguins entre descendants d'un taureau d'IA Charolais (mutant de *EDAR* ; Bourneuf et al. 2017). Toutefois la génétique inverse présente un taux de faux positifs élevé comme l'ont montré Charlier et al. (2016) dans le cadre d'une étude qui leur a permis d'identifier et de confirmer neuf mutations létales à l'état embryonnaire. Ce faisant ils ont estimé le taux de faux positifs à 85% pour les variants tronquants (décalages du cadre de lecture et gains de codon stop) et 94% pour les faux-sens prédits comme non tolérés.

Ce taux de faux positifs, l'investissement important en moyen requis pour les vérifications et le phénotypage des animaux en ferme, et le besoin d'avoir accès à de nombreux jeux de données explique pourquoi seulement 4 études ont été consacrées à cette approche dans les années 2016-2017, deux par mon équipe d'accueil et deux par l'équipe de Carole Charlier à l'Université de Liège, en Belgique (Li et al. 2016; Michot et al. 2016; Charlier et al. 2016; Bourneuf et al. 2017) Toutefois les travaux de génétique inverse ont repris ces dernières années à la faveur d'un effort de séquençage de plus de 500 bovins Français de différentes races dans le cadre du projet SeqOccIn (2014-2020) financé par le Fonds Européen de Développement Régional et la région Occitanie.

L'étude de Seqoccin (2016) a abouti à la découverte de 1 548 nouveaux variants avec une annotation délétère (Boussaha et al. 2023b). Certains variants ont été confirmés et caractérisés,

avec une validation à grande échelle et l'identification d'individus d'intérêt par le biais de génotypages de puces SNP (Boussaha et al. 2023b).

Ces études suggèrent l'intérêt de séquencer systématiquement les candidats à la reproduction. Elle présenterait divers avantages (identification de tous les variants génétiques, y compris les plus rares, imputation plus précise au niveau des séquences, identification facilitée des variants candidats de QTL) et, concernant notre sujet, elle permettrait une identification précoce des nouveaux défauts génétiques potentiels. Nous en reparlerons à la fin de ce manuscrit dans la discussion générale.

7.2.3. Détection d'anomalies chromosomiques par étude de déséquilibre de liaison

L'analyse des déséquilibres de liaison entre chromosomes offre une méthode innovante pour détecter des anomalies chromosomiques, comme démontré par Jourdain et al. (2023). Cette approche novatrice a été spécifiquement conçue pour identifier des déséquilibres de liaison anormaux intra famille. Pour être efficace, elle nécessite au moins 30 veaux génotypés de même père.

Le déséquilibre de liaison (DL) constitue le fondement de cette méthode, représentant une mesure inverse du rapprochement physique entre deux *locus*. Il caractérise l'écart entre la fréquence d'un couple d'allèles de deux *locus* et le produit des fréquences alléliques correspondantes. Il est souvent défini sous forme de corrélation (r) ou de r^2 . Un DL nul reflète une absence d'association et donc de déséquilibre, tandis qu'un r^2 élevé traduit une forte association.

En théorie, les chromosomes ségrègent indépendamment les uns des autres à la méiose et le DL entre *loci* de différents chromosomes est supposé nul ou faible. La détection de DL significativement différent de zéro entre des marqueurs situés sur des chromosomes différents témoigne d'une liaison physique entre ces chromosomes. L'analyse approfondie de 5 571 familles de père comptant au moins 30 descendants, a permis à Jourdain et al. (2023) de mettre en lumière 12 réarrangements chromosomiques. Ces résultats ont été validés par des analyses cytogénétiques et de séquençage *long reads*, conférant robustesse et crédibilité à cette approche de détection des anomalies chromosomiques au sein des populations bovines.

Tableau I-3 : Liste des anomalies récessives détectées dans les populations bovines Holstein, Montbéliarde et Normande avec un variant causal connu.

Sont renseignés la race affectée, le phénotype, la méthode qui a permis la cartographie, le nom de l'haplotype ou de l'anomalie, le gène associé, le lien vers la description dans la base de données OMIA, le variant causal, le type de mutation, et la référence bibliographique décrivant pour la première fois le variant causal.

Race	Phénotype	Méthode de détection	Haplotype	Gene	OMIA	Variant causal	Type de variant	Référence pour le variant
Holstein	Mort embryonnaire	Déficit en homozygotes	HH1	APAF1	1	chr5:g.62810245C>T	SNV(nonsense)	(Adams et al. 2016)
Holstein	Mort embryonnaire	Déficit en homozygotes	HH2	IFT80	1823	chr1:g.107172616delT	délétion, petite (<=20)	(Y. Yang et al. 2022)
Holstein	Avortement	Déficit en homozygotes	HH3	SMC2	1824	chr8:g.93753358T>C	SNV(missense)	(McClure et al. 2014)
Holstein	Mort embryonnaire	Déficit en homozygotes	HH25	RIOX1	2547	chr10:g.84938408_84938437del	30bp délétion (in-frame deletion)	(Häfliger et al. 2022)
Holstein	Mortalité	Déficit en homozygotes	HH13	KIR2DS1	1836	chr18:g.62758881G>A	SNV(nonsense)	(Häfliger et al. 2022)
Holstein	Mortalité	Déficit en homozygotes	HH35	PCDH15	2548	chr26:g.5325675C>G	SNV(missense)	(Häfliger et al. 2022)
Holstein	Avortement	Déficit en homozygotes	HH4	GART	1826	chr1:g.1997582A>C	SNV(missense)	(Fritz et al. 2013)
Holstein	Avortement	Déficit en homozygotes	HH5	TFB1M	1941	chr9:93,233kb to 93,371kb_del	gross deletion	(Schütz et al. 2016)
Holstein	Mort embryonnaire	Déficit en homozygotes	HH6	SDE2	2149	chr16:g.29020700A>G	SNV(Start lost)	(Fritz et al. 2018)
Holstein	Mort embryonnaire	Déficit en homozygotes	HH7	CENPU	1830	chr27:g.15123637_15123640del	5bp délétion (splice site)	(Hozé et al. 2020)
Holstein	Faiblesse musculaire	Déficit en homozygotes	Recumbency	CACNA1S	2819	chr16:g.79613592C>T	SNV(missense)	(Al-Khudhair et al. 2024)
Holstein	Mortalité	GWAS	CDH	APOB	1965	chr11:c.398_399insNSTd119:g.1_1299	Insertion 1299 pb	(Menzi et al. 2016)

Holstein	Avortement	Cartographie par homozygotie	CVM	SLC35A3	1340	chr3:g.43261945C>A	SNV(missense)	(Thomsen et al. 2006)
Holstein	Syndactylie	Cartographie par homozygotie	HHM	LRP4	963	chr15:g.76800972_76800973delinsAT	Délétion+insertion, petite (<=20)	(Duchesne et al. 2006)
Holstein	Parakératose	Cartographie par homozygotie	BHZD	SLC39A4	593	chr14:g.537516G>A	SNV(missense)	(Yuzbasiyan-Gurkan et al. 2006)
Holstein	Ehlers-Danlos	Cartographie par homozygotie		EPYC	1716	chr5:g.20856381C>A	SNV(missense)	(Tajima et al. 1999)
Holstein	Mort embryonnaire	Cartographie par homozygotie puis déficit	HHD (DUMPS)	UMPS	262	chr1:g.69151931C>T	SNV(stop-gain)	(Schwenger et al. 1993)
Holstein	Mortalité	GWAS proxy		FGD4	2374	chr5:g.77262490C>T	SNV(splicing)	(E. G. M. Reynolds et al. 2021)
Holstein	Mortalité	Cartographie par homozygotie puis déficit	HHB (BLAD)	ITGB2	595	chr1:g.144770078T>C	SNV(missense)	(Shuster et al. 1992)
Holstein	Tetradysmelia	Cartographie par homozygotie		RSPO2	2297	chr14:g.56451029_56501201delinsTGACAA	délétion+insertion, petite (<=20)	(Becker et al. 2020)
Holstein	Mort embryonnaire	Génétique inverse (WGS)		RABGGTB	2037	chr3:g.69060748T>C	SNV(missense)	(Charlier et al. 2016)
Holstein	Mort embryonnaire	Génétique inverse (WGS)		RNF20	2038	chr8:g.91297797A>T	SNV(missense)	(Charlier et al. 2016)

Holstein	Mortalité+ Avortement	Déficit en homozygotes	HH0 (brachypina)	FANCI	151	chr21:g.20773899_20777226del	Délétion, grande (>20)	(Charlier et al. 2012)
Holstein	Mortalité	HHED (cf Chapitre V)	BLIRD	ITGB7		chr5:g.26807079G>A	SNV(missense)	(Besnard et al. 2023)
Holstein		HHED	H29P38	BBS1		chr29:g.44558958C>T	SNV(stop-gained)	(Besnard et al. 2023)
Holstein - Charolais - Normande	Cécité	Génétique inverse (WGS)		RP1	2029	chr14:g.22340665_22340666insA	Insertion, petite (<=20)	(Michot et al. 2016)
Holstein	Retard de croissance	GWAS proxy		GALNT2	2375	chr28:g.2281801G>A	SNV(missense)	(E. G. M. Reynolds et al. 2021)
Montbeliarde	Avortement	Déficit en homozygotes	MH1	PFAS	1827	chr19:g.27895397C>T	SNV(missense)	(Michot et al. 2017)
Montbeliarde	Avortement	Déficit en homozygotes	MH2	SLC37A2	1828	chr29:g.28510651C>T	SNV(nonsense)	(Fritz et al. 2013)
Montbéliarde	Avortement	Génétique inverse (WGS)		SLC35A3	1340	chr3:g.43268369G>T	SNV(missense)	(Bourneuf et al. 2017)
Montbéliarde	Mortalité	Cartographie par homozygotie	SHGC	CEP250	1502	chr13:g.64710424C>T	SNV(stop-gain)	(Floriot et al. 2015)
Montbéliarde		HHED	M2P118	NYAP2		chr2:g.113733613A>T	SNV(missense)	(Besnard et al. 2023)
Montbéliarde	Mortalité	HHED	MTCP	NOA1A		chr6:g.72359797_72359798insG	SNV(frameshift variant)	(Besnard et al. 2023)
Normande	Avortement	Déficit en homozygotes	NH7	CAD	2201	chr11:g.72409143T>C	SNV(missense)	(Mesbah- Uddin et al. 2019)
Normande		HHED	N10P65	DMXL2		chr10:g.58850499 G>A	SNV(missense)	(Besnard et al. 2023)

Normande		HHED	N15P64	DEPDC7		chr15:g.63682108T>A	SNV(missense)	(Besnard <i>et al.</i> 2023)
Normande	Mortalité	HHED	ALOP	RFC5		chr17:g.57188407_57188409del	inframe_deletion	(Besnard <i>et al.</i> 2023)
Normande		HHED	N29P34	SNX19		chr29:g.36943623 G>T	SNV(missense)	(Besnard <i>et al.</i> 2023)

Chapitre II. Exploration phénotypique de la mortalité

1. Introduction

Dans le cadre de ma thèse, j'ai entrepris une analyse de la mortalité juvénile au sein des trois principales races bovines laitières en France. Cette exploration avait pour but de dégager des tendances et d'identifier les facteurs clés qui influencent la survie des veaux de façon à les prendre en compte dans les différentes études.

J'ai choisi de diviser la mortalité juvénile en quatre grandes phases : la période périnatale ou néonatale (0 à 2 jours), la période post-natale (3 à 14 jours), la période pré-sevrage (15 à 55 jours) et la période dite post-sevrage (56 à 365 jours). Notons que ces termes sont abusifs car le sevrage est souvent postérieur à 55 jours en France et la période de post-sevrage inclut également la seconde moitié de la période pré-sevrage. D'un point de vue zootechnique, cette distinction se justifie par le fait que la fonction ruminale commence à se mettre en place avec une distribution d'aliments partiellement solides, et que la plupart des épisodes de diarrhées ont déjà eu lieu.

J'ai également limité mon étude aux femelles élevées dans leur ferme de naissance afin d'éviter des biais dus à des changements de conditions environnementales. Le choix d'exclure les veaux mâles de l'étude s'explique par les spécificités de la conduite de ces animaux. En effet, la grande majorité des mâles de race laitière quittent leur élevage de naissance après 2 à 3 semaines de vie pour être élevés et valorisés en veaux de boucherie. La faible valeur économique de ces animaux face aux coûts vétérinaires peut conduire à des défauts de soins et leur allotement peut être la source de challenges immunitaires très variés en fonction des pathogènes présents dans leurs élevages d'origine et dans leur élevage de destination. Les conditions d'élevage sont donc très variables d'un animal à l'autre et l'interprétation des mortalités est quasiment impossible.

2. Matériel et méthodes

Les données utilisées dans cette étude sont issues des données nationales du SNIG (voir le 3.1.1 paragraphe du Chapitre I). Elles ont subi une série de filtres mis en place dans le cadre du projet Morphée (Leclerc et *al.* 2016) afin d'éliminer les individus issus d'élevages avec une proportion trop élevée de données aberrantes.

L'Identification Pérenne Généralisée (IPG) permet d'obtenir pour chaque animal les informations sur les ascendants, sur la date de naissance ainsi que sur la cause et la date de mort.

Ces informations nous permettent de calculer un taux de mortalité pour une période donnée et d'étudier sa variation en fonction de différents facteurs (race, année, saison ou encore famille pour les taureaux ayant 100 filles ou plus). Le taux de mortalité est le rapport entre le nombre de veaux morts lors de la période étudiée et le nombre de veaux vivants au début de cette période. Les données sont tronquées de sorte que la date d'extraction moins la date de naissance soit supérieure à la période étudiée pour tous les animaux. En d'autres termes si l'on s'intéresse à la survie sur une période d'un an, il faut que les animaux étudiés soient nés au moins un an avant la dernière extraction des données. Comme nous l'avons vu dans l'introduction, les analyses porteront exclusivement sur les femelles de races Holstein, Montbéliarde et Normande. Les effectifs des femelles disposant d'informations exploitables par race et par année de naissance sont présentés dans le Tableau II-1.

Tableau II-1 : Effectifs des veaux femelles étudiés par année de naissance pour les trois principales races laitières françaises.

Année de naissance	Montbéliarde	Normande	Holstein
2000	225263	239747	1114115
2001	226070	230563	1102828
2002	228178	224716	1076251
2003	224393	211741	1022352
2004	218970	200455	971871
2005	221213	191992	962597
2006	218613	180688	935003
2007	216569	172727	907591
2008	231672	174376	979738
2009	235641	162280	961269
2010	230876	159630	965488
2011	237190	154842	981662
2012	241039	150986	996910
2013	245557	147232	1017664
2014	253796	146736	1053116

2015	258635	140900	1076817
2016	250114	135540	1065753
2017	232963	127749	1004355
2018	230514	125054	1005110
2019	217153	117559	943840
2020	212217	112846	923195
2021	211991	107214	886376
2022	204320	100492	841066

3. Résultats et discussion

Dans un premier temps, je me suis intéressé à l'évolution annuelle de la mortalité juvénile totale (calculée de 0 à 365 jours après la naissance) pour les races Holstein, Montbéliarde et Normande (Figure II-1).

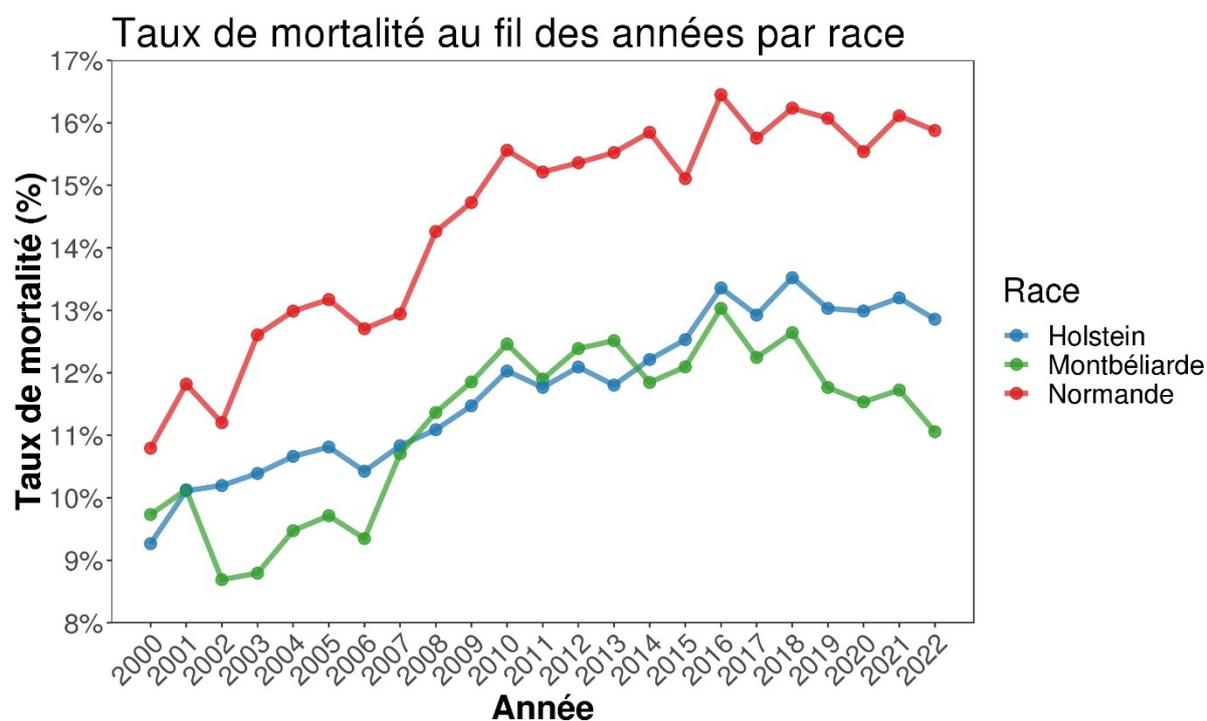


Figure II-1: Evolution annuelle du taux de mortalité juvénile totale des veaux femelles de races Holstein, Montbéliarde et Normande de l'année 2000 à 2022.

La mortalité juvénile des veaux femelles n'a fait que croître entre 2000 et 2016 dans les trois races, avant de se stabiliser de 2017 à 2022 en Normandie et Holstein, et même de diminuer faiblement en Montbéliarde sur cette période. Ce changement d'inflexion à partir de 2016 concorde avec le déploiement sur le terrain de logiciels d'accouplements prenant en compte le taux de consanguinité des produits à naître et les risques de naissances d'animaux homozygotes pour les anomalies génétiques caractérisées notamment en race Montbéliarde (p.ex. Brochard et *al.* 2018). Malgré tout, en 2022, près d'un veau femelle sur 6 meurt de causes naturelles avant d'atteindre l'âge d'un an en Normandie, 1 veau sur 8 en Holstein, et 1 veau sur 9 en Montbéliarde. Ces chiffres soulignent l'ampleur de ce problème, ayant un impact négatif sur la rentabilité économique des élevages laitiers et sur la prise en compte du bien-être animal en élevage.

J'ai ensuite étudié l'évolution annuelle des taux de mortalité moyens pour chaque sous-période (périnatale, post-natale, pré-sevrage, post-sevrage) et chaque race (Figure II-2).

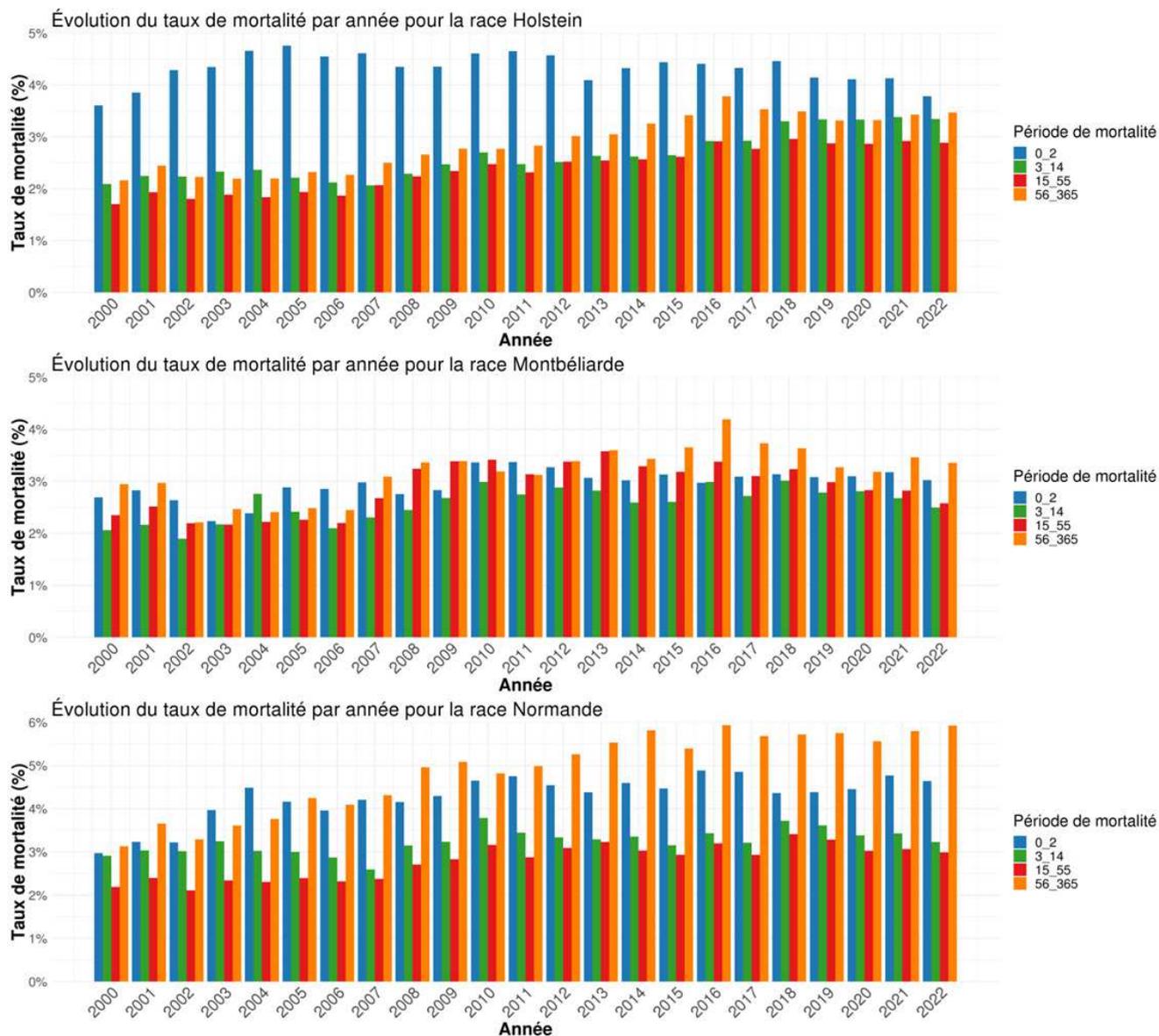


Figure II-2 : Evolution des taux de mortalité en pourcentage de 2000 à 2022 pour chaque période (0_2 périnatale, 3_14 post-natale, 15_55 pré-sevrage, 56_365 post-sevrage) pour les races Holstein, Montbéliarde et Normande.

Tout d’abord, on peut noter que la part des mortalités liées à chaque période est différente d’une race à l’autre : elles apparaissent beaucoup plus équilibrées par exemple en race Montbéliarde qu’en race Holstein, pour laquelle la mortalité périnatale est la classe de loin la plus représentée (avec une tendance à la diminution sur les dernières années).

L’augmentation des mortalités juvéniles de 2000 à 2016, commune à toutes les races, cache des réalités différentes. En race Montbéliarde, elle est surtout due à une augmentation des mortalités durant les périodes post-natale, pré-sevrage et post-sevrage alors que les mortalités périnatales sont restées stables. En race Normande, on observe une augmentation sur toutes les périodes

mais plus prononcée sur la période post-sevrage de 55 à 365 jours et de 0 à 3 jours. En race Holstein, pour les années de naissance les plus anciennes l'augmentation est due une augmentation de la mortalité périnatale, alors qu'elle est surtout associée à une augmentation des mortalités durant la période post-sevrage pour les veaux nés entre 2007 et 2016.

Après 2016, si la mortalité totale semble stagner, elle évolue différemment entre catégories. En race Holstein, la mortalité périnatale semble se réduire et celle des autres périodes semble se dégrader. En 2022, les quatre périodes semblent équilibrées en race Montbéliarde, avec des taux faibles et une légère hausse de la proportion des morts liées à la période post-sevrage. Pour la race Normande, les taux de mortalité sur les périodes post-natale et pré-sevrage sont proches de ceux de la race Montbéliarde (3% contre 2,8 % en Montbéliarde). Au contraire, les mortalités en périodes périnatale et post-sevrage sont bien plus élevées : elles avoisinent respectivement 4,6% et presque 6%. Enfin, la race Holstein semble avoir réduit la mortalité périnatale et possède des taux de mortalités plutôt équilibrés d'environ 3% mais qui restent un peu au-dessus de la race Montbéliarde.

Après m'être intéressé aux évolutions annuelles des taux de mortalité par période je me suis penché sur la variation du taux de mortalité annuel en fonction de la saison de naissance du veau (Figure II-3). Pour des raisons climatiques, de concentration des naissances, de gestion du troupeau, etc... les animaux nés en hiver affichent une mortalité plus élevée tandis que la saison la plus favorable est l'été. Les saisons d'automne et de printemps sont intermédiaires. Ces différences assez fortes entre les saisons devront être prises en compte dans la construction du modèle que nous présenterons au dernier chapitre de ce manuscrit.

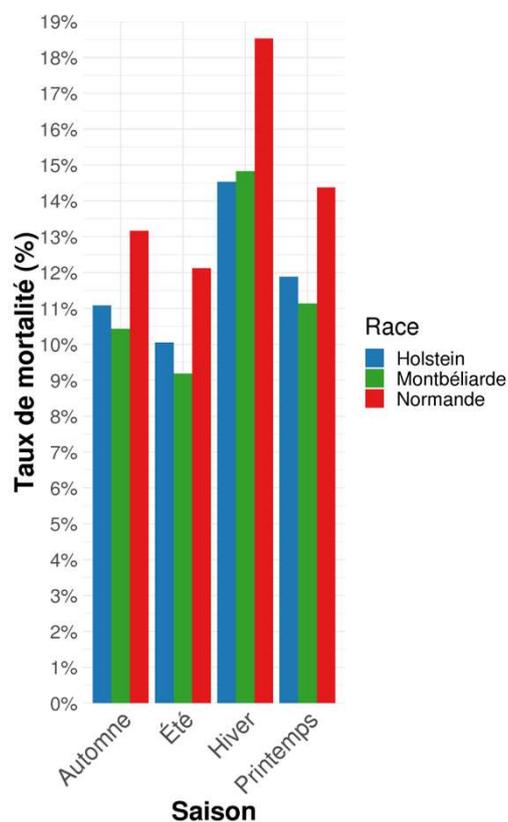


Figure II-3 : Histogramme montrant les différences de taux de mortalité sur la période de 0 à 365 jours pour les veaux femelles de races Holstein, Montbéliarde et Normande selon la saison de naissance.

Enfin, j’ai calculé les taux de mortalité par famille de taureau et par période (cf. Figure II-4). En effet, la structure de population dans les races bovines nous permet d’obtenir de nombreuses familles de taureaux avec plus de 100 filles. Sur ces descendance, le calcul de taux de mortalité est donc pertinent pour illustrer la part de la mortalité juvénile potentiellement d’origine génétique. Comme j’avais pu l’observer en race Holstein dans le cadre de mon premier article (Besnard et *al.* 2022), ces taux présentent une distribution proche de la normalité (ce qui est logique compte tenu du nombre de descendants), avec certaines différences entre races.

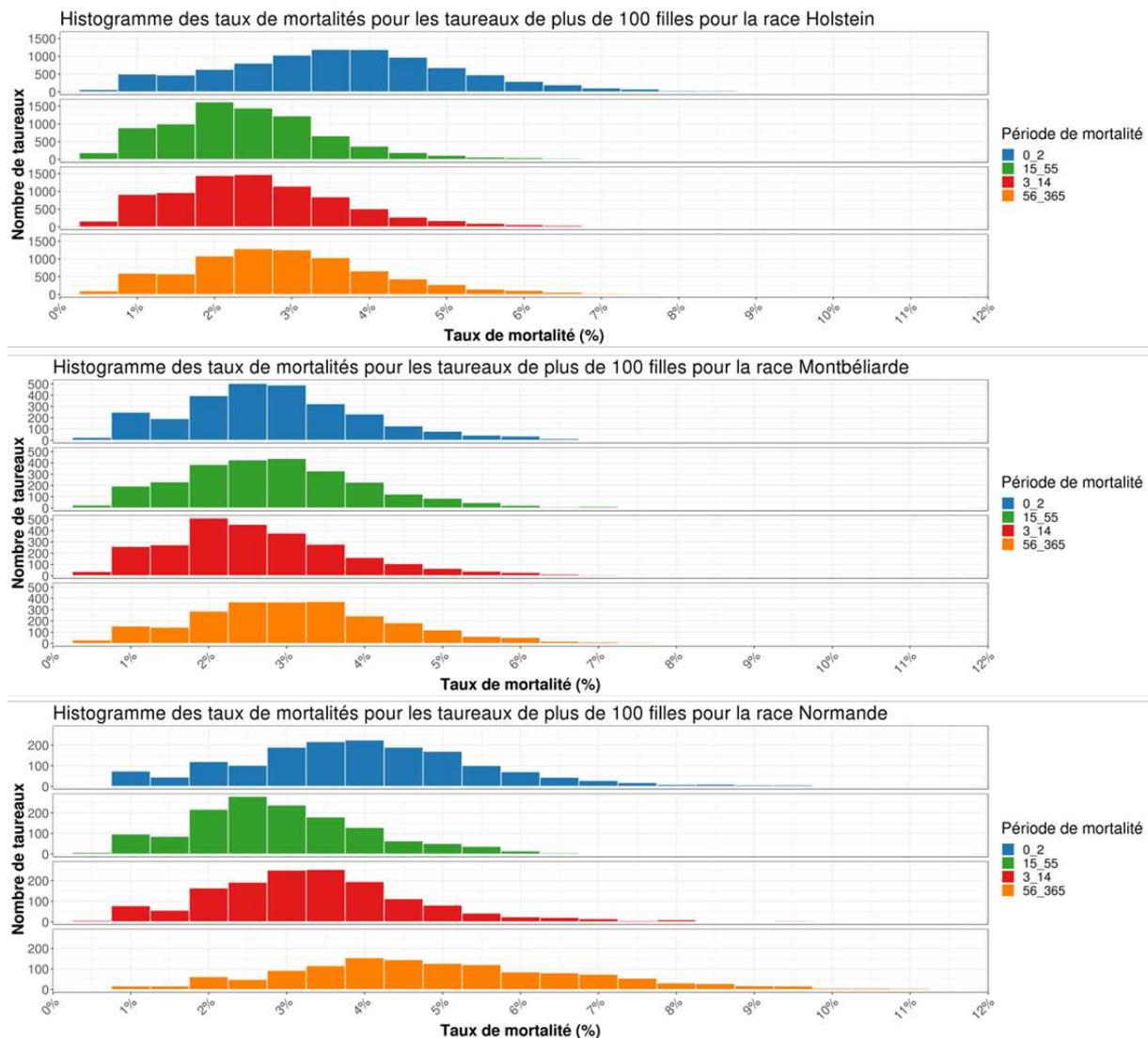


Figure II-4 : Histogramme des taux de mortalité pour les taureaux avec plus de 100 filles pour 4 périodes (périnatale, post-natale, pré-sevrage et post-sevrage).

En race Montbéliarde, comme nous l'avons vu précédemment, les taux de mortalité sont assez similaires entre périodes de vie, et une partie importante des taureaux ont des taux de mortalité proche de zéro pour une période donnée. Pour la Normande, les distributions des taux de mortalité au cours des périodes périnatale (0-2 jours) et post-sevrage (56-365 jours) sont beaucoup plus étendues, suggérant une variabilité plus élevée entre les taureaux. Il en est de même pour la Holstein, où le taux de mortalité pendant la période périnatale montre une variabilité plus élevée que pour les autres périodes.

Enfin, j'ai calculé les taux de mortalité juvénile sur l'ensemble de la première année pour les taureaux des trois races (Figure II.5). On retrouve logiquement les résultats déjà obtenus

précédemment à savoir un taux de mortalité global plus élevé en race Normande et la distribution plus étendue des mortalités par famille de père, déjà observés dans les Figure II.1 Figure II-4.

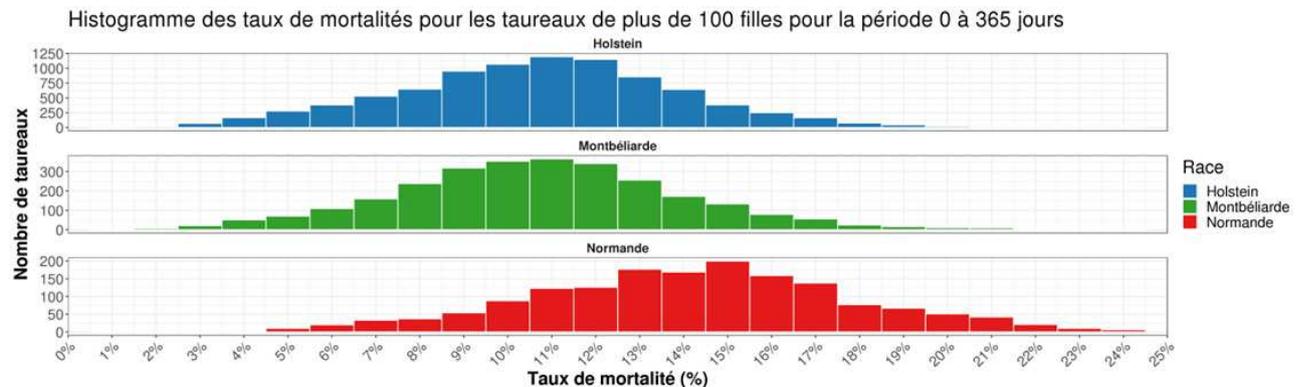


Figure II-5 : Histogramme des taux de mortalité juvénile au cours de la première année (0 à 365 jours) pour l'ensemble des taureaux des races Holstein, Montbéliarde et Normande.

4. Conclusion

Il existe des périodes distinctes sur lesquelles on peut caractériser la mortalité juvénile en fonction de plusieurs facteurs, et avec des réalités biologiques différentes. Nous avons mis en évidence que la mortalité était variable selon l'année de naissance, la race, la saison et même la famille des animaux. Malgré les améliorations que pourrait apporter une meilleure gestion de la mortalité en élevage laitier, ce caractère reste peu sélectionné en raison de sa faible héritabilité. Seule la mortinatalité (mortalité 0-2 jours) fait l'objet d'une évaluation génétique, mais ce caractère n'est pas pris en compte dans l'objectif de sélection. L'utilisation de ces index se limite à des conseils sur le choix des taureaux lors des accouplements, en particulier sur les génisses. La mortalité juvénile n'est pas considérée du tout. Après cette présentation générale des caractéristiques de mortalité, le chapitre suivant présentera la détection et la caractérisation de taureaux extrêmes sur les périodes de mortalités décrites précédemment. Nous essayerons à travers différents chapitres de répondre à la question de recherche suivante :

Comment la sélection génétique basée sur des index de mortalité juvénile, en tenant compte des analyses phénotypiques et génétiques, y compris les anomalies génétiques dominantes et récessives, ainsi que les loci influençant la létalité embryonnaire et la survie juvénile, peut-elle améliorer la gestion de la mortalité en élevage laitier, et quelles sont les implications de l'intégration de la mortalité juvénile dans les objectifs de sélection génomique ?

Chapitre III. Identifications d'anomalies génétiques dominantes

1. Introduction

Dans le cadre de cette étude, nous avons entamé une exploration des cas atypiques, ou "outliers". Pour ce faire, nous avons concentré notre attention sur les taureaux Holstein présentant des taux de mortalité extrêmement élevés dans leur descendance.

Les distributions de mortalité moyenne par descendance présentées dans le chapitre précédent montrent quelques taureaux peu nombreux mais avec des résultats extrêmes. Dans un premier temps, avant d'analyser la variabilité présente entre la plupart des autres taureaux, nous avons d'abord ciblé ces cas extrêmes qui correspondent sans doute à des situations spécifiques à chacun de ces taureaux et sans doute à un déterminisme dominant. Il est à noter que dans ces descendance, les mortalités élevées n'ont pas fait l'objet de remontées à l'ONAB, ce qui montre que l'absence de signe d'appel induit souvent une absence de détection par les plateformes de surveillance. Cette étude constitue une preuve de concept puisque l'étude de deux descendance a permis de décrire deux nouveaux syndromes dominants chez les bovins dus à des mutations de novo. Ces résultats démontrent que l'approche d'analyse de descendance extrême est efficace et se révèle complémentaire, en l'absence de symptômes visibles spécifiques, des approches basées sur la surveillance des cas. Par contre, la structure de l'ONAB a permis de lancer une alerte pour faire remonter des cas, ce qui nous a permis une analyse fonctionnelle complète. Nous avons ainsi pu caractériser deux anomalies autosomiques dominantes, une mutation mosaïque pour un taureau et un réarrangement chromosomique pour le second taureau. Ces travaux sont présentés dans un article scientifique publié dans la revue *Journal of Dairy Science* et qui a fait l'objet de la distinction de Landmark sur le site OMIA. Ce travail a aussi été présenté au congrès mondial de génétique animale (WCGALP) à Rotterdam (Besnard et al. 2022).

J'ai également participé à la caractérisation des mortalités dans la descendance de taureaux atteints d'anomalies chromosomiques et ainsi contribué à un article publié dans la revue *Genome Research* (Jourdain et al. 2023), et faisant sa couverture. Un autre article (Corbeau et al. 2024) est publié dans *Genetics Selection Evolution* sur la caractérisation d'une mutation dominante dans le gène *COL1A1*, responsable d'ostéogénèse imparfaite, étude à laquelle j'ai

participé pour la cartographie de la mutation. Ce travail a aussi fait l'objet d'une présentation orale au congrès mondial WCGALP (Grohs et *al.* 2022).

2. Analyse de deux anomalies dominantes pour des taureaux avec des taux de mortalité extrêmes

Journal of Dairy Science
Official Publication of the American Dairy Science Association

Articles Publish Topics About Contact

RESEARCH | VOLUME 106, ISSUE 1, P439-451, JANUARY 2023 [Download Full Issue](#)

Detailed analysis of mortality rates in the female progeny of 1,001 Holstein bulls allows the discovery of new dominant genetic defects

F. Besnard  • H. Leclerc • M. Boussaha • ... S. Mattalia • D. Boichard • A. Capitan  • 

Show all authors

Open Access • Published: November 01, 2022 • DOI: <https://doi.org/10.3168/jds.2022-22365>

Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)

Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges

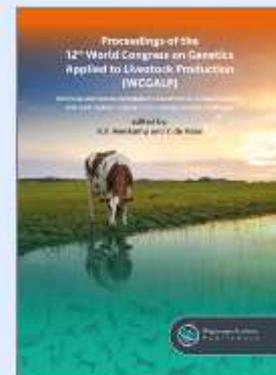
Editors: R.F. Veerkamp and Y. de Haas

Published: 2022 Pages: 3364

eISBN: 978-90-8686-940-4

<https://doi.org/10.3920/978-90-8686-940-4>

Book Type: Conference Proceedings



525. Detailed analysis of mortality rates in the female progeny of Holstein bulls allows the discovery of new dominant genetic defects

F. Besnard , H. Leclerc , N. Cesbron , M. Boussaha , C. Grohs , N. Jewell , A. Pinton , H. Barasc , M. Femenia , L. Dorso , B. Strugnelli , T. Floyd , C. Danchin , R. Guatteo , X. Hubin , S. Mattalia , D. Boichard , A. Capitan 

Pages: 2178 - 2181

OPEN ACCESS

https://doi.org/10.3920/978-90-8686-940-4_525

Published Online: February 09, 2023



Detailed analysis of mortality rates in the female progeny of 1,001 Holstein bulls allows the discovery of new dominant genetic defects

F. Besnard,^{1,2*} H. Leclerc,^{2,3} M. Boussaha,² C. Grohs,² N. Jewell,⁴ A. Pinton,⁵ H. Barasc,⁵ J. Jourdain,^{2,3} M. Femenia,² L. Dorso,⁶ B. Strugnelli,⁷ T. Floyd,⁸ C. Danchin,¹ R. Guatteo,⁶ D. Cassart,⁹ X. Hubin,¹⁰ S. Mattalia,^{1,2} D. Boichard,² and A. Capitan^{2,3*}

¹Institut de l'élevage, 75012 Paris, France

²Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

³ELIANCE Fédération d'entreprises du conseil et du service en élevage, 75012 Paris, France

⁴Animal and Plant Health Agency, Carmarthen Veterinary Investigation Centre, Carmarthen, SA31 3EZ, United Kingdom

⁵GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet Tolosan, France

⁶Oniris, INRAE, BIOEPAR, 44307 Nantes, France

⁷Farm Post Mortems Ltd., Hamsterley, Bishop Auckland, DL13 3QF, United Kingdom

⁸Animal and Plant Health Agency, Weybridge, KT15 3NB, United Kingdom

⁹Faculté de Médecine Vétérinaire, Liège Université, 4000 Liège, Belgium

¹⁰Association Wallonne des Eleveurs, 5590 Ciney, Belgium

ABSTRACT

Reducing juvenile mortality in cattle is important for both economic and animal welfare reasons. Previous studies have revealed a large variability in mortality rates between breeds and sire progeny groups, with some extreme cases due to dominant mutations causing various syndromes among the descendants of mosaic bulls. The purpose of this study was to monitor sire-family calf mortality within the French and Walloon Holstein populations, and to use this information to detect genetic defects that might have been overlooked by lack of specific symptoms. In a population of heifers born from 1,001 bulls between 2017 and 2020, the average sire-family mortality rates were of 11.8% from birth to 1 year of age and of 4.2, 2.9, 3.1, and 3.2% for the perinatal, postnatal, preweaning, and postweaning sub-periods, respectively. After outlining the 5 worst bulls per category, we paid particular attention to the bulls Mo and Pa, because they were half-brothers. Using a battery of approaches, including necropsies, karyotyping, genetic mapping, and whole-genome sequencing, we described 2 new independent genetic defects in their progeny and their molecular etiology. Mo was found to carry a de novo reciprocal translocation between chromosomes BTA26 and BTA29, leading to increased embryonic and juvenile mortality because of aneuploidy. Clinical examination of 2 calves that were monosomic for a large proportion of BTA29, including an ortholo-

gous segment deleted in human Jacobsen syndrome, revealed symptoms shared between species. In contrast, Pa was found to be mosaic for a dominant de novo non-sense mutation of GATA 6 binding protein (*GATA6*), causing severe cardiac malformations. In conclusion, our results highlight the power of monitoring juvenile mortality to identify dominant genetic defects due to de novo mutation events.

Key words: bovine genetic defects, juvenile mortality, heart defects, *GATA6*, chromosomal rearrangements

INTRODUCTION

Juvenile mortality has a severe impact on the cattle industry, because calves are the major output in beef production and are necessary for replacement in dairy production. Beyond economics, juvenile mortality affects the environmental impact of cattle breeding and addresses serious animal welfare concern (Østerås et al., 2007; Uetake 2013; Knapp et al., 2014). For all these reasons, juvenile mortality is an increasingly important field of research.

Several cross-sectional studies have been carried out worldwide, generally focusing on different periods to accurately monitor juvenile mortality (Reiten et al., 2018; Hyde et al., 2020; Dachrodt et al., 2021). Causes of death change during the course of the first year of life, the main ones being calving conditions or problems of fetal maturity, insufficient colostrum intake, digestive troubles, and respiratory diseases for the perinatal, postnatal, preweaning and postweaning periods, respectively.

Studies conducted in Holstein cattle have found mortality rates of 6.8 to 7.3% in France and 8.8% in the

Received June 3, 2022.

Accepted July 29, 2022.

*Corresponding authors: florian.besnard@idele.fr and aurelien.capitan@inrae.fr

US (Johanson et al., 2011; Raboisson et al., 2013) from birth to the second day of life, and of 12.9% from d 3 to 365 in the Chinese population (Zhang et al., 2019). In Norway, aggregating data from various dairy breeds, the mortality rate was only 7.8% over the whole first year of life in 2005 (Gulliksen et al., 2009), suggesting an influence of genetic components and farming systems on calves' survival. This assumption has been further supported by a comprehensive analysis of juvenile mortality in 19 French cattle breeds, which highlighted the breed purpose (beef or dairy), breed, sex, and sire progeny groups as the main factors influencing juvenile mortality across periods (e.g., Leclerc et al., 2016).

Heritability estimates for juvenile mortality are lower than 10% (Fuerst-Waltl and Sørensen 2010; van Pelt et al., 2012), reflecting the multiplicity of factors involved and the predominant effect of environment. However, as frequently observed for low-heritability traits, the genetic variability is large, with a genetic standard deviation around 5%, corresponding to a huge genetic coefficient of variation of 50%. The corresponding genetic variability is difficult to characterize biologically because the largest part of the mortalities results from common infections with undetermined, and probably underestimated, genetic components. Nevertheless, some genetic factors have been identified in the past, such as bovine leukocyte adhesion deficiency (Shuster et al., 1992), but most of them are recessive and explain a very small proportion of the mortalities. In some rare situations, their determinism is dominant and can explain large mortality rates in the progeny of carrier sires. These dominant conditions can be transmitted by carrier animals either because of incomplete penetrance or of mosaicism for de novo mutations (Bourneuf et al., 2017). One possible way to seek for dominant deleterious mutations is to analyze mortality rates in the progeny of artificial insemination sires used in multiple farms, assuming that extreme values hide congenital anomalies.

In this context, the purpose of this study was two-fold: to finely monitor calf mortality at the level of sire families within the French and Walloon Holstein populations, and to use this information to detect genetic defects that might have been overlooked by lack of specific externally visible symptoms.

MATERIALS AND METHODS

Mortality Rates in Sire Families at Different Ages

Data on the pedigree, sex, date of birth, date of death, and cause of death (either natural or slaughtering) of Holstein animals were recovered from the bovine French and Walloon databases. The data set included

calves born from 2017 to 2020. To focus on the most reliable data, only female calves that remained on their farm of birth until death or during their whole first year of life were selected. Sire families with fewer than 100 female progeny were disregarded. Accordingly, the final data set comprised 2.25 million daughters from 1,001 sires (with a mean of 2,246 and a maximum of 35,375 females per sire family). Natural mortality rates were computed during the first year of life and for 4 sub-periods known to correspond to distinct predominant causes of death: perinatal (d 0–2), postnatal (d 3–14), preweaning (d 15–55), and postweaning (d 56–365) mortality (Santman-Berends et al., 2019; Dachrodt et al., 2021). Mortality rates were calculated as the number of calves that died of natural causes during a window of time divided by the number of calves alive at the start date.

Then, we paid particular attention to the 5 bulls showing the highest mortality rates for each period, to identify sires potentially transmitting unreported dominant genetic defects to their progeny. Among them, 2 sires (Mo and Pa) were selected for subsequent analyses because they were half-brothers and potentially transmitted a common genetic defect.

Clinical Examination

Two affected calves of Mo (2 females) and 8 of Pa (4 females, 4 males) were necropsied by trained veterinarians in France, Belgium, and the UK. By “affected calves” we mean animals that have been reported by breeders as suffering from unexplained weakness, diminished growth rates, and often spontaneous death despite intensive care. Gross phenotypic description was also available for 5 additional clinically affected calves of Mo (see Supplemental Notes S1 and S2 for information on the age and symptoms of all calves examined; https://figshare.com/projects/Besnard_JDS_Supplementary_material/140747; Besnard, 2022). At the time of the study, biological material was still available for 2 affected calves of Mo and 8 of Pa.

Karyotyping

Giemsa-stained karyotype of sire Mo and of 2 affected daughters of Pa were obtained from blood lymphocytes as described in Ducos et al. (1998). Of note, Pa was dead at time of the study and thus not available for sampling.

Analysis of Semen Quality and Fertility

Because chromosomal rearrangements can negatively affect spermatogenesis, 5 different traits were analyzed

for a cohort of 50 bulls, including Mo, that had their semen collected on a routine basis in the same artificial insemination center: the mean volume of the ejaculate (in mL measured by weighting), its concentration (in million spermatozooids per milliliter, measured by spectrophotometry), fresh mass motility and individual motility (in score and percentage, respectively, based on microscope observation), as well as post-freezing mean motility and progressive motility, measured with computer-assisted sperm analysis and IVOS II (O'Meara et al., 2022). The number of records per bull and trait ranged from 2 to 39. In addition, 2 fertility traits were calculated for the initial cohort of 1,001 sires mentioned previously. The nonreturn rate at 56 d corresponds to the percentage of cows inseminated with the semen of a given sire that were not reinseminated within the following 56 d, and the conception rate corresponds to the percentage of inseminations that led to the birth of a calf.

Analysis of Illumina SNP Array Genotypes

The bull Mo, 15 of his progeny and 1 of their dams, as well as Pa, 203 of his progeny and 89 of their dams, and finally Mogul (sire of both bulls) were genotyped with various Illumina arrays over time (Bovine SNP50, EuroG10K, and EuroGMD). Genotypes were phased and imputed to the Bovine SNP50 using FImpute3 (Sargolzaei et al., 2014) in the framework of the French genomic evaluation, as described in Mesbah-Uddin et al. (2019).

Following the detection by karyotyping of a chromosomal rearrangement in Mo, we analyzed along chromosomes BTA26 and BTA29 which of the paternal or maternal phases of this bull were transmitted to offspring, to detect recombination events. In parallel, we also mined the raw genotypes of affected daughters for increased rates of Mendelian transmission errors (a sign of monosomy) or increased rates of markers with null genotypes (“-/-”; a sign of trisomy).

For Pa, no chromosomal rearrangement was identified, and other investigations were carried out. Assuming a dominant inheritance with somatic mosaicism, we performed transmission disequilibrium tests for 16,487 informative markers, for 14 progeny that died during the preweaning period (including 5 already necropsied at that time), and 189 half-sib controls still alive at 2 years of age. The proportion of each of the paternal alleles transmitted to the case and control groups were compared using a Fisher test with Bonferroni correction. Finally, ggplot2 and Rcolorbrewer were used for data visualization with the R software (R version 4.1.2).

After the discovery of the causative mutation in the *GATA6* gene (see the Results section), we used allele transmission proportions for 2 flanking informative markers within the control population to estimate the proportion of mosaicism in Pa's germ cells. Given the deleterious consequences of the *GATA6* mutation on heart development, we expect that control calves carrying the at-risk haplotype inherited the ancestral version of this haplotype (i.e., predating the mutation event). The proportion of affected gametes was calculated as $(nHb - nHa)/(2 \times nHb)$, with nHa the number of carriers of the at-risk haplotype among half-sib controls and nHb the number of carriers of the alternative paternal haplotype within the same population. Finally, we used a chi-squared goodness-of-fit test to compare the observed proportion of affected gametes with those expected assuming mosaicism rates of 1/2, 1/4, 1/8, and 1/16 in Pa's germ cells.

Gene Content and Comparative Genomics

The gene content of specific regions was extracted from the bovine ARS-UCD1.2 and human GRCh38.p13 genome assemblies using the BioMart tool (Ensembl release 106; <https://www.ensembl.org/biomart/martview/>). In parallel, we used the synteny tool from Ensembl to identify conserved blocks between bovine and human chromosomes (https://www.ensembl.org/Bos_taurus/Location/Synteny/). Then we compiled the list of genes in common between the BTA29 segment deleted in Mo's affected calf and the core HSA11 deletion responsible for Jacobsen syndrome in human.

Analysis of Whole-Genome Sequences

The genome of 1 affected calf of Pa was sequenced at a coverage of 19.4× on an Illumina HiSeq3000 HWI-J00173 platform with 150 bp paired-end reads, after library preparation with an average insert size of 440 bp using the NEXTflex PCR-Free DNA Sequencing Kit (Bioo Scientific). The whole genome sequence data are available under the study accession no. ERR9669242 at the European Nucleotide Archive (www.ebi.ac.uk/ena). Reads were aligned on the ARS-UCD1.2 bovine genome assembly and processed in accordance with the guidelines of the 1000 Bull Genomes Project (Hayes and Daetwyler 2019) for the detection of SNPs and small InDels. Assuming that the causative mutation is dominant and occurred de novo, we retained only heterozygous variants that were (1) absent from 5,116 control genomes from run 9 of the 1000 Bull Genomes Project and (2) located within the mapping interval (positions 19,505,558 to 37,877,867 bp on BTA24).

The remaining variants were annotated using Variant Effect Predictor (Ensembl release 106; www.ensembl.org/Tools/VEP). In addition, we detected structural variants within the mapping interval using Pindel (Ye et al., 2009), Delly (Rausch et al., 2012), and Lumpy software (Layer et al., 2014), and applied the same filters after comparison with analogous data of 62 control genomes (Boussaha et al., 2015).

Genotyping of the GATA6 Candidate Variant

DNA samples from Pa (extracted from semen), 3 affected calves, and 3 controls carrying the same paternal haplotype but in the nonmutated version, as well as their 6 dams, were genotyped for variant g.34,187,181T > A on BTA24 using PCR and Sanger sequencing. A segment of 321 bp was PCR amplified in a Mastercycler Pro thermocycler (Eppendorf) using primers CAGTGGGCGCTAAACTACC and AGACCTGCTGGAGGACCTG and the Go-Taq Flexi DNA Polymerase (Promega), according to the manufacturer's instructions. Amplicons were purified and bidirectionally sequenced by Eurofins MWG (Hilden, Germany) using conventional Sanger sequencing, before analysis with NovoSNP software for variant detection (Weckx et al., 2005).

RESULTS AND DISCUSSION

Analysis of Mortality Rates at Different Stages in the Progeny of Individual Sires

The natural mortality rate of heifers during their first year of life was 11.8% on average in the population of 1,001 bulls analyzed, with 4.2% for perinatal, 2.9% for postnatal, 3.1% for preweaning, and 3.2% for postweaning mortalities. These rates were lower than most of those reported in the literature (e.g., Johanson et al., 2011; Raboisson et al., 2013; Leclerc et al., 2016; Zhang et al., 2019), probably because we considered only females. Sex is known to have a significant effect on juvenile mortality (Raboisson et al., 2013; Hyde et al., 2020), notably because females receive more care than males, due their higher financial value. The addition of vitality at birth in the French Holstein total merit index in 2009 may also have contributed to a reduction of perinatal mortality through selection.

Interestingly, natural mortality rates per period and per half-sib family showed approximately normal distribution, suggesting quantitative inheritance (Figure 1). Yet we observed outlier families with possible mono- or oligogenic inheritance of excess mortality and focused on the 5 worst sires per category (Table 1). Among

them, the 2 bulls Mo and Pa, ranked number 1 and number 5 for mortality rate over the first year of life, were half-brothers sired by the popular bull Mogul (HOLUSAM003006972816). Although they displayed distinct profiles (with, for example, 16.7 and 5.3% perinatal mortality versus 2.4 and 6.3% preweaning mortality, respectively), their close relationship raised the question of a common underlying pathophysiology, and therefore they were selected for further analysis.

Identification and Characterization of a Reciprocal Translocation Between BTA26 and BTA29 in Mo

Genetic Analyses of Mo and His Progeny. To gain insights into the causes of increased mortality within the Mo and Pa sire families, we carried out a series of investigations, starting with cytogenetic analyses. Although the karyotypes of 2 affected daughters of Pa were apparently normal (not shown), we observed a reciprocal translocation between chromosomes BTA26 and BTA29 in Mo [t(26;29)(q11;q19); Figure 2A].

Subsequent analysis of Illumina SNP array genotypes from Mo, his own sire Mogul, and 15 of Mo's progeny enabled us to define the approximate borders of chromosomal break and fusion points, and to determine that the affected chromosomes originated from Mogul (Figure 2B, C; Supplemental Figure S1, https://figshare.com/projects/Besnard_JDS_Supplementary_material/140747; Besnard, 2022). Considering that Mogul was extensively used as a bull sire and did not display abnormal juvenile mortality rates, these results suggest that the rearrangement occurred in the germ cells of Mogul during the meiosis that gave the spermatozoon at Mo's conception.

In addition, we demonstrated that 2 affected daughters of Mo with DNA samples available were monosomic for approximately the first 70% of BTA29 (767 markers, 36.7 Mb; Supplemental Figure S1). Interestingly, comparative genomics revealed synteny between part of the hemizygous region and the monosomy of the telomeric region of chromosome 11q responsible for Jacobsen syndrome (Figure 2C; Mattina et al., 2009). Both segments share a common set of 69 orthologous protein coding genes out of the 318 affected by monosomy in Mo's progeny and the ~100 of the core Jacobsen deletion (Supplemental Table S1, https://figshare.com/projects/Besnard_JDS_Supplementary_material/140747, Besnard, 2022; Rodríguez-López et al., 2021).

Phenotypic Characterization of Mo's Calves and Mo's Semen Characteristics. In human, Jacobsen syndrome has been extensively studied, with 200 cases compiled in the Human Phenotype Ontology da-

Table 1. Selection of the 5 bulls showing the highest mortality rates for each period and over the whole first year (bold number in table) versus average value of population; the bulls are sorted in descending order on the entire first-year mortality rate¹

Bull	No. of daughters	Mortality rates per period, %				
		Perinatal	Postnatal	Preweaning	Postweaning	First year
Mo ²	108	16.7	6.7	2.4	7.3	30.6
Lp	110	3.6	4.9	6.2	8.8	20.9
Ce	126	1.6	9.0	6.3	5.8	20.6
Sg	130	5.4	8.6	3.8	4.9	20.0
Pa ²	2,297	5.3	6.1	6.3	4.8	19.9
Mn	122	9.0	2.0	3.1	8.4	19.7
Jo	103	1.9	5.1	3.2	8.8	17.5
Ja	1,196	5.2	6.1	4.2	3.9	17.5
Wt	124	2.4	5.9	6.3	3.9	16.9
Pg	315	2.9	4.4	2.5	8.7	16.8
Ip	203	5.9	1.7	2.8	8.1	16.8
Co	1,749	9.4	3.9	3	2.2	16.5
Io	815	8.3	2.5	2.9	3.9	15.8
Me	117	2.6	4.5	6.6	3	15.4
Ln	2,399	8.1	2.6	3	2.7	14.9
Average value for 1,001 bulls	2,250	4.2	2.9	3.1	3.2	11.8

¹Each rate is calculated based on the calves alive at the start of the period. For this reason, summing up the mortality rates for the 4 subperiods does not give exactly the same number as the mortality rate for the whole period.

²Bulls are paternal half-brothers.

tabase (<http://human-phenotype-ontology.github.io>). As with the progeny of Mo, most of human cases are due to translocations between HSA11 and other chro-

mosomes (Basinko et al., 2011). The clinical features of Jacobsen syndrome include various symptoms that are more or less expressed depending of the patient,

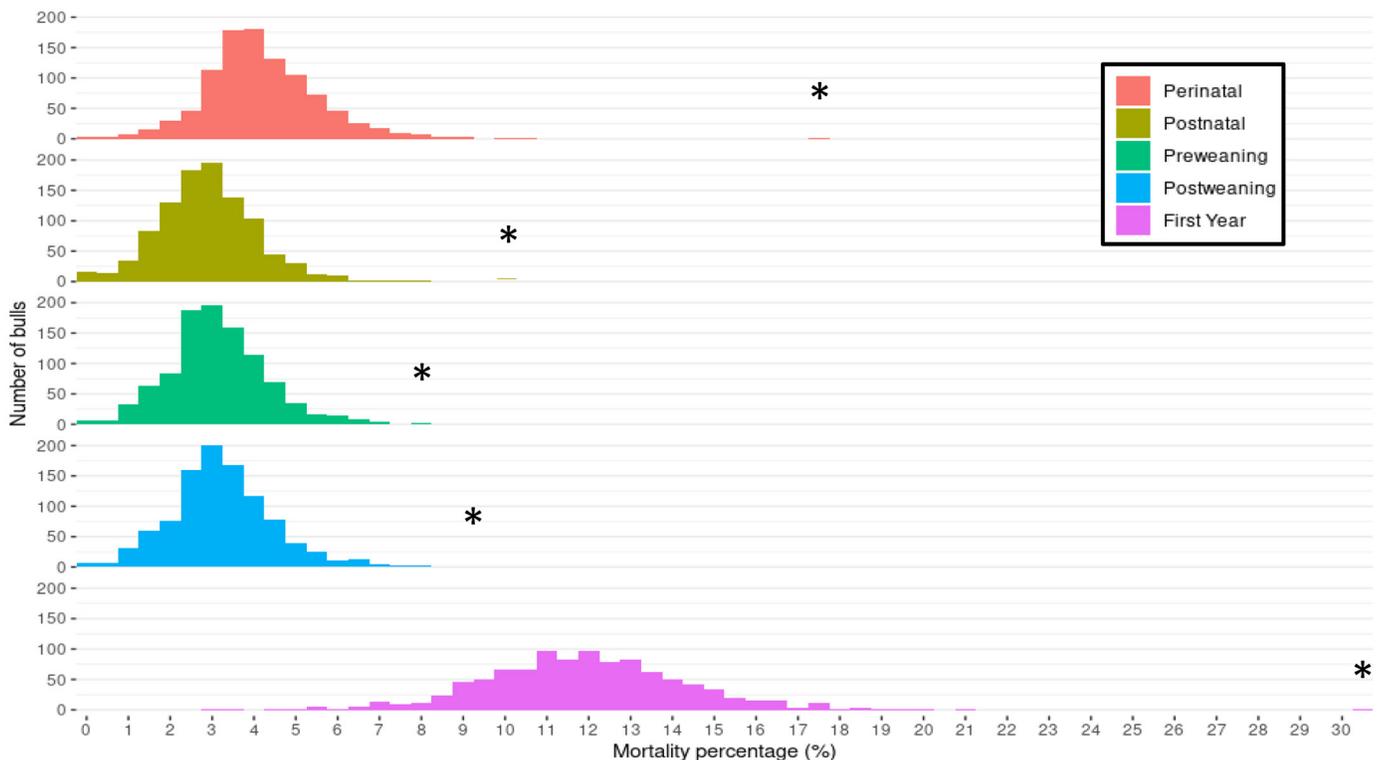


Figure 1. Distribution of daughters' mortality rates of 1,001 bulls for 5 periods of death. Asterisks highlight barely visible outliers.

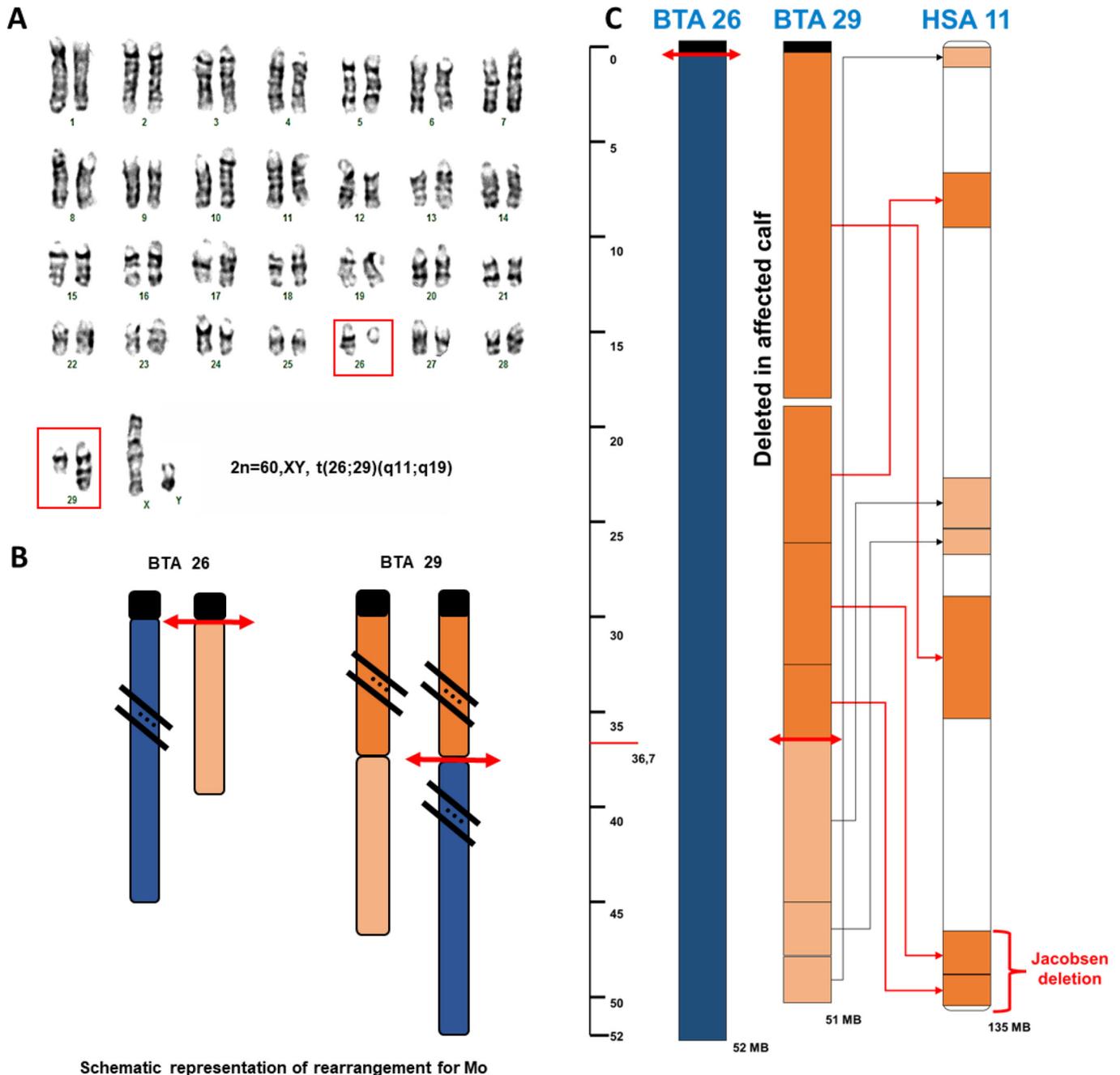


Figure 2. Karyotype of sire Mo and details on the chromosomal segments involved. (A) Giemsa-stained karyotype of Mo showing a reciprocal translocation between BTA26 and BTA29. (B) Schematic representation of the chromosomal rearrangement with mutant chromosomes on the right of each pair. (C) Approximate localization of the breakpoints (based on the analysis of SNP array genotypes from 15 progeny of Mo; Supplemental Figure S1, https://figshare.com/projects/Besnard_JDS_Supplementary_material/140747; Besnard, 2022) and synteny between BTA29 and human HSA11 chromosomes.

such as Paris-Trousseau thrombocytopenia, growth rate reduction, and psychomotor impairment, as well as cardiac, craniofacial, gastrointestinal, renal, genitourinary, ophthalmic, and orthopedic anomalies (<https://www.omim.org/entry/147791>). In agreement with the

observations made in humans, clinical examination of the 2 calves partially monosomic for BTA29 and of 5 additional cases for which no DNA was available revealed very similar symptoms (Figure 3; Supplemental Note S1).

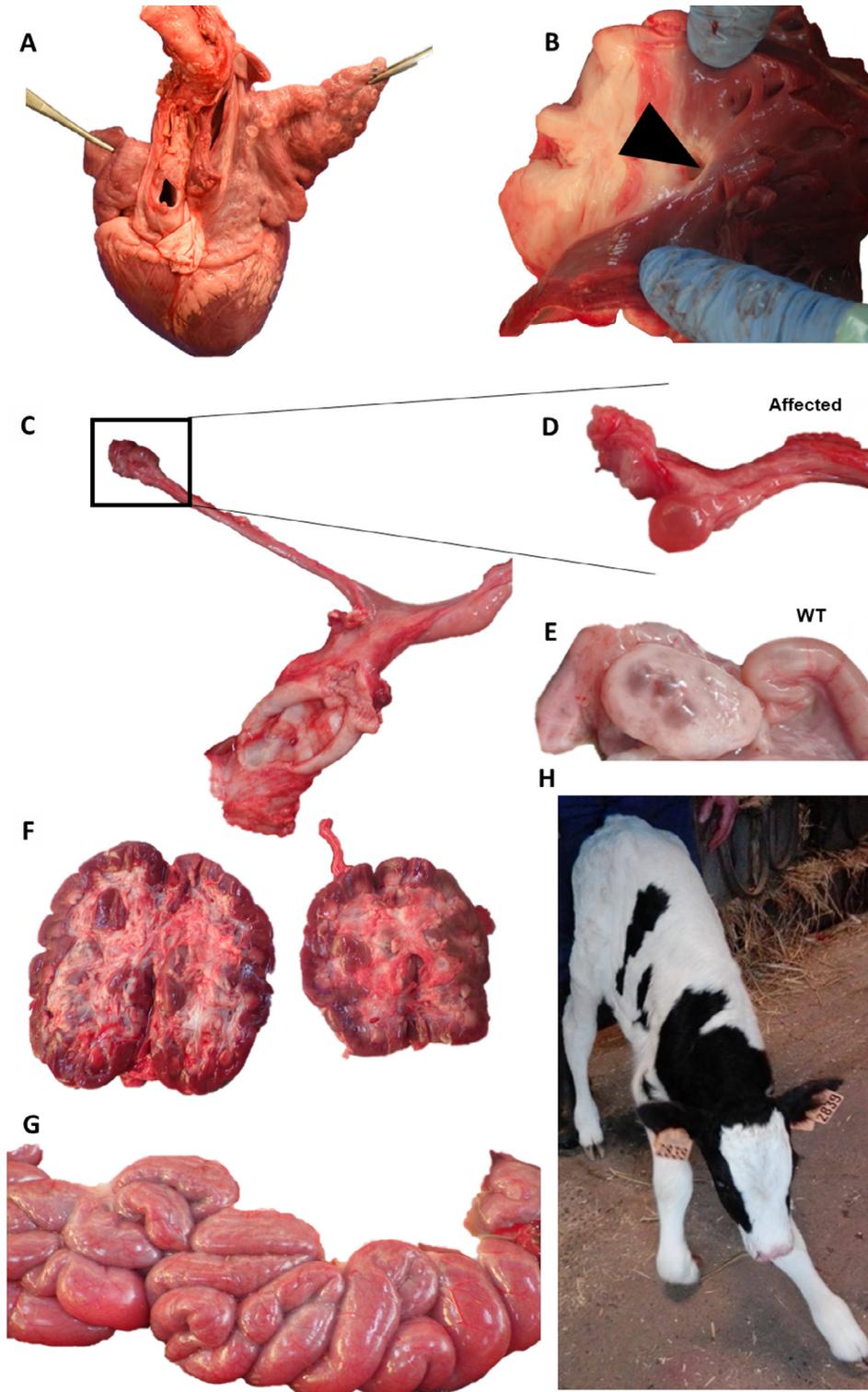


Figure 3. Clinical features observed in the progeny of Mo. (A–G) Symptoms displayed by 2 calves partially monosomic for BTA29 who died at birth (male case 1, A) or was euthanized at 3.5 mo (female case 2, B–G). (A) Heart with tetralogy of Fallot. (B) Open heart, with black arrow pointing to an interventricular septal defect of 1-cm diameter located high under the sigmoid valves. (C) Blind and hypoplastic uterine horn ending with an atrophied and cystic ovary. (D) Detail of the affected ovary in comparison with a matched control (E). (F) Moderate hypogenesis of the left kidney, whose volume is 2/3 of the right kidney. (G) Abnormally hydrated content in the colon and rectum. (H) Picture of female case 3, showing articular defects and difficulties standing. For further information, see Supplemental Note S1 (https://figshare.com/projects/Besnard_JDS_Supplementary_material/140747; Besnard, 2022).

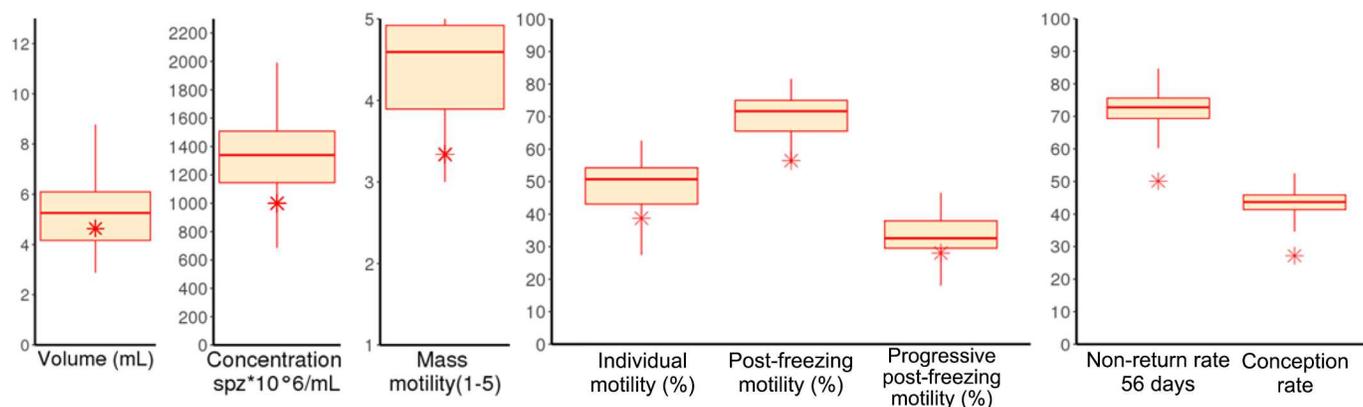


Figure 4. Analysis of the quality and fertility of Mo's semen compared with control bulls. Boxplots of the distribution of 50 bulls for 6 traits measured in the same artificial insemination center and of 1,001 bulls for the nonreturn rate at 56 d and the conception rate. Mo is indicated with a red asterisk. Traits are as follows, from left to right: average volume of ejaculate (mean = 5.21 mL, Mo = 4.64), average spermatozoid concentration in ejaculate [mean = 1,329 million spermatozooids (spz)/mL, Mo = 998], average mass motility score is manually assessed by artificial insemination operators from 1 to 5 (mean = 4.36, Mo = 3.34), average individual motility (mean = 70%, Mo = 56), average post-freezing motility (mean = 49%, Mo = 39), 56-d nonreturn rate (mean = 71%, Mo = 50), and conception rate (mean = 43%, Mo = 27). Boxplots should be read as follows: the horizontal line contained within each box marks the median value (second quartile) of the data. Lower and upper lines correspond to the first and third quartiles. Whiskers correspond to the first and third quartiles ± 1.5 interquartile range, respectively. The units of y-axes are indicated below each figure.

Instances of reciprocal translocations are rare in cattle, with only 20 reports counted in a recent review of literature by Iannuzzi and coauthors, none of which affected chromosome 29 (Iannuzzi et al., 2021).

Regarding aneuploidies affecting BTA29, only 1 complete trisomy has been reported before this study, in a stillborn Braunvieh calf showing preterm delivery, dwarfism, and severe craniofacial malformations (Häfliger et al., 2020). The absence of other reports on trisomy for BTA29 despite the segregation of a BTA1–29 Robertsonian fusion in various cattle breeds (Gustavsson, 1979), as well as the lack of human patients trisomic for the Jacobsen segment on HSA11 orthologous to part of BTA29 (e. g. Pylyp et al., 2018), suggest that this condition would lead to embryonic death in both species.

Because chromosomal abnormalities affect not only the viability of conceptuses but also meiosis and gametogenesis (Raudsepp and Chowdhary, 2016), we investigated several traits related to semen volume, quality, and fertility in Mo and 2 groups of Holstein bulls (Figure 4). Among 50 bulls reared and sampled in the same artificial insemination center, Mo showed normal average volume of ejaculate but low semen quality, with average concentration of semen and fresh and post-freezing motility trait records in the lowest quartile. The influence of the chromosomal rearrangement was even more severe with regard to fertility, Mo ranking as the worst sire for nonreturn rate at 56 d and the fifth worst for conception rate among our cohort of 1,001 Holstein sires. This major degradation of fertility,

observed on both early and late indicators of insemination success, is most probably the result of the premature death of a substantial proportion of aneuploid conceptuses throughout the gestation.

Thus, we report the first large animal model for Jacobsen syndrome in humans, and the first instance of partial monosomy for BTA29 in cattle, to our knowledge.

Identification of a Mosaic GATA6 Nonsense Mutation in Pa

Despite their close relationship, a different etiology was suspected for the excess of mortality observed among the daughters of Pa, because the peak of mortality occurred later in life than for Mo's offspring. This assumption was rapidly confirmed by clinical examination and karyotyping of Pa's descendants.

Clinical Examination of Pa's Progeny. A survey of French and British veterinarians allowed us to collect phenotypic information on Pa's descendants, among which 8 showed symptoms compatible with severe heart defects either leading to premature death or justifying euthanasia on humane grounds (Supplemental Note S2). Autopsies gave results strikingly similar to the systematic observation of a persistent *truncus arteriosus* (TA; i.e., a malformation of the large vessels at the base of the heart, characterized by the development of a single arterial trunk straddling the 2 ventricles, above a large interventricular communication, which gives rise to the aorta and the 2 branches of the pulmonary

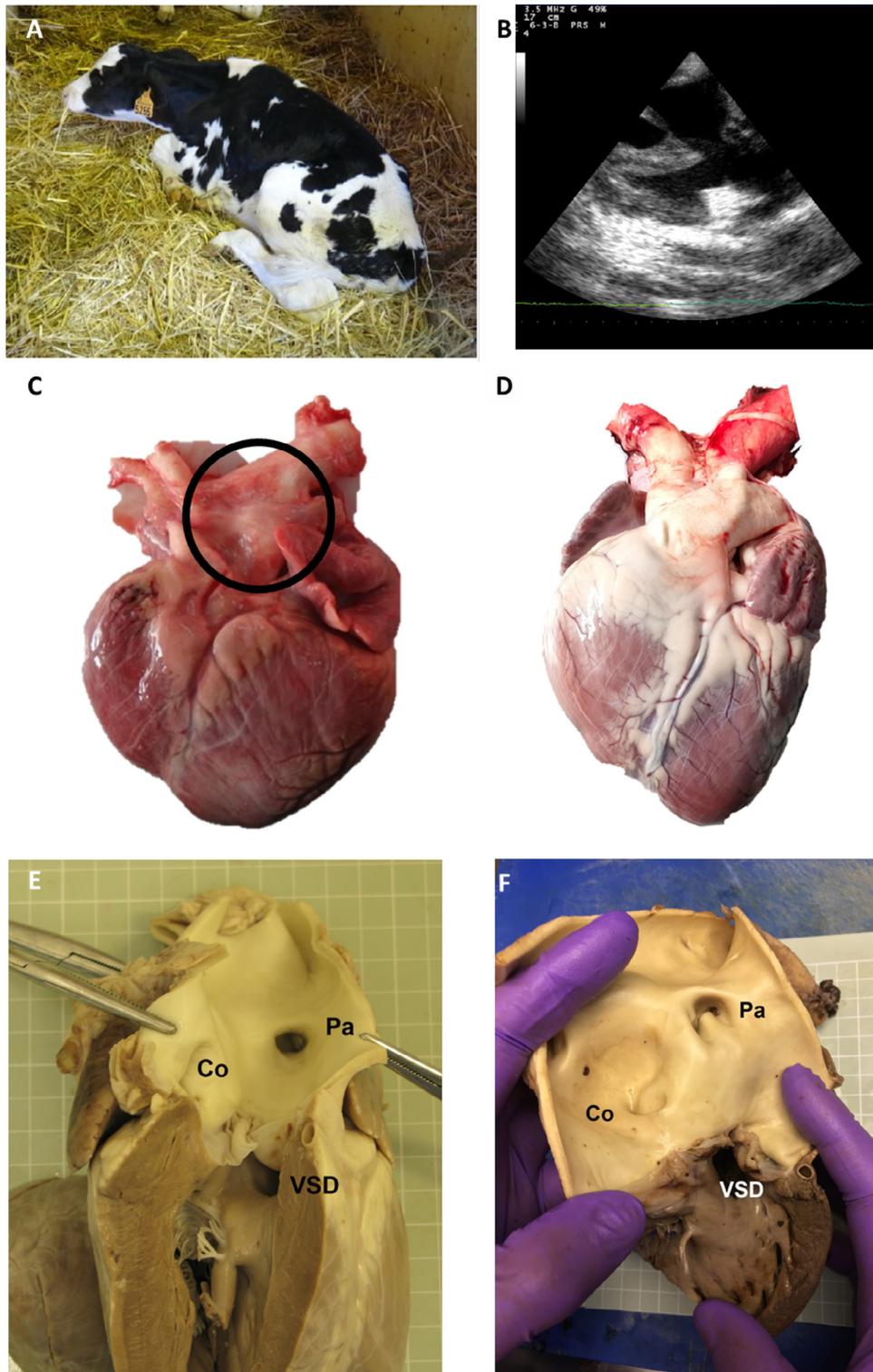


Figure 5. Clinical findings in Pa calves. (A–C) Pictures of case 1, euthanatized at 42 d of age. (A) Live calf on farm. (B) Ultrasonography showing communications between the auricles and the common arterial trunk. (C, D) Hearts of case 1 and of a matched control, respectively. Note the persistent *truncus arteriosus* (circle) and the modification of the general shape of the heart in C vs. D. (E) Right lateral view of the dissected right ventricular outflow tract and common arterial trunk of case 6. Note the thickened right ventricular wall and ventricular septal defect (VSD); the *truncus arteriosus* is situated over the ventricular septum and has a single common arterial valve with 3 leaflets; the coronary arteries arise from an ostium on the left side of this vessel (Co); and the pulmonary arteries arise from a common ostium on the right side (Pa). (F) Right dorsolateral view of the open *truncus arteriosus* in case 7 showing the VSD and ostia of the coronary arteries (Co) and pulmonary arteries (Pa).

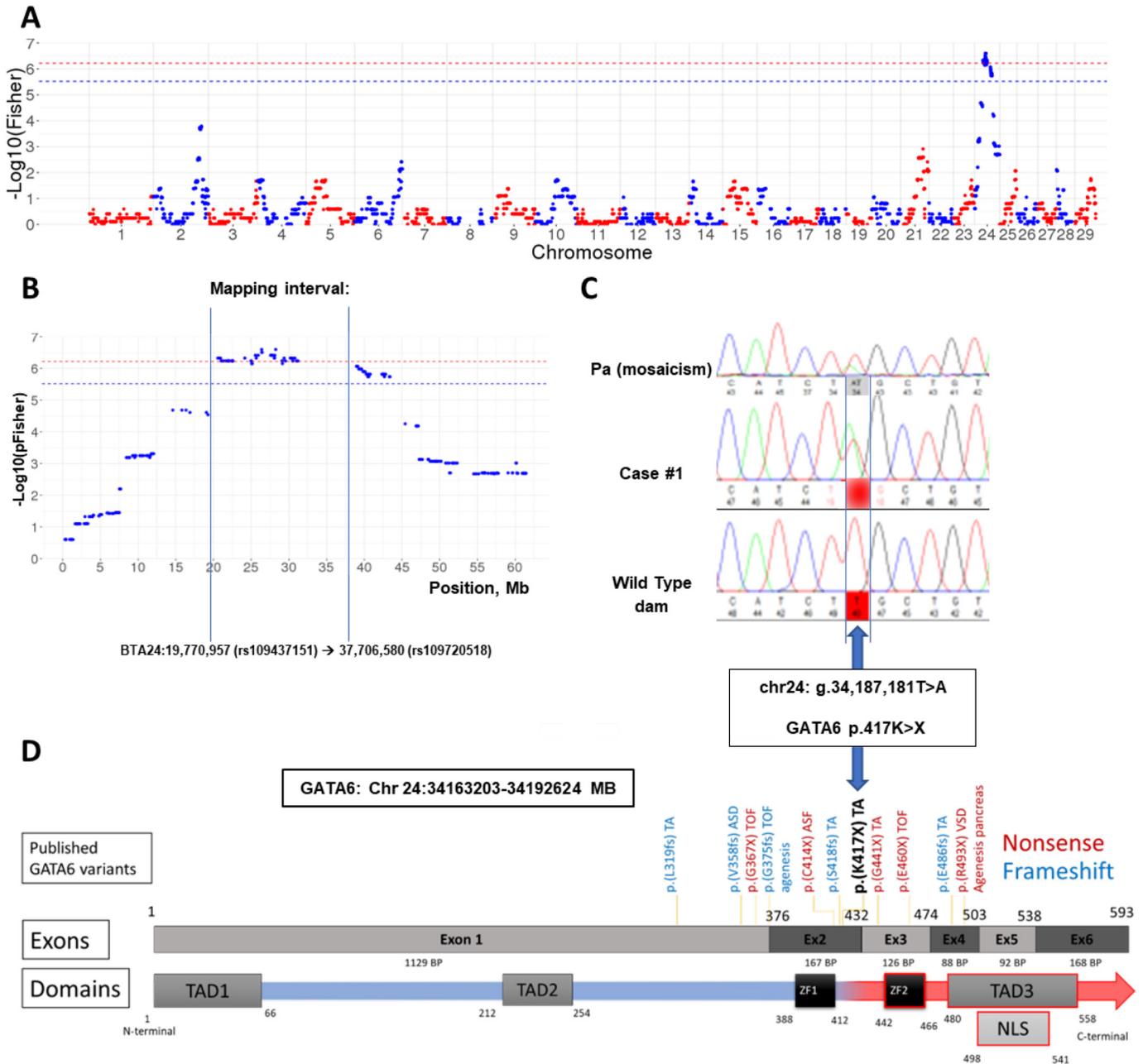


Figure 6. Mapping and identification of a de novo nonsense mutation of GATA6. (A, B) Manhattan plot of the results of mapping of the *truncus arteriosus* locus using transmission disequilibrium test, with a zoom on BTA24 (B). The blue and red lines represent the significance threshold for $P < 0.05$ and $P < 0.01$ after Bonferonni correction for multiple testing. (C) Electropherogram of the sire Pa, one affected calf, and its nonaffected dam, for a segment of BTA24 encompassing variant g.34,187,181T > A. Note the small proportion of allele A versus T in Pa as compared with case 1, supporting mosaicism. (D) Domain and region information for GATA6, with phenotype information for the bovine mutation p.417K > X and truncating variants reported in the orthologous protein in human. TAD: transcription activation domain; ZF: zinc finger domain; NLS: nuclear localization signal domain; TOF: tetralogy of Fallot; ASD: atrial septal defect; VSD: ventricular septal defect. Information obtained from the from the UniProt database (<http://www.uniprot.org/>; accession numbers A0A4W2FXQ7 and Q92908).

artery), sometimes associated with additional heart septation defects (Figure 5).

Mapping and Identification of the Causative Mutation. Given the fact that Pa was apparently un-

affected and that TA has never been reported outside of his progeny among thousands of genetic defects reported to the French National Observatory for Bovine Abnormalities (Grohs et al., 2016) over the last 20

years, we assumed a dominant inheritance associated with germline or somatic mosaicism, or both, in the sire. Therefore, we analyzed SNP array genotypes of 14 progeny that died during the preweaning period (including 5 necropsied) and 189 half-sib controls still alive at 2 years of age, via transmission disequilibrium test. We mapped the TA locus on BTA24 between positions 19,505,558 (rs453420861) and 37,877,878 (rs723126921) bp on the ARS-UCD1 assembly. Then we sequenced the genome of one TA-affected animal with Illumina technology and used up to 5,116 genomes from run 9 of the 1000 Bull Genomes Project (Hayes and Daetwyler 2019) as controls.

Filtering for heterozygous SNP, InDels, and structural variations that were absent from controls yielded only 29 positional candidates within the interval (Supplemental Table S2, https://figshare.com/projects/Besnard_JDS_Supplementary_material/140747; Besnard, 2022). Only one of them appeared as a bona fide functional candidate variant: a thymine-to-adenine substitution in exon 2 of *GATA6* predicted to introduce a premature stop codon (chr24: g.34,187,181T > A; *GATA6* p.K417X). If translated, the mutant protein would be shortened by approximately 30% and would lack 3 domains essential for the proper function of this transcription factor, controlling heart development in vertebrates (Brewer and Pizzey 2006; Lentjes et al., 2016; Figure 6D). Experiments in mice have demonstrated that the conditional inactivation of *GATA6* in heart progenitor cells causes embryonic lethality due to interrupted aortic arch and persistent TA (Lentjes et al., 2016). In humans about 80 dominant mutations of *GATA6* have been described to date, which cause various heart or pancreatic development anomalies depending on their nature and location (for a review see Škorić-Milosavljević et al., 2019). Remarkably, the 2 orthologous human truncating mutations located closest to the present bovine nonsense variant (pS418fs and pG441X) have been reported to cause exactly the same phenotype, that is, persistent TA, supporting the causality of the latter mutation (Figure 6D).

Validation of the Causality of the *GATA6* Mutation. For verification, we genotyped this *GATA6* nonsense variant by PCR and Sanger sequencing in Pa, 3 affected calves, and 3 controls carrying the same paternal haplotype but supposedly in the nonmutated version, as well as their 6 dams. As expected, the mutant allele was found in the heterozygous state only in the 3 cases and in Pa's semen, thus confirming the de novo nature and therefore the causality of the mutation (Figure 6C).

Then we analyzed the segregation distortion for 2 markers adjacent to the mutation among the 189 control calves of Pa that were still alive at 2 years of

age. We found 57 controls carrying the same paternal haplotype as the affected animals but presumably in its ancestral version (i.e., without the de novo mutation) and 132 with the second paternal haplotype. From this 57:132 ratio, we estimated a proportion of 28.4% of mutant spermatozooids $[(132 - 57)/(2 \times 132)]$ and thus 56.8% of mutant germ cells. Comparing the proportion observed in controls with those expected for various degrees of mosaicism using a chi-squared goodness-of-fit test, we demonstrated that this distortion was compatible with a degree of mosaicism of 1/2 ($P = 0.28$) and rejected lower levels of mosaicism (proportions of 1/4, 1/8, and 1/16; $P = 0.00012$ and lower). These results suggest that the mutation occurred either early in the germline progenitor cells of Pa, or possibly at the first division of the egg cell. Unfortunately, Pa was dead at time of the study, and we did not have access to tissues other than semen to answer this question.

CONCLUSIONS

With a few exceptions, we observed a nearly normal distribution of juvenile mortality rates among the daughters of 1,001 Holstein sires. By focusing on the progeny of 2 outlier bulls, we identified 2 de novo mutations consisting of a balanced translocation between chromosomes 26 and 29, and a mosaic nonsense mutation of *GATA6* (see Online Mendelian Inheritance in Animals entries OMIA 002558-9913, <https://www.omia.org/OMIA002558/9913/>, and 002559-9913, <https://www.omia.org/OMIA002559/9913/>). Furthermore, we described the first large animal models for human Jacobsen syndrome and persistent *truncus arteriosus* due to *GATA6* haploinsufficiency, to our knowledge. These results demonstrate the suitability of our approach to reveal genetic defects that are hardly detectable with traditional heredo-surveillance in the absence of specific externally visible symptoms. Beyond this proof of concept, the calculation of mortality rates at different ages for the whole population of bulls paves the way for future detection of QTL influencing juvenile mortality.

ACKNOWLEDGMENTS

We are grateful to the staff of the Faculty of Veterinary Medicine, University of Liège (Liège, Belgium), for conducting necropsies on Mo's calves and to the partners of the 1000 Bull Genomes consortium (Melbourne, Australia) for providing control whole genome sequence data. We are thankful to Inovéo and E. Henrotte for sharing anonymous data on semen quality and fertility. We thank LABOGENA DNA for giving close support and reactivity for the analysis of the genotypes as well as GeT-PlaGe for their cooperation in the analysis of the

chromosomal rearrangement, with the support of Anne Calgaro and Nathalie Mouney (Cytogene team, UMR GenPhySE). Finally, we express our special thanks to Nora Cesbron [Laboratoire de l'Environnement et de l'Alimentation de la Vendée (LEAV), La Roche Sur Yon, France] for her efforts in clinical examination and for her expertise in cardiology. F. Besnard is a recipient of a CIFRE PhD grant from IDELE, with the financial support of the Association Nationale de la Recherche et de la Technologie and APIS-GENE (Paris, France). Surveillance for livestock disease conducted by the APHA is funded by the UK Department for the Environment, Food, and Rural Affairs (London, England) and the devolved governments of Scotland and Wales. The authors have not stated any conflicts of interest.

REFERENCES

- Basinko, A., S. Audebert-Bellanger, N. Douet-Guilbert, J. Le Franc, P. Parent, S. Quemener, P. La Selve, C. Bovo, F. Morel, M.-J. Le Bris, and M. De Braekeleer. 2011. Subtelomeric monosomy 11q and trisomy 16q in siblings and an unrelated child: Molecular characterization of two der(11)t(11;16). *Am. J. Med. Genet. A.* 155A:2281–2287. <https://doi.org/10.1002/ajmg.a.34162>.
- Besnard, F. 2022. Supplementary material. Figshare. https://figshare.com/projects/Besnard_JDS_Supplementary_material/140747.
- Bourneuf, E., P. Otz, H. Pausch, V. Jagannathan, P. Michot, C. Grohs, G. Piton, S. Ammermüller, M. A. Colle, C. Klopp, and D. Esquerré. C. Wurmsler, K. Flisikowski, H. Schwarzenbacher, J. Burgstaller, M. Brüggemann, E. Dietschi, N. Rudolph, M. Freick, S. Barbey, G. Fayolle, C. Danchin-Burge, L. Schibler, B. Bed'Hom, B. J. Hayes, H. D. Daetwyler, R. Fries, D. Boichard, D. Pin, C. Drögemüller, and A. Capitan. 2017. Rapid discovery of de novo deleterious mutations in cattle enhances the value of livestock as model species. *Sci. Rep.* 7:11466. <https://doi.org/https://doi.org/10.1038/s41598-017-11523-3>.
- Boussaha, M., D. Esquerré, J. Barbieri, A. Djari, A. Pinton, R. Letaief, G. Salin, F. Escudié, A. Roulet, S. Fritz, F. Samson, C. Grohs, M. Bernard, C. Klopp, D. Boichard, and D. Rocha. 2015. Genome-wide study of structural variants in bovine Holstein, Montbéliarde and Normande dairy breeds. *PLoS One* 10:e0135931. <https://doi.org/10.1371/journal.pone.0135931>.
- Brewer, A., and J. Pizzey. 2006. GATA factors in vertebrate heart development and disease. *Expert Rev. Mol. Med.* 8:1–20. <https://doi.org/10.1017/S1462399406000093>.
- Dachrodt, L., H. Arndt, A. Bartel, L. M. Kellermann, A. Tautenhahn, M. Volkman, K. Birnstiel, P. Do Duc, A. Hentzsch, K. C. Jensen, M. Klawitter, P. Paul, A. Stoll, S. Woudstra, P. Zuz, G. Knubben, M. Metzner, K. E. Müller, R. Merle, and M. Hoedemaker. 2021. Prevalence of disorders in preweaned dairy calves from 731 dairies in Germany: A cross-sectional study. *J. Dairy Sci.* 104:9037–9051. <https://doi.org/10.3168/jds.2021-20283>.
- Ducos, A., H. M. Berland, A. Pinton, E. Guillemot, A. Seguela, M. F. Blanc, A. Darre, and R. Darre. 1998. Nine new cases of reciprocal translocation in the domestic pig (*Sus scrofa domestica* L.). *J. Hered.* 89:136–142. <https://doi.org/10.1093/jhered/89.2.136>.
- Fuerst-Waltl, B., and M. K. Sørensen. 2010. Genetic analysis of calf and heifer losses in Danish Holstein. *J. Dairy Sci.* 93:5436–5442. <https://doi.org/10.3168/jds.2010-3227>.
- Grohs, C., A. Duchesne, S. Floriot, A. Ducos, C. Danchin-Burge, and Université Paris-Saclay. 2016. The national observatory of bovine defects: actions and results for an efficient management of genetic abnormalities (in French). *INRAE Prod. Anim.* 29:307–318. <https://doi.org/10.20870/productions-animales.2016.29.5.2998>.
- Gulliksen, S. M., K. I. Lie, T. Løken, and O. Østerås. 2009. Calf mortality in Norwegian dairy herds. *J. Dairy Sci.* 92:2782–2795. <https://doi.org/10.3168/jds.2008-1807>.
- Gustavsson, I. 1979. Distribution and effects of the 1/29 Robertsonian translocation in cattle. *J. Dairy Sci.* 62:825–835. [https://doi.org/10.3168/jds.S0022-0302\(79\)83334-2](https://doi.org/10.3168/jds.S0022-0302(79)83334-2).
- Häfliger, I. M., F. Seefried, and C. Drögemüller. 2020. Trisomy 29 in a stillborn Swiss Original Braunvieh calf. *Anim. Genet.* 51:483–484. <https://doi.org/10.1111/age.12929>.
- Hayes, B. J., and H. D. Daetwyler. 2019. 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: Applications and outcomes. *Annu. Rev. Anim. Biosci.* 7:89–102. <https://doi.org/10.1146/annurev-animal-020518-115024>.
- Hyde, R. M., M. J. Green, V. E. Sherwin, C. Hudson, J. Gibbons, T. Forshaw, M. Vickers, and P. M. Down. 2020. Quantitative analysis of calf mortality in Great Britain. *J. Dairy Sci.* 103:2615–2623. <https://doi.org/10.3168/jds.2019-17383>.
- Iannuzzi, A., P. Parma, and L. Iannuzzi. 2021. Chromosome abnormalities and fertility in domestic bovids: A review. *Animals (Basel)* 11:802. <https://doi.org/10.3390/ani11030802>.
- Johanson, J. M., P. J. Berger, S. Tsuruta, and I. Misztal. 2011. A Bayesian threshold-linear model evaluation of perinatal mortality, dystocia, birth weight, and gestation length in a Holstein herd. *J. Dairy Sci.* 94:450–460. <https://doi.org/10.3168/jds.2009-2992>.
- Knapp, J. R., G. L. Laur, P. A. Vadas, W. P. Weiss, and J. M. Tricarico. 2014. Invited Review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. *J. Dairy Sci.* 97:3231–3261. <https://doi.org/10.3168/jds.2013-7234>.
- Layer, R. M., C. Chiang, A. R. Quinlan, and I. M. Hall. 2014. LUMPY: A probabilistic framework for structural variant discovery. *Genome Biol.* 15:R84. <https://doi.org/10.1186/gb-2014-15-6-r84>.
- Leclerc, H., R. Lefebvre, M. Douguet, F. Phocas, and S. Mattalia. 2016. Mortality of calves: Phenotypic and genetic analysis [in French]. In *Proc. 23rd Renc. Rech. Rum.*, Paris.
- Lentjes, M. H. F. M., H. E. C. Niessen, Y. Akiyama, A. P. de Bruïne, V. Melotte, and M. van Engeland. 2016. The emerging role of GATA transcription factors in development and disease. *Expert Rev. Mol. Med.* 18:e3. <https://doi.org/10.1017/erm.2016.2>.
- Mattina, T., C. S. Perrotta, and P. Grossfeld. 2009. Jacobsen syndrome. *Orphanet J. Rare Dis.* 4:9. <https://doi.org/10.1186/1750-1172-4-9>.
- Mesbah-Uddin, M., C. Hoze, P. Michot, A. Barbat, R. Lefebvre, M. Boussaha, G. Sahana, S. Fritz, D. Boichard, and A. Capitan. 2019. A missense mutation (p.Tyr452Cys) in the CAD gene compromises reproductive success in French Normande cattle. *J. Dairy Sci.* 102:6340–6356. <https://doi.org/10.3168/jds.2018-16100>.
- O'Meara, C., E. Henrotte, K. Kupisiewicz, C. Latour, M. Broekhuijs, A. Camus, L. Gavin-Plagne, and E. Sellem. 2022. The effect of adjusting settings within a computer-assisted sperm analysis (CASA) system on bovine sperm motility and morphology results. *Anim. Reprod.* 19:e20210077. <https://doi.org/10.1590/1984-3143-ar2021-0077>.
- Østerås, O., M. S. Gjestvang, S. Vatn, and L. Sølverød. 2007. Perinatal death in production animals in the Nordic countries—Incidence and costs. *Acta Vet. Scand.* 49(Suppl. 1):S14. <https://doi.org/10.1186/1751-0147-49-S1-S14>.
- Pylyp, L. Y., L. O. Spynenko, N. V. Verhoglyad, A. O. Mishenko, D. O. Mykytenko, and V. D. Zukin. 2018. Chromosomal abnormalities in products of conception of first-trimester miscarriages detected by conventional cytogenetic analysis: A review of 1000 cases. *J. Assist. Reprod. Genet.* 35:265–271. <https://doi.org/10.1007/s10815-017-1069-1>.
- Raboisson, D., F. Delor, E. Cahuzac, C. Gendre, P. Sans, and G. Allaire. 2013. Perinatal, neonatal, and rearing period mortality of dairy calves and replacement heifers in France. *J. Dairy Sci.* 96:2913–2924. <https://doi.org/10.3168/jds.2012-6010>.
- Raudsepp, T., and B. P. Chowdhary. 2016. Chromosome aberrations and fertility disorders in domestic animals. *Annu. Rev. Anim.*

- Biosci. 4:15–43. <https://doi.org/10.1146/annurev-animal-021815-111239>.
- Rausch, T., T. Zichner, A. Schlattl, A. M. Stutz, V. Benes, and J. O. Korbel. 2012. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28:i333–i339. <https://doi.org/10.1093/bioinformatics/bts378>.
- Reiten, M., T. Rousing, P. T. Thomsen, N. D. Otten, B. Forkman, H. Houe, J. T. Sørensen, and M. K. Kirchner. 2018. Mortality, diarrhea and respiratory disease in Danish dairy heifer calves: Effect of production system and season. *Prev. Vet. Med.* 155:21–26. <https://doi.org/10.1016/j.prevetmed.2018.04.007>.
- Rodríguez-López, R., F. Gimeno-Ferrer, E. Montesinos, I. Ferrer-Bo-lufer, C. G. Luján, D. Albuquerque, C. M. Cataluña, V. Ballesteros, and M. A. Pérez-Gramunt. 2021. Immune deficiency in Jacobson syndrome: Molecular and phenotypic characterization. *Genes (Basel)* 12:1197. <https://doi.org/10.3390/genes12081197>.
- Santman-Berends, I. M. G. A., Y. H. Schukken, and G. van Schaik. 2019. Quantifying calf mortality on dairy farms: Challenges and solutions. *J. Dairy Sci.* 102:6404–6417. <https://doi.org/10.3168/jds.2019-16381>.
- Sargolzaei, M., J. P. Chesnais, and F. S. Schenkel. 2014. A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15:478. <https://doi.org/10.1186/1471-2164-15-478>.
- Shuster, D. E., M. E. Kehrlı, M. R. Ackermann, and R. O. Gilbert. 1992. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proc. Natl. Acad. Sci. USA* 89:9225–9229. <https://doi.org/10.1073/pnas.89.19.9225>.
- Škorić-Milosavljević, D., F. V. Y. Tjong, J. Barc, Ad P. C. M. Backx, S.-A. B. Clur, K. Spaendonck-Zwarts, R.-J. Oostra, N. Lahrouchi, L. Beekman, R. Bökenkamp, D. Q. C. M. Barge-Schaapveld, B. J. Mulder, E. M. Lodder, C. R. Bezzina, and A. V. Postma. 2019. GATA6 mutations: Characterization of two novel patients and a comprehensive overview of the GATA6 genotypic and phenotypic spectrum. *Am. J. Med. Genet. A* 179:1836–1845. <https://doi.org/10.1002/ajmg.a.61294>.
- Uetake, K. 2013. Newborn calf welfare: A review focusing on mortality rates. *Anim. Sci. J.* 84:101–105. <https://doi.org/10.1111/asj.12019>.
- van Pelt, M. L., H. Eding, P. Vessies, and G. de Jong. 2012. Developing a genetic evaluation for calf survival during rearing in the Netherlands. *Interbull Bull.* 46:61–65.
- Weckx, S., J. Del-Favero, R. Rademakers, L. Claes, M. Cruets, P. De Jonghe, C. Van Broeckhoven, and P. De Rijk. 2005. NovoSNP, a novel computational tool for sequence variation discovery. *Genome Res.* 15:436–442. <https://doi.org/10.1101/gr.2754005>.
- Ye, K., M. H. Schulz, Q. Long, R. Apweiler, and Z. Ning. 2009. Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25:2865–2871. <https://doi.org/10.1093/bioinformatics/btp394>.
- Zhang, H., Y. Wang, Y. Chang, H. Luo, L. F. Brito, Y. Dong, R. Shi, Y. Wang, G. Dong, and L. Liu. 2019. Mortality-culling rates of dairy calves and replacement heifers and its risk factors in Holstein cattle. *Animals (Basel)* 9:730. <https://doi.org/10.3390/ani9100730>.

3. Discussion

Les mutations dominantes délétères sont très fortement contre-sélectionnées, en particulier quand elles induisent la mort de l'animal. Celles qu'on observe sont donc fréquemment *de novo*. Dans certains cas cependant, des taureaux reproducteurs porteurs d'une anomalie parviennent à disséminer leurs gamètes avant d'être retirés ou détectés comme porteurs d'une anomalie. Plusieurs facteurs peuvent expliquer qu'ils soient passés inaperçus, notamment un faible impact de la mutation sur les phénotypes visibles, un traitement préférentiel pour les taureaux en centre d'insémination, un environnement plus favorable et moins de pression infectieuse, l'absence de phénotype délétère chez le taureau lui-même, ou encore un phénotype observable uniquement chez les femelles.

Au cours de ma thèse, j'ai observé plusieurs cas de taureaux porteurs de mutations délétères ayant transmis leur anomalie à de nombreux descendants par IA. Certains de ces cas ont été discutés dans des articles antérieurs, mettant en lumière des mutations mosaïques ou des translocations chromosomiques sans impact sur le phénotype du taureau porteur. Dans les deux cas, le phénotype n'est donc visible que chez les descendants. Dans le cas mosaïque, la mutation a eu lieu au cours du développement et l'individu ne porte qu'une petite proportion de cellules mutées. Cette situation n'est pas si rare, puisque l'équipe qui m'a accueilli a pu l'observer plusieurs fois, avec des mutations *de novo* dans *COL2A1* (syndrome Bulldog, Daetwyler et al. 2014) ou dans *CHD7* (syndrome de CHARGE, Bourneuf et al. 2017). Dans le cas du réarrangement, le taureau porte l'anomalie mais de façon équilibrée, il est donc viable et phénotypiquement normal. Une partie de ses gamètes est déséquilibrée, induisant des mortalités à des stades variés. Mais une autre partie est équilibrée, de sorte qu'il transmet l'anomalie qui se maintient sur plusieurs générations.

De multiples réarrangements chromosomiques ont été découverts par Jourdain et al. (2023), soulignant les répercussions de la transmission de translocations. Grâce à cette étude, de plus en plus d'entreprises de sélection mettent en place un caryotype systématique de leurs taureaux pour détecter ces réarrangements en amont.

Une mutation dominante sans mosaïsme chez un taureau d'insémination est d'autant plus grave que le taureau ne développe pas le phénotype délétère, et seuls les veaux en élevage commercial sont affectés. Par exemple, une famille de taureaux d'insémination porteurs d'une mutation dans *STAT1*, associée à la susceptibilité aux infections mycobactériennes, a été identifiée (travaux en cours au sein de l'UMT eBis). Cependant, seul l'un des trois taureaux porteurs a succombé avant

la fin de sa période de collecte de semence, ce qui soulève la possibilité d'une mauvaise immunité masquée dans un environnement sain de centre d'insémination. Les descendants de ces taureaux sont exposés à une pression infectieuse plus élevée, développant ainsi pour la plupart le phénotype délétère.

En conclusion, une analyse minutieuse des remontés à l'ONAB ou encore des taureaux présentant des taux de mortalité élevés ou des déséquilibres de liaison inhabituels peut permettre la découverte d'anomalies génétiques. D'autres caractères, tels que le poids des veaux ou les carnets sanitaires, peuvent également être utilisés pour détecter des anomalies. Cependant, la généralisation de ces méthodes au niveau national est une tâche lourde et de longue haleine.

Chapitre IV. Analyse d'association (GWAS) des taux de mortalité dans les descendance des taureaux

1. Introduction

Après avoir recherché des mutations dominantes responsables de taux de mortalité extrêmes mais limités à la descendance de certains taureaux de race Holstein, nous avons réalisé des études d'association afin de mettre en évidence des QTL (ou *Quantitative Trait Loci*) impactant la survie des veaux à l'échelle populationnelle. Comme détaillé dans la suite de ce chapitre j'ai utilisé comme performances les taux de mortalité calculés sur la descendance femelle de chaque taureau pour la période 0 à 365 jours et pour les quatre mêmes sous-périodes que dans l'article précédent, que j'ai corrigés pour des effets de milieu. J'ai mené cette étude dans les trois principales races laitières françaises car elles seules disposaient de suffisamment de taureaux avec des performances et des génotypes disponibles sur puce à SNP ($n > 1000$), ainsi que suffisamment de génomes séquencés ($n > 100$) afin de permettre une imputation des génotypes d'une densité de 50 000 marqueurs environ vers l'ensemble des variants du génome. L'imputation à l'échelle de la séquence du génome complet permet de détecter directement les variants candidats des QTL, qui sont la plupart du temps dans des régions non codantes du génome. Ces *loci* mis en évidence ont des effets de plus faible amplitude que les *loci* Mendéliens ou gènes majeurs. Les effets mis en évidence par GWAS sont des coefficients de régression sur le génotype, ils ont donc une signification statistique additive. Cela ne préjuge pas de la nature de l'effet biologique. Cette étude m'a permis d'acquérir de nouvelles compétences et de me familiariser avec de nouveaux logiciels et langage de programmation.

2. Matériel et méthodes

2.1. Construction du phénotype corrigé de mortalité

Pour chaque femelle née entre le 01/01/2000 et le 04/08/2021 (soit 1 an avant la date de début de l'étude), l'identifiant de son père, sa date de naissance, son élevage de naissance, sa date et cause de mort (mort naturelle/abattage) et dates de mouvements éventuels (vente à un autre

troupeau) ont été extraits de la base nationale. Tout veau femelle mort à la ferme pendant la période spécifique de 0 à 1 an s'est vu attribuer un phénotype de 100, les autres veaux se sont vu attribuer un phénotype de 0 s'ils sont restés vivants sur l'exploitation de naissance et ont atteint l'âge d'un an. Le même phénotype a été construit pour la période périnatale (de 0 à 2 jours d'âge), la période postnatale (de 3 à 14 jours d'âge), la période de pré-sevrage (de 15 à 55 jours d'âge) et la période de post-sevrage (de 56 à 365 jours d'âge). Quand un individu a changé d'élevage au cours de sa première année, les variables impactées n'ont pas été calculées, afin d'éviter d'éventuels biais liés à un stress dû au transport ou à un challenge immunitaire causé par la rencontre de nouveaux pathogènes.

Une correction de ce phénotype a ensuite été effectuée pour l'effet de la combinaison élevage de naissance-année-saison. Cet effet a été estimé avec un modèle père. Les calculs ont été réalisés avec le programme BLUPF90 (Misztal et al. 2014).

Pour chaque sous-période spécifique, un taux de mortalité par taureau a été calculé comme la moyenne des performances corrigées des filles du taureau (Santman-Berends, et al. 2019). Les taux de mortalité corrigés sont calculés pour les taureaux disposant d'un minimum de 100 génisses nées et restées dans leur élevage de naissance jusqu'à l'âge d'un an ou jusqu'à leur mort, si elle est intervenue avant cet âge. Par exemple, un des taureaux avait une descendance de 400 veaux femelles au début de la période de pré-sevrage (15 jours). Pendant cette période (jusqu'à 55 jours), 3 d'entre elles ont été abattues, 25 génisses sont mortes dans l'exploitation et 2 d'entre elles ont été envoyées dans une autre exploitation. En supposant que les corrections sont nulles pour les effets troupeau-année-saison, le taux de mortalité est de : 25/395 car nous retirons les 5 génisses qui n'ont pas été présente dans l'élevage sur toute la période (les 3 abattues et les 2 vendues). Après avoir conservé les taureaux disposant de génotypages sur puce à SNP, les jeux de données comprenaient 1482 taureaux Normands, 2371 Montbéliards, et 6613 Holsteins.

2.2. Imputation des génotypes de puces à SNP vers la séquence complète du génome

Ces taureaux ont été génotypés avec la puce Illumina Bovine SNP50 (50K ; Illumina Inc., San Diego, CA). En utilisant l'assemblage ARS-UCD1.2 du bovin, les génotypes de tous les taureaux ont été imputés sur la séquence. L'imputation est réalisée en deux étapes pour améliorer la précision (van Binsbergen et al. 2014): d'abord de la 50K à la HD avec le logiciel Fimpute, puis de la HD à la séquence avec le logiciel Minimac (Howie et al. 2012). La première étape a été réalisée intra race en utilisant un panel de référence d'ancêtres clés (530 Montbéliards, 551 Normands et

788 Holstein) génotypés avec la puce HD (777 000 SNP, puce Illumina Bovine HD ; Illumina Inc., San Diego, CA). La seconde étape a été réalisée intra race mais avec un panel de référence multiracial de 2712 animaux de races françaises incluant 1059 Holstein, 63 Montbéliards et 45 Normands. Ces séquences proviennent des projets de l'équipe et d'animaux du run9 du consortium *1000 Bull Genomes*. Les variants de la séquence ont été sélectionnés selon le protocole défini dans le projet *1000 Bull Genome* (Daetwyler et al. 2014). Puis les variants ont été filtrés selon le protocole de construction de la collection de séquences bovines françaises et annotés avec les logiciels SNPEff et VEP (Cingolani et al. 2012; Boussaha et al. 2016; McLaren et al. 2016).

2.3. Analyses d'association sur séquences

Nous avons réalisé des analyses d'association intra-race pour tous les variants polymorphes de $MAF > 0,02$ et avec une précision d'imputation $R^2 > 0,25$ sur les phénotypes de mortalité. Après filtres, sur un total de 29 577 118 SNP et petits InDels, 12 062 394, 11 732 041 et 11 801 159 ont été respectivement sélectionnés sur les races Montbéliarde, Normande et Holstein.

Toutes les analyses d'association ont été effectuées en utilisant l'option mlma du logiciel GCTA, qui applique un modèle linéaire mixte (Yang et al. 2011) à tous les variants tel que :

$$\mathbf{y} = \mathbf{1}\mu + \boldsymbol{\alpha}\mathbf{b} + \mathbf{u} + \mathbf{e}$$

où \mathbf{y} représente le vecteur des valeurs phénotypiques corrigées, μ est une moyenne, $\boldsymbol{\alpha}$ est le vecteur des doses de génotypes imputées, \mathbf{b} est l'effet fixe du variant, \mathbf{u} est le vecteur des effets polygéniques (prenant en compte la structure génétique de la population) et \mathbf{e} est l'erreur résiduelle. Le seuil de significativité a été fixé à 10^{-6} , soit $-\log_{10}(p)=6$ et pour chaque région le variant le plus significatif dit variant « cime » a été sélectionné.

2.4. Recherche de variants candidats

Pour chaque QTL les variants présentant une valeur de $-\log_{10}(p)$ supérieure ou égale à 2/3 de la $-\log_{10}(p)$ du variant « cime » et affectant les régions codantes du génome d'après les annotations disponibles ont été considérés. Leurs conséquences phénotypiques ont ensuite été prédites en étudiant les conséquences de mutations délétères affectant des gènes orthologues chez l'Homme et la souris (Kolishovski et al. 2019; Amberger et al. 2019).

3. Résultats

3.1. Résultats sur les GWAS

Au total 25 QTL correspondant à 17 régions uniques ont été détectés comme significatifs dans au moins une des 15 analyses réalisées (3 races x 5 périodes). Parmi eux 12 sont significatifs pour une seule période de mortalité dans une seule race, 3 sont significatifs pour plusieurs périodes dans une seule race et 2 sont partagés entre races (voir Tableau IV-1). Les fréquences des variants « cime » varient considérablement, allant de 2,6% à 87,8%, tandis que les cologarithmes des valeurs de p ($-\log_{10}(p)$) varient entre 6,2 et 36,7. Les effets estimés varient de 0,15 à 1,62%. Estimés sur les moyennes de père, ils doivent être multipliés par deux pour refléter l'effet à l'échelle individuelle. Ils varient donc de 0,3 à 3,2%.

C'est en race Holstein que la population de taureaux est la plus grande, c'est donc ce dispositif qui est le plus puissant et qui, logiquement, fournit le plus grand nombre de QTL. Onze QTL ont été détectés dans les GWAS sur séquences des différentes périodes (Figure IV-1). Cinq QTL significatifs sont trouvés pour la période périnatale (a.), 1 QTL dans chacune des périodes post-natale, pré-sevrage et post-sevrage (b.,c.,d.) et 3 QTL dans la période globale. Trois QTL baptisés MH16P64, MH23P25 et MH5P95 (avec des $-\log(p)$ allant de 10,1 à 36,7) sont communs aux cartographies s'intéressant à la période globale et à au moins une autre période.

En race Montbéliarde, 9 QTL ont été découverts : 2 pour la mortalité périnatale (Figure IV-2), 1 pour la période post-natale (b.), 1 pour la période pré-sevrage (c.), 2 pour le post-sevrage (d.) et 3 pour la mortalité sur toute la première année (e.). Il est intéressant de noter qu'un QTL revient dans 3 périodes différentes, sur le chromosome 6, avec une position « cime » à 72 567 590 et un $-\log(P)$ variant de 12,7 à 21,2. Un autre QTL nommé MM11P50 est détecté dans les périodes périnatale et globale (a. et e.) avec un $-\log(P)$ variant de 6,9 à 8,2. Il est intéressant de noter que le même SNP « cime » (en position 49 640 162 paires de bases sur le chromosome 11) est également détecté en race Holstein (nommé MH11P50 dans cette race) dans la cartographie portant sur la mortalité périnatale, et les effets sont très proches dans les deux races (0,50 et 0,52%).

Enfin 6 QTL ont été détectés en Normande (Figure IV-3) notamment 1 pour les mortalités périnatales (a.), 1 en post-natale (b.) 2 en post-sevrage (d.) et 1 pour la période complète (e.). Contrairement à la Montbéliarde et à la Holstein, aucun QTL n'est significatif pour deux périodes différentes. Par contre un QTL de taux de mortalité périnatal (MN23P25) est identique à un QTL

déecté en race Holstein pour le même caractère et affiche des effets proches (0,7 et 0,8%, respectivement).

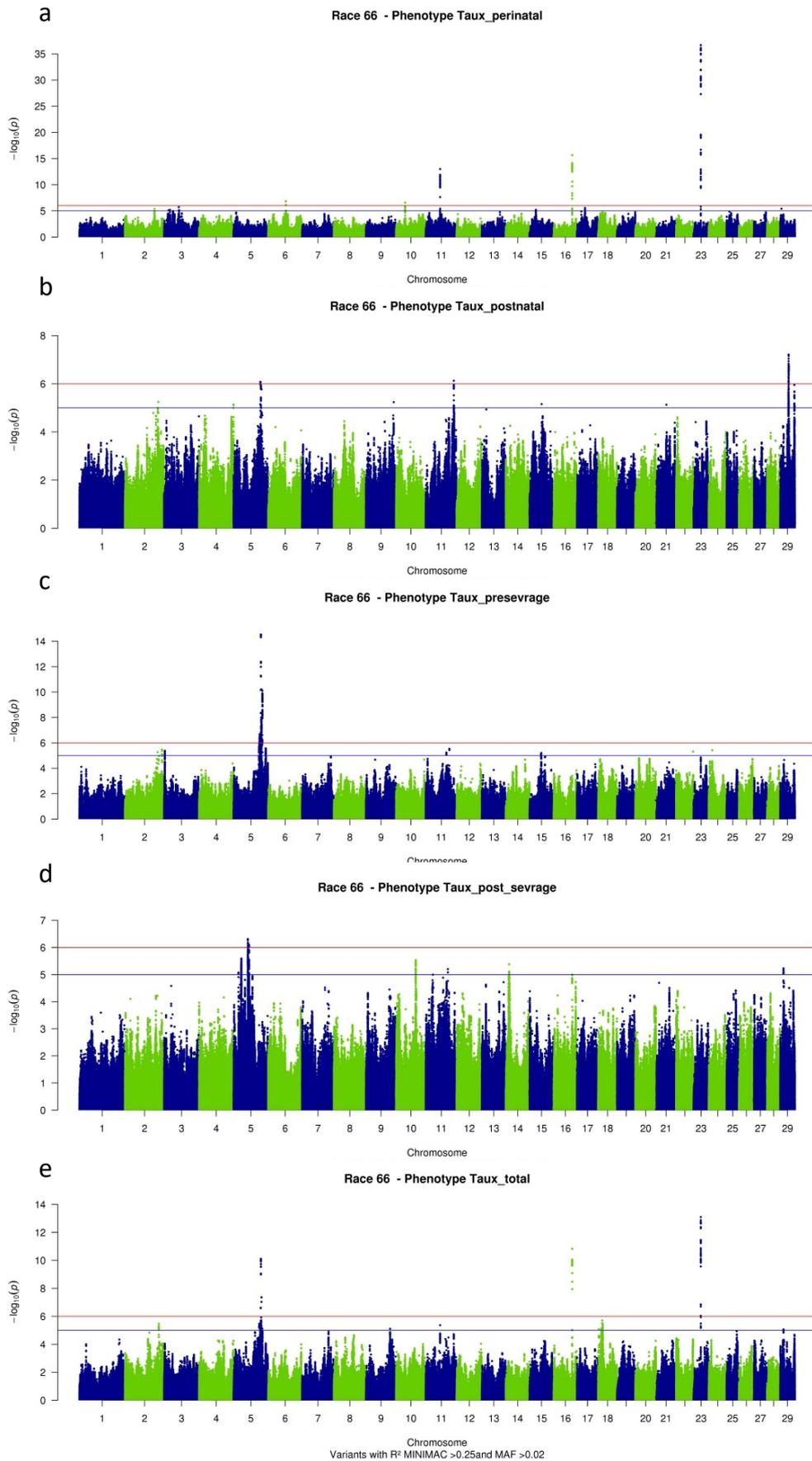


Figure IV-1 : Manhattan Plots de l'analyse GWAS des 5 caractères de mortalité pour la race Holstein (66).

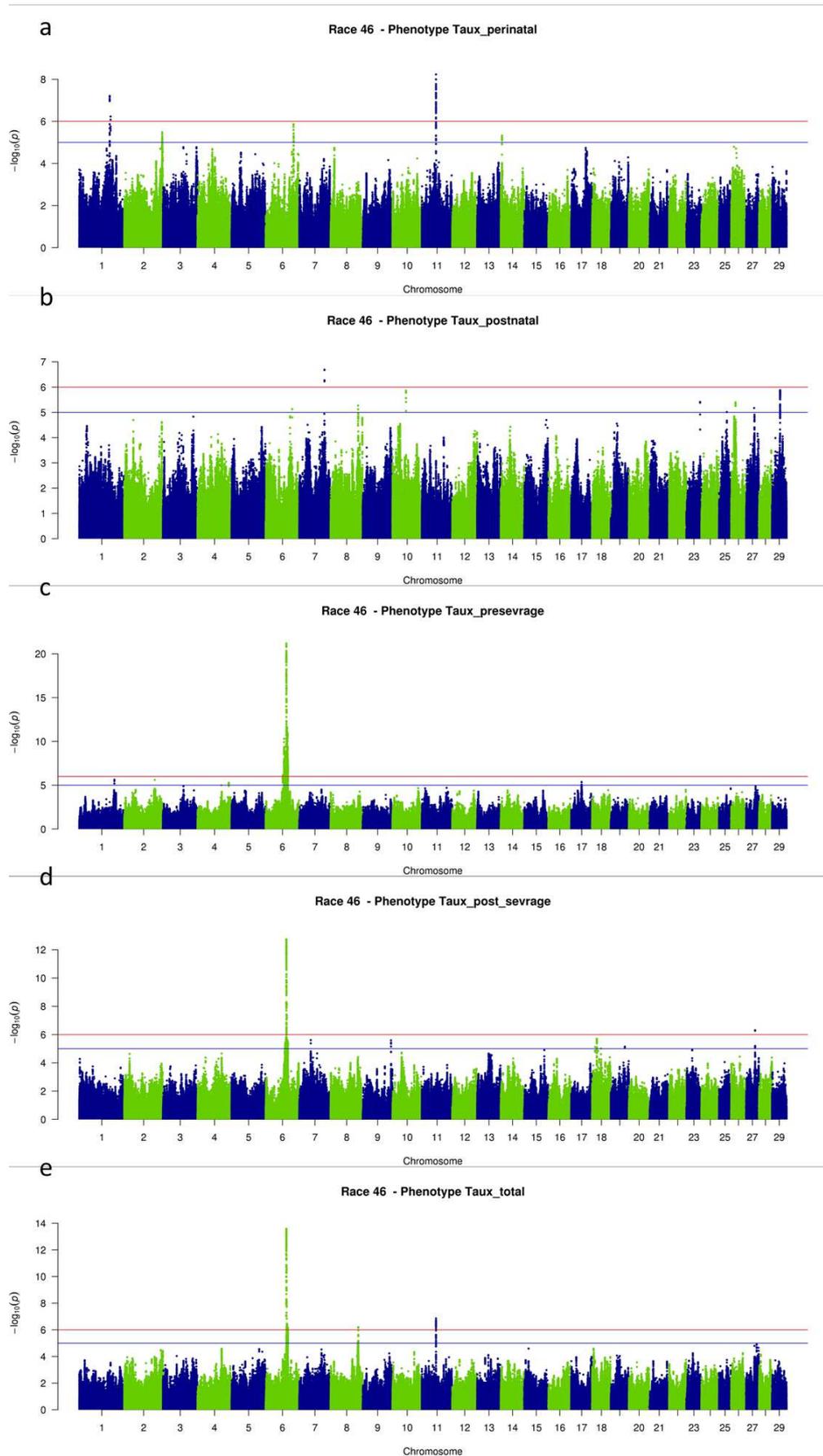


Figure IV-2 : Manhattan Plots de l'analyse GWAS des 5 caractères de mortalité pour la race Montbéliarde (46).

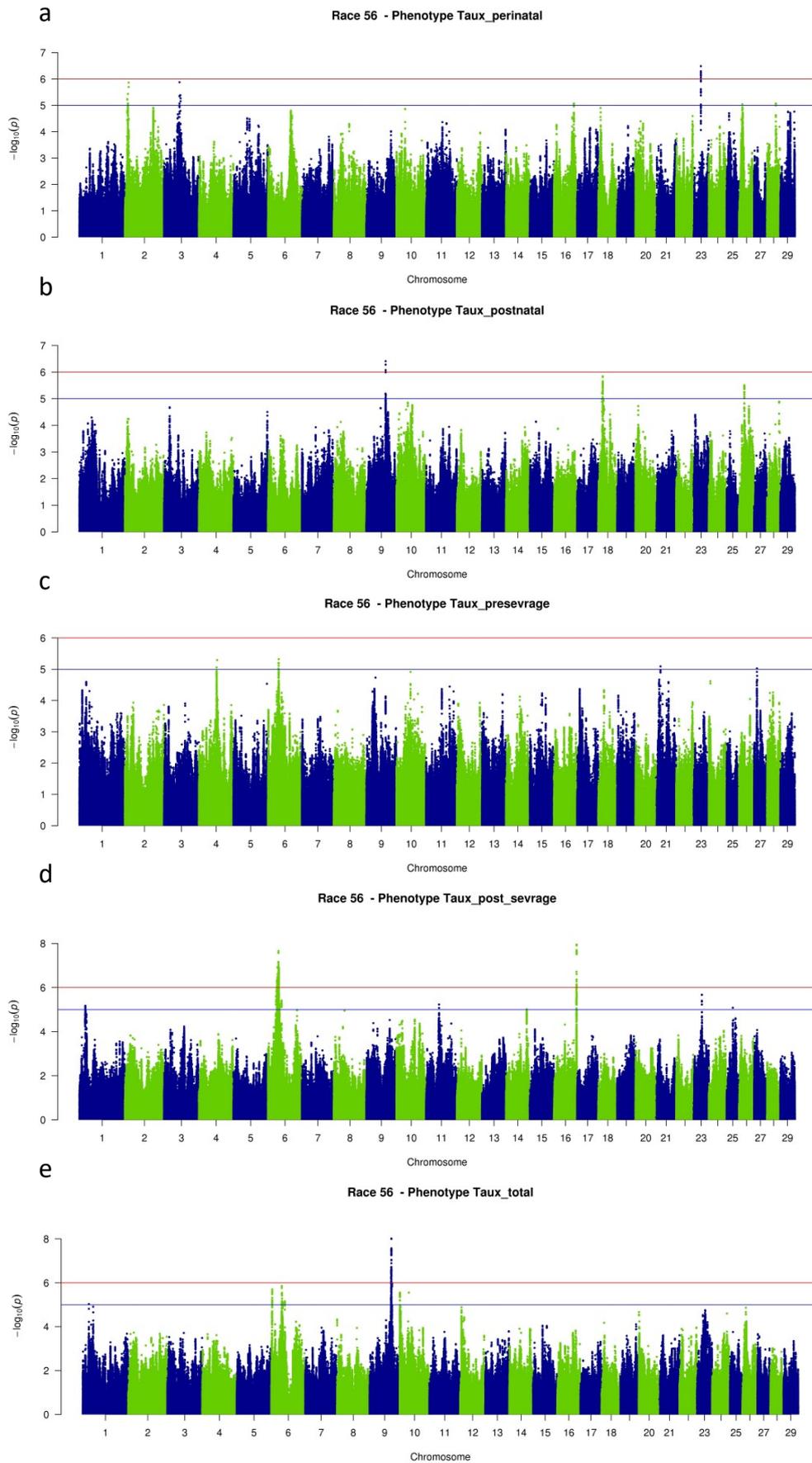


Figure IV-3 : Manhattan Plots de l'analyse GWAS des 5 caractères de mortalité pour la race Normande (46).

Tableau IV-1 : Liste des QTL significatifs et de leurs variants « cime » pour les 3 races et 5 caractères de mortalité juvénile étudiés.

QTL_ID: nom donné au QTL, **Chr:** chromosome, **b:** Effet allélique associé à l'allèle alternatif du SNP « cime » (effet sur la moyenne de la descendance ; il représente la moitié de l'effet de substitution), **Freq_alt:** fréquence de l'allèle alternatif du SNP « cime ». Les QTL ayant un effet significatif sur plusieurs périodes sont mis en évidence en gras et ceux partagés entre races sont soulignés.

Race	Caractère	QTL_ID	Chr	Position	-Log(P)	b (%)	Ref	Freq_alt (%)
Holstein	Taux post sevrage	MH5P49	5	49476273	6,31	-0,15	C	50,3
Holstein	Taux pré-sevrage	MH5P95	5	94876545	14,54	0,71	G	2,6
Holstein	Taux total	MH5P95	5	94876545	10,11	1,27	G	2,6
Holstein	Taux périnatal	MH6P61	6	61317377	6,83	0,44	G	3,1
Holstein	Taux périnatal	MH10P31	10	31468535	6,56	-0,23	C	13,5
Holstein	Taux périnatal	MH11P50	11	49640162	13	0,26	C	34
Holstein	Taux périnatal	MH16P64	16	64188174	15,65	0,3	T	36,3
Holstein	Taux total	MH16P64	16	64188174	10,83	0,41	T	36,3
Holstein	Taux périnatal	MH23P25	23	24922231	36,69	0,4	C	32,5
Holstein	Taux total	MH23P25	23	24922231	13,1	0,38	C	32,5
Holstein	Taux post-natale	MH29P29	29	29367341	7,21	0,23	A	9,8
Montbéliarde	Taux périnatal	MM1P107	1	107238246	7,2	-0,4	T	87,8
Montbéliarde	Taux pré-sevrage	MM6P73	6	72583729	21,17	1,04	C	5,4
Montbéliarde	Taux post sevrage	MM6P73	6	72567590	12,74	0,6	A	8,1
Montbéliarde	Taux total	MM6P73	6	72583040	13,58	1,62	T	5,4
Montbéliarde	Taux post-natale	MM7P89	7	88707238	6,69	0,3	T	17,2
Montbéliarde	Taux total	MM8P97	8	97246347	6,2	-0,42	G	51,2
Montbéliarde	Taux périnatal	MM11P50	11	49640162	8,23	0,25	C	47,9
Montbéliarde	Taux total	MM11P50	11	49640162	6,86	0,44	C	47,9
Montbéliarde	Taux post sevrage	MM27P32	27	32271077	6,31	-0,22	C	53,1
Normande	Taux post sevrage	MN6P36	6	36404086	7,65	0,93	A	15,1
Normande	Taux post-natale	MN9P67	9	67130897	6,41	-0,44	G	77,4
Normande	Taux total	MN9P76	9	75586734	8,01	-0,74	ATGAT	37,8
Normande	Taux post sevrage	MN16P79	16	78976775	7,95	-0,44	GCA	46,1
Normande	Taux périnatal	MN23P25	23	24922231	6,48	0,35	C	31,4

3.2. Résultats sur les variants candidats

Globalement, les quatre QTL ayant les effets alléliques les plus marqués sur les taux de mortalité sont MM6P73, MH5P95, MN6P36 et MN11P50/MH11P50. Ces différents QTL sont retrouvés dans les trois races étudiées et sont les seuls QTL associés à des variants candidats avec un score SIFT délétère (Tableau IV-2).

Un variant avec pour conséquence un *frameshift* (décalage dans la lecture du codant) est associé au QTL MM6P73 pour les trois périodes. Ce type de variant entraîne des conséquences importantes sur la traduction de la protéine et ce variant sera discuté plus en détail dans le Chapitre V.

Pour le QTL MN6P36, un variant localisé dans le gène *FAM13A* est associé avec le QTL avec une fréquence importante chez les taureaux étudiés de la race Normande. Ce variant est le seul associé au QTL qui est prédit comme délétère selon le score SIFT, à vérifier plus en détail car il est annoté « *low_confidence* ». Des mutations sur *FAM13A* ont été associées à des maladies pulmonaires interstitielles (Fingerlin et al. 2013). Pour les QTL MN11P50/MH11P50, un *splice_donor_variant* a été associé avec les QTL. Ce SNP est localisé dans un transcrit de *ELMOD3*, qui existe seulement chez le bovin (source UCSC). Il sera intéressant de récupérer des ARN chez des contrôles et des cas pour voir s'il y a réellement une différence de transcription.

Pour les autres QTL, notamment MH29P29, MH5P49 et MM1P107, 3 variants candidats ont été trouvés, tous ayant un score SIFT toléré allant de 0,13 à 0,18, ils ont un effet allélique sur la mortalité de 0,34, -0,30 et -0,80% respectivement

Tableau IV-2 : Liste des variants candidats affectant une région codante du génome

REF et ALT: allèle de référence au niveau de l'assemblage du génome bovin (ARS-UCD1.2/bosTau9 ; (Rosen et al. 2020)) et allèle alternatif, respectivement. Freq_alt : fréquence de l'allèle alternatif. Transcrit : transcrit selon l'annotation Ensembl release 111 (Martin et al. 2023). SIFT : score SIFT calculé par le logiciel VEP (Ng et Henikoff 2003; McLaren et al. 2016)

Race	QTL	Période	Chr	Position	REF	ALT	Freq_alt	b	Log_p	Transcrit	Protéine/conséquence	SIFT
Montbéliarde	MM1P107	Périnatale	1	107237713	G	T	87,8%	-0,40%	7,20	ENSBTAT00000106519.1	C1H3orf80 p.20 L>M	0,18
Holstein	MH5P49	Post-sevrage	5	49438405	C	G	51,9%	-0,15%	5,58	ENSBTAT00000053671.4	C5H12orf56 p.398T>R	0,13
Normande	MN6P36	Post-sevrage	6	35975293	G	A	41,0%	0,75%	5,21	ENSBTAT00000076308.2	FAM13A p.535R>K	1
Normande	MN6P36	Post-sevrage	6	36028197	A	G	18,0%	0,64%	6,32	ENSBTAT00000076308.2	FAM13A p.1077R>G	0,04
Normande	MN6P36	Post-sevrage	6	37410605	T	C	18,4%	0,65%	6,25	ENSBTAT00000121583.1	LCORL p.458I>V	0,82
Montbéliarde	MM6P73	Pré-sevrage	6	72359797	C	CG	5,8%	0,92%	17,56	ENSBTAT00000025792.5	NOA1A p.D400RfsX9	-
Montbéliarde	MM6P73	Post-sevrage	6	72359797	C	CG	5,8%	0,79%	11,68	ENSBTAT00000025792.5	NOA1A p.D400RfsX9	-
Montbéliarde	MM6P73	Totale	6	72359797	C	CG	5,8%	1,49%	12,05	ENSBTAT00000025792.5	NOA1A p.D400RfsX9	-
Holstein	MH29P29	Post-natale	29	29778035	G	A	12,9%	0,17%	5,11	ENSBTAT00000066792.2	KIRREL3 p.280T>I	0,17

4. Discussion

Le fait que seul les QTL ayant le plus d'effet sur une hausse de la mortalité ont des variants délétères dans le codant montre que certains variants avec des effets moins élevés sur la mortalité et donc modérés sur le codant peuvent passer inaperçus lors de la recherche de variant causaux.

Les QTL MM11P50/MH11P50, ainsi que les QTL MH23P25/MN23P25, présentent des pics centrés sur des SNP qui sont inclus dans la puce 50k+, F1111660 et F2304960 respectivement. En revanche, le pic MM6P73 (QTL qui sera validé dans le Chapitre V) ne présente pas cette caractéristique.

En effet, les QTL 23P25 de la Holstein et de la Normande ont été invalidés à la suite de l'analyse des fréquences des différentes puces, qui a révélé une inversion d'allèles entre les versions 50k v1 et v2 par rapport aux autres versions de la puce (L+, LX, M+, HD, 5+). En effet, pour certains marqueurs, les premiers typages réalisés au Centre National de Génotypage différaient des typages réalisés par la suite. Cette inversion s'est traduite ensuite par un artefact lors de l'analyse d'association : la mortalité qui a augmenté au cours du temps a été associée à tort à ces variants dont la fréquence s'est inversée durant la même période. Par ailleurs, cette erreur a également conduit à une très mauvaise imputation dans la région, de sorte que les pics artefact correspondaient aux typages vrais. Ces résultats ont ensuite été éliminés.

A la suite de cette étude, 164 marqueurs associés aux 25 QTL de mortalité ont été ajoutés sur la puce de génotypage (*cf.* Annexe 1 : Tableau contenant les SNP mis sur la puce EuroGMD V12). Ces marqueurs permettront de tester leur effet sur la mortalité à partir de typages vrais et non imputés. Certains seront validés dans la suite du manuscrit (notamment *NOA1* dans le Chapitre V), tandis que d'autres feront partie d'une étude prochaine. En effet, il est nécessaire de collecter environ 2 à 3 ans de données pour obtenir assez de génotypes et de phénotypes pour distinguer des différences significatives et ainsi confirmer ou non les variants candidats.

5. Conclusion

Ce chapitre dédié aux analyses d'association a permis d'identifier 17 QTL distincts associés à la mortalité, variant en fréquences dans notre échantillon. Pour quatre de ces QTL, nous avons mis en évidence des variants candidats délétères, dont l'un d'eux a été invalidé ultérieurement. Nous

n'avons recherché les variants candidats que dans les gènes. Il est tout à fait possible que ce choix soit trop restrictif et ces candidats doivent impérativement être confirmés, au moins statistiquement sur des données indépendantes.

Ce chapitre ouvre la voie à plusieurs perspectives prometteuses pour la recherche future. En premier lieu, une nouvelle étape d'imputation sur séquence est envisagée, impliquant l'utilisation d'un panel de référence actualisé et la correction des SNP problématiques, dans le but de corriger et d'améliorer la première phase d'imputation de la puce 50K à la puce HD.

En parallèle, une extension de la sélection des individus à inclure dans les analyses est prévue, en englobant tous les taureaux ayant engendré plus de 100 ou 1000 descendants. Cette démarche permettra de conduire de nouvelles analyses d'association génomique (GWAS), visant à accroître la puissance, potentiellement aboutissant à l'identification supplémentaire de QTL. De plus, l'élargissement de la période d'observation à 720 jours, ainsi que l'inclusion des données de boucherie, sont envisagés afin de réaliser une analyse plus exhaustive de la mortalité.

En somme, ce chapitre a permis une exploration approfondie du phénotype de mortalité, ainsi que des premières investigations sur la causalité des variants. La prochaine étape dans le chapitre suivant consistera à appliquer une autre méthode de cartographie génétique, afin d'identifier de nouveaux variants candidats et de valider, à l'aide d'études phénotypiques détaillées, trois variants causaux.

Chapitre V. Détection des anomalies récessives chez les bovins laitiers : une approche intégrant l'analyse des trajectoires de vie et le déficit et l'enrichissement en homozygotes

1. Introduction

Après avoir identifié les anomalies dominantes et recherché des QTL à l'aide de GWAS, une proportion significative de la mortalité demeure inexpliquée. Dans cette perspective, nous avons élaboré une méthode novatrice pour repérer les anomalies récessives qui auraient pu échapper à la détection par les méthodes conventionnelles discutées dans la synthèse bibliographique.

Cette approche repose sur une analyse minutieuse des trajectoires de vie des animaux, combinée à la recherche d'un enrichissement et d'un déficit en homozygotes (*Homozygous Haplotype Enrichment / Depletion*, HHED) au sein des trois principales races laitières. Les méthodes traditionnelles de détection des anomalies génétiques récessives présentent leurs limites, notamment dans leur capacité à dépister les défauts chez les animaux vivants à un instant précis (cf. I-7.2.1). De plus, ces méthodes sont souvent confrontées à des problèmes de déséquilibre ou de pénétrance incomplètes, rendant la détection des anomalies plus compliquée.

Dans cette optique, nous avons recherché un enrichissement en homozygotes parmi les animaux décédés. La méthodologie repose sur l'utilisation conjointe des données des animaux génotypés à des fins d'évaluation génomique, et de leurs trajectoires de vie telles qu'on peut l'appréhender à partir de l'informations des bases de données nationales. Cette analyse se concentre ici sur la mortalité juvénile, mais elle peut être adaptée à d'autres stades physiologiques et pathologies. Par la suite, les génotypes des puces SNP sont étudiés pour détecter des haplotypes délétères, et des tests sur haplotypes sont utilisés pour estimer le statut des individus génotypés.

Cette méthodologie innovante est validée par l'analyse des trajectoires de vie de millions d'animaux issus de parents génotypés. L'origine de ces haplotypes est déterminée et l'évolution de leur fréquence au cours du temps estimée. Enfin, les séquences de milliers d'animaux sont

analysées pour identifier des variants candidats, ce qui permet une caractérisation fonctionnelle et une estimation des effets de ces variants sur la survie et divers caractères au niveau de la population. L'article soumis à Genome Biology est court et très synthétique, en format « Letter ». Afin d'en faciliter la lecture, je vous propose de consulter le fichier des données supplémentaires en parallèle du manuscrit (Données supplémentaires <https://www.biorxiv.org/content/10.1101/2023.09.22.558782v1.supplementary-material>).

2. Article soumis : “Massive detection of cryptic recessive genetic defects in cattle mining millions of life histories “



HOME | SUBMIT

New Results

Follow this preprint

Massive detection of cryptic recessive genetic defects in cattle mining millions of life histories

Florian Besnard, Ana Guintard, Cécile Grohs, Laurence Guzylack-Piriou, Margarita Cano, Clémentine Escoufflaire, Chris Hozé, Hélène Leclerc, Thierry Buronfosse, Lucie Dutheil, Jeanlin Jourdain, Anne Barbat, Sébastien Fritz, Marie-Christine Deloche, Aude Remot, Blandine Gaussères, Adèle Clément, Marion Bouchier, Elise Contat, Anne Relun, Vincent Plassard, Julie Rivière, Christine Péchoux, Marthe Vilotte, Camille Eche, Claire Kuchly, Mathieu Charles, Arnaud Boulling, Guillaume Viard, Stéphanie Minéry, Sarah Barbey, Clément Birbes, Coralie Danchin-Burge, Frédéric Launay, Sophie Mattalia, Aurélie Allais-Bonnet, Bérangère Ravary, Yves Millemann, Raphaël Guatteo, Christophe Klopp, Christine Gaspin, Carole Lampietro, Cécile Donnadieu, Denis Milan, Marie-Anne Arcangioli, Mekki Boussaha, Gilles Foucras, Didier Boichard, Aurélien Capitan

doi: <https://doi.org/10.1101/2023.09.22.558782>

Massive detection of cryptic recessive genetic defects in cattle mining millions of life histories

Florian Besnard^{1,2,*}, Ana Guintard^{1,3}, Cécile Grohs¹, Laurence Guzylack-Piriou⁴, Margarita Cano¹, Clémentine Escouflaire^{1,3}, Chris Hozé^{1,3}, Hélène Leclerc^{1,3}, Thierry Buronfosse⁵, Lucie Dutheil⁴, Jeanlin Jourdain^{1,3}, Anne Barbat¹, Sébastien Fritz^{1,3}, Marie-Christine Deloche^{1,3}, Aude Remot⁶, Blandine Gaussères⁴, Adèle Clément⁴, Marion Bouchier⁵, Elise Contat⁵, Anne Relun⁷, Vincent Plassard⁸, Julie Rivière^{1,9}, Christine Péchoux¹, Marthe Vilotte¹, Camille Eche¹⁰, Claire Kuchly¹⁰, Mathieu Charles¹, Arnaud Boulling¹, Guillaume Viard^{3,11}, Stéphanie Minéry², Sarah Barbey¹², Clément Birbes¹³, Coralie Danchin-Burge², Frédéric Launay¹², Sophie Mattalia², Aurélie Allais-Bonnet^{3,11}, Bérangère Ravary⁸, Yves Millemann⁸, Raphaël Guatteo⁷, Christophe Klopp¹³, Christine Gaspin¹³, Carole Iampietro¹⁰, Cécile Donnadiou¹⁰, Denis Milan¹⁴, Marie-Anne Arcangioli⁵, Mekki Boussaha¹, Gilles Foucras⁴, Didier Boichard¹ and Aurélien Capitan^{1,3*}

1 Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

2 IDELE, 75012 Paris, France

3 ELIANCE, 75012 Paris, France

4 IHAP, Université de Toulouse, INRAE, ENVT, 31076 Toulouse, France

5 VetAgro Sup, Université Lyon1, 69280 Marcy-l'Etoile, France

6 INRAE, Université de Tours, ISP, 37380 Nouzilly, France

7 Oniris, INRAE, BIOEPAR, 44300 Nantes, France

8 ENVA, 94700 Maisons-Alfort, France

9 Université Paris-Saclay, INRAE, AgroParisTech, MICALIS, 78350 Jouy-en-Josas, France

10 INRAE, US 1426, GeT-PlaGe, Genotoul, France Génomique, Université Fédérale de Toulouse, 31320 Castanet-Tolosan, France

11 Université Paris-Saclay, UVSQ, INRAE, BREED, 78350 Jouy-en-Josas, France

12 Unité Expérimentale du Pin, INRAE, 61310 Le Pin au Haras, France

13 Université Fédérale de Toulouse, INRAE, BioinfOmics, GenoToul Bioinformatics facility, 31320 Castanet-Tolosan, France

14 GenPhySE, Université Fédérale de Toulouse, INRAE, INPT, ENVT, 31320 Castanet-Tolosan, France

* Corresponding authors:

Aurélien Capitan

Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France

aurelien.capitan@inrae.fr

Florian Besnard

IDELE, 149 rue de Bercy, 75012 Paris, France

florian.besnard@idele.fr

Abstract

We present a data-mining framework designed to detect recessive defects in livestock that have been previously missed due to a lack of specific signs, incomplete penetrance, or incomplete linkage disequilibrium. This approach leverages the massive data generated by genomic selection. Its basic principle is to compare the observed and expected numbers of homozygotes for sliding haplotypes in animals with different life histories. Within three cattle breeds, we report 33 new loci responsible for increased risk of juvenile mortality and present a series of validations based on large-scale genotyping, clinical examination, and functional studies for candidate variants affecting the *NOA1*, *RFC5*, and *ITGB7* genes. In particular, we describe disorders associated with *NOA1* and *RFC5* mutations for the first time in vertebrates. The discovery of these many new defects will help to characterize the genetic basis of inbreeding depression, while their management will improve animal welfare and reduce losses to the industry.

Keywords: data science, recessive genetic defects, livestock, large-scale genotyping, whole-genome sequencing, inbreeding depression, life history.

Main

Cattle breeds are populations of limited effective size, subject to recurrent outbreaks of recessive defects. Historically, surveillance networks have been established to allow early detection of animals with distinctive clinical features and positional cloning of the locus. With the tremendous development of high-throughput genotyping and sequencing technologies, alternative strategies

have been devised to detect genetic conditions that may easily go unnoticed due to a lack of specific signs. These include the search for homozygous haplotype deficiency (VanRaden et al. 2011, Fritz et al. 2013), reverse genetics (e.g. ; Charlier et al. 2016) and non-additive association analysis using proxy phenotypes (Reynolds et al. 2021), which have been particularly effective in mapping loci responsible for embryonic or perinatal death and growth retardation. However, these methods are not suitable in cases of incomplete penetrance or incomplete linkage between the causative variant and surrounding markers, which is typically the case for emerging defects due to recent mutations.

To address this issue, we have developed an integrative approach, which is summarized in Fig. 1. The first steps are to list animals genotyped with SNP arrays for routine genomic evaluation, and to mine databases for information on events that shaped their lives to define groups of individuals with distinct life histories (Fig. 1a,b). As a proof of concept, we considered 16,655 heifers that died of natural causes during the rearing period and 488,967 adult cows from the Holstein, Montbeliarde, and Normande breeds (Supplementary Table 1). We analyzed phased Illumina BovineSNP50 genotypes, searching for sliding haplotypes of 20 markers, with at least 25% enrichment in dead heifers and 25% depletion in cows in terms of observed versus expected homozygous animals based on the genotypes of their paternal and maternal ancestors. This approach, which we named “homozygous haplotype enrichment/depletion” (HHED) mapping, yielded numerous candidate regions (Fig. 1c and 2; Supplementary Tables 2,3). For further analysis, we selected the top 20 peak haplotypes per breed and examined the life histories of 8.8 million females born before 2016 whose sires and maternal grandsires (MGS) were genotyped (Fig. 1e,d). After comparing the number of animals in three categories (“dead heifers”, “cows”, and “other”; see Methods) between the offspring of at-risk (carrier sires and carrier MGS) and control mating, we validated 34 of the 60 haplotypes (13, 11, and 10 in the three breeds, respectively) with frequencies ranging from 1.5 to 7.6% (Benjamini-Hochberg adjusted chi-square p -value ≤ 0.05 ; Supplementary Tables 4,5). We then analyzed 5 million recent insemination records with sire and dam genotyped. We found 3.4, 3.6 and 8.0 % of at-risk mating in the three breeds and, by extrapolation, estimated that more than 293 thousand calves were born homozygous for at least one locus of increased risk of juvenile mortality in France during the last 10 years (Supplementary Table 6).

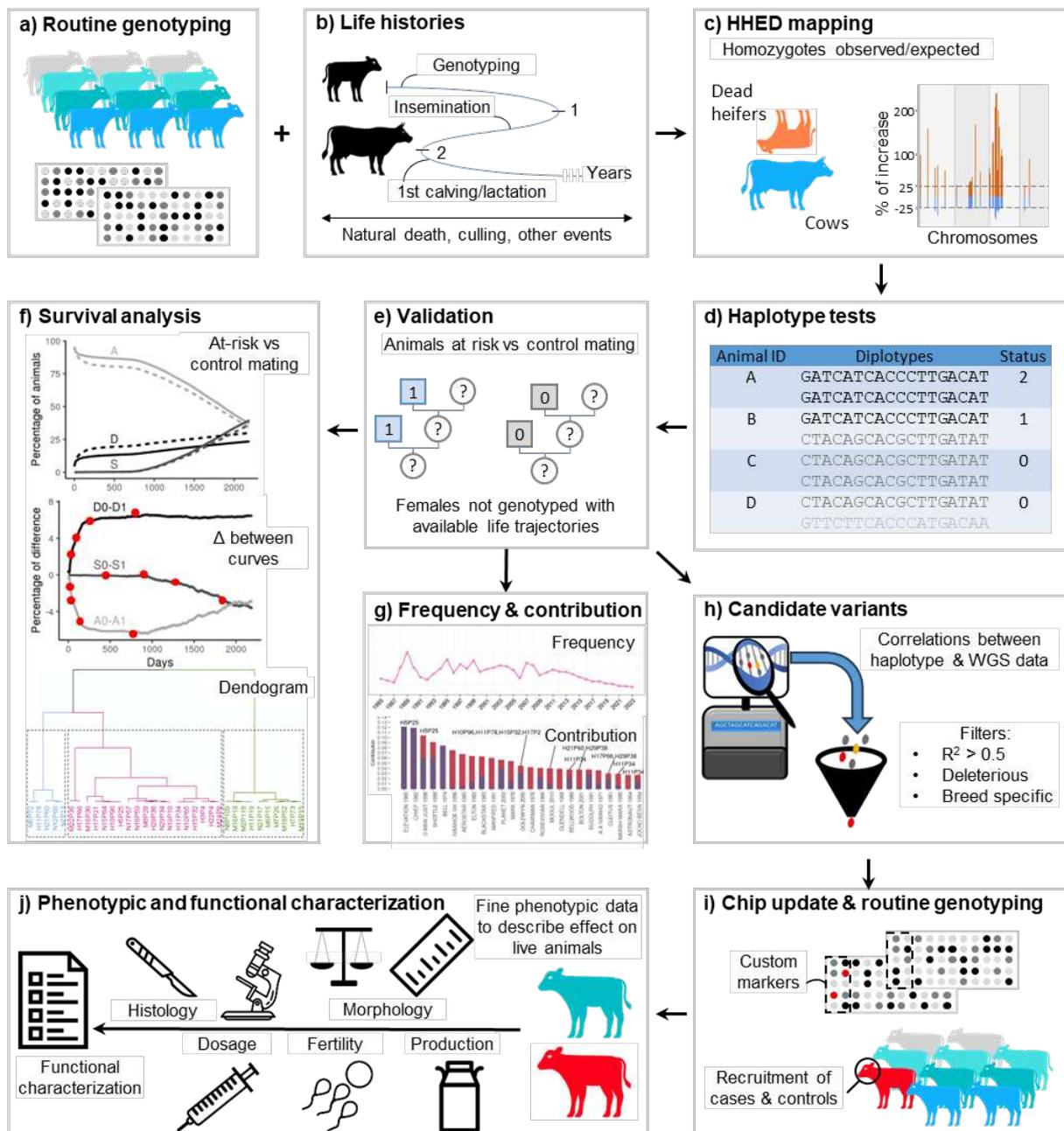


Figure 1. Overview of the analysis plan developed in this study to detect and characterize cryptic recessive genetic defects in cattle.

Graphs are shown for illustrative purposes only. Detailed information is provided throughout the manuscript. The first two steps consist of listing animals genotyped for genomic evaluations that have genotypes available for their ancestors (a), and extracting information from national databases to form groups of thousands to tens of thousands of individuals with different life histories (b). Here we focused on juvenile mortality, but the strategy is applicable to other physiological stages and pathologies (premature culling due to infertility or poor production, pre- or post-partum death of dams, etc.). Then phased SNP array genotypes are searched for sliding haplotypes of 20 markers, with at least 25% enrichment in the affected group and 25% depletion in the unaffected group in terms of observed versus expected homozygotes based on the genotypes of their paternal and maternal ancestors (c). Haplotype tests are set up to provide a status (0, 1, or 2 copies) for all genotyped individuals for a selection of the most promising loci (here $n=20$ per breed; d). The life history of millions of animals born from at-risk or control mating (which are mostly not genotyped) is analyzed and the proportions of animals meeting the inclusion criteria in the affected and unaffected groups are considered for validation (e). These data are also used to characterize survival curves over time using various parameters, principal component analysis, and hierarchical clustering (f). The evolution of haplotype frequencies over time and their causes (e.g. founder effect, heterozygous advantage) are also studied (g). Whole genome sequences from thousands of animals with available haplotype status are then analyzed to identify candidate variants (h). Some

of these variants may have been included as markers in the SNP arrays used for genomic evaluation in the framework of reverse genetics studies and have genotypes available for large populations over the last decade (i). These data are used to recruit case and control animals for pathophysiological and functional characterization and to estimate the effects of these variants on survival and a variety of traits at the population level (j). Icons are from www.thenounproject.com.

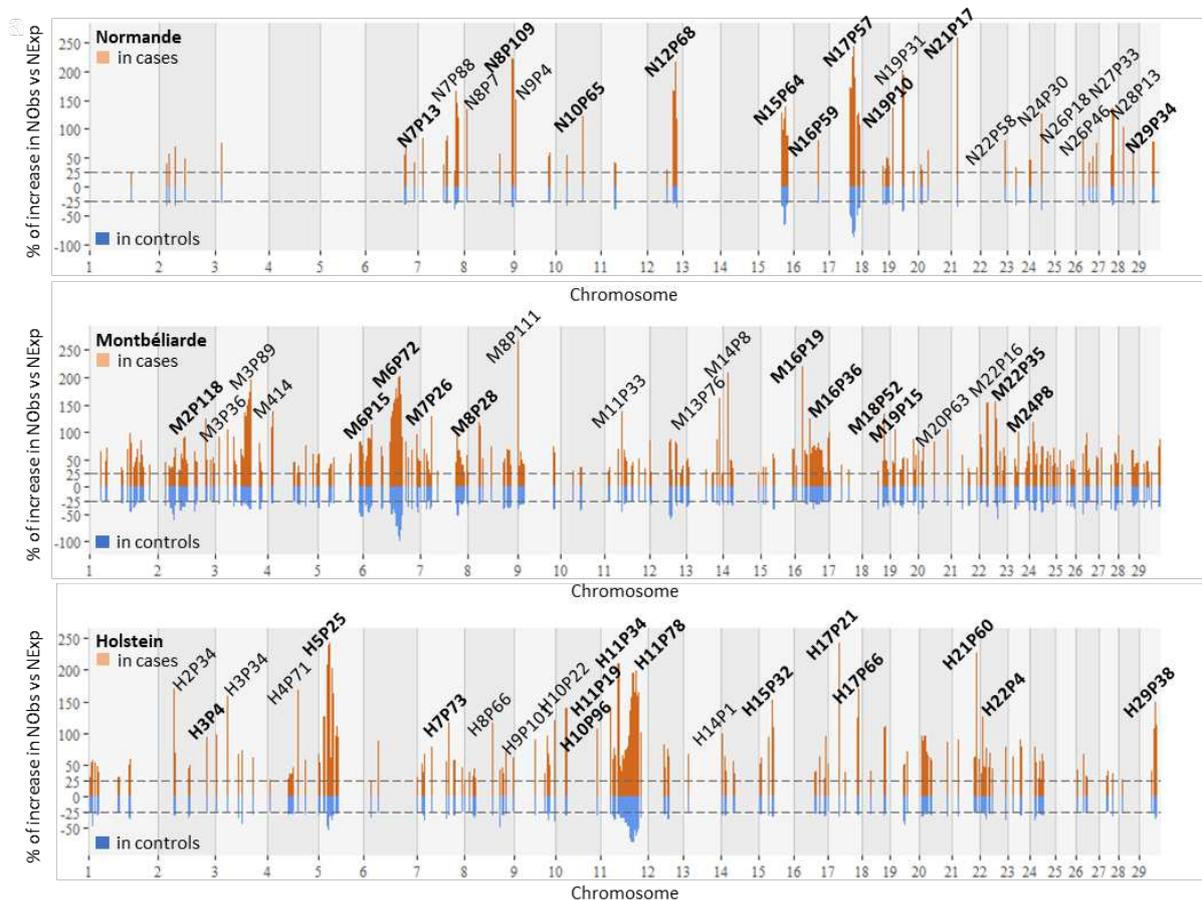


Figure 2. Results of HHED mapping.

NObs and NExp: Number of observed and expected homozygous animals, respectively for sliding haplotypes of 20 markers (see Methods). Data supporting the graphs and details of the haplotypes are presented in Supplementary Table 3. The top 20 loci per breed are highlighted (also see Supplementary Table 4), and those validated after analysis of the life histories of 8.8 million females born from at-risk and control mating (see below) are bolded (Benjamini-Hochberg adjusted chi-squared p -value < 0.05 ; Supplementary Table 5). Note that the haplotype names are based on the breed (first character), chromosome (first number), and position (P) in Mb (last number).

To better characterize the effects of the 34 validated haplotypes on female survival, we then calculated the daily difference in proportions between at-risk (1) and control (0) mating for animals that died of natural causes (D), were slaughtered (S), or were still alive (A) over six years (Supplementary Tables 7,8). We then scored the days on which 25, 50, 75, and 100 % of the maximum deviation between the proportion differences were reached (Supplementary Table 9). Using these 12 parameters, we performed a principal component analysis and a hierarchical

clustering (Fig. 3a,b), which distinguished three categories of haplotypes according to age and cause of death: i) early juvenile mortality, ii) late juvenile mortality, and iii) increased mortality and premature culling throughout life (Fig. 3c,d, e,f and g,h, respectively). It is worth noting that although the penetrance of juvenile mortality was incomplete for the latter group of loci, we were still able to detect them with our approach. We would also like to highlight that only one haplotype (H11P78) was in linkage disequilibrium with a deleterious variant previously reported in the literature, namely a transposable element insertion in the gene encoding the apolipoprotein B (*APOB*; $R^2=0.66$ based on 721,006 genotypes; Supplementary Table 10). *APOB*-deficient Holstein calves have been reported to suffer from cholesterol deficiency (CD) and die within the first weeks or months of life (Kipp et al. 2016; Menzi et al. 2016; Mock et al. 2016), which is in perfect agreement with our results and provides a validation of our approach (Fig. 3d).

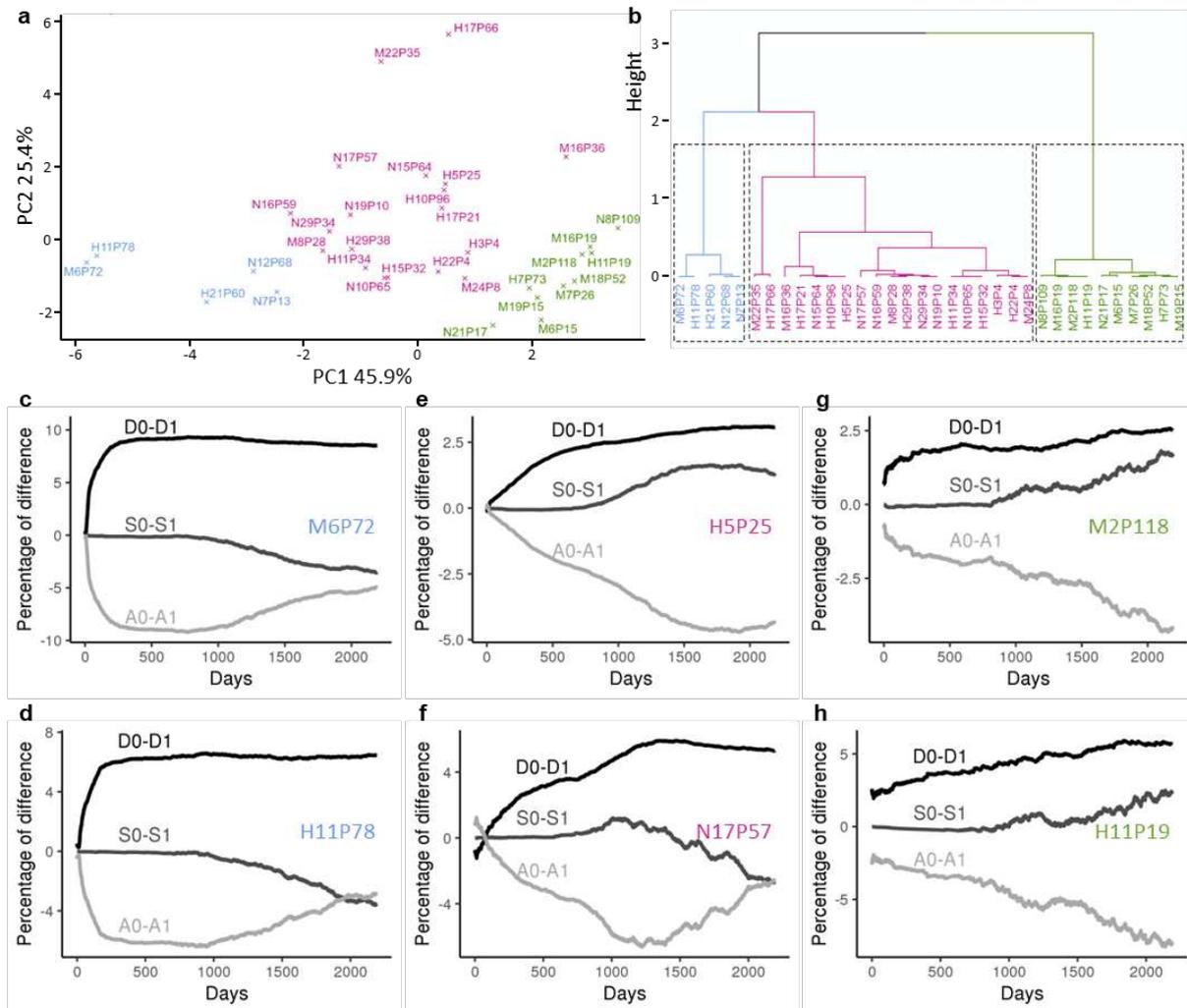


Figure 3. Effects of the 34 validated haplotypes on female survival.

a,b) Results of principal component analysis (PCA; a) and hierarchical clustering (HC; b) for 34 haplotypes x 12 parameters (Supplementary Table 9). PC1 and PC2: Proportion of variance explained by the principal components for the first two PCA dimensions. c-h) Daily difference in proportions between at-risk (1) and control (0) mating for animals that died of natural causes (D), were slaughtered (S), or were still alive (A) over six years (Supplementary Table 8). Results are shown for two haplotypes for each of the three PCA-HC clusters (see color code). Note that they include H11P78, which corresponds to the CDH locus (see below), and three other haplotypes that were subjected to clinical and functional investigations later in the article (H5P25, M6P72, and N17P57).

As a next step, we examined the evolution of haplotype frequencies over time (Fig. 4a; Supplementary Tables 11,12). We observed sudden increases in frequency associated with the overuse of influential ancestors, and then slight and regular decreases, most likely due to the massive dissemination of semen from new elite bulls, diluting the genetic contribution of previous ones to the population. The sire GOLDWYN caught our attention, as he is the main source of seven of the 13 haplotypes validated in Holstein cattle, including H11P78/CD (Fig. 4b). This bull was very popular in the late 2000s because of the outstanding functional qualities of his daughters and because he was less related to the population than other champions at that time. This example reminds us of the importance of a more balanced use of breeding stock, as new lines introduce genetic variability, but also potentially new deleterious recessive variants.

In parallel, we estimated the effects of the 34 haplotypes in the heterozygous state on 14 traits routinely collected for selection purposes (see Methods; Supplementary Table 13; Fig. 4c). Due to the high statistical power provided by the large size of our cohorts ($1,930 \leq n \leq 509,258$), we found significant effects in 30% of the analyses (Benjamini-Hochberg adjusted Student's t-test p-value ≤ 0.01). However, most of these were of small magnitude and we conclude that, overall, the new haplotypes identified in this study do not confer any heterozygous advantage. Interestingly, H11P78 even showed substantial negative effects on several traits (Fig. 4c), echoing previous reports of some CD-clinically affected individuals who were only heterozygous for the *APOB* insertion (Häfliger et al. 2019), and of alterations in lipid metabolism in healthy heterozygotes versus controls (Gross et al. 2016). Therefore, our population-level findings shed new light on CD and support the idea that it is a co-dominantly inherited metabolic disorder.

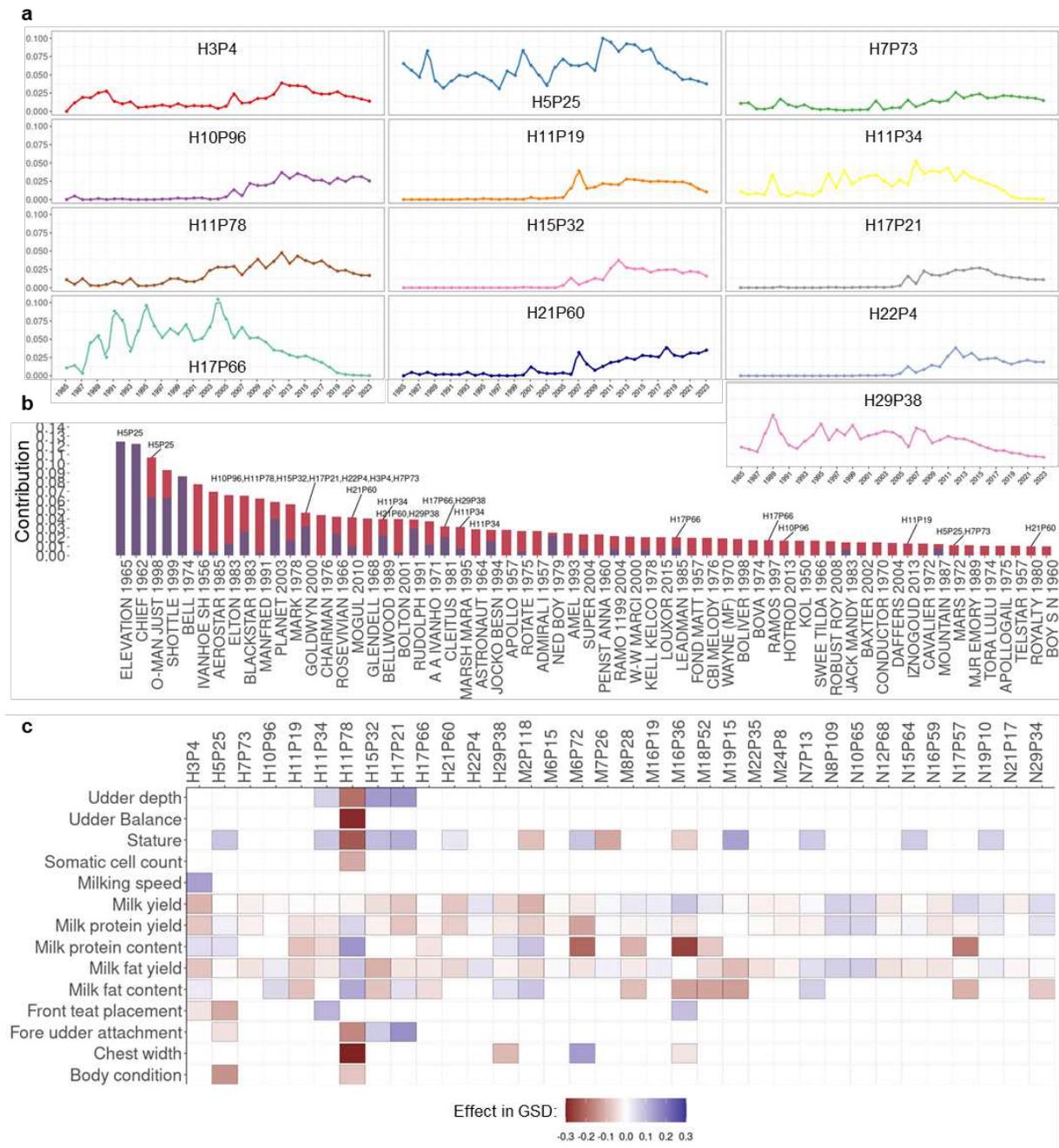


Figure 4. Evolution of haplotype frequencies, pedigree analyses, and estimation of haplotype effects on recorded traits.

a) Evolution of frequency for 13 Holstein haplotypes between 1985 and 2023 (Supplementary Table 11). b) Representation of the marginal (purple) and raw genetic contribution of the most used sire (more than 1% of raw contribution) in the French Holstein female population born within the period 2019-2022 (3,058,756 animals) based on pedigree information (see Methods; Supplementary Table 12). c) Effects of the heterozygosity for each Holstein, Montbeliarde, and Normande haplotype on 14 traits expressed in genetic standard deviation (GSD; Supplementary Table 13). Only significant effects are shown (Benjamini-Hochberg adjusted Student's t-test p-value ≤ 0.01). Note the significant negative effects observed for H11P78/CDH on several traits, including chest width and stature (-0.30 and -0.23 GSD, respectively), which support reduced growth in heterozygous carriers.

To gain insight into the molecular mechanisms associated with the 33 novel genetic determinants of increased juvenile mortality, we analyzed the whole genome sequences of 247 Holstein, 160 Montbeliarde, and 118 Normande cattle, as well as 1,344 controls from more than 70 breeds (Supplementary Tables 14,15). By focusing on breed-specific variants that were predicted to be detrimental to protein function and had a genotype-haplotype square correlation (R^2) of 0.50 or greater, we prioritized 8 candidates (Supplementary Table 16). The identification rate (8/33) may seem somewhat low, and we recognize that the causative variants may affect non-coding regions or segregate in more than one breed, but our goal was to first identify the strongest candidate variants, and then recruit case and control individuals for subsequent phenotypic characterization. As part of reverse genetics studies (e.g., ; Bourneuf et al. 2017), six of these candidate variants had been included in the design of the SNP arrays used for genomic evaluation in France, and genotypes were available for 0.4 to 1.8 million animals from 15 breeds.

After verifying that these large-scale genotyping data yielded results consistent with those of whole-genome sequence analysis (Supplementary Table 18), we selected one variant per breed and performed a series of complementary analyses to provide a proof of concept.

The candidate for Holstein haplotype H5P25 is a point mutation of the $\beta 7$ integrin (ITGB7 p.G375S) affecting a residue conserved among 128 vertebrate orthologs located at the contact interface with the $\alpha 4$ integrin (Fig. 5a; Supplementary Table 19). The $\alpha 4\beta 7$ integrin heterodimer is a major cell adhesion molecule expressed on the surface of CD4 T lymphocytes, that mediates their trafficking from the bloodstream to the digestive tissues (Fig. 5b; Gorfou, Rivera-Nieves, and Ley 2009). This essential process of intestinal immunity could be affected by the ITGB7 substitution, which, based on interaction modeling, is predicted to cause a clash between the two chains and reduce their binding affinity (Fig. 5a). To investigate this hypothesis, we necropsied three homozygous mutant heifers. Although we did not find any specific gross lesions, we observed a complete absence of $\alpha 4\beta 7^{\text{pos}}$ CD45RO^{pos} memory CD4 T cells in the lamina propria of the jejunum (Fig. 5c, Supplementary Table 20). We also examined 13 case-control pairs of heifers on farms. Homozygous mutants showed stunted growth with a 27% reduction in average daily gain and significant alterations in hematological parameters, such as outstanding WBC and lymphocyte counts (Fig. 5c,d; Supplementary Tables 21,22). Taken together, these results support a deficit in CD4 T cell homing and retention in the gut, as well as other serious immune alterations that require further investigation. For these reasons, we have decided to name this new recessive defect "Bovine Lymphocyte Intestinal Retention Defect" (BLIRD).

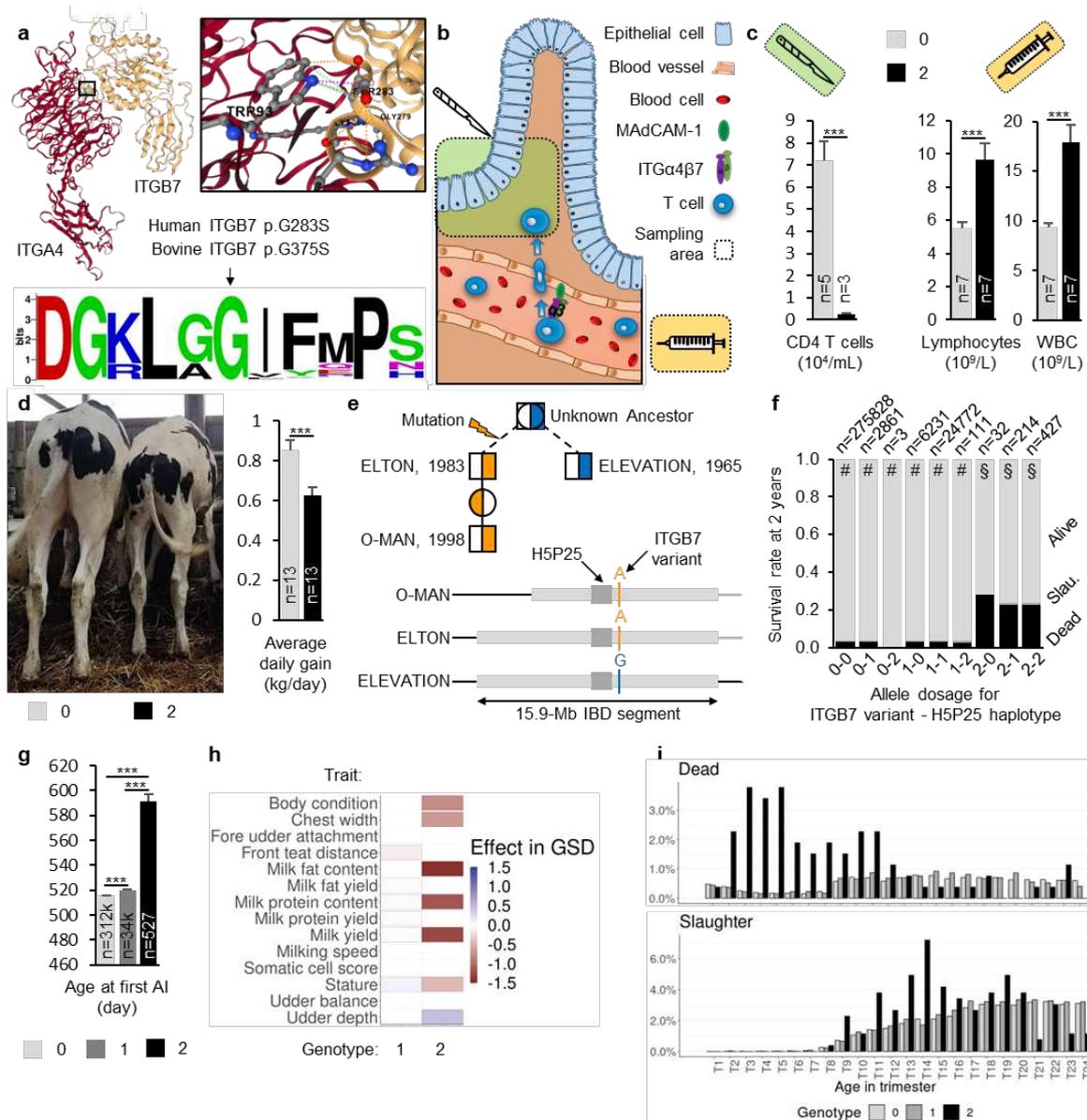


Figure 5. Characterization of the ITGB7 substitution.

a top) Modeling of the 3D structure of the ITGA4 β 7 heterodimer using crystallography in humans, and evaluation of the consequences of glycine to serine substitution at amino acid position 283 in humans (375 in bovine). The purple dashed lines illustrate the clash with a predicted 0.8 kcal/mol reduction in binding affinity (see Methods for further details). a, down) Weblogo representation of multiple protein alignments showing perfect conservation of the mutated glycine residue among 128 vertebrate ITGB7 orthologs (Supplementary Table 19). b) Schematic representation of the ITGA4 β 7-mediated transmigration of T lymphocytes (T cells) from blood vessels to the lamina propria of the jejunum, and details of the sampling area for subsequent analyses. c) Dosage of α 4 β 7pos CD45ROpos memory CD4 T cells in the lamina propria of the jejunum and of immune cells in the circulating blood of homozygous wild-type (0) and mutant (2) heifers (Supplementary Tables 20,21). ***: Student t test p-value ≤ 0.001 . d) Pictures of 247-day-old wild-type and 245-day-old homozygous mutant heifers, illustrating growth retardation in the latter; and analysis of average daily gain in 13 cases and 13 matched control heifers aged six months to two years (Supplementary Table 22). e) Origin of H5P25 haplotype and ITGB7 mutation (see Supplementary Table 23 for details). f) Proportion of animals that died, were slaughtered, or are still alive at 2 years of age for each combination of ITGB7 variant genotype and H5P25 status. Different symbols indicate significant differences (Fisher p-value ≤ 0.01) between genotype-haplotype combinations in the proportions of animals within each of the three categories (Supplementary Table 24). g) Analysis of age at first AI for the three genotypes at the ITGB7 variant (Supplementary Table 25). ***: Student t test p-value ≤ 0.001 . Effects of heterozygosity and homozygosity at the ITGB7 variant on 14 recorded traits (Supplementary Table 26). Only significant effects are shown (Benjamini-Hochberg adjusted Student's t-test p-value ≤ 0.01). i) Trimester repartition of the proportion of

females that died or were slaughtered before reaching six years of age by genotype group at the *ITGB7* substitution (Supplementary Table 27).

Based on pedigree and genotype information, we determined that ELEVATION, the most influential ancestor of the Holstein breed, and ELTON, the grandsire of the third most influential ancestor, O-MAN, shared an identical-by-descent segment of 15.9 Mb encompassing the H5P25 haplotype and that the *ITGB7* mutation occurred in an ancestor of ELTON (Fig. 5e; Supplementary Table 23). Taking advantage of this incomplete LD, we examined the two-year survival of the different haplotype x genotype combinations (see Methods) and observed an excess of mortality only in homozygous mutants, regardless of their haplotype status, further confirming the causality of the *ITGB7* substitution (Fig. 5f; Supplementary Table 24). Finally, population-level analyses revealed delayed age at first insemination (Fig. 5g; Supplementary Table 25), consistent with reduced growth (e.g., Jourdain et al. 2023), adverse effects on production traits (Fig. 5h; Supplementary Table 26), and premature culling in homozygous mutants compared to other genotypes (Fig. 5i; Supplementary Table 27). As a recessive defect that (i) causes juvenile mortality with incomplete penetrance, (ii) is due to a mutation in incomplete LD with surrounding markers, and (iii) has not been previously detected despite segregating for more than 40 years in Holstein cattle, BLIRD provides a textbook example to validate our approach.

We then studied two variants affecting proteins for which no living homozygous mutants have been described so far in animals: an inframe deletion of the Replication Factor C Subunit 5 affecting a residue conserved in 633 eukaryotic orthologs (RFC5 p.E369del; $R^2=1$ with haplotype N17P57; Fig. 6a; Supplementary Tables 16;28), and a frameshift insertion in the Nitric Oxide-Associated Protein 1 (NOA1 p.D400RfsX9; $R^2=0.55$ with M6P72; Fig. 6i; Supplementary Table 16).

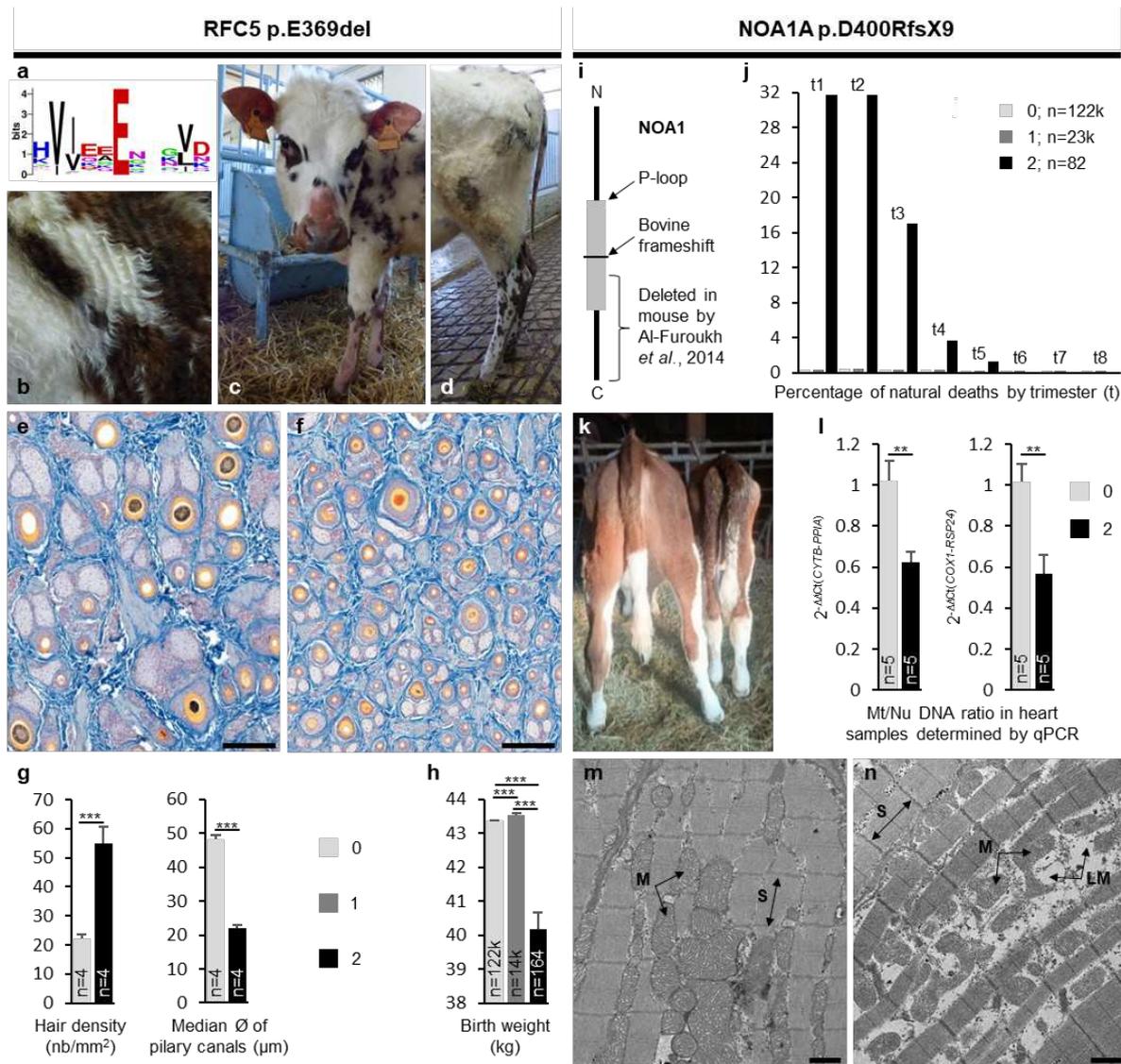


Figure 6. Pathophysiological characterization of RFC5 and NOA1 mutants in Normande and Montbeliarde cattle, respectively.

a) Weblogo representation of multiple protein alignments showing perfect conservation of the glutamic acid residue at position 369 amino acids in bovine among 633 eukaryotic RFC5 orthologs (Supplementary Table 28). b) Detail of the shoulder and base of the neck of a homozygous mutant heifer showing a thin and wavy hair coat. c,d) Images of the same animal showing alopecia of the ears, the top of the snout, extremities of the limbs, and tail. e,f) Skin sections of shoulder skin samples from wild-type (e) and RFC5 homozygous mutant (f) 1.5-year-old heifers stained with a Roan solution. Scale bars = 200 μ M. g) Analysis of shoulder skin sections from 4 cases and 4 control heifers for hair density in a randomly selected square of 1 mm², and median size of 50 adjacent pilary canals (Supplementary Table 29). h) Analysis of birth weight for the three genotypes at the RFC5 deletion (Supplementary Table 30). i) Schematic representation of the bovine NOA1 protein (Ensembl ID: ENSBTAP00000025792, 731 aa). If expressed, the p.D400RfsX9 frameshifted bovine protein would lack half of the P-loop containing nucleoside triphosphate hydrolase domain (P-loop) and a conserved C-terminal region, the deletion of which causes impaired mitochondrial import of NOA1 in cultured mouse myoblasts (Al-Furoukh et al. 2014). j) Percentage of natural deaths by trimester for 145 thousand females genotyped for the NOA1 frameshift variant. Note that only the first two years of a six-year study are shown (Supplementary Table 35). k) Wild-type (left) and NOA1 homozygous mutant (right) 4.5-month-old heifers. l) Relative changes in mitochondrial-to-nuclear (Mt/Nu) DNA ratio in heart samples determined by qPCR (Supplementary Tables 37,38). m,n) Transmission electron microscopy images of myocardial samples from wild-type (m) and NOA1 homozygous mutant (n) four-month-old heifers. Scale bars = 1 μ M. S: Sarcomere, the basic unit of myofibrils. M: Mitochondria. LM: Lysed mitochondria. Note the presence of empty spaces between myofibrils due to increased mitochondrial cell death in (n). ** and ***: Student t test p-value ≤ 0.01 and ≤ 0.001 , respectively. 0, 1 and 2: wild-type, heterozygous and homozygous genotypes for the RFC5 inframe deletion or NOA1 frameshift insertion, respectively.

RFC5 is one of the five subunits of the replication factor C complex, which is essential for DNA replication and cell proliferation in eukaryotes (Cullmann et al. 1995; Furukawa et al. 2003; Li et al. 2018). Clinical examination of six homozygous mutant heifers revealed stunted growth, chronic diarrhea with no gross lesions at necropsy, abnormally thin and wavy hair coat, and alopecia of the body extremities (Fig. 6b-d). Breeders reported that hairless patches appeared after the juvenile molt and peaked in size each winter, suggesting that low temperatures impair hair growth in homozygotes. This observation is consistent with the temperature sensitivity reported for two RFC subunit mutants in yeast, which is associated with blockage of DNA replication (N. Reynolds, Fantès, and MacNeill 1999). Histological analysis of shoulder skin samples showed increased hair density and decreased hair diameter in case versus control heifers, demonstrating that the homozygosity at the *RFC5* inframe deletion also affects hair follicle differentiation during embryogenesis (Fig. 6e-g, Supplementary Table 29). Finally, phenotypic characterization at the population level revealed reduced birth weight in homozygous mutants supporting *in-utero* growth retardation, and elevated rates of late juvenile mortality and premature culling (Fig. 6h; Supplementary Tables 30,31; Supplementary Fig. 1).

NOA1 is a nuclear-encoded protein required for mitochondrial protein translation and respiration, and the p.D400RfsX9 mutation is predicted to alter its function and mitochondrial import (Kolanczyk et al. 2011; Al-Furoukh et al. 2014; Fig. 6i). While *NOA1* knockout in mice causes midgestation lethality (Kolanczyk et al. 2011), we estimated that only 24.4% of the homozygous mutants died during pregnancy (Supplementary Table 32). An additional 50.7% died before reaching the age of genotyping (Supplementary Table 33), and the 24.9% that were genotyped died in most cases before one year of age (Supplementary Tables 34,35; Fig. 6j). Thus, the eight two- to four-month-old genotyped homozygotes (one male, seven females) that we followed clinically were among the longest-lived. The females developed ill-thrift between three and 12 months of age and were euthanatized for ethical reasons. Hematological and immune analyses revealed neutrophilia, indicating the presence of inflammation but no anomaly of reactive oxygen species (ROS) production by neutrophils (Supplementary Table 36, Supplementary Note). In addition, *NOA1* mutants showed abnormal blood biochemical parameters suggesting a metabolic disorder and extensive mitochondrial apoptosis, as revealed by electron microscopy and relative quantification of mitochondrial and nuclear DNA in myocardial samples (Supplementary Tables 36-38; Fig. 6l-n). Finally, the only homozygous male never showed clinical signs until the end of its follow-up at one year. In contrast to mice, the distribution of deaths over a long period and access to large populations whose genetic background and rearing conditions are not standardized offer interesting prospects for the future

identification of genetic or environmental factors that might compensate for NOA1 loss-of-function in cattle.

In conclusion, using a data science-based approach, we have identified numerous recessive loci responsible for increased risk of juvenile mortality in cattle that had previously been overlooked. The data-mining framework described in this paper is readily applicable to other physiological stages and any population that benefits from large datasets generated by genomic evaluation. The management of these new genetic defects will have a direct impact on animal breeding, helping to reduce animal suffering and economic loss to the industry. Finally, our approach also offers exciting prospects for basic research, by identifying large animal models for immune or metabolic disorders, some of which involve understudied genes, that can be characterized at a population level.

Methods

Animals and data sets

A large number of animal populations and data sets were considered in this study. These are detailed below for each analysis. Briefly, the most important of these are: (i) pedigree and life history information on millions of female cattle (date of birth; lifespan; cause of death; dates of insemination and bull IDs; dates of calving and onset of lactation) extracted from the French national bovine database; (ii) phased and imputed Illumina BovineSNP50 array genotypes generated as part of the routine bovine genomic evaluation; (iii) performance for various traits corrected for non-genetic factors as estimated in the national genomic evaluation; (iv) whole genome sequences of 1869 cattle; (v) SNP array genotypes for candidate variants in large populations; and (vi) animals recruited from commercial farms for physiopathological characterization and functional analyses.

Mapping of recessive loci, validation, and analysis of survival curves

Pedigree and life history information was extracted from the French national bovine database for 5.96 million Holstein, 1.63 million Montbéliarde, and 1.24 million Normande females whose sires and maternal grandsires (MGS) were genotyped. The discovery population was restricted to genotyped females and included 8,203, 6,198 and 2,254 “dead heifers” (females that died of natural causes before 3 years of age and were never inseminated) and 291,529, 141,343 and

56,095 cows (females that calved and started a first lactation) from the three breeds, respectively (Supplementary Table 1). These animals and their ancestors were genotyped using various Illumina SNP arrays over time (LD, ~7 K SNPs; custom LD, ~10 K to 20 K; BovineSNP50, ~50 K; EuroGMD, ~63 K; and HD, ~777K). Raw genotypes were imputed and phased for 44,596 autosomal SNPs by FImpute (Sargolzaei et al. 2014) as part of the French routine genomic evaluation of cattle, as described by Mesbah-Uddin et al. 2019; Supplementary Table 2). We considered sliding haplotypes of 20 markers, and counted the number of homozygotes observed (NObs) within each group of genotyped individuals. In parallel, we estimated the expected number of homozygotes (NExp) using within-family transmission probability. We filtered haplotypes satisfying the following criteria: Nobs \geq 10 in cases, increase in homozygotes (i.e. (Nobs-Nexp)/Nexp) \geq 25% in cases and \leq -25% in controls (Supplementary Table 3). Among stretches of consecutive haplotypes, we selected the one showing the highest increase in homozygotes in cases as the “peak haplotype” (Supplementary Table 4). For validation, we compared the proportions of animals belonging to three categories (“dead heifers”, “cows”, and “others”) using a chi-squared test with Benjamini-Hochberg correction (p-value \leq 0.05) among the descendants of at-risk mating (“1”; carrier sire and carrier MGS) or control mating (“0”; noncarrier sire and MGS). Note that the “other” category includes females that do not meet the criteria retained for the “dead heifers” and “cows” groups (e.g. heifers that were slaughtered, heifers that died after 3 years of age, etc.). This analysis was performed on a validation population of 5.96 million Holstein, 1.63 million Montbeliarde and 1.24 million Normande individuals born between 2000 and 2015, most of whom were not genotyped themselves (Supplementary Table 5). Then, for 34 validated haplotypes, we calculated the daily proportion of animals that died of a natural causes (D), were slaughtered (S) or were still alive (A) over a period of 6 years for mating types 0 and 1 (Supplementary Table 7). We also calculated the D0-D1, S0-S1 and A0-A1 differences in proportion on a daily basis and scored the days on which 25, 50, 75 and 100 % of the maximum deviation between each proportion difference was reached (Supplementary Table 8). We then used these 12 parameters to perform a principal component analysis and a hierarchical clustering using the RStudio package Factoshiny v.1.2.5033 (Supplementary Table 9).

Estimation of the number of calves born homozygous for one or more validated haplotypes in the last ten years

We considered 55.5 million inseminations performed during the period 2013-2022, of which 2.0 million involved females and males that were both genotyped. To estimate the number of homozygous calves (NH), we took into account the total number of inseminations (NAI), the proportion of at-risk mating observed within genotyped couples (PR), the Mendelian probability (0.25), the average conception rate in each breed (CR), and finally the proportion of females bred by AI (%AI): $NH = NAI * PR * CR * 0.25 / \%AI$ (Supplementary Table 6; Escouflaire and Capitan 2021).

Evolution of haplotype frequencies over time

The frequency of the 34 validated haplotypes was calculated on an annual basis considering 1,185,446 Holstein, 591,294 Montbeliarde and 180,722 Normande individuals of any sex available in the French bovine national genomic evaluation database born between 1985 and 2023 (Supplementary Table 11).

Genetic contribution of the ancestors of the actual female populations

To identify the main ancestors of the current female populations and calculate their raw and marginal genetic contributions, we analyzed the pedigrees of 3,058,756 Holstein, 738,333 Montbeliarde and 333,793 Normande females born within the period 2019-2022 with at least sire and dam information available (Supplementary Table 12) using the PEDIG software (Boichard 2002).

Effects of haplotypes on recorded traits

The effects of the 34 haplotypes in the heterozygous state were estimated on 14 morphological and production traits routinely recorded for selection purposes. To remove the various environmental factors affecting these phenotypes, we used yield deviation data, i.e. records adjusted for the non-genetic effects included in the genomic evaluation models. Yield deviations are a by-product of the official genomic evaluations carried out by GenEval on behalf of the French breeding organizations (for details on the models used, see <https://www.geneval.fr/english>). Yield deviations were analyzed with a mixed model including the fixed effect of the haplotype studied (0 versus 1 copy), the fixed effect of the year of recording, and the individual random polygenic effect. Calculations were performed with BLUPF90 software (Misztal et al. 2014). The sample size ranged from 1930 to 509,258 cows per group

(Supplementary Table 13). Student's t-test was used to compare the means between groups for each trait and adjusted using the Benjamini-Hochberg method. To allow comparisons between traits, effects were converted to genetic standard deviations (GSD) based on the genetic parameters estimated for the national genomic evaluations.

Estimating effects in homozygotes did not seem relevant to us because at this stage we cannot know whether the causal variants are in complete linkage disequilibrium with the causative variant. These effects may be strongly biased if, among animals homozygous for the haplotype with performance records, there is a large increase in the proportion of animals that are not homozygous for the causative variant as a result of natural counterselection of double homozygotes (for both the haplotype and the causative variant).

DNA extraction

Genomic DNA was extracted from semen straws, blood or myocardium samples using the DNeasy Blood and Tissue Kit (Qiagen). DNA quality was controlled by electrophoresis and quantified using a Nanodrop spectrophotometer.

Whole genome sequencing and variant calling

We analyzed the whole genome sequences of 1,869 cattle from more than 70 different breeds generated by Illumina technology (Supplementary Table 14). Of these, 1,137 were obtained from public databases and 732, consisting mainly of influential AI bulls of French breeds, were sequenced at the GeT-PlaGe facility during the last decade (<http://genomique.genotoul.fr/>). Paired-end libraries with insert sizes ranging from 300 to 500 bp were generated using the NEXTflex PCR-Free DNA Sequencing Kit (Biooscientific) or the TruSeq Nano DNA Library Prep Kit (Illumina). Libraries were quantified using the KAPA Library Quantification Kit (Cliniscience), checked using a High Sensitivity DNA Chip (Agilent) or the Fragment Analyzer (Agilent), and sequenced on Illumina machines (HiSeq2500, HiSeq3000, HiSeq10X, or NovaSeq 6000) with 2 × 100-bp or 2 × 150-bp read lengths, according to the manufacturer's protocol. Reads were aligned to the bovine ARS-UCD1.2 reference genome sequence using the Burrows–Wheeler aligner (BWA-v0.6.1-r104; H. Li and Durbin 2009). SNPs and small InDels were identified using the GATK-HaplotypeCaller software (McKenna et al. 2010) as previously described (Daetwyler et al. 2014; Boussaha et al. 2016), while structural variations (SVs) were detected using the Pindel (Ye et al. 2009), Delly (Rausch et al. 2012), and Lumpy (Layer et al. 2014) software.

Variant filtering and annotation

R²-correlations between haplotype status and genotypes for variants located in a 20 Mb region centered on each of the 33 haplotypes were calculated for 247 Holstein, 160 Montbeliarde, and 118 Normande WGS. The number of haplotype carrier alleles ranged from one to 22 in this data set (Supplementary Table 15). Variants with an R² score ≥ 0.5 were selected, and those segregating in more than one breed represented in the full set of WGS were filtered out, assuming that the causative mutations occurred after the creation of the breeds. The remaining variants (consisting only of SNPs and small InDels) were annotated using Variant Effect Predictor (Ensembl release 110; <https://www.ensembl.org/info/docs/tools/vep/index.html>), and only those predicted to be deleterious were considered (i.e., missense variants with a SIFT score ≤ 0.05 , stop gain or loss, premature start, start loss, frameshift, inframe insertion or deletion, and variants affecting splice donor or acceptor sites). Finally, the consistency of these annotations was verified using independent resources: some variants predicted by VEP to be deleterious and located in non-coding regions (i.e., with no evidence of bovine expressed sequence tag or orthologous transcripts in mouse or human) according to the USC genome browser (<http://genome.ucsc.edu/>) were further removed. Information on the pathophysiological features associated with mutations affecting orthologous genes in humans and mice was extracted from the Online Mendelian Inheritance in Man (OMIM; <https://www.omim.org>) and Mouse Genome Informatics (MGI; <https://www.informatics.jax.org>) databases (Supplementary Table 16). The putative amino acid sequence of the mutant NOA1 protein was obtained using the ExPASy translate tool (<https://web.expasy.org/translate/>) after insertion of the frameshift allele into the cDNA of the Ensembl transcript ENSBTAT00000025792. Information on "domains & features" was obtained from the eponymous Ensembl "transcript-based display". To analyze amino acid conservation at the ITGB7 and RFC5 mutation sites, "1-to-1" orthologous proteins from numerous animal species were extracted from the "orthologues" "gene-based display" available for Ensembl entries ENSBTAG00000018993 and ENSBTAG00000007137 and aligned using ClustalW software version 2.1 (Thompson, Higgins, and Gibson 1994; <https://www.genome.jp/tools-bin/clustalw>). Finally, a sequence logo was generated using WebLogo (Crooks et al. 2004; <http://weblogo.berkeley.edu>). Note that RFC5 in animals is orthologous to RFC3 in fungi and plants, as supported by orthology comparisons with the yeast (*S. cerevisiae*) protein entry YNL290W that are available in Ensembl release 110 (www.ensembl.org/), Ensembl Fungi release 57 (<http://fungi.ensembl.org/>), and Ensembl Plants release 57 (<http://plants.ensembl.org/index.html>). "1-to-1" orthologs of yeast RFC3 were

extracted from the latter two databases and added to the animal orthologs of the bovine RFC5 protein for analysis. Information on protein IDs, species, and amino acid sequence around the mutation sites is provided in Supplementary Tables 19 and 28.

Analysis of the 2/3D structure of the mutant ITGB7 protein

The effect of the bovine p.G375S point mutation on the ITGB7 protein structure was evaluated using the mCSM-PPI2 server (https://biosig.lab.uq.edu.au/mcsm_ppi2/; Rodrigues et al. 2019). For this purpose, the crystal structure of the human ITGA4/ITGB7 dimer interacting with the Act-1 mAb was used as a reference structure model (accession number 3V4P in the Worldwide Protein Data Bank; www.wwpdb.org). The orthologous glycine residue is located at position 283 in the human ITGB7 protein used for modeling.

Large-scale genotyping of candidate variants and analysis of LD with haplotypes

The Illumina SNP arrays used for genomic evaluation in France have a custom section to which probes for genotyping thousands of deleterious variants have been added over time through various forward and reverse genetic projects. Genotypes were available for the APOB insertion responsible for recessive CD in Holstein cattle and for six candidate variants for new loci associated with increased juvenile mortality discovered in this study. Information on the probes and the number of genotypes available in 15 breeds can be found in Supplementary Tables 10, 16, 17 and 18. These data were used to calculate allele frequencies, to verify the breed specificity of the markers, to compute contingency tables between haplotype status and genotypes, and to calculate R^2 square correlations. Finally, they were also used to investigate the effect of the ITGB7, RFC5, and NOA1 candidate variants on various traits (see related sections).

Cause of incomplete LD between some haplotypes and their candidate variants

To determine whether the incomplete LD was due to a relatively recent de novo mutation or to an ancient founder effect followed by recombination events between the haplotype and the variant, we sorted carriers of (i) both the haplotype and the variant, (ii) the haplotype but not the variant, and (iii) the variant but not the haplotype. We searched for IBD segments between these animals and their ancestors using both genotype and pedigree information. Illumina BovineSNP50 array

genotypes were available for most of the AI bulls used in France since 1985. This analysis was performed for H5P25/*ITGB7* (Supplementary Table 23) and M6P72/*NOA1* (Supplementary Fig. 2)

Imputation of *ITGB7* and *RFC5* variants using long size haplotypes

As the *ITGB7* and *RFC5* candidate variants were included in the arrays used for genomic evaluations in France in early 2019, we have developed an approach to impute them in animals genotyped before that date, in order to allow survival studies over 6 years and to increase cohort sizes for studying different traits. For the *ITGB7* substitution, we considered a 34 Mb segment (476 markers from markers ARS-BFGL-NGS-69702 to ARS-BFGL-NGS-7850) centered on the 15.9 IBD segment shared by the bulls ELEVATION and ELTON, which were (i) heterozygous carrier of the H5P25 haplotype but wild type for the variant for the former and (ii) double heterozygous for the latter (Figure 5; Supplementary Table 23). For the *RFC5* inframe deletion, which was in complete LD with N17P57, we arbitrarily considered a 5 Mb segment centered on the inframe deletion (105 markers from Hapmap34428-BES2_Contig387_701 to BTB-00682446). See Supplementary Table 2 for information on the marker map. Using 272,326 Holstein cattle genotyped for the *ITGB7* variant and 53,263 Normande cattle genotyped for the *RFC5* variant, we created a bank of long-size haplotypes associated with either the ancestral or the mutant allele. If a long-size haplotype was not associated with only one allele, it was considered dubious and eliminated. We then proceeded to genotype imputation for 789,594 and 127,783 animals, respectively. Animals with one or two long-size haplotypes unknown in the haplotype bank were not considered. This procedure has been designed to reduce imputation errors to a level close to zero. Finally, we generated a database of 585,671 *ITGB7* genotypes in Holstein and 164,291 *RFC5* genotypes in Normande cattle with 1.15 and 2.08 ratios of imputed/real genotypes, respectively. Imputation was not required for the *NOA1* frameshift insertion as it has been genotyped since 2013.

Analysis of survival based on genotype for candidate

To verify the causality of the *ITGB7* and *NOA1* variants, a first analysis of the proportions of animals still alive at 2 years of age or that died or were slaughtered before was performed for each combination of variant genotype x haplotype status (Supplementary Tables 24,34). This analysis was not performed for the *RFC5* inframe deletion because of its complete LD with the N17P57 haplotype. In addition, for the three genotypes of each variant, we calculated the proportions of

animals still alive at 6 years of age or that died or were slaughtered before, and their repartition per trimester (Supplementary Tables 27,31,35).

Effects of the *ITGB7*, *RFC5* and *NOA1* on various traits

The effects of heterozygosity or homozygosity at the *ITGB7* substitution were estimated for 14 traits routinely recorded for evaluation purposes using the same model as for the haplotypes (see above; Supplementary Table 26), for age at first insemination and age at first calving (Supplementary Table 25), and for birth weight (Supplementary Table 30). Only the latter phenotype was analyzed for *RFC5* and *NOA1* due to insufficient numbers of homozygous mutant females with insemination and production record available.

Estimation of the proportion of *NOA1* homozygous mutants that died during embryonic development and between birth and genotyping

The conception rate (CR; i.e. the proportion of successful AI) was calculated for mating between males and females heterozygous for the *NOA1* frameshift variant (1*1) and for mating between wild type parents (0*0) at two developmental stages (heifers and primiparous cows; Supplementary Table 32). The CR was expressed as a proportion of the CR observed in control mating for each female category (Centered_CR). The proportion of embryonic lethality among homozygous mutant conceptuses was calculated as $(\text{Centered_CR}_{0*0} - \text{Centered_CR}_{1*1})/0.25$, assuming that 25% of the conceptuses from at-risk mating are expected to be homozygous mutants, and full penetrance of embryonic lethality in the latter group. In parallel, we counted the proportion of each genotype observed in the offspring of heterozygous parental pairs and estimated the proportion of homozygous mutants genotyped out of the proportion expected according to Mendelian rules. These two values were used to derive the proportion of homozygous mutants that were born but died before reaching the age of genotyping (see Supplementary Table 33 for more information).

Pathophysiological examination

ITGB7, *RFC5*, and *NOA1* homozygous mutant cattle aged two months to three years old and their matched controls were examined on farms. Blood samples were collected for serum and cell content analysis. Some of them were hospitalized for clinical follow-up at the bovine clinic of one

of the four French National Veterinary Schools (Ecoles Nationales Vétérinaires de France, ENVF). During this time, routine blood analyses (hematology and biochemistry) and specific dosages (beta-hydroxybutyric acid, non-esterified fatty acids, lactate, etc.; see Supplementary Table 36) were performed at the central clinical pathology laboratory of each school. Euthanasia and necropsy were performed to identify gross lesions and to obtain tissue samples for further analysis to confirm hypothesized mechanisms of pathophysiology. Details on the number and age of animals considered are provided for each analysis in the figures in the Results section and in the Supplementary Tables.

Hemogram

Blood was drawn from the jugular vein into 4-mL K 3-EDTA tubes (Venosafe, Terumo, France) and gently mixed by inversion. Tubes were immediately refrigerated at ~4°C until hematologic analysis at the central laboratory of each school. Air-dried blood smears were prepared and stained with a May-Grünwald/Giemsa automated stainer (Aerospray hematology slide stainer cytocentrifuge 7150, Wescor, USA) for microscopic evaluation. Measurements were performed on a Sysmex XT-2000iV analyzer as recommended by the manufacturer, using settings for bovine blood (Sysmex XT-2000iV software v.00-13) (Herman et al. 2018). Analyzer-measured variables included red blood cell (RBC) count by impedance (RBC-I) and optical (RBC-O) measurements, hemoglobin concentration (HGB), and white blood cell (WBC) count. Neutrophil, lymphocyte, monocyte, eosinophil, and basophil counts were determined from 100 leukocytes counted per oil immersion field (1,000×). The percentages of each cell type were determined and the corresponding cell counts were calculated from the WBC.

Metabolite analysis

Anticoagulated tubes were centrifuged within half an hour to prevent further exchange of analytes between blood cells and plasma. One mL of EDTA blood was mixed with 6% (w/v) perchloric acid for the determination of lactate. Biochemical analyses were performed on a Konelab 30 (Thermo Fisher Scientific inc., USA) using reagents from the same company, except for the determination of plasma lactate (Diasys Diagnostic Systems, Germany) or GLDH (Roche Diagnostic, Switzerland). Plasma troponin I concentration was determined by Immulite 2000 (Siemens, Germany). The HClO₄ extracts were neutralized with 20% (w/v) KOH before analysis according to

Baird and Heitzman 1970. Biological parameters were compared with reference samples prepared from healthy animals according to ASVCP recommendations (Friedrichs et al. 2012).

Quantification of CD4⁺ mem a⁺ b7⁺ T cells in the jejunal lamina propria using flow cytometry

One gram of small intestinal wall in the jejunal loop from heifers (n=3 homozygous mutant and 5 wild type) aged 1.5 to 3 years was washed in cold PBS, cut into 0.5 cm pieces, incubated four times in 30 mL of PBS 3 mM EDTA (Sigma-Aldrich), and digested in 20 mL of DMEM added with 20% FCS and 100 U/ml of collagenase (Sigma-Aldrich) for 40 min at 37 °C. Jejunum LP mononuclear cells were isolated on a 40–80% Percoll gradient after centrifugation at 1800 g for 15 min at room temperature. Then 1-2 x 10⁶ Mononuclear cells were resuspended in HBSS, 0.5% BSA, and 10 mM Hepes. Cell viability was assessed using Viability 405/520 Fixable Dye (Miltenyi Biotec, Germany). Incubation with the antibodies was performed at 4°C for 30 min in the dark. The antibodies used were: CD4-PB (clone CC8, Biorad, Hercules, U.S), CD45-FITC (clone 1.11.32, Biorad), CD45R0-A647 (clone IL-A116, Biorad), Beta7-RPE (clone FIB27, Biolegend, San Diego, U.S.), Alpha4-PECy7 (clone 9F10, Biolegend). Data were collected on a MACSQuant[®] Analyzer (Miltenyi Biotec) and analyzed using Flowlogic software (Miltenyi Biotec).

Light microscopy

Shoulder skin samples from four heifers homozygous for the *RFC5* inframe deletion and four matched controls, were fixed in 10% neutral buffered formalin for 24 h at +4°C, and then dehydrated in a graded ethanol series (30% to 100%), cleared in xylene, and embedded in paraffin wax. Longitudinal microtome sections (5 µm, Leica RM2245) were mounted on adhesive slides (Klinipath-KP-PRINTER ADHESIVES), deparaffinized, and stained with a Roan solution (nuclear fast red, orange G, and aniline blue). Slides were digitized with the Panoramic Scan 150 and analyzed with CaseViewer 2.4 software (3D Histech). The number of hair follicles in a randomly selected 1 mm² square and the diameter of the pilary canal of 50 adjacent hair follicles were measured for each animal (Supplementary Table 29).

Transmission electron microscopy

Left ventricular heart samples from three heifers homozygous for the *NOA1* frameshift variant and three matched controls were fixed with 2% glutaraldehyde in 0.1 M Na cacodylate buffer pH 7.2, for 4 hours at room temperature. The specimens were then contrasted with Oolong Tea Extract (OTE) 0.2% in cacodylate buffer, postfixed with 1% osmium tetroxide containing 1.5% potassium cyanoferrate, dehydrated in a graded ethanol series (30% to 100%), and embedded in Epon, after the ethanol was gradually replaced by ethanol-Epon mixtures. Thin sections (70 nm) were collected on 200 mesh copper grids, and counterstained with lead citrate. The grids were

examined with a Hitachi HT7700 electron microscope operated at 80kV (Milexia, France), and images were acquired with a charge-coupled device camera (AMT).

Analysis of mitochondrial/nuclear DNA ratio by quantitative PCR

After preliminary testing focused on specificity and efficiency of amplification, two sets of mitochondrial and nuclear genes were selected with primer pairs showing similar slope (*CYTB* and *PPIA*, and *COX1* and *RSP24*, respectively; Supplementary Tables 37,38). Quantitative PCR was performed in triplicate on a QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific) for five homozygous carriers of the *NOA1* frameshift insertion and five matched controls. The reaction mixture contained 10 μ L 2X SYBR Green PCR Master Mix (Thermo Fisher Scientific), 1 ng total DNA extracted from myocardial samples, and 300 nM forward and reverse primers, in a total volume of 20 μ L. For each animal the Δ Ct(*CYTB-PPIA*) and Δ Ct(*COX1-RSP24*) values were calculated based on the mean cycle threshold (Ct) of the triplicates. Finally the $2^{-\Delta\Delta\text{Ct}(\text{CYTB-PPIA})}$ and $2^{-\Delta\Delta\text{Ct}(\text{COX1-RSP24})}$ were used as two different indicators to measure the relative changes in the ratio of mitochondrial to nuclear DNA between the case and control groups.

Ethics statement

All experiments reported in this work comply with the ethical guidelines of the French National Research Institute for Agriculture, Food and Environment (INRAE) and its French research partners. Blood samples from affected and unaffected animals were collected by licensed veterinarians. No animal was procreated purposely for this study. The owners of the animals had consented to the diagnostic examination for research purposes and to the inclusion of their animals in this study. Invasive sampling was performed post-mortem only. As a consequence, no ethical approval was required for this study. Finally, all the data analyzed in the present study were obtained with the permission of breeders, breeding organizations and research group providers.

Data Availability

Raw sequencing data from 732 cattle reported in this study have been deposited in the European Nucleotide Archive (ENA, <https://www.ebi.ac.uk/ena>) under accession numbers PRJEB64022 and PRJEB64023. Sequences from 1137 cattle from previous studies are available in the NCBI

BioProject and ENA databases under the accession numbers listed in Supplementary Table 14. Source data for the other analyses are provided in the Supplementary Table file with this article.

References

- Adams, Heather A., Tad S. Sonstegard, Paul M. VanRaden, Daniel J. Null, Curt P. Van Tassell, Denis M. Larkin, and Harris A. Lewin. 2016. "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* 99 (8): 6693–6701. <https://doi.org/10.3168/jds.2015-10517>.
- Agreste. 2023. "Graph'Agri 2023." 2023. <https://agreste.agriculture.gouv.fr/agreste-web/disaron/GraFra2023Integral/detail/>.
- Aguilar, I., I. Misztal, D.L. Johnson, A. Legarra, S. Tsuruta, and T.J. Lawlor. 2010. "Hot Topic: A Unified Approach to Utilize Phenotypic, Full Pedigree, and Genomic Information for Genetic Evaluation of Holstein Final Score." *Journal of Dairy Science* 93 (2): 743–52. <https://doi.org/10.3168/jds.2009-2730>.
- Al-Furoukh, Natalie, Julia R. Kardon, Marcus Krüger, Marten Szibor, Tania A. Baker, and Thomas Braun. 2014. "NOA1, a Novel ClpXP Substrate, Takes an Unexpected Nuclear Detour Prior to Mitochondrial Import." Edited by Vimal Selvaraj. *PLoS ONE* 9 (7): e103141. <https://doi.org/10.1371/journal.pone.0103141>.
- Al-Khudhair, A., P. M. VanRaden, D. J. Null, M. Neupane, M. C. McClure, and C. D. Dechow. 2024. "New Mutation within a Common Haplotype Is Associated with Calf Muscle Weakness in Holsteins." *Journal of Dairy Science*, January. <https://doi.org/10.3168/jds.2023-24121>.
- Amberger, Joanna S, Carol A Bocchini, Alan F Scott, and Ada Hamosh. 2019. "OMIM.Org: Leveraging Knowledge across Phenotype–Gene Relationships." *Nucleic Acids Research* 47 (D1): D1038–43. <https://doi.org/10.1093/nar/gky1151>.
- Baird, G. D., and R. J. Heitzman. 1970. "Gluconeogenesis in the Cow. The Effects of a Glucocorticoid on Hepatic Intermediary Metabolism." *Biochemical Journal* 116 (5): 865–74. <https://doi.org/10.1042/bj1160865>.
- Barbat, Anne, Helene Leclerc, M Philippe, Sebastien Fritz, J B Daviere, and F Guillaume. 2016. "GénoSanté: améliorer la santé productive des vaches laitières par la sélection génomique et la conduite d'élevage: une première étape avec l'acétonémie."
- Becker, Doreen, Rosemarie Weikard, Christoph Schulze, Peter Wohlsein, and Christa Kühn. 2020. "A 50-Kb Deletion Disrupting the RSPO2 Gene Is Associated with Tetradysmelia in Holstein Friesian Cattle." *Genetics Selection Evolution* 52 (1): 68. <https://doi.org/10.1186/s12711-020-00586-y>.
- Bérodier, Marie, Mickaël Brochard, Didier Boichard, Stephanie Minery, and Vincent Ducrocq. 2020. "Plans d'accouplements sur la base d'informations génomiques: mieux conseiller les éleveurs en utilisant mieux toute l'information disponible."
- Besnard, F., H. Leclerc, M. Boussaha, C. Grohs, N. Jewell, A. Pinton, H. Barasc, et al. 2022. "Detailed Analysis of Mortality Rates in the Female Progeny of 1,001 Holstein Bulls Allows the Discovery of New Dominant Genetic Defects." *Journal of Dairy Science*, November, S0022030222006373. <https://doi.org/10.3168/jds.2022-22365>.
- Besnard, F., H. Leclerc, N. Cesbron, M. Boussaha, C. Grohs, N. Jewell, A. Pinton, et al. 2022. "525. Detailed Analysis of Mortality Rates in the Female Progeny of Holstein Bulls Allows the Discovery of New Dominant Genetic Defects." In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*, 2178–81. Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-940-4_525.

- Besnard, Florian, Ana Guintard, Cécile Grohs, Laurence Guzylack-Piriou, Margarita Cano, Clémentine Escouflaire, Chris Hozé, et al. 2023. “Massive Detection of Cryptic Recessive Genetic Defects in Cattle Mining Millions of Life Histories.” bioRxiv. <https://doi.org/10.1101/2023.09.22.558782>.
- Binsbergen, Rianne van, Marco Cam Bink, Mario Pl Calus, Fred A. van Eeuwijk, Ben J. Hayes, Ina Hulsegge, and Roel F. Veerkamp. 2014. “Accuracy of Imputation to Whole-Genome Sequence Data in Holstein Friesian Cattle.” *Genetics, Selection, Evolution: GSE* 46 (1): 41. <https://doi.org/10.1186/1297-9686-46-41>.
- Boichard, D., L. Maignel, and É Verrier. 1997. “The Value of Using Probabilities of Gene Origin to Measure Genetic Variability in a Population.” *Genetics Selection Evolution* 29 (1): 5. <https://doi.org/10.1186/1297-9686-29-1-5>.
- Boichard, Didier. 2002. “Pedig: A Fortran Package for Pedigree Analysis Suited for Large Populations.” https://www6.jouy.inra.fr/gabi_eng/Our-resources/Tool-development/Pedig.
- . 2020. “Nouvelles Modalités de Diffusion de La Génétique.” In *Génétique Des Animaux d'élevage: Diversité et Adaptation Dans Un Monde Changeant*, 1:193–203. Quae. <https://www.quae.com/produit/1635/9782759231003/genetique-des-animaux-d-elevage>.
- Boichard, Didier, Hoyoung Chung, Romain Dassonneville, Xavier David, André Eggen, Sébastien Fritz, Kimberly J. Gietzen, et al. 2012. “Design of a Bovine Low-Density SNP Array Optimized for Imputation.” Edited by Zhanjiang Liu. *PLoS ONE* 7 (3): e34130. <https://doi.org/10.1371/journal.pone.0034130>.
- Boichard, Didier, Vincent Ducrocq, Pascal Croiseau, and Sébastien Fritz. 2016. “Genomic Selection in Domestic Animals: Principles, Applications and Perspectives.” *Comptes Rendus Biologies, Trajectories of genetics, 150 years after Mendel / Trajectoire de la génétique, 150 après Mendel Guest Editors / Rédacteurs en chef invités : Bernard Dujon, Georges Pelletier*, 339 (7): 274–77. <https://doi.org/10.1016/j.crv.2016.04.007>.
- Boichard, Didier, and Eduardo Manfredi. 1994. “Genetic Analysis of Conception Rate in French Holstein Cattle.” *Acta Agriculturae Scandinavica, Section A—Animal Science* 44 (3): 138–45. <https://doi.org/10.1080/09064709409410890>.
- Bonaiti, B., D. Boichard, E. Verrier, V. Ducrocq, A. Barbat, and M. Briend. 1990. “La méthode française d'évaluation génétique des reproducteurs laitiers.” *INRAE Productions Animales* 3 (2): 83–92. <https://doi.org/10.20870/productions-animales.1990.3.2.4363>.
- Boubet, Boris. 2021. “DOSSIER-PATHOLOGIES-RESPIRATOIRES.” <https://www.gdscreuse.fr/Wp-Content/Uploads/2017/10/2021-01-DOSSIER-PATHOLOGIES-RESPIRATOIRES.Pdf>. January 2021. <http://www.gdscreuse.fr/wp-content/uploads/2017/10/2021-01-DOSSIER-PATHOLOGIES-RESPIRATOIRES.pdf>.
- Bourneuf, E., P. Otz, H. Pausch, V. Jagannathan, P. Michot, C. Grohs, G. Piton, et al. 2017. “Rapid Discovery of De Novo Deleterious Mutations in Cattle Enhances the Value of Livestock as Model Species.” *Scientific Reports* 7 (1): 11466. <https://doi.org/10.1038/s41598-017-11523-3>.
- Boussaha, Mekki, Arnaud Boulling, Valérie Wolgust, Lorraine Bourgeois-Brunel, Pauline Michot, Cécile Grohs, Nicolas Gaiani, et al. 2023. “Integrin Alpha 6 Homozygous Splice-Site Mutation Causes a New Form of Junctional Epidermolysis Bullosa in Charolais Cattle.” *Genetics Selection Evolution* 55 (1): 40. <https://doi.org/10.1186/s12711-023-00814-1>.
- Boussaha, Mekki, Camille Eché, Clémentine Escouflaire, Cécile Grohs, Carole Iampietro, Aurélien Capitan, Denis Milan, et al. 2023. “Search for New Mutations in Cattle by Systematic Whole Genome Resequencing.” In . <https://hal.inrae.fr/hal-04170956>.
- Boussaha, Mekki, Pauline Michot, Rabia Letaief, Chris Hozé, Sébastien Fritz, Cécile Grohs, Diane Esquerré, et al. 2016. “Construction of a Large Collection of Small Genome Variations in French Dairy and Beef Breeds Using Whole-Genome Sequences.” *Genetics Selection Evolution* 48 (1): 87. <https://doi.org/10.1186/s12711-016-0268-z>.

- Brochard, Mickaël, Didier Boichard, Aurelien Capitan, Guillaume Guillaume, Sebastien Fritz, L Nicod, and Jean-Marc Vacelet. 2018. "pANO, Le Risque d'anomalie Létale Pour Les Produits d'accouplements: Principe et Utilisation En Race Montbéliarde Sur La Zone Gen'IA Test," December, 5.
- Bursted, Bruna, Malú Zamariolli, Fernanda Teixeira Bellucco, and Maria Isabel Melaragno. 2022. "Mechanisms of Structural Chromosomal Rearrangement Formation." *Molecular Cytogenetics* 15 (1): 23. <https://doi.org/10.1186/s13039-022-00600-6>.
- Capitan, Aurélien, Aurélie Allais-Bonnet, Alain Pinton, Brigitte Marquant-Le Guienne, Daniel Le Bourhis, Cécile Grohs, Stéphan Bouet, et al. 2012. "A 3.7 Mb Deletion Encompassing ZEB2 Causes a Novel Polled and Multisystemic Syndrome in the Progeny of a Somatic Mosaic Bull." *PLoS One* 7 (11): e49084. <https://doi.org/10.1371/journal.pone.0049084>.
- Capitan, Aurélien, Cécile Grohs, Mathieu Gautier, and André Eggen. 2009. "The Scurs Inheritance: New Insights from the French Charolais Breed." *BMC Genetics* 10 (1): 33. <https://doi.org/10.1186/1471-2156-10-33>.
- Cavanagh, Julie A. L., Imke Tammen, Peter A. Windsor, John F. Bateman, Ravi Savarirayan, Frank W. Nicholas, and Herman W. Raadsma. 2007. "Bulldog Dwarfism in Dexter Cattle Is Caused by Mutations in ACAN." *Mammalian Genome* 18 (11): 808–14. <https://doi.org/10.1007/s00335-007-9066-9>.
- Charlier, Carole, Jorgen Steen Agerholm, Wouter Coppieters, Peter Karlskov-Mortensen, Wanbo Li, Gerben de Jong, Corinne Fasquelle, et al. 2012. "A Deletion in the Bovine FANCI Gene Compromises Fertility by Causing Fetal Death and Brachyspina." *PLOS ONE* 7 (8): e43085. <https://doi.org/10.1371/journal.pone.0043085>.
- Charlier, Carole, Wouter Coppieters, Frédéric Rollin, Daniel Desmecht, Jorgen S Agerholm, Nadine Cambisano, Eloisa Carta, et al. 2008. "Highly Effective SNP-Based Association Mapping and Management of Recessive Defects in Livestock." *Nature Genetics* 40 (4): 449–54. <https://doi.org/10.1038/ng.96>.
- Charlier, Carole, Wanbo Li, Chad Harland, Mathew Littlejohn, Wouter Coppieters, Frances Creagh, Steve Davis, et al. 2016. "NGS-Based Reverse Genetic Screen for Common Embryonic Lethal Mutations Compromising Fertility in Livestock." *Genome Research* 26 (10): 1333–41. <https://doi.org/10.1101/gr.207076.116>.
- Christensen, Ole F, and Mogens S Lund. 2010. "Genomic Prediction When Some Animals Are Not Genotyped." *Genetics Selection Evolution* 42 (1): 2. <https://doi.org/10.1186/1297-9686-42-2>.
- Cingolani, Pablo, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. 2012. "A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain W1118; Iso-2; Iso-3." *Fly* 6 (2): 80–92. <https://doi.org/10.4161/fly.19695>.
- Cole, John B. 2015. "A Simple Strategy for Managing Many Recessive Disorders in a Dairy Cattle Breeding Program." *Genetics Selection Evolution* 47 (1): 94. <https://doi.org/10.1186/s12711-015-0174-9>.
- Colleau, J., S. Fritz, F. Guillaume, A. Baur, D. Dupassieux, L. Journaux, A. Eggen, and D. Boichard. 2015. "Simulation des potentialités de la sélection génomique chez les bovins laitiers." *INRAE Productions Animales* 28 (3): 251–58. <https://doi.org/10.20870/productions-animales.2015.28.3.3030>.
- Corbeau, Julien, Cécile Grohs, Jeanlin Jourdain, Mekki Boussaha, Florian Besnard, Anne Barbat, Vincent Plassard, et al. 2024. "A Recurrent de Novo Missense Mutation in COL1A1 Causes Osteogenesis Imperfecta Type II and Preterm Delivery in Normande Cattle." *Genetics Selection Evolution* 56 (1): 39. <https://doi.org/10.1186/s12711-024-00909-3>.
- Crooks, Gavin E., Gary Hon, John-Marc Chandonia, and Steven E. Brenner. 2004. "WebLogo: A Sequence Logo Generator." *Genome Research* 14 (6): 1188–90. <https://doi.org/10.1101/gr.849004>.

- Cullmann, Gerhard, Karen Fien, Ryuji Kobayashi, and Bruce Stillman. 1995. "Characterization of the Five Replication Factor C Genes of *Saccharomyces Cerevisiae*." *Molecular and Cellular Biology* 15 (9): 4661–71. <https://doi.org/10.1128/MCB.15.9.4661>.
- Cuttance, Emma, and Richard Laven. 2019. "Perinatal Mortality Risk Factors in Dairy Calves." *Veterinary Journal (London, England: 1997)* 253 (November):105394. <https://doi.org/10.1016/j.tvjl.2019.105394>.
- Daetwyler, Hans D, Aurélien Capitan, Hubert Pausch, Paul Stothard, Rianne Van Binsbergen, Rasmus F Brøndum, Xiaoping Liao, et al. 2014. "Whole-Genome Sequencing of 234 Bulls Facilitates Mapping of Monogenic and Complex Traits in Cattle." *Nature Genetics* 46 (8): 858–65. <https://doi.org/10.1038/ng.3034>.
- Danchin-Burge, C., S. Danvy, D. Laloë, and E. Verrier. 2017. "Création d'un observatoire de la VARIabilité génétique des RUMinants et des Equidés (VARUME). Innovations Agronomiques 55, 235-245." <https://doi.org/10.15454/1.5137783633347627E12>.
- Dempster, Everett R., and I. Michael Lerner. 1950. "Heritability of Threshold Characters." *Genetics* 35 (2): 212–36.
- "Des indicateurs génomiques de résistance à la paratuberculose disponibles en race Normande." 2024. INRAE Institutionnel. April 2024. <https://www.inrae.fr/actualites/indicateurs-genomiques-resistance-paratuberculose-disponibles-race-normande>.
- Duchesne, A., M. Gautier, S. Chadi, C. Grohs, S. Floriot, Y. Gallard, G. Caste, A. Ducos, and A. Eggen. 2006. "Identification of a Doublet Missense Substitution in the Bovine LRP4 Gene as a Candidate Causal Mutation for Syndactyly in Holstein Cattle." *Genomics* 88 (5): 610–21. <https://doi.org/10.1016/j.ygeno.2006.05.007>.
- Duchesne, Amandine, Anne Vaiman, Johan Castille, Christian Beauvallet, Pauline Gaignard, Sandrine Floriot, Sabrina Rodriguez, et al. 2017. "Bovine and Murine Models Highlight Novel Roles for SLC25A46 in Mitochondrial Dynamics and Metabolism, with Implications for Human and Animal Health." Edited by Edmund Kunji. *PLOS Genetics* 13 (4): e1006597. <https://doi.org/10.1371/journal.pgen.1006597>.
- Duchesne, Amandine, Anne Vaiman, Magali Frah, Sandrine Floriot, Sabrina Legoueix-Rodriguez, Anne Desmazières, Sébastien Fritz, et al. 2018. "Progressive Ataxia of Charolais Cattle Highlights a Role of KIF1C in Sustainable Myelination." Edited by Carole Charlier. *PLOS Genetics* 14 (8): e1007550. <https://doi.org/10.1371/journal.pgen.1007550>.
- Escouflaire, C., and A. Capitan. 2021. "Analysis of Pedigree Data and Whole-genome Sequences in 12 Cattle Breeds Reveals Extremely Low Within-breed Y-chromosome Diversity." *Animal Genetics* 52 (5): 725–29. <https://doi.org/10.1111/age.13104>.
- Escouflaire, Clémentine, Emmanuelle Rebours, Mathieu Charles, Sébastien Orellana, Margarita Cano, Julie Rivière, Cécile Grohs, Hélène Hayes, and Aurélien Capitan. 2019. "A de Novo 3.8-Mb Inversion Affecting the EDA and XIST Genes in a Heterozygous Female Calf with Generalized Hypohidrotic Ectodermal Dysplasia." *BMC Genomics* 20 (1): 715. <https://doi.org/10.1186/s12864-019-6087-1>.
- Falconer, Douglas Scott. 1996. *Introduction to Quantitative Genetics*. Pearson Education India.
- Fasquelle, Corinne, Arnaud Sartelet, Wanbo Li, Marc Dive, Nico Tamma, Charles Michaux, Tom Druet, et al. 2009. "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 Genetics* 5 (9): e1000666. <https://doi.org/10.1371/journal.pgen.1000666>.
- Fernando, Rohan L., Hao Cheng, and Dorian J. Garrick. 2016. "An Efficient Exact Method to Obtain GBLUP and Single-Step GBLUP When the Genomic Relationship Matrix Is Singular." *Genetics Selection Evolution* 48 (1): 80. <https://doi.org/10.1186/s12711-016-0260-7>.
- Fingerlin, Tasha E., Elissa Murphy, Weiming Zhang, Anna L. Peljto, Kevin K. Brown, Mark P. Steele, James E. Loyd, et al. 2013. "Genome-Wide Association Study Identifies Multiple Susceptibility Loci for Pulmonary Fibrosis." *Nature Genetics* 45 (6): 613–20. <https://doi.org/10.1038/ng.2609>.

- Fishman, Lila, and Mariah McIntosh. 2019. "Standard Deviations: The Biological Bases of Transmission Ratio Distortion." *Annual Review of Genetics* 53 (1): 347–72. <https://doi.org/10.1146/annurev-genet-112618-043905>.
- Floriot, Sandrine, Christine Vesque, Sabrina Rodriguez, Florence Bourgain-Guglielmetti, Anthi Karaiskou, Mathieu Gautier, Amandine Duchesne, et al. 2015. "C-Nap1 mutation affects centriole cohesion and is associated with a Seckel-like syndrome in cattle." *Nature Communications* 6 (1): 6894. <https://doi.org/10.1038/ncomms7894>.
- Foulley, J. L., D. Gianola, and R. Thompson. 1983. "Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening." *Génétique Sélection Évolution* 15 (3): 401–24. <https://doi.org/10.1051/gse:19830305>.
- Frappeau, Véronique. 2023. "Chiffres Clés Du Dispositif Génétique Français 2022." Idele.Fr. February 2023. https://idele.fr/?eID=cmis_download&oid=workspace%3A%2F%2FSpacesStore%2F5147239d-fcb0-443e-aa88-9923f7dcdd60&cHash=931fca920c222b56261bd72385043302.
- Fresco, S., D. Boichard, S. Fritz, R. Lefebvre, S. Barbey, M. Gaborit, and P. Martin. 2023. "Comparison of Methane Production, Intensity, and Yield throughout Lactation in Holstein Cows." *Journal of Dairy Science* 106 (6): 4147–57. <https://doi.org/10.3168/jds.2022-22855>.
- Friedrichs, Kristen R., Kendal E. Harr, Kathy P. Freeman, Balazs Szladovits, Raquel M. Walton, Kirstin F. Barnhart, and Julia Blanco-Chavez. 2012. "ASVCP Reference Interval Guidelines: Determination of de Novo Reference Intervals in Veterinary Species and Other Related Topics." *Veterinary Clinical Pathology* 41 (4): 441–53. <https://doi.org/10.1111/vcp.12006>.
- Fritz, Sébastien. 2018. "An Initiator Codon Mutation in SDE2 Causes Recessive Embryonic Lethality in Holstein Cattle" 101 (7): 12.
- Fritz, Sébastien, Aurelien Capitan, Anis Djari, Sabrina C. Rodriguez, Anne Barbat, Aurélie Baur, Cécile Grohs, et al. 2013. "Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG and SLC37A2." Edited by Reiner Albert Veitia. *PLoS ONE* 8 (6): e65550. <https://doi.org/10.1371/journal.pone.0065550>.
- Fuerst-Waltl, B., and M.K. Sørensen. 2010. "Genetic Analysis of Calf and Heifer Losses in Danish Holstein." *Journal of Dairy Science* 93 (11): 5436–42. <https://doi.org/10.3168/jds.2010-3227>.
- Furukawa, Tomoyuki, Toyotaka Ishibashi, Seisuke Kimura, Hiroshi Tanaka, Junji Hashimoto, and Kengo Sakaguchi. 2003. "Characterization of All the Subunits of Replication Factor C from a Higher Plant, Rice (*Oryza Sativa* L.), and Their Relation to Development." *Plant Molecular Biology* 53 (1/2): 15–25. <https://doi.org/10.1023/B:PLAN.0000009258.04711.62>.
- Gonzalez-Peña, D., N. Vukasinovic, J.J. Brooker, C.A. Przybyla, and S.K. DeNise. 2019. "Genomic Evaluation for Calf Wellness Traits in Holstein Cattle." *Journal of Dairy Science* 102 (3): 2319–29. <https://doi.org/10.3168/jds.2018-15540>.
- Gorfu, Gezahegn, Jesús Rivera-Nieves, and Klaus Ley. 2009. "Role of B7 Integrins in Intestinal Lymphocyte Homing and Retention." *Current Molecular Medicine* 9 (7): 836–50. <https://doi.org/10.2174/156652409789105525>.
- Gourreau, Jean-Marie. 2011. *Guide pratique des maladies des bovins*. Agriproduction. Paris: Éd. France agricole Institut de l'élevage.
- Grobet, L., L. J. Martin, D. Poncelet, D. Pirottin, B. Brouwers, J. Riquet, A. Schoeberlein, et al. 1997. "A Deletion in the Bovine Myostatin Gene Causes the Double-Muscléd Phenotype in Cattle." *Nature Genetics* 17 (1): 71–74. <https://doi.org/10.1038/ng0997-71>.
- Grohs, C., J. Corbeau, J. Jourdain, M. Boussaha, V. Plassard, A. Barbat, C. Hamelin, et al. 2022. "521. A de Novo Missense Mutation of COL1A1 Causes Osteogenesis Imperfecta Type 2

- and Premature Delivery in Normande Cattle.” In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*, 2162–65. Rotterdam, the Netherlands: Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-940-4_521.
- Grohs, C, A Duchesne, S Floriot, A Ducos, C Danchin-Burge, and Université Paris-Saclay. 2016. “L’Observatoire National des Anomalies Bovines, son action et ses résultats pour une aide efficace à la gestion des anomalies génétiques,” 11.
- Gross, J. J., A.-C. Schwinn, F. Schmitz-Hsu, F. Menzi, C. Drögemüller, C. Albrecht, and R. M. Bruckmaier. 2016. “Rapid Communication: Cholesterol Deficiency–Associated APOB Mutation Impacts Lipid Metabolism in Holstein Calves and Breeding Bulls¹.” *Journal of Animal Science* 94 (4): 1761–66. <https://doi.org/10.2527/jas.2016-0439>.
- Häfliger, Irene M., Mirjam Spengeler, Franz R. Seefried, and Cord Drögemüller. 2022. “Four Novel Candidate Causal Variants for Deficient Homozygous Haplotypes in Holstein Cattle.” *Scientific Reports* 12 (1): 5435. <https://doi.org/10.1038/s41598-022-09403-6>.
- Häfliger, Irene Monika, Sonja Hofstetter, Thomas Mock, Manuela Hanna Stettler, Mireille Meylan, Kemal Mehinagic, Nadine Stokar-Regenscheit, and Cord Drögemüller. 2019. “APOB-associated Cholesterol Deficiency in Holstein Cattle Is Not a Simple Recessive Disease.” *Animal Genetics* 50 (4): 372–75. <https://doi.org/10.1111/age.12801>.
- Hazel, L. N. 1943. “The Genetic Basis for Constructing Selection Indexes.” *Genetics* 28 (6): 476–90.
- Henderson, C. R. 1975. « Best Linear Unbiased Estimation and Prediction under a Selection Model». *Biometrics* 31 (2): 423-47. <https://doi.org/10.2307/2529430>.
- Herman, Nicolas, Catherine Trumel, Anne Geffré, Jean-Pierre Braun, Marion Thibault, François Schelcher, and Nathalie Bourgès-Abella. 2018. “Hematology Reference Intervals for Adult Cows in France Using the Sysmex XT-2000iV Analyzer.” *Journal of Veterinary Diagnostic Investigation* 30 (5): 678–87. <https://doi.org/10.1177/1040638718790310>.
- Howie, Bryan, Christian Fuchsberger, Matthew Stephens, Jonathan Marchini, and Gonçalo R. Abecasis. 2012. “Fast and Accurate Genotype Imputation in Genome-Wide Association Studies through Pre-Phasing.” *Nature Genetics* 44 (8): 955–59. <https://doi.org/10.1038/ng.2354>.
- Hozé, C., C. Escouflaire, M. Mesbah-Uddin, A. Barbat, M. Boussaha, M.C. Deloche, D. Boichard, S. Fritz, and A. Capitan. 2020. “Short Communication: A Splice Site Mutation in CENPU Is Associated with Recessive Embryonic Lethality in Holstein Cattle.” *Journal of Dairy Science* 103 (1): 607–12. <https://doi.org/10.3168/jds.2019-17056>.
- Hoze, Chris, Sebastien Fritz, Aurélia Baur, Cécile Grohs, Coralie Danchin-Burge, and Didier Boichard. 2018. “Prise en compte des gènes d’intérêt dans les objectifs de sélection en bovins laitiers.” In . <https://hal.inrae.fr/hal-02735939>.
- Huang, Lam Opal, Aurélie Labbe, and Claire Infante-Rivard. 2013. “Transmission Ratio Distortion: Review of Concept and Implications for Genetic Association Studies.” *Human Genetics* 132 (3): 245–63. <https://doi.org/10.1007/s00439-012-1257-0>.
- Ibi, Takayuki, Alexander K Kahi, and Hiroyuki Hirooka. 2014. “Genetic Parameters of Postnatal Mortality and Birth Weight in Japanese Black Calves.” *Animal Science Journal*, 6.
- Id-Lahoucine, S., A. Cánovas, C. Jatón, F. Miglior, P.A.S. Fonseca, M. Sargolzaei, S. Miller, F.S. Schenkel, J.F. Medrano, and J. Casellas. 2019. “Implementation of Bayesian Methods to Identify SNP and Haplotype Regions with Transmission Ratio Distortion across the Whole Genome: TRDscan v.1.0.” *Journal of Dairy Science* 102 (4): 3175–88. <https://doi.org/10.3168/jds.2018-15296>.
- Id-Lahoucine, S., J. Casellas, D. Lu, M. Sargolzaei, S. Miller, and A. Cánovas. 2023. “Distortion of Mendelian Segregation across the Angus Cattle Genome Uncovering Regions Affecting Reproduction.” *Scientific Reports* 13 (1): 13393. <https://doi.org/10.1038/s41598-023-37710-z>.

- Jamrozik, J., J. Fatehi, G. J. Kistemaker, and L. R. Schaeffer. 2005. "Estimates of Genetic Parameters for Canadian Holstein Female Reproduction Traits." *Journal of Dairy Science* 88 (6): 2199–2208. [https://doi.org/10.3168/jds.S0022-0302\(05\)72895-2](https://doi.org/10.3168/jds.S0022-0302(05)72895-2).
- Jourdain, Jeanlin, Harmonie Barasc, Thomas Faraut, Anne Calgaro, Nathalie Bonnet, Camille Marcuzzo, Amandine Suin, et al. 2023. "Large-Scale Detection and Characterization of Interchromosomal Rearrangements in Normozoospermic Bulls Using Massive Genotype and Phenotype Data Sets." *Genome Research* 33 (6): 957–71. <https://doi.org/10.1101/gr.277787.123>.
- Kipp, S., D. Segelke, S. Schierenbeck, F. Reinhardt, R. Reents, C. Wurmser, H. Pausch, et al. 2016. "Identification of a Haplotype Associated with Cholesterol Deficiency and Increased Juvenile Mortality in Holstein Cattle." *Journal of Dairy Science* 99 (11): 8915–31. <https://doi.org/10.3168/jds.2016-11118>.
- Kishimoto, Keishi, Ayumi Kato, Shigehiro Osada, Makoto Nishizuka, and Masayoshi Imagawa. 2010. "Fad104, a Positive Regulator of Adipogenesis, Negatively Regulates Osteoblast Differentiation." *Biochemical and Biophysical Research Communications* 397 (2): 187–91. <https://doi.org/10.1016/j.bbrc.2010.05.077>.
- Kishimoto, Keishi, Makoto Nishizuka, Daiki Katoh, Ayumi Kato, Shigehiro Osada, and Masayoshi Imagawa. 2013. "FAD104, a Regulatory Factor of Adipogenesis, Acts as a Novel Regulator of Calvarial Bone Formation." *The Journal of Biological Chemistry* 288 (44): 31772–83. <https://doi.org/10.1074/jbc.M113.452961>.
- Kolanczyk, Mateusz, Markus Pech, Tomasz Zemojtel, Hiroshi Yamamoto, Ivan Mikula, Maria-Antonietta Calvaruso, Mariël Van Den Brand, et al. 2011. "NOA1 Is an Essential GTPase Required for Mitochondrial Protein Synthesis." Edited by M. Bishr Omary. *Molecular Biology of the Cell* 22 (1): 1–11. <https://doi.org/10.1091/mbc.e10-07-0643>.
- Kolishovski, Georgi, Anna Lamoureux, Paul Hale, Joel E. Richardson, Jill M. Recla, Omoluyi Adesanya, Al Simons, Govindarajan Kunde-Ramamoorthy, and Carol J. Bult. 2019. "The JAX Synteny Browser for Mouse-Human Comparative Genomics." *Mammalian Genome* 30 (11–12): 353–61. <https://doi.org/10.1007/s00335-019-09821-4>.
- Layer, Ryan M, Colby Chiang, Aaron R Quinlan, and Ira M Hall. 2014. "LUMPY: A Probabilistic Framework for Structural Variant Discovery." *Genome Biology* 15 (6): R84. <https://doi.org/10.1186/gb-2014-15-6-r84>.
- Leclerc, H., R. Lefebvre, M. DOUGUET, F. PHOCAS, and S. MATTALIA. 2016. "Mortalité des veaux : analyse phénotypique et étude de la composante génétique." <https://hal.archives-ouvertes.fr/hal-01608259/>.
- Legarra, A., and I. Misztal. 2008. "Technical Note: Computing Strategies in Genome-Wide Selection." *Journal of Dairy Science* 91 (1): 360–66. <https://doi.org/10.3168/jds.2007-0403>.
- Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform." *Bioinformatics* 25 (14): 1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
- Li, Wanbo, Arnaud Sartelet, Nico Tamma, Wouter Coppieters, Michel Georges, and Carole Charlier. 2016. "Reverse Genetic Screen for Loss-of-Function Mutations Uncovers a Frameshifting Deletion in the Melanophilin Gene Accountable for a Distinctive Coat Color in Belgian Blue Cattle." *Animal Genetics* 47 (1): 110–13. <https://doi.org/10.1111/age.12383>.
- Li, Yanling, Sijie Gan, Lin Ren, Long Yuan, Junlan Liu, Wei Wang, Xiaoyu Wang, et al. 2018. "Multifaceted Regulation and Functions of Replication Factor C Family in Human Cancers." *American Journal of Cancer Research* 8 (8): 1343–55.
- Li, Yun, Cristen Willer, Serena Sanna, and Gonçalo Abecasis. 2009. "Genotype Imputation." *Annual Review of Genomics and Human Genetics* 10:387–406. <https://doi.org/10.1146/annurev.genom.9.081307.164242>.

- Lynch, C., E. M. Leishman, F. Miglior, D. Kelton, F. Schenkel, and C. F. Baes. 2024. "Review: Opportunities and Challenges for the Genetic Selection of Dairy Calf Disease Traits." *Animal*, March, 101141. <https://doi.org/10.1016/j.animal.2024.101141>.
- Maigret, Corinne. 2024. "Rapport Bovins 2023 Productions Lait et Viande." https://idele.fr/inosys-reseaux-elevage/?eID=cmis_download&oID=workspace%3A%2F%2FSpacesStore%2F5173106d-ab95-464e-8ce4-073ec3c92720&cHash=6da68eec8bbb3ed5cd604af16949809e.
- Martin, F. J, M Ridwan Amode, Alisha Aneja, Olanrewaju Austine-Orimoloye, Andrey G Azov, If Barnes, Arne Becker, et al. 2023. "Ensembl 2023." *Nucleic Acids Research* 51 (D1): D933–41. <https://doi.org/10.1093/nar/gkac958>.
- Mattalia, Sophie, and Denis Laloë. 2009. "UMT Evaluation Génétique Des Bovins : Bilan Des Travaux Menés En 2008; Projets Pour 2009." Technical Report. Institut de l'Elevage. <https://hal.science/hal-01193622>.
- Matukumalli, Lakshmi K., Cynthia T. Lawley, Robert D. Schnabel, Jeremy F. Taylor, Mark F. Allan, Michael P. Heaton, Jeff O'Connell, et al. 2009. "Development and Characterization of a High Density SNP Genotyping Assay for Cattle." Edited by Amanda Ewart Toland. *PLoS ONE* 4 (4): e5350. <https://doi.org/10.1371/journal.pone.0005350>.
- McClure, Matthew C., Derek Bickhart, Dan Null, Paul VanRaden, Lingyang Xu, George Wiggans, George Liu, et al. 2014. "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. <https://doi.org/10.1371/journal.pone.0092769>.
- McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome Research* 20 (9): 1297–1303. <https://doi.org/10.1101/gr.107524.110>.
- McLaren, William, Laurent Gil, Sarah E. Hunt, Harpreet Singh Riat, Graham R. S. Ritchie, Anja Thormann, Paul Flicek, and Fiona Cunningham. 2016. "The Ensembl Variant Effect Predictor." *Genome Biology* 17 (1): 122. <https://doi.org/10.1186/s13059-016-0974-4>.
- McVean, Gil A., David M. Altshuler (Co-Chair), Richard M. Durbin (Co-Chair), Gonçalo R. Abecasis, David R. Bentley, Aravinda Chakravarti, Andrew G. Clark, et al. 2012. "An Integrated Map of Genetic Variation from 1,092 Human Genomes." *Nature* 491 (7422): 56–65. <https://doi.org/10.1038/nature11632>.
- Menzi, F., N. Besuchet-Schmutz, M. Fragnière, S. Hofstetter, V. Jagannathan, T. Mock, A. Raemy, et al. 2016. "A Transposable Element Insertion in *APOB* Causes Cholesterol Deficiency in Holstein Cattle." *Animal Genetics* 47 (2): 253–57. <https://doi.org/10.1111/age.12410>.
- Mesbah-Uddin, M., C. Hoze, P. Michot, A. Barbat, R. Lefebvre, M. Boussaha, G. Sahana, S. Fritz, D. Boichard, and A. Capitan. 2019. "A Missense Mutation (p.Tyr452Cys) in the CAD Gene Compromises Reproductive Success in French Normande Cattle." *Journal of Dairy Science* 102 (7): 6340–56. <https://doi.org/10.3168/jds.2018-16100>.
- Meuwissen, T H, B J Hayes, and M E Goddard. 2001. "Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps." *Genetics* 157 (4): 1819–29.
- Michot, Pauline, Sabine Chahory, Andrew Marete, Cécile Grohs, Dimitri Dagios, Elise Donzel, Abdelhak Aboukadiri, et al. 2016. "A Reverse Genetic Approach Identifies an Ancestral Frameshift Mutation in RP1 Causing Recessive Progressive Retinal Degeneration in European Cattle Breeds." *Genetics Selection Evolution* 48 (1): 56. <https://doi.org/10.1186/s12711-016-0232-y>.
- Michot, Pauline, Oscar Fantini, Régis Braque, Aurélie Allais-Bonnet, Romain Saintilan, Cécile Grohs, Johanna Barbieri, et al. 2015. "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." *Genetics, Selection, Evolution: GSE* 47 (1): 37. <https://doi.org/10.1186/s12711-015-0110-z>.

- Michot, Pauline, Sébastien Fritz, Anne Barbat, Mekki Boussaha, Marie-Christine Deloche, Cécile Grohs, Chris Hoze, et al. 2017. “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* 100 (10): 8176–87. <https://doi.org/10.3168/jds.2017-12579>.
- Ministère de l’agriculture de l’agroalimentaire et de la forêt. 2016. *Le règlement zootechnique européen*. <https://agriculture.gouv.fr/le-reglement-zootechnique-europeen-un-nouveau-cadre-reglementaire-europeen-pour-la-genetique>.
- Misztal, Ignacy, Daniela Lourenco, Ignacio Aguilar, Andres Legarra, and Zulma Vitezica. 2014. “Manual for BLUPF90 Family of Programs.”
- Mock, T., K. Mehinagic, F. Menzi, E. Studer, A. Oevermann, M.H. Stoffel, C. Drögemüller, M. Meylan, and N. Regenscheit. 2016. “Clinicopathological Phenotype of Autosomal Recessive Cholesterol Deficiency in Holstein Cattle.” *Journal of Veterinary Internal Medicine* 30 (4): 1369–75. <https://doi.org/10.1111/jvim.13976>.
- Neupane, M., J.L. Hutchison, C.P. Van Tassell, and P.M. VanRaden. 2021. “Genomic Evaluation of Dairy Heifer Livability.” *Journal of Dairy Science*, May, S0022030221005968. <https://doi.org/10.3168/jds.2020-19687>.
- Ng, Pauline C., and Steven Henikoff. 2003. “SIFT: Predicting Amino Acid Changes That Affect Protein Function.” *Nucleic Acids Research* 31 (13): 3812–14.
- Nishizuka, Makoto, Keishi Kishimoto, Ayumi Kato, Masahito Ikawa, Masaru Okabe, Ryuichiro Sato, Hiroyuki Niida, Makoto Nakanishi, Shigehiro Osada, and Masayoshi Imagawa. 2009. “Disruption of the Novel Gene Fad104 Causes Rapid Postnatal Death and Attenuation of Cell Proliferation, Adhesion, Spreading and Migration.” *Experimental Cell Research* 315 (5): 809–19. <https://doi.org/10.1016/j.yexcr.2008.12.013>.
- Pedersen, Jørn, Morten Kargo, Anders Fogh, Jukka Pösö, Jan-Åke Eriksson, Ulrik Sander Nielsen, and Gert Pedersen Aamand. 2015. “Note on Economic Value of Young Stock Survival,” August, 11.
- Pelt, M L van, H Eding, P Vessies, and G de Jong. 2012. “Developing a Genetic Evaluation for Calf Survival during Rearing in The Netherlands.” . . *Cork*, no. 46, 5.
- Quaas, R. L. 1988. “Additive Genetic Model with Groups and Relationships.” *Journal of Dairy Science* 71 (June):91–98. [https://doi.org/10.1016/S0022-0302\(88\)79986-5](https://doi.org/10.1016/S0022-0302(88)79986-5).
- Rausch, T., T. Zichner, A. Schlattl, A. M. Stutz, V. Benes, and J. O. Korbel. 2012. “DELLY: Structural Variant Discovery by Integrated Paired-End and Split-Read Analysis.” *Bioinformatics* 28 (18): i333–39. <https://doi.org/10.1093/bioinformatics/bts378>.
- Rendel, J. M., and Alan Robertson. 1950. “Estimation of Genetic Gain in Milk Yield by Selection in a Closed Herd of Dairy Cattle.” *Journal of Genetics* 50 (1): 1–8. <https://doi.org/10.1007/BF02986789>.
- Reynolds, Edwardo G. M., Catherine Neeley, Thomas J. Lopdell, Michael Keehan, Keren Dittmer, Chad S. Harland, Christine Couldrey, et al. 2021. “Non-Additive Association Analysis Using Proxy Phenotypes Identifies Novel Cattle Syndromes.” *Nature Genetics* 53 (7): 949–54. <https://doi.org/10.1038/s41588-021-00872-5>.
- Reynolds, N., P. A. Fantès, and S. A. MacNeill. 1999. “A Key Role for Replication Factor C in DNA Replication Checkpoint Function in Fission Yeast.” *Nucleic Acids Research* 27 (2): 462–69. <https://doi.org/10.1093/nar/27.2.462>.
- Rodrigues, Carlos H M, Yoochan Myung, Douglas E V Pires, and David B Ascher. 2019. “mCSM-PPI2: Predicting the Effects of Mutations on Protein–Protein Interactions.” *Nucleic Acids Research* 47 (W1): W338–44. <https://doi.org/10.1093/nar/gkz383>.
- Rosen, Benjamin D, Derek M Bickhart, Robert D Schnabel, Sergey Koren, Christine G Elsik, Elizabeth Tseng, Troy N Rowan, et al. 2020. “De Novo Assembly of the Cattle Reference Genome with Single-Molecule Sequencing.” *GigaScience* 9 (3): g1aa021. <https://doi.org/10.1093/gigascience/g1aa021>.

- Santman-Berends, I.M.G.A., Y.H. Schukken, and G. van Schaik. 2019. "Quantifying Calf Mortality on Dairy Farms: Challenges and Solutions." *Journal of Dairy Science* 102 (7): 6404–17. <https://doi.org/10.3168/jds.2019-16381>.
- Schütz, Ekkehard, Christin Wehrhahn, Marius Wanjek, Ralf Bortfeld, Wilhelm E. Wemheuer, Julia Beck, and Bertram Brenig. 2016. "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* 11 (4): e0154602. <https://doi.org/10.1371/journal.pone.0154602>.
- Schwenger, Barbara, Stephan Schöber, and Dettlef Simon. 1993. "DUMPS Cattle Carry a Point Mutation in the Uridine Monophosphate Synthase Gene." *Genomics* 16 (1): 241–44. <https://doi.org/10.1006/geno.1993.1165>.
- Segelke, D., H. Täubert, F. Reinhardt, and G. Thaller. 2016. "Considering Genetic Characteristics in German Holstein Breeding Programs." *Journal of Dairy Science* 99 (1): 458–67. <https://doi.org/10.3168/jds.2015-9764>.
- Sellier, Pierre, Didier Boichard, and Etienne Verrier. 2019. "La génétique animale à l'INRA: soixante ans d'une histoire scientifique en prise avec le monde de la sélection et riche en rebondissements technologiques." *Histoire de la recherche contemporaine*, no. Tome VIII-n°1 (June), 86–97. <https://doi.org/10.4000/hrc.3222>.
- Shuster, D E, M E Kehrli, M R Ackermann, and R O Gilbert. 1992. "Identification and Prevalence of a Genetic Defect That Causes Leukocyte Adhesion Deficiency in Holstein Cattle." *Proceedings of the National Academy of Sciences* 89 (19): 9225–29. <https://doi.org/10.1073/pnas.89.19.9225>.
- Stevenson, J. S., and J. H. Britt. 2017. "A 100-Year Review: Practical Female Reproductive Management." *Journal of Dairy Science* 100 (12): 10292–313. <https://doi.org/10.3168/jds.2017-12959>.
- Tajima, M., S. Miyake, K. Takehana, A. Kobayashi, O. Yamato, and Y. Maede. 1999. "Gene Defect of Dermatan Sulfate Proteoglycan of Cattle Affected with a Variant Form of Ehlers-Danlos Syndrome." *Journal of Veterinary Internal Medicine* 13 (3): 202–5. [https://doi.org/10.1892/0891-6640\(1999\)013<0202:gdodsp>2.3.co;2](https://doi.org/10.1892/0891-6640(1999)013<0202:gdodsp>2.3.co;2).
- Thompson, Julie D., Desmond G. Higgins, and Toby J. Gibson. 1994. "CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice." *Nucleic Acids Research* 22 (22): 4673–80. <https://doi.org/10.1093/nar/22.22.4673>.
- Thomsen, Bo, Per Horn, Frank Panitz, Emøke Bendixen, Anette H. Petersen, Lars-Erik Holm, Vivi H. Nielsen, Jørgen S. Agerholm, Jens Arnbjerg, and Christian Bendixen. 2006. "A Missense Mutation in the Bovine SLC35A3 Gene, Encoding a UDP-N-Acetylglucosamine Transporter, Causes Complex Vertebral Malformation." *Genome Research* 16 (1): 97–105. <https://doi.org/10.1101/gr.3690506>.
- Vaiman, Anne, Sébastien Fritz, Christian Beauvallet, Mekki Boussaha, Cécile Grohs, Nathalie Daniel-Carlier, Anne Relun, Didier Boichard, Jean-Luc Vilotte, and Amandine Duchesne. 2022. "Mutation of the MYH3 Gene Causes Recessive Cleft Palate in Limousine Cattle." *Genetics Selection Evolution* 54 (1): 71. <https://doi.org/10.1186/s12711-022-00762-2>.
- VanRaden, P.M. 2008. "Efficient Methods to Compute Genomic Predictions." *Journal of Dairy Science* 91 (11): 4414–23. <https://doi.org/10.3168/jds.2007-0980>.
- VanRaden, P.M., K.M. Olson, D.J. Null, and J.L. Hutchison. 2011. "Harmful Recessive Effects on Fertility Detected by Absence of Homozygous Haplotypes." *Journal of Dairy Science* 94 (12): 6153–61. <https://doi.org/10.3168/jds.2011-4624>.
- Veit, H. P., and R. L. Farrell. 1978. "The Anatomy and Physiology of the Bovine Respiratory System Relating to Pulmonary Disease." *The Cornell Veterinarian* 68 (4): 555–81.
- Weller, Joel Ira. 2021. "Genetic and Environmental Analysis of Female Calf Survival in the Israel Holstein Cattle Population" 104 (3): 14.

- Wright, Sewall. 1922. "Coefficients of Inbreeding and Relationship." *The American Naturalist* 56 (645): 330–38.
- Yang, Jian, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. 2011. "GCTA: A Tool for Genome-Wide Complex Trait Analysis." *The American Journal of Human Genetics* 88 (1): 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
- Yang, Y., J. Si, X. Lv, D. Dai, L. Liu, S. Tang, Y. Wang, S. Zhang, W. Xiao, and Y. Zhang. 2022. "Integrated Analysis of Whole Genome and Transcriptome Sequencing Reveals a Frameshift Mutation Associated with Recessive Embryonic Lethality in Holstein Cattle." *Animal Genetics* 53 (1): 137–41. <https://doi.org/10.1111/age.13160>.
- Ye, Kai, Marcel H. Schulz, Quan Long, Rolf Apweiler, and Zemin Ning. 2009. "Pindel: A Pattern Growth Approach to Detect Break Points of Large Deletions and Medium Sized Insertions from Paired-End Short Reads." *Bioinformatics* 25 (21): 2865–71. <https://doi.org/10.1093/bioinformatics/btp394>.
- Yuzbasiyan-Gurkan, Vilma, and Elizabeth Bartlett. 2006. "Identification of a Unique Splice Site Variant in *SLC39A4* in Bovine Hereditary Zinc Deficiency, Lethal Trait A46: An Animal Model of Acrodermatitis Enteropathica." *Genomics* 88 (4): 521–26. <https://doi.org/10.1016/j.ygeno.2006.03.018>.

Acknowledgements

We are grateful to L. Balberini (Auriva), M. Chambrial (Origenplus), M. Courdier (Evajura), G. Fayolle (Umotest), C. Hamelin (Innoval), M. Philippe (Synetics), S. Patey (Genes Diffusion), N. Cesbron (Oniris), and the many breeders, veterinarians and agricultural technicians involved in this study for providing access to animals and samples. We also thank M. Bernard, C. Bevilacqua, E. Doz-Deblauwe, M. Femenia, C. Fouéré, N. Gaiani, J. Ros and N. Winter (INRAE) for their assistance and GenEval for providing data on yield deviations and genetic standard deviations. F.B. is supported by a CIFRE PhD grant from IDELE, with the financial support of the Association Nationale de la Recherche et de la Technologie and APIS-GENE (Paris, France). This study was also supported by the Effitness and Welcow projects funded by APIS-GENE, by the Cartoseq (ANR10-GENM-0018) and Bovano (ANR-14-CE19-0011) projects co-funded by the Agence Nationale de la Recherche and APIS-GENE, and by the SeqOccln project co-funded by the European Union and the Occitania Region (FEDER-FSE MIDI-PYRENEES ET GARONNE 2014-2020).

Author contributions

Conceived and coordinated the project: A. Capitan. Designed the experiments: A. Capitan, F.B., G.F., D.B. and M.-A.A. Performed HHED mapping: A.G. and A. Capitan. Set up haplotype tests: S.F. Analyzed survival curves: A.G., A. Capitan and C. Escouflaire. Analyzed pedigree information: F.B. and S. Minéry. Extracted DNA: C. Grohs and M.-C.D. Were involved in library preparation and

whole genome sequencing: C. Eche and C.I. Processed WGS data: M. Boussaha, C. B., C. Klopp, M. Charles and C. Kuchly. Filtered and annotated variants: F.B., A. Capitan and A.G. Analyzed large-scale genotyping data: F.B., A. Capitan and C. Escouflaire. Estimated the effect of haplotypes or deleterious variants on a variety of traits: F.B., A. Capitan, J.J. and A. Barbat. Performed the pathophysiological analyses and necropsy examinations: G.F., L.G.-P., M.-A.A., A. Clément., L.D., B.G., E.C., M. Bouchier, T.B., A. Remot, V.P., A. Relun, B.R., Y.M., R.G. Assisted in sampling: A. Capitan, C. Grohs, M-C.D., M. Cano. Performed histological and ultrastructure analyses: M. Cano, C.P., A. Capitan, J.R. and M.V. Performed qPCR analysis: C. Grohs. Contributed reagents/materials/analysis tools: C.H., H.L., A. Boulling, S.B., C. Grohs, C.D.-B., F.L., S. Mattalia, A.A.B., G.V., C. Gaspin, C.D. and D.M. Wrote the manuscript: A. Capitan and F.B. with contributions from G.F., L.G.-P., D.B. J.J. and M.-A.A.

Competing interests

The authors declare no competing interests.

3. Discussion

Cette étude fait suite à la méthode présentée dans la partie 7.2.1 Déficiences en homozygotes et propose un perfectionnement de la détection des anomalies génétiques récessives dans une population. Par cette méthode, nous avons pu détecter 34 *loci*, proposer 8 variants candidats et décrire l'impact de trois mutations causales récessives ségrégant dans les trois grandes races laitières. Ces mutations ont un impact considérable sur la reproduction et la mortalité des bovins. Cet impact est d'autant plus important du fait de l'utilisation massive de certains taureaux dans les races laitières. En effet, comme nous l'avons vu dans la partie « 5. Structure des populations bovines et émergence régulière d'anomalies génétiques », les méthodes de sélection actuelles favorisent la dissémination d'anomalies récessives dans la population.

Nous avons aussi pu mettre en évidence une problématique liée à la détection par haplotype. En effet, une des mutations en Holstein (sur *ITGB7*) est sur un haplotype très courant car porté par le plus grand ancêtre commun de la race (ELEVATION). La mutation a été détectée chez des taureaux apparentés (ELTON puis O-MAN). L'haplotype coexiste donc sous deux formes, avec et sans le variant causal, ce qui limite le déséquilibre de liaison entre l'haplotype et le variant causal. Cela complexifie la détection dans la cartographie, car beaucoup de veaux porteurs de l'haplotype ne sont pas porteurs de la mutation et donc le signal par HHED est réduit. Ainsi, l'apparition d'une anomalie dans un haplotype très présent dans la population est difficile à détecter avant qu'un nouveau taureau porteur de l'anomalie et recombinant à proximité de la mutation ne dissémine un nouvel haplotype original très présent. Au contraire, lorsque la mutation est présente sur plusieurs haplotypes (l'haplotype ancestral et des haplotypes recombinants) comme le cas de *NOA1* (mitochondropathie en race Montbéliarde), alors l'effet est dilué sur plusieurs haplotypes. Dans le cas de cette anomalie, le taureau DEBOUT (1972, deuxième contributeur de la race) est le premier porteur connu de la mutation et l'a disséminée dans la population, et son descendant le taureau GARDIAN, né en 1991 et huitième contributeur de la race, a eu une recombinaison et ne porte plus l'haplotype qui a permis de détecter la mutation. Ces deux différents scénarios ont un impact sur les analyses supplémentaires, notamment sur les corrélations des haplotypes avec les variants causaux (0,57 pour *ITGB7* et 0,55 pour *NOA1*). Il y a donc des améliorations possibles dans la détection des mutations à partir d'haplotypes. On pourrait ainsi combiner différents haplotypes dans le cas de la mitochondropathie ou retirer certains porteurs de l'haplotype sur la base de l'information pedigree dans le cas de BLIRD (*Bovine Lymphocyte Intestinal Retention Defect*) en race Holstein.

Plus généralement, on pourrait calculer une probabilité d'identité sur la base de l'information marqueurs et généalogie et utiliser cette probabilité pour calculer la corrélation.

Une autre anomalie a été détectée en race Holstein aux Etats-Unis appelée *recumbency*, avec un ancêtre commun appelé SOUTHWIND, très présent dans le pool génétique des Etats-Unis (Al-Khudhair et al. 2024) mais avec une contribution de moins de 1% dans la population française. Comme nous l'avons vu dans la partie « Structure des populations bovines et émergence régulière d'anomalies génétiques », les mutations récessives détectés dans les populations ont très souvent un ancêtre commun qui est un fort contributeur du pool génétique de la race, nous réaborderons ce point dans le Chapitre VIII Discussion générale et perspectives.

La méthode demande de grandes ressources de données, tant au niveau génétique, au niveau des pedigrees et des phénotypes, qui sont disponibles dans les races bovines françaises. Cette contrainte limite malheureusement les possibilités d'étude dans les populations de plus petite taille ou avec peu d'animaux génotypés. Dans ces populations, il est plus compliqué de détecter des anomalies sans signe d'appel, alors que leur effectif génétique faible laisse suggérer qu'elles sont aussi atteintes que les populations de plus grande taille.

La détection d'anomalies récessives dans les populations bovines permet l'étude fonctionnelle de mutations sur des modèles mammifères de grande taille, montrant l'intérêt du bovin dans la caractérisation des gènes communs aux mammifères. Grâce à la collaboration avec différents vétérinaires, cette étude a pu intégrer des analyses poussées de l'impact fonctionnel de la mutation sur l'immunité des veaux, et différentes études sont en cours, notamment sur le gène *ITGB7* pour une meilleure caractérisation.

Chapitre VI. Détection de *loci* influençant la létalité embryonnaire et la survie juvénile avec un déterminisme autosomique dominant à pénétrance incomplète.

1. Introduction

Le présent chapitre présente une méthode de détection de variants délétères dominants, voire surdominants, caractérisés par une pénétrance incomplète à l'état hétérozygote au sein des populations bovines. Comme nous l'avons décrit dans le Chapitre I, la structure de ces populations est caractérisée par de grandes familles de père. Les variants recherchés ici, dont les effets néfastes ne se manifestent qu'en présence de pathogènes spécifiques ou dans des conditions environnementales particulières, sans que l'animal ne soit nécessairement homozygote, soulèvent des problématiques complexes en matière d'analyse génétique.

Aussi, les morts n'étant fréquemment pas typés, cela induit un biais d'échantillonnage sur les animaux génotypés qui ne sont pas représentatifs de la globalité de la population. Dans ce contexte, notre étude se concentre sur la recherche de distorsion de ségrégation haplotypique que nous nommons *Haplotype Segregation Distorsion* ou HSD. Cet outil permet l'identification d'anomalies dominantes à pénétrance incomplète (Id-Lahoucine et al. 2019; 2023). Cette approche repose sur la nécessité d'exploiter au mieux les informations issues de données manquantes, notamment celles associées à la mortalité embryonnaire et juvénile, même antérieures au processus de génotypage (cf. Figure VI.1).

A noter que cette recherche ne cible pas forcément les variants délétères mais permet aussi de chercher des variants améliorateurs, pouvant par exemple apporter une résistance à certains pathogènes ou pouvant améliorer l'implantation ou le développement embryonnaire.

Pour ce faire, nous explorons la distorsion de la transmission des haplotypes portés à l'état hétérozygote par des taureaux dans leur descendance. L'hypothèse d'une transmission mendélienne postule que le taureau transmet ses deux haplotypes à sa descendance avec la même probabilité de 0,5. Lorsque cela n'est pas le cas, nous détectons une distorsion chez les animaux sains typés. Des écarts au schéma de transmission mendélienne peuvent se produire à différentes étapes biologiques, notamment la production de gamètes, le stockage des gamètes,

la capacité spermatique (motilité, viabilité, etc.), la fécondation, le développement embryonnaire, et elle peut dépendre de phénomènes tels que la méiose, les interactions gamétiques (Fishman et McIntosh 2019; Huang et al 2013), et enfin la fitness reproductive ou la survie. Ces différentes étapes sont présentées en Figure VI.1.

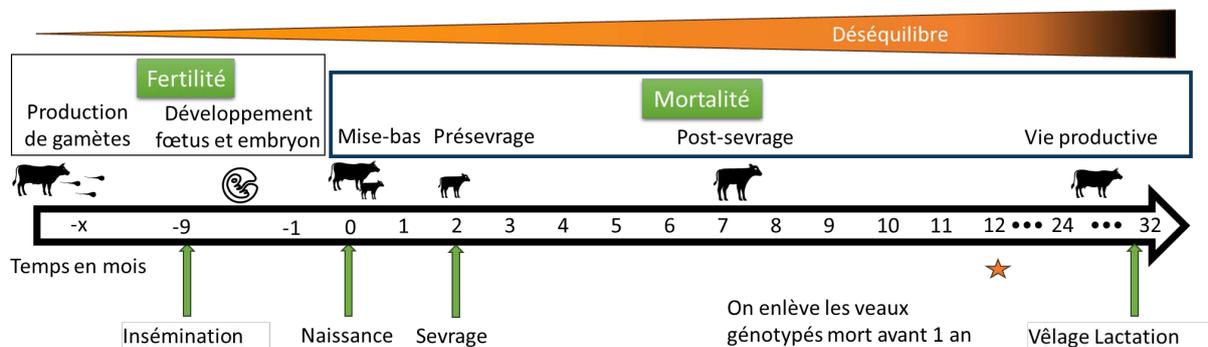


Figure VI-1 : Schéma résumant les différentes causes possibles de déséquilibres de transmissions de segments chromosomiques observés entre un taureau et ses descendants vivants à un temps « t ».

Ces distorsions de ségrégation peuvent être dues notamment à une sélection des gamètes au niveau du tractus femelle ou de la fécondation, ou encore à la mort embryonnaire, fœtale ou postnatale de certains génotypes en fonction de divers facteurs. L'étoile symbolise la durée de vie minimale pour la sélection des génotypes dans le cadre de cette étude.

2. Matériel et Méthodes

2.1. Animaux et données

Cette étude utilise trois jeux de données principaux :

- Les génotypes de 748 263 bovins Holstein, 397 200 Montbéliarde et 115 090 Normande ont été phasés et imputés pour 44 596 SNP autosomiques dans le cadre de l'évaluation génomique de routine des bovins en France (Boichard et al. 2012) La liste des marqueurs considérés est identique à celle présentée dans le tableau Supplémentaire ST2 de Besnard et al. (2023, <https://www.biorxiv.org/content/10.1101/2023.09.22.558782v1>). Ces génotypes ont été obtenus initialement à partir de différentes puces à SNP commercialisées par l'entreprise Illumina (« LD » comprenant environ 7 000 marqueurs ; LD customs : de 10 000 à 20 000 marqueurs ; BovineSNP50, environ 50 000 ; EuroGMD, environ 63 000 ; et HD, environ 777 000). L'imputation est très précise, particulièrement quand le père est typé, ce qui est le cas ici.

- Les informations sur la généalogie, le sexe et l'historique de vie (date de naissance, durée de vie, cause de décès, dates d'insémination et identifiants des taureaux) de millions d'individus disponibles dans la base de données nationale bovine française.

- Les variants de séquence annotés (SNP, InDels et variations structurales) présents dans les génomes de 1869 bovins dont la séquence est alignée sur l'assemblage ARS-UCD1.2, et analysés dans le cadre de l'étude antérieure (Tableau ST14, Besnard *et al.* 2023). Cette population comprend 247 individus Holstein, 160 Montbéliarde et 118 Normande, ainsi que 1 344 animaux issus de plus de 70 autres races.

2.2. Recherche de distorsions de transmission d'haplotypes dans des familles de demi-sœurs de père

Un total de 1 726 familles de taureaux Holstein, 974 Montbéliarde et 479 Normande, comprenant de 50 à 14 704 filles typées, encore en vie à un an d'âge et réparties dans au moins cinq troupeaux, ont été incluses dans l'étude (moyenne et écart type = 398 +/- 642 filles par famille). Au sein de chaque famille, j'ai calculé le nombre de porteurs de chacun des deux haplotypes paternels pour des fenêtres coulissantes de 20 marqueurs (~1 Mb) le long du génome, en ignorant les fenêtres pour lesquelles les taureaux étaient homozygotes. Dans la descendance de chaque taureau hétérozygote, j'ai comparé à chaque position les comptages observés pour un haplotype donné et pour son haplotype alternatif aux comptages attendus en supposant un ratio de distribution de 50-50, en utilisant un test du χ^2 . J'ai conservé uniquement les haplotypes observés à l'état hétérozygote chez au moins dix taureaux et avec une valeur p inférieure à 0,01 après correction de Bonferroni (en tenant compte de 2 488 067 haplotypes testés : 939 326 haplotypes en race Holstein, 890 745 en Montbéliarde et 657 996 en Normande). Enfin pour chacun des *loci* dépassant le seuil de significativité j'ai retenu l'haplotype le plus significatif, dit haplotype « cime ».

2.3. Analyse du taux de conception et de la mortalité juvénile dans une population de validation

Dans un second temps j'ai déterminé le statut (non porteur, hétérozygote ou homozygote) à chacun des 179 haplotypes « cime » retenus pour l'ensemble des animaux génotypés à ma disposition (Annexe 2). Les taux de conception et de mortalité juvénile ont été calculés pour la descendance d'accouplements entre taureaux hétérozygotes et vaches non porteuses, d'une part, et la descendance d'accouplements entre parents non porteurs, d'autre part. Le nombre

d'observations par haplotype « cime » et type d'accouplement variait de 2 988 à 3 082 426 inséminations et de 2 481 à 607 472 veaux femelles pour l'étude des mortalités. J'ai appliqué un test du χ^2 pour comparer les proportions d'inséminations amenant à un veau et celles ayant échoué, ainsi que les proportions de femelles de la descendance mortes avant l'âge d'un an ou toujours en vie à ce moment-là, dans les deux types d'accouplements. Enfin, j'ai considéré comme « validés sur cette population indépendante » les haplotypes avec une valeur p ajustée selon la méthode de Bonferroni inférieure à 0,01 et un effet sur au moins un des deux caractères cohérent avec le déficit ou l'enrichissement détecté dans la recherche initiale de distorsion de transmission.

2.4. Recherche de variants candidats et prédiction de leurs effets phénotypiques

J'ai ensuite recherché des variants candidats pour 40 de ces *loci* validés. Pour ce faire j'ai calculé les corrélations R^2 entre le statut sur haplotypes de 1869 individus dont le génome a été séquencé dans le cadre d'une étude précédente (cf. Chapitre V) et leurs génotypes aux variants situés dans une région de 10 Mb centrée sur chaque haplotype. Afin de prioriser les meilleurs variants candidats parmi des dizaines de milliers de variants, j'ai retenu des variants situés dans la séquence codante, présentant un score de corrélation avec l'haplotype de $R^2 \geq 0,5$, prédits comme délétères et ne ségrégant que dans la race dans laquelle l'haplotype a été détecté. Cette dernière décision repose sur l'observation que les mutations causales sont le plus souvent race-spécifiques car survenues après la formation des races en raison des effets fondateurs considérables décrits dans le Chapitre I de ce manuscrit de thèse.

Seuls ceux prédits comme étant délétères ont été pris en compte (c'est-à-dire que j'ai considéré comme potentiellement délétères : les variants faux-sens avec un score SIFT $\leq 0,05$, les gains ou pertes de codon stop, les gains ou pertes de sites d'initiation de la transcription, les décalages du cadre de lecture, les insertions ou délétions d'un multiple de trois nucléotides dans le cadre de lecture, et les variants affectant les sites donneurs ou accepteurs d'épissage, et les variants structuraux entraînant la délétion de tout ou partie de la région codante d'un gène). Enfin, j'ai vérifié la cohérence entre les annotations des variants du génome bovin réalisées par Ensembl (release 110; <https://www.ensembl.org/info/docs/tools/vep/index.html>) et par le NCBI, disponible dans le navigateur génomique de l'UCSC (<http://genome.ucsc.edu/>). Cela m'a conduit à supprimer certains variants prédits délétères par l'outil VEP d'Ensembl mais se trouvant dans des régions non codantes d'après la seconde annotation (c'est-à-dire sans preuve

d'ARNm bovin exprimé ou de transcrits orthologues chez la souris ou l'Homme à la même position).

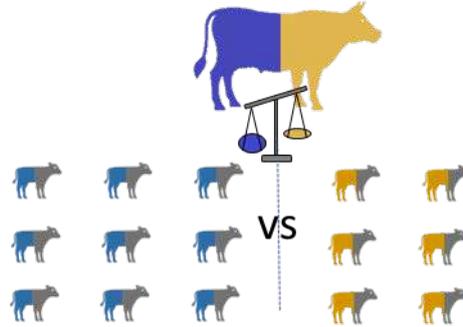
Deux variants ont fait l'objet d'analyses complémentaires. Afin d'étudier la conservation des nucléotides au niveau des sites de mutation Chr1 g.95275103G>A et Chr23 g.28371447G>A, j'ai récupéré les protéines orthologues "1-à-1" à partir des données brutes des orthologues disponibles sur le site de Ensembl en effectuant l'alignement à l'aide du programme Clustal (<https://www.ebi.ac.uk/jdispatcher/msa/clustalo>) et j'ai sélectionné la région de 21 acides aminés centrée sur le variant d'intérêt et généré une représentation logo à l'aide du programme WebLogo (Crooks et al. 2004; <http://weblogo.berkeley.edu>).

J'ai également exploité les informations sur les caractéristiques physiopathologiques associées aux mutations affectant les gènes orthologues chez l'Homme (Online Mendelian Inheritance in Man ; OMIM ; <https://www.omim.org>) et la souris (Mouse Genome Informatics ; MGI ; <https://www.informatics.jax.org>) pour essayer de prédire les conséquences chez le bovin des variants candidats retenus.

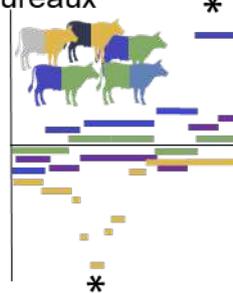
La procédure résumée est décrite dans la Figure VI.2 et les effectifs associés sont résumés dans le Tableau VI-1.

0. Analyse des données

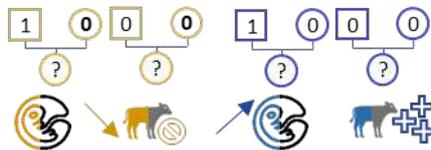
1. Calcul des transmissions d'un locus du taureau à sa descendance



2. Regroupement des loci pour l'ensemble des taureaux *



3. Validation sur les caractères par accouplement à risque



4. Recherche de variants candidats

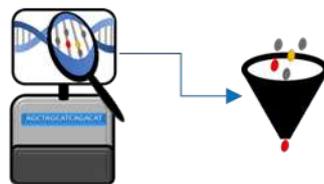


Figure VI-2 : Description de l'approche utilisée pour la détection et la validation d'haplotypes responsables de mortalité juvénile ou de mort embryonnaire.

Tableau VI-1 : Résumé des jeux de données étudiés à chaque étape de l'étude HSD et de leurs effectifs pour les trois races.

Etape	Nombre:	Holstein	Montbéliarde	Normande	Total
0	Animaux génotypés	748 263	397 200	115 090	1 260 553

0	Familles de taureaux étudiées	1 726	974	479	3 179
1	Haplotypes avec distorsion de ségrégation	939 326	890 745	657 996	2 488 067
2	Nombre d'haplotypes	90	66	23	179
3	Haplotypes avec un effet significatif sur le taux de conception	13	16	2	31
3	Haplotypes avec un effet significatif sur la mortalité	4	6	0	10
4	Séquences de génomes complets	247	160	118	1 869

3. Résultats

L'étude de 3179 familles de pères de trois races différentes a permis d'identifier un total de 179 *loci* (90 en Holstein, 66 en Montbéliarde et 23 en Normande) présentant une distorsion de transmission significative parmi les femelles vivantes à 1 an (Annexe 2, 179 haplotypes validés). Parmi eux, 165 haplotypes « cimes » affichaient un déficit de transmission ($0,49 \leq \text{NObs}/\text{NAtt} \leq 0,98$) avec NObs et NAtt les nombres respectivement observés et attendus de phases paternelles porteuses de l'haplotype étudié) tandis que 14 affichaient un excès de transmission par rapport à l'attendu ($1,01 \geq \text{NObs}/\text{NAtt} \geq 1,16$). Les haplotypes paternels sous-représentés et donc potentiellement associés à des variants génétiques entraînant une surmortalité étaient donc nettement majoritaires avec une proportion de 92,2%, globalement ou dans chaque race prise individuellement (95,6% en Holstein, 92,4 en Montbéliarde et 78,3 % en Normande) par rapport aux haplotypes paternels sur-représentés et qui favoriseraient la survie. Pour ces haplotypes significatifs le nombre de taureaux porteurs hétérozygotes présentant un déséquilibre varie de 10 à 483 taureaux. Dans la plupart des cas, la distorsion, même significative, est faible et nécessite donc validation.

Pour cela, nous avons comparé la mortalité juvénile dans les accouplements considérés comme à risque sur la base du statut haplotypique, 6 et 4 haplotypes ont un effet significatif sur la mortalité en Montbéliarde et en Holstein respectivement (Tableau VI-2). Sur les dix haplotypes validés, 9 ont un effet négatif avec des effets de 0,7 à 3,0% de mortalité en plus sur les accouplements à risque. Un seul haplotype réduit la mortalité juvénile pour la Montbéliarde, avec un effet positif de 1,7%. Il est intéressant de noter que cet haplotype est localisé à proximité immédiate de l'anomalie MTCP de la Montbéliarde, sur le chromosome 6 en position 70MB. Une

hypothèse plausible est que cet effet positif reflète simplement que cet haplotype est non porteur de MTCP. Un des haplotypes HSDM3P64 est validé à la fois pour le taux de conception et pour le taux de mortalité.

Puis nous avons comparé les taux de conception liés aux accouplements à risque en se basant sur les statuts des haplotypes « cime » dans la population de validation. Nous avons trouvé 13, 16 et 2 haplotypes significatifs dans les races Holstein, Montbéliarde et Normande respectivement, avec 29 haplotypes ayant un effet négatif faible à fort (entre -1% et -10% d'IA réussies, cf. Tableau VI-3). Seuls deux haplotypes ont un effet positif sur les taux de conception, notamment en Holstein et en Montbéliarde, avec des effets positifs de 0,8% et 1,4% respectivement. Cela correspond à la proportion des haplotypes trouvés en déficit dans la descendance des taureaux, avec une proportion de 93% toutes races confondues.

Tableau VI-2 : Résultats significatifs des taux de mortalité liés aux haplotypes « cime » dans la population de validation.

Le tableau présente les résultats significatifs de l'analyse des taux de mortalité associés aux 179 haplotypes « cime » dans la population de validation. La catégorie « A risque » correspond aux conceptus issus de taureaux hétérozygotes et de vaches non porteuses ; la catégorie « Contrôle » à ceux issus de parents non porteurs. La p-value corrigée est ajustée pour le nombre de tests selon la méthode de Bonferroni.

Race	Identifiant de l'haplotype	Taux de mortalité		Nombre de veaux nés		Veaux morts avant l'âge d'un an		Veaux en vie à l'âge d'un an		p-value		Effet	Différence
		Contrôle	A risque	Contrôle	A risque	Contrôle	A risque	Contrôle	A risque	Brute	Corrigée		
Montbéliarde	HSDM2P61	10,12%	11,04%	160065	29091	16196	3212	143869	25879	2,28E-06	4,08E-04	Positif	-0,92%
Montbéliarde	HSDM9P90	10,19%	11,10%	178804	29247	18223	3246	160581	26001	2,78E-06	4,97E-04	Positif	-0,91%
Montbéliarde	HSDM24P12	10,25%	11,39%	224984	21195	23051	2415	201933	18780	2,22E-07	3,97E-05	Positif	-1,15%
Montbéliarde	HSDM6P70	10,53%	8,83%	252335	20790	26580	1835	225755	18955	2,58E-15	4,62E-13	Négatif	1,70%
Montbéliarde	HSDM27P22	10,31%	12,07%	259223	9580	26723	1157	232500	8423	4,70E-08	8,41E-06	Positif	-1,76%
Montbéliarde	HSDM3P64	10,34%	13,36%	271192	3421	28034	457	243158	2964	3,04E-08	5,45E-06	Positif	-3,02%
Holstein	HSDH7P68	10,99%	11,66%	525473	43510	57741	5072	467732	38438	2,13E-05	3,82E-03	Positif	-0,67%
Holstein	HSDH13P79	10,98%	12,16%	547121	28090	60078	3416	487043	24674	1,26E-09	2,25E-07	Positif	-1,18%
Holstein	HSDH5P22	10,98%	11,93%	549398	40541	60310	4835	489088	35706	5,82E-09	1,04E-06	Positif	-0,95%
Holstein	HSDH8P105	11,01%	14,07%	607472	2481	66894	349	540578	2132	2,67E-06	4,78E-04	Positif	-3,06%

Tableau VI-3 : Résultats significatifs des taux de conception liés aux haplotypes « cime » dans la population de validation.

Ce tableau présente les résultats significatifs de l'analyse des taux de conception associés aux 179 haplotypes « cime » dans la population de validation. . La catégorie « A risque » correspond aux inséminations entre taureaux hétérozygotes et vaches non porteuses ; la catégorie « Contrôle » à celles entre parents non porteurs. La p-value corrigée est ajustée pour le nombre de tests selon la méthode de Bonferroni

Race	Identifiant de l'haplotype	Taux de conception		Nombre d'accouplements		Réussite à l'IA		Echec de l'IA		p-value		Effet	Différence
		Contrôle	A risque	Contrôle	A risque	Contrôle	A risque	Contrôle	A risque	Brute	Corrigée		
Holstein	HSDH4P107	0,510	0,474	3082426	16071	1573410	7614	1509016	8457	1,78E-20	3,19E-18	Négatif	0,037
Holstein	HSDH2P5	0,510	0,491	3062332	14053	1563276	6895	1499056	7158	2,68E-06	4,80E-04	Négatif	0,020
Holstein	HSDH16P44	0,510	0,494	3054617	26948	1559268	13309	1495349	13639	6,08E-08	1,09E-05	Négatif	0,017
Holstein	HSDH6P37	0,511	0,499	3024726	44764	1544228	22316	1480498	22448	4,57E-07	8,18E-05	Négatif	0,012
Holstein	HSDH15P77	0,511	0,499	2921825	150838	1492577	75203	1429248	75635	1,49E-20	2,66E-18	Négatif	0,012
Holstein	HSDH16P54	0,511	0,501	2850187	96283	1455770	48211	1394417	48072	8,85E-10	1,58E-07	Négatif	0,010
Holstein	HSDH13P15	0,511	0,501	2774819	233533	1417187	117097	1357632	116436	5,22E-18	9,35E-16	Négatif	0,009
Holstein	HSDH5P43	0,510	0,503	2690788	162775	1371137	81892	1319651	80883	4,03E-07	7,21E-05	Négatif	0,006
Holstein	HSDH12P23	0,511	0,505	2447323	408712	1251016	206411	1196307	202301	3,35E-13	6,00E-11	Négatif	0,006
Holstein	HSDH29P14	0,511	0,507	2325083	476473	1187197	241458	1137886	235015	1,34E-06	2,40E-04	Négatif	0,004
Holstein	HSDH6P110	0,511	0,507	2176928	590828	1112151	299430	1064777	291398	2,57E-08	4,61E-06	Négatif	0,004
Holstein	HSDH1P98	0,512	0,503	2169436	390072	1110121	196229	1059315	193843	2,52E-23	4,51E-21	Négatif	0,009
Holstein	HSDH2P128	0,510	0,519	2025607	438920	1033830	227671	991777	211249	1,44E-23	2,57E-21	Positif	0,008

Montbéliarde	HSDM4P44	0,562	0,466	1189097	2988	668083	1391	521014	1597	5,97E-26	1,07E-23	Négatif	0,096
Montbéliarde	HSDM28P3	0,562	0,537	1182813	8966	664494	4819	518319	4147	3,92E-06	7,02E-04	Négatif	0,024
Montbéliarde	HSDM11P43	0,562	0,530	1174999	23620	660715	12507	514284	11113	1,06E-23	1,89E-21	Négatif	0,033
Montbéliarde	HSDM1P91	0,562	0,533	1162062	11197	653060	5965	509002	5232	5,91E-10	1,06E-07	Négatif	0,029
Montbéliarde	HSDM7P48	0,562	0,542	1160540	17460	652379	9457	508161	8003	6,31E-08	1,13E-05	Négatif	0,020
Montbéliarde	HSDM3P64	0,562	0,533	1158438	14651	651318	7813	507120	6838	2,45E-12	4,39E-10	Négatif	0,029
Montbéliarde	HSDM2P118	0,562	0,541	1149870	19432	646461	10517	503409	8915	5,45E-09	9,76E-07	Négatif	0,021
Montbéliarde	HSDM7P68	0,563	0,538	1123063	52400	632196	28184	490867	24216	1,65E-29	2,96E-27	Négatif	0,025
Montbéliarde	HSDM23P33	0,563	0,538	1119704	65696	630505	35329	489199	30367	6,20E-37	1,11E-34	Négatif	0,025
Montbéliarde	HSDM26P25	0,562	0,549	1103363	52667	619952	28897	483411	23770	2,60E-09	4,66E-07	Négatif	0,013
Montbéliarde	HSDM1P54	0,562	0,553	1085765	48999	610403	27076	475362	21923	2,81E-05	5,02E-03	Négatif	0,010
Montbéliarde	HSDM7P39	0,562	0,553	1061151	68693	596293	37981	464858	30712	3,96E-06	7,08E-04	Négatif	0,009
Montbéliarde	HSDM5P48	0,560	0,574	1005857	86767	563374	49792	442483	36975	4,26E-15	7,63E-13	Positif	0,014
Montbéliarde	HSDM11P61	0,563	0,556	969217	133450	545934	74160	423283	59290	1,84E-07	3,30E-05	Négatif	0,008
Montbéliarde	HSDM20P24	0,562	0,556	924198	179594	519630	99927	404568	79667	4,99E-06	8,92E-04	Négatif	0,006
Montbéliarde	HSDM4P54	0,563	0,555	898743	159856	506379	88772	392364	71084	1,79E-09	3,20E-07	Négatif	0,008
Normande	HSDN24P30	0,522	0,506	459886	23923	240012	12114	219874	11809	2,87E-06	5,15E-04	Négatif	0,016
Normande	HSDN14P5	0,523	0,511	343675	60584	179879	30952	163796	29632	1,38E-08	2,46E-06	Négatif	0,013

Pour ces 40 haplotypes confirmés, nous avons cherché des variants candidats corrélés aux 1 869 séquences à disposition. Après avoir effectué les différents filtres décrits, seulement deux de ces haplotypes ont des variants candidats corrélés dans notre base de données WGS, tous deux en race Montbéliarde (Tableau VI-4). Les deux variants trouvés en position Chr1 g.95275103G>A et Chr23 g.28371447G>A ont des scores de corrélation R^2 de 1 avec les haplotypes.

Le premier haplotype étudié cause un déficit de 2,5% de taux de conception. Le variant candidat est situé sur le chromosome 23 en position 28371447 paires de bases et correspond à la substitution d'une guanine (G) par une adénine (A) au niveau de l'ADN et d'une proline par une leucine au niveau du 96^{ème} acide aminé de la protéine MRPS18B (p.96P>L). Il possède un score SIFT très délétère de 0,01 suggérant qu'il impacte fortement la fonction de la protéine. Néanmoins, le Weblogo construit à partir de l'alignement des différents orthologues avec Clustal ne montre qu'une conservation partielle de la proline à cette position (Figure VI.3). En effet, différentes espèces, et notamment l'alpaga, sont aussi porteuses d'une leucine dans leur protéine à la même position. Ainsi, ce variant peut probablement être éliminé du fait de sa conservation seulement partielle chez des espèces plus ou moins proches du bovin.

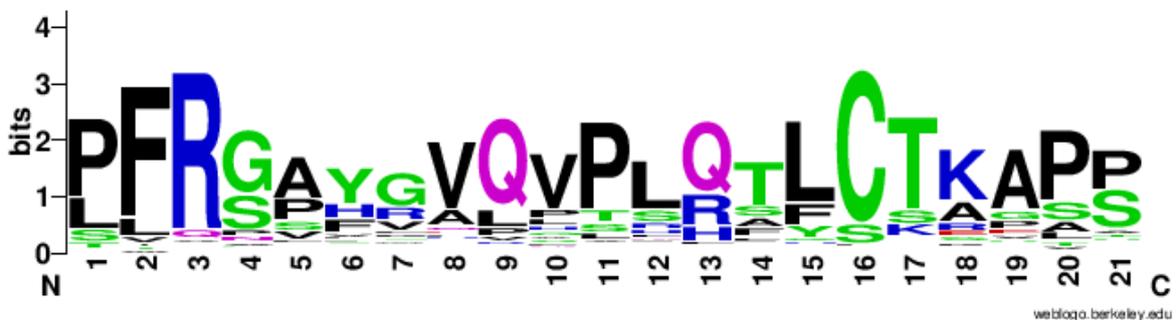


Figure VI-3 : Weblogo montrant la conservation de la séquence protéique de MRPS18B autour de la substitution p.96P>L +/- 10 acides aminés.

Cet alignement a été réalisé avec les protéines orthologues de 170 espèces disponibles dans Ensembl.

Le deuxième variant, corrélé avec l'haplotype HSDM1P91 qui a un effet validé sur les taux de conception de 3%, est situé sur le chromosome 1 en position 95 275 103. Il correspond à une substitution de G par A dans le gène *FNDC3B*. Ce SNP induit un changement de l'acide aminé proline en position 125 de la séquence protéique par une leucine (p.12P>L). Ce variant a un score SIFT de 0,01, modifiant donc un acide aminé très conservé chez différentes espèces plus ou moins proches génétiquement.

Pour l'ensemble des 195 espèces avec un gène orthologue 1-to-1 avec le gène *FNDC3B* humain, toutes possèdent une proline à cette position (Figure VI.4 en position 11). Cette conservation complète suggère fortement un impact délétère d'une mutation dans cette protéine et à cette position. Kishimoto *et al.* (2010) ont montré que la protéine avait un effet dans la différenciation des adipocytes. Le gène a un impact sur la régulation de l'expression de certains facteurs de croissance et les KO chez les souris induisent de la mortalité à la naissance avec des anomalies pulmonaires (Nishizuka *et al.* 2009) ainsi qu'une ossification prématurée de l'os calvarial de type craniosténose (Kishimoto *et al.* 2013 ; [OMIM](#), [MGI](#)).



Figure VI-4 : Weblogo montrant la conservation de la séquence protéique de FNDC3B autour de la substitution p.125P>L +/-10 acides aminés.

Cet alignement a été réalisé avec les protéines orthologues de 195 espèces disponibles dans Ensembl

Tableau VI-4 : Informations sur les variants candidats et leurs conséquences phénotypiques présumées.

Les chromosomes (Chr) et les positions font référence à l'assemblage du génome bovin ARS-UCD1.2. R² : Corrélacion entre l'haplotype intra-race et le génotype basée sur des individus séquencés à l'échelle du génome entier. Ref et Alt : allèles de référence et alternatif, respectivement. Pour chaque variant candidat, un résumé des annotations obtenues à l'aide du Variant Effect Predictor (Ensembl release 110 ; <https://www.ensembl.org/info/docs/tools/vep/index.html>) est fourni.

RACE	Haplotype ID	Variant candidat	Chr	Position	Ref	Alt	R ²	Ensembl transcript ID	Protéine	Conséquence	Catégorie	Score SIFT	Taille de la protéine
Montbéliarde	HSDM1P91	Chr1 g.95275103G>A	1	95275103	G	A	1	ENSBTAT00000025816.7	FNDC3B p.125P>L	missense_variant	MODERATE	0,01	1207aa
Montbéliarde	HSDM23P33	Chr23 g.28371447G>A	23	28371447	G	A	1	ENSBTAT00000009107.5	MRPS18B p.96P>L	missense_variant	MODERATE	0,01	326aa

4. Discussion

Ainsi, cette méthode n'a permis d'identifier qu'un seul variant candidat passant tous les filtres mis en place. Afin de détecter plus de variants candidats il faudrait rechercher les variants structuraux ainsi que les variants non délétères, notamment pour les haplotypes ayant un effet positif. En effet, des haplotypes ayant un déséquilibre de transmission ne sont pas à pénétrance complète, les effets des variants candidats ne sont donc pas systématiquement annotés délétères pour le codant, entraînant des modifications moins impactantes pour la protéine. Par ailleurs, nous avons ignoré les variants candidats potentiels hors séquences codantes, du fait d'une annotation insuffisante ou absente.

La mutation dans *FNCD3B* pourrait avoir un effet sur la croissance de l'embryon en dérégulant l'expression de facteurs de croissance entraînant des létalités fœtales, expliquant la différence entre les taux de conception pour les inséminations artificielles contrôle et à risque.

De plus, nous avons trouvé que la méthode de détection était très sensible aux erreurs de génotypage, d'imputation et de phasage. Certains SNP peuvent être transmis entre 70% et 90% dans la descendance d'un taureau alors que les haplotypes qui les entourent sont transmis de façon équilibrée. Différentes causes sont possibles, notamment l'imputation des puces à faible densité sur la puce 50K, le phasage des régions à fort taux d'homozygotie entraînant des difficultés pour associer les phases aux parents. Enfin, nous avons constaté que pour une trentaine de SNP dont la MAF est proche de 50%, les allèles ont été inversés entre les typages les plus anciens (2008) et les suivants. Ces échanges ont un impact important car les typages concernent les premières populations de référence constituées de taureaux d'IA (voir GWAS Chapitre IV).

Il est à noter que l'accumulation des résultats entre familles est également une source de perte de puissance. En effet, il est possible que le même haplotype de marqueurs ne porte pas toujours le variant délétère chez tous les taureaux, surtout si cet haplotype est fréquent.

L'exploration des haplotypes par famille de taureaux permet d'éliminer la majorité des faux positifs en sélectionnant seulement des haplotypes communs avec un déséquilibre pour différents taureaux. Une autre possibilité est d'avoir une étape intermédiaire de sélection des SNP qui n'ont pas de déséquilibre de transmission. En enlevant les SNP ayant une transmission trop déséquilibrée dans la descendance de multiples taureaux non apparentés, nous réduisons les erreurs. Les SNP étant choisis initialement pour différencier les individus dans le maximum de races, la probabilité d'enlever des SNP avec un fort DL avec une mutation causale est très

réduite. En effectuant cette étape, nous pouvons considérer que la majorité des haplotypes trouvés en déséquilibre significatif de transmission chez les taureaux vont être réellement associés à des réalités biologiques et non à des erreurs de génotypage sur un SNP.

Ainsi, pour perfectionner la méthode, nous proposons d'utiliser ces deux étapes, et en parallèle d'avoir une étape de sélection de grandes familles de taureaux apparentés. Ces groupes de familles sont à constituer sur la base du pedigree et les déséquilibres de transmission sont recherchés indépendamment dans chaque groupe. Ces lignées devront avoir l'ancêtre commun séquencé pour espérer découvrir une mutation candidate. Pour chacun des haplotypes transmis de l'ancêtre aux plus jeunes taureaux, un déséquilibre de transmission sera observé s'il est relié à une mutation à pénétrance incomplète. Certains taureaux deviennent des pères à taureaux et peuvent avoir des milliers voire des dizaines de milliers de descendants, notamment en Holstein. Il sera intéressant d'étudier indépendamment ces lignées, et de le faire génération par génération pour éviter de considérer ensemble les haplotypes à la fois porteurs et non porteurs du variant causal. Citons par exemple *ITGB7*, dont l'haplotype a une fréquence très élevée mais dont seule une faible partie est porteuse de la mutation (cf Chapitre V).

Données supplémentaires

Annexe 2: Résultats de l'analyse génomique pour la distorsion de ségrégation des haplotypes dans 1 726 familles de taureaux Holstein, 974 familles de Montbéliarde et 479 familles de Normande.

ID_1er_mrq : identifiant interne du premier marqueur de l'haplotype (Voir ST1 de la publication de (Florian Besnard et *al.* 2023) pour la corrélation avec l'ID Illumina). Pos_1er_mrq : position du premier marqueur de l'haplotype en Mb. Nb_htz_pere : nombre de taureaux porteurs hétérozygotes. NObs et NAtt : nombres observés et attendus de descendants porteurs de la phase paternel. P-value : pour un haplotype donné, valeur p du test du Chi2 d'équilibre de transmission intra père. Adj_p-value : valeur p ajustée selon la méthode de Bonferroni. Seuls les haplotypes observés à l'état hétérozygote chez au moins dix taureaux et avec une valeur p inférieure à 0,01 après correction de Bonferroni ont été conservés (en tenant compte de 2 488 067 haplotypes testés).

Chapitre VII. Développement d'une évaluation génomique sur la mortalité juvénile

1. Introduction

Les approches précédentes ont cherché à détecter des variants génétiques responsables de mortalité. On peut aussi réaliser une analyse globale qui permet d'estimer la part d'origine génétique dans la variabilité de l'expression du caractère, avec une approche multi-marqueurs pour prendre en compte les régions du génome les plus importantes. C'est l'objectif de ce chapitre qui aborde la construction du modèle d'évaluation génomique de type Single Step et ses résultats.

La construction d'une évaluation génétique (ou génomique) implique un certain nombre de choix opérationnels : définition du ou des caractères évalués ; méthode statistique ; sources d'information utilisées ; facteurs de variation du modèle. Ce chapitre développe ces différents aspects.

Il existe une bibliographie sur le sujet, que nous présentons ici. Le premier point concerne la définition des caractères concernés. On peut définir la mortalité globale comme une variable binaire mort/vivant sur une période d'âge donnée, entre zéro et un an par exemple. Cela présente l'avantage statistique d'augmenter l'incidence des veaux morts et donc l'informativité des données, mais l'inconvénient de mélanger des causes de mortalité très différentes et donc des déterminismes génétiques différents. A l'inverse, on peut considérer que les mortalités à des stades différents constituent autant de caractères différents : le déterminisme de chaque caractère est plus homogène, l'interprétation biologique est sans doute plus aisée, mais l'incidence à chaque période est plus faible. Il est généralement admis que la mortalité à la naissance constitue un caractère à part, fortement dépendant des conditions de naissance, et donc avec des causes très différentes des mortalités ultérieures. Pour la période 3 jours – un an, les arguments varient entre auteurs, selon leur fibre biologique ou statistique.

Pedersen *et al.* (2015) définissent les caractères au travers des périodes spécifiques de mortalité, différenciant les deux sexes et les phases de la croissance, notamment de 1 à 30 jours et de 31 à 458 jours pour les génisses. Mais la plupart des auteurs (van Pelt *et al.* (2012) aux Pays-Bas,

Gonzalez-Peña et al. (2019) et Neupane et al. 2021 aux Etats-Unis, Weller et al. (2021) en Israël) considèrent un seul caractère.

Même si les causes de mortalité varient, il est intéressant de considérer la survie globale qui mesure la résilience de l'animal et son aptitude à surmonter les aléas. Le choix peut également dépendre de la taille des troupeaux : la précision des évaluations va dépendre du nombre de décès dans le groupe de contemporains. Avec des élevages de taille réduite en France, le découpage par élevage peut s'avérer très consommateur d'information et induire une forte baisse de précision.

Un deuxième aspect concerne la distribution binaire du caractère mortalité. Les hypothèses usuelles du modèle linéaire sont l'indépendance et la normalité des résidus du modèle et, plus généralement, des différents effets aléatoires. Avec un caractère binaire, ces hypothèses ne sont pas respectées, et le sont d'autant moins que l'incidence est faible. En effet, avec une incidence p , la variance phénotypique est $p(1-p)$ et varie donc en fonction de p et ne peut être considérée comme relativement constante que lorsque p s'écarte des extrêmes 0 et 1. Le modèle à seuil (par ex : Foulley et al. 1983) est souvent proposé comme une alternative adaptée. Il suppose l'existence d'une variable sous-jacente non observée, normalement distribuée, et l'existence d'un seuil (si deux catégories, $n-1$ seuils si n catégories). La variable sous-jacente est soumise au modèle linéaire avec ses effets fixés et aléatoires tandis que le phénotype est 0 ou 1 selon que la variable est inférieure ou supérieure au seuil. En pratique, différentes études ont comparé modèle linéaire et modèle à seuil pour des caractères binaires (Boichard et al. 1994, sur la fertilité; Jamrozik et al. 2005; Lynch et al. 2024). Les résultats d'index sont très proches entre les deux types de modèles. Il est souvent montré que l'héritabilité de la variable sous-jacente est supérieure à celle de la variable binaire, et ce d'autant plus que l'incidence est faible. Sur la population danoise, Fuerst-Waltl et Sørensen (2010) ont trouvé des héritabilités de 0,042 et 0,076 respectivement avec les modèles linéaires et à seuil pour la période de la naissance au vêlage. Dempster et Lerner (1950) ont montré que le rapport d'héritabilité est de $z^2 / p(1-p)$ avec z l'ordonnée de la fonction de densité normale au seuil définissant les deux catégories. Toutefois, la variable sous-jacente n'est pas observée et la précision des index, au final, est souvent peu différente. D'un point de vue calculatoire, le modèle à seuil est non linéaire et il est donc plus complexe à implémenter. Sa convergence n'est pas totalement assurée, en particulier si l'incidence est nulle ou très faible pour certains niveaux de facteur. Compte tenu de ces inconvénients, le modèle à seuil est rarement utilisé en routine dans des modèles d'évaluation dans de grandes populations. L'évaluation génétique de la mortalité en France était basée sur un modèle à seuil incluant les effets génétiques du père et du grand-père maternel.

L'évolution vers le Single Step s'est aussi traduit par passage au modèle linéaire, avec effets directs et maternels corrélés.

L'évaluation génétique requiert la connaissance des paramètres génétiques. Les coefficients d'héritabilité sont toujours faibles dans la littérature avec des périodes étudiées hétérogènes. Pour la période 3 à 365 jours, van Pelt et al. (2012) estiment l'héritabilité à 0,011. L'héritabilité aux différentes périodes semble encore plus faible (0,006 entre 3 et 14 jours, 0,005 entre 15 et 180 jours), probablement du fait de niveaux de mortalité plus réduits. Gonzalez-Peña et al. (2019) ont estimé une héritabilité de 0,06 pour le phénotype de mortalité de 2 à 365 jours chez les veaux Holstein américains. En Israël Weller et al. (2021) ont estimé une héritabilité de 0,007 de la survie des génisses à 305 jours. Neupane et al. (2021) ont trouvé une héritabilité de 0,004 pour la survie à 18 mois aux Etats-Unis. En dépit de cette faible héritabilité, ils estiment un écart type génétique de 2,5 à 3% pour la période entre 3 à 365 jours. Compte tenu du niveau moyen de mortalité, le coefficient de variation génétique est très important.

En conclusion, l'héritabilité est généralement très faible. La variabilité génétique est cependant importante et offre le potentiel d'une réduction sensible des mortalités. Un modèle à seuil est en théorie meilleur mais il est plus complexe à mettre en œuvre en routine, pour un gain final réduit, c'est pourquoi nous avons choisi d'étudier un modèle linéaire.

2. Matériel, Méthodes et Résultats intermédiaires.

La mise en place d'évaluations nécessitant plusieurs étapes d'exploration des données, j'ai choisi ici de présenter les résultats intermédiaires en même temps que le matériel et méthodes pour gagner en clarté. Pour respecter la chronologie des travaux et clarifier la lecture la partie suivante introduira des améliorations potentielles, notamment grâce à la prise en compte de « *Runs of Homozygosity* » (ROH).

2.1. Choix des périodes

Pour le choix du modèle et l'estimation des paramètres génétiques dans les trois principales races laitières françaises, quatre périodes ont été étudiées dans ma thèse, en m'appuyant sur les connaissances biologiques sur les causes principales de mortalité juvénile à un âge donné

(voir Chapitre I) et sur les résultats statistiques des taux de mortalité à un âge donné, illustrés dans la Figure VII-1 avec le cas de la race Holstein. Ces résultats ont été obtenus à partir des dates de naissance et de morts déclarés par les éleveurs dans le cadre de l'IPG et enregistrés dans le SIG, pour tous les veaux nés en France entre 2015 et 2022 et avec une filiation certifiée.

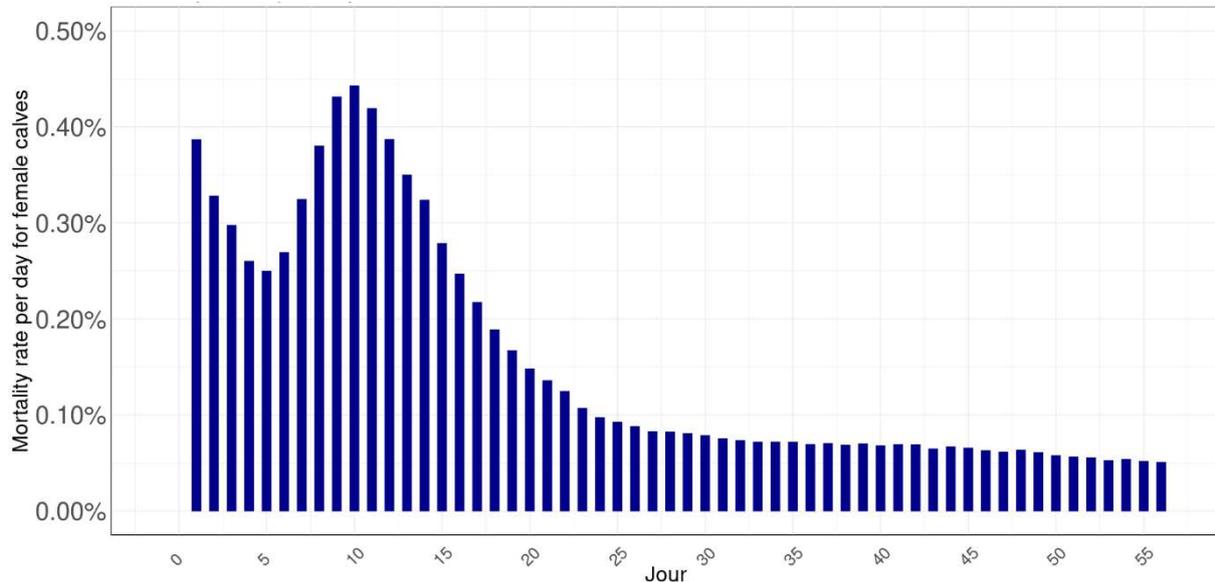


Figure VII-1 : Histogramme de la distribution des taux de mortalité entre les jours 1 et 56 en race Holstein.

La première période entre 0 et 2 jours caractérise la vitalité à la naissance des veaux (ou, à l'inverse, les mortalités liées aux dystocies ou à une faiblesse intrinsèque du veau). Comme évoqué dans le chapitre I, deux caractères (la vitalité à la naissance du veau dénommée VIN et la capacité de la mère à assurer la survie de son veau dénommée VIV pour vitalité au vêlage) sont déjà évalués en routine dans les races laitières depuis août 2008. Les évaluations pour ces deux caractères sont réalisées avec les mêmes données et un seul modèle incluant à la fois les effets génétiques directs et maternels, potentiellement corrélés. En 2022, le modèle a évolué vers le Single Step modèle animal.

La seconde période s'étend de 3 à 30 jours, et correspond au pic de mortalité lié en majorité à des maladies infectieuses de types diarrhées. La sélection contre la mortalité à ce stade ciblera donc l'augmentation des résistances ou la diminution des susceptibilités aux pathogènes les plus fréquents, notamment la cryptosporidiose et les rotavirus. La troisième période 31 à 365 jours permet de cibler la résistance des animaux aux maladies infectieuses respiratoires. Enfin,

une période globale de 3 jours à 1 an de vie permet de considérer une résistance globale aux infections.

Nous avons donc choisi de considérer les périodes 0 à 2 jours, 3 à 30 jours, 31 à 365 jours et 3 à 365 jours pour caractériser les différentes causes de mortalité et obtenir une évaluation globale de la mortalité juvénile dans les trois principales races laitières françaises. Les périodes sont donc différentes de celles étudiées dans les chapitres précédents (0-2, 3-14, 15-55 et 56-365 jours) et leur nombre a été réduit. En effet, comme nous le verrons dans cette partie, le découpage en périodes sur un caractère à faible incidence en plus de l'utilisation conjointe d'un effet fixe basé sur les groupes de contemporains est très consommateur d'information. Pour obtenir une meilleure prédiction, nous avons donc choisi la période globale et deux périodes sous-jacentes.

2.2. Choix des facteurs de variation pour le modèle d'évaluation génétique

Nous supposons que les facteurs ayant le plus d'impact sur la mortalité sont les mêmes pour les trois périodes choisies et pour les trois races laitières, pour simplifier la compréhension du modèle et sa mise en œuvre. Comme indiqué précédemment (Chapitre II. Exploration phénotypique de la mortalité), les données des veaux mâles sont difficiles à interpréter et nous ne considérons donc que les données des veaux femelles. Les veaux nés jumeaux ne sont pas considérés dans cette analyse, même si cette option pourra être reconsidérée.

Les facteurs pris en compte sont les suivants :

- La parité de la mère, les veaux issus de nullipares ayant en moyenne des conditions de mise bas plus difficiles et une mortalité plus élevée dans les premiers jours suivant la naissance (Figure VII.2).

Mortality rates per day for female calves born between 2015 and 2021, 1 is for heifers and 2 for multiparous cow

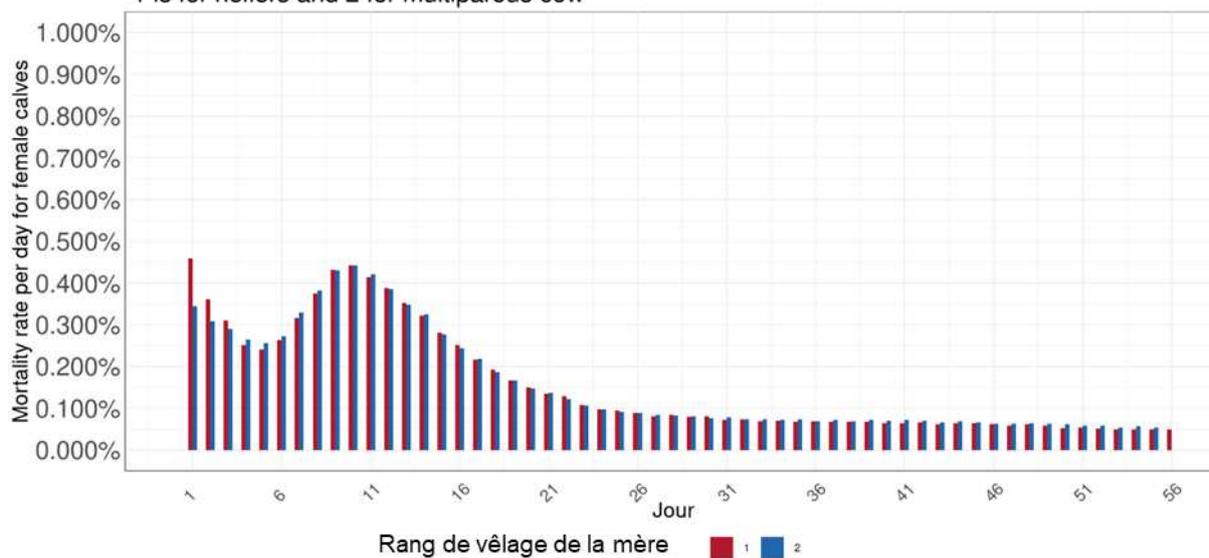


Figure VII-2 : Histogramme montrant la différence des taux de mortalité entre les primipares et les vaches pour les veaux femelles nés entre 2015 et 2021 de la race Holstein.

Rang de vêlage de la mère : la valeur 1 correspond aux génisses et la valeur 2 aux vaches.

- Les conditions de naissance. Comme le montre la Figure VII-3, les conditions de naissance difficiles augmentent encore le taux de mortalité, même au-delà du 2^{ème} jour de vie. Cet effet est encore visible durant tout le premier mois, même s'il tend à s'estomper (Figure VII.3). Trois groupes sont considérés, les vêlages sans assistance (1), les vêlages avec faible assistance (2) et les vêlages avec assistance difficile et césariennes (3). On considère donc ici les conditions de naissance comme un facteur non génétique, ce qui constitue une approximation qui nous semble acceptable.

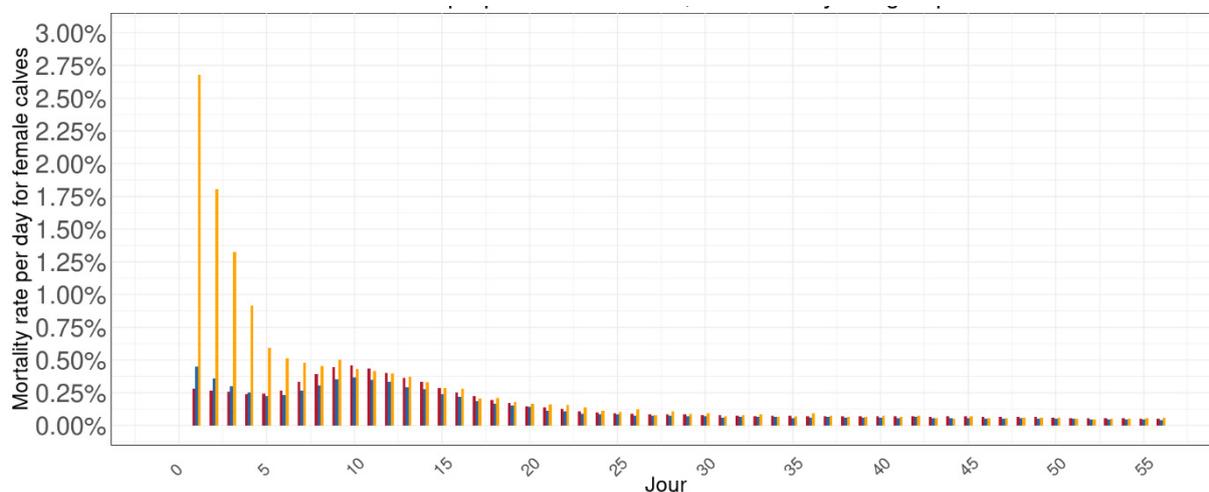


Figure VII-3 : Histogramme montrant les taux de mortalité journaliers pour les veaux femelles nés entre 2015 et 2021 de la race Holstein en fonction de leurs conditions de naissance.

Sont représentés en rouge les vêlages sans assistance (code 1), en bleu ceux avec assistance légère (code 2) et en jaune le regroupement des vêlages difficile avec assistance (code 3) et des césariennes (code 4, qui sont très peu nombreuses).

- Enfin, tout modèle génétique suppose une comparaison de performances dans un groupe homogène dit groupe de contemporains. Nous avons choisi la combinaison troupeau x année x saison de naissance, avec les quatre saisons, « printemps » « été » « automne » et « hiver » (cf Chapitre II., Figure II.3 : Histogramme montrant les différences de mortalités pour les veaux femelles selon la saison de naissance pour les trois races sur la période de 0 à 365 jours). Cet effet capte la variabilité entre différentes années et les saisons dans un même troupeau ainsi que les différences de mortalité entre élevages traduisant différents modes de conduite et niveaux de maîtrise sanitaire. Notons que cet effet prend aussi en compte des différences de mortalités en fonction de la taille de l'élevage, le taux de mortalité augmentant en général de concert (cf. Figure VII.4). Le fait de mettre l'élevage en effet fixe réduit fortement la variabilité étudiée mais elle permet d'éliminer les biais d'échantillonnage et les biais d'effets environnementaux qui ont lieu sur les exploitations. C'est la méthode utilisée en France dans les évaluations en routine et elle est utilisable exclusivement lorsque l'on travaille avec beaucoup de données et avec la structure de population bovine actuelle, qui bénéficie d'une bonne connexion entre troupeaux grâce à un fort taux de pénétration de l'insémination animale dans les élevages laitiers. Les reproducteurs mâles sont distribués dans différentes exploitations, ce qui permet de différencier correctement la variabilité liée à la génétique du reproducteur de la variabilité entre élevages.

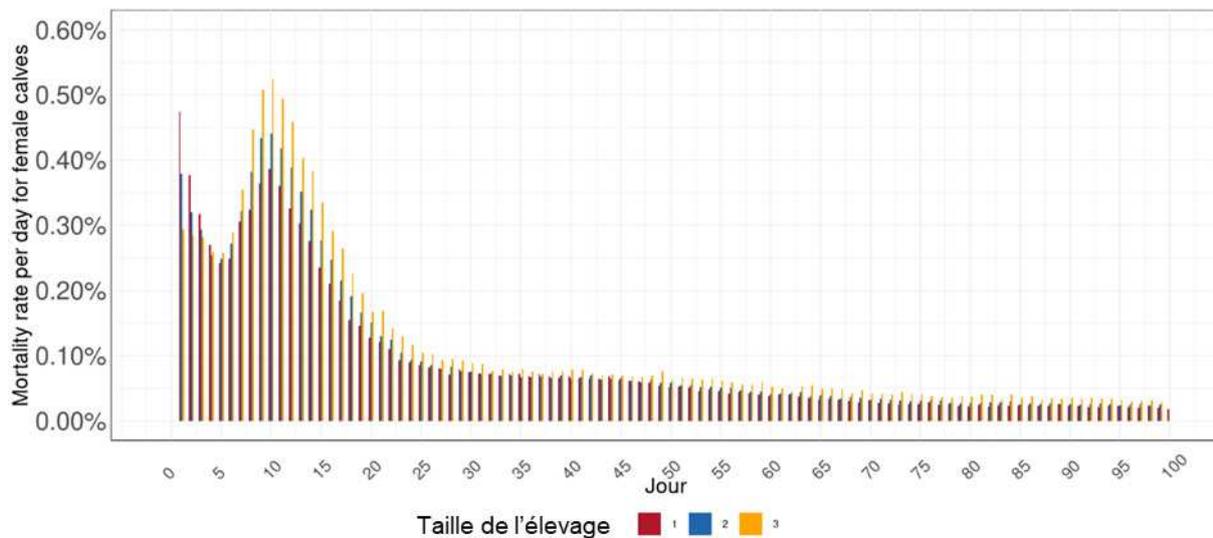


Figure VII-4 : Histogramme montrant les taux de mortalité journaliers pour les veaux femelles nés entre 2015 et 2021 de la race Holstein en fonction de la taille des élevages.

Sont représentés en rouge les plus petits élevages avec moins de 20 naissances femelles par an, en bleu ceux avec 20 à 45 femelles nées par an et en jaune ceux au-dessus.

2.3. Paramètres génétiques

Les valeurs génétiques sont par nature des effets aléatoires, de variance égale à la variance génétique et de covariance proportionnelle à l'apparentement entre individus. Le modèle d'évaluation suppose donc la connaissance des paramètres génétiques.

Pour estimer précisément une variabilité génétique a priori très faible, il est nécessaire d'analyser un très grand jeu de données. Le modèle idéal est un modèle animal, mais il est difficile à mettre en œuvre sur des millions de données. Dans le cas présent, un modèle père est une bonne approximation. Dans un modèle père, les mères sont supposées non apparentées entre elles et avec les pères, et l'effet du père est le seul effet génétique apparaissant dans le modèle. Les effets de la mère et de l'aléa de méiose du veau sont regroupés et contribuent à la variance résiduelle. Ce modèle estime la variance père, supposée égale au quart de la variance génétique, et la variance résiduelle. Au-delà de 3 jours, l'effet maternel est supposé dilué et n'est pas considéré dans le modèle.

L'étude a porté sur l'ensemble des génisses Holstein, Montbéliards et Normands nés de 2015 à 2022 disposant d'une parenté certifiée et avec un père connu. Nous avons, pour chaque caractère, sélectionné les groupes de contemporains (Troupeau x Année x Saison) pour lesquels nous disposons d'au moins 3 animaux avec performances. Seuls les veaux pour lesquels les

pères avaient des informations pour au moins 100 veaux dans 5 élevages différents ont été considérés. Après sélection, les analyses ont porté sur 558 462 génisses Montbéliarde, 242 840 Normande et 2 436 791 Holstein (caractère 3-365 jours).

Tableau VII-1 : Estimation des paramètres génétiques dans les races Montbéliarde, Normande et Holstein.

Colonne 2 : variance (écart type d'erreur entre parenthèses), colonne 3 : variance résiduelle (écart-type d'erreur entre parenthèses) ; colonnes 4 à 7 : héritabilités sur la diagonale, corrélations génétiques au-dessus de la diagonale et corrélations résiduelles en dessous de la diagonale.

Montbéliarde	Gen. Var.	Res. Var.	0_2	3_30	31_365	3_365
0_2	0,27 (0,09)	277,93 (1,10)	0,39%	22,8%	6,1%	18,1%
3_30	1,03 (0,25)	430,23 (1,73)	-5,0%	0,95%	50,4%	89,1%
31_365	0,83 (0,21)	392,40 (1,62)	-4,1%	-7,0%	0,85%	84,1%
3_365	2,68 (0,56)	752,85 (3,03)	-6,7%	70,7%	65,7%	1,42%
Normande	Gen. Var.	Res. Var.	0_2	3_30	31_365	3_365
0_2	0,68 (0,23)	383 (2,22)	0,71%	6,3%	1,7%	3,1%
3_30	1,28 (0,38)	432 (2,57)	-5,9%	1,18%	30,7%	70,1%
31_365	3,76 (0,85)	565 (3,46)	-5,8%	-8,6%	2,64%	89,4%
3_365	5,82 (1,31)	895 (5,33)	-8,6%	62,7%	72,2%	2,59%
Holstein	Gen. Var.	Res. Var.	0_2	3_30	31_365	3_365
0_2	0,57 (0,09)	374 (0,70)	0,61%	18,5%	22,9%	19,6%
3_30	0,84 (0,11)	406 (0,74)	-5,2%	0,83%	29,4%	83,1%
31_365	0,82 (0,11)	400 (0,75)	-4,6%	-6,9%	0,81%	77,6%
3_365	1,98 (0,26)	740 (1,43)	-7,2%	69,2%	67,2%	1,07%

Les analyses sont réalisées avec les programmes de la famille BLUPF90 (Misztal et al. 2014). Des analyses mono- et bi-caractères sont conduites pour estimer les coefficients d'héritabilité et de corrélations génétiques entre caractères. Les résultats sont présentés dans le Tableau VII-1 pour les 3 races laitières et les 4 caractères.

Au cours de la période 0 à 2 jours, on observe une héritabilité très réduite au sein de la population bovine Montbéliarde (0,39%), possiblement liée au fait que cette race présente une mortalité périnatale plus faible (entre 2 et 3 % en moyenne) en comparaison avec les races Normande et Holstein, qui affichent une mortalité de l'ordre de 5 %. Sur la plage 3 à 30 jours, une héritabilité d'environ 1% est constatée dans les trois races. Entre 31 et 365 jours, l'héritabilité est estimée à

0,8% pour les races Holstein et Montbéliarde, et contraste avec une valeur de 2,6% pour la race Normande.

En ce qui concerne le caractère global sur l'ensemble de la période, les héritabilités se révèlent légèrement plus élevées, s'établissant à 1,4% pour la Montbéliarde, 1,1% pour la Holstein, et 2,6% pour la Normande, résultant d'une incidence globale plus forte et donc d'une meilleure expression des différences génétiques.

Globalement, les héritabilités sont faibles, comme attendu. Cela dit les paramètres génétiques montrent aussi qu'il existe une variabilité génétique importante, avec par exemple des écarts-types génétiques sur la période de 3 à 365 jours représentant 1,4 à 2,4% de mortalité, ce qui est encourageant pour la sélection de ces caractères. Elles sont systématiquement plus élevées en race Normande que dans les deux autres races laitières, ce qui peut probablement s'expliquer au moins en partie par une fréquence plus élevée de la mortalité juvénile dans cette race. Si une évaluation et une sélection sur la mortalité juvénile est souhaitable dans les trois races, c'est assurément en race Normande que c'est le plus important du fait d'une mortalité et d'une héritabilité plus élevées.

Les corrélations génétiques sont faibles entre mortinatalité et mortalité juvénile, ce qui justifie de considérer ces caractères séparément. La mortalité précoce et la mortalité tardive apparaissent positivement corrélées mais correspondent cependant à des caractères différents. La mortalité globale est très fortement corrélée aux deux mortalités partielles et constitue donc un bon compromis.

3. Résultats de l'évaluation pilote par Single step

3.1. Modèle animal single step

Le modèle inclut les mêmes effets que pour l'estimation des paramètres génétiques. Toutefois, un modèle animal est choisi plutôt qu'un modèle père. En effet, pour l'évaluation génétique à composantes de variance connues, les contraintes calculatoires sont moins fortes et ne justifient pas le recours à un modèle père. Par ailleurs, un modèle animal permet d'éviter toute approximation, de mieux prendre en compte une éventuelle évolution génétique et la sélection éventuelle sur ce caractère. Il permet aussi de tirer pleinement profit des génotypes sur les veaux, qui auraient dû être négligés dans un modèle père, et donc de réduire les incertitudes liées à l'aléa de méiose.

L'approche Single Step est un type d'évaluation génomique permettant d'analyser conjointement et simultanément les performances des animaux génotypés et non génotypés. Elle évite le biais de sélection associé au modèle à deux étapes, et prend en compte toute l'information disponible de façon optimale.

Le programme utilisé ici doit être compatible avec une très grande population et de très nombreux animaux génotypés. Nous avons eu recours au programme HSSGBLUP, également utilisé en routine pour les évaluations nationales bovines pour tous les caractères. Il s'agit d'un modèle hybride estimant explicitement les valeurs génétiques des animaux non typés et les effets des SNP, les valeurs génétiques estimés des animaux typés étant ensuite déduites des effets des SNP. Il s'agit de l'implémentation de la méthode proposée par Fernando et al. (2016).

Nous présentons ici les résultats obtenus lors de la validation d'un modèle basé sur le caractère correspondant à la période 3 à 365 jours, en raison de la plus grande héritabilité constatée au sein des trois races laitières ainsi que de sa corrélation élevée avec les autres caractères.

Les résultats d'évolutions phénotypique et génétique sont présentés dans la Figure VII.5 pour les trois races. Les évolutions phénotypiques de 2000 à 2022, exprimées en taux de mortalité, sont présentées en traits pleins (échelle de gauche) et montrent une augmentation régulière des mortalités dans les trois races, au moins jusqu'en 2016, les tendances variant ensuite selon les races comme nous l'avons vu dans le Chapitre II. Les évolutions génétiques (courbes en pointillé, échelle de droite) sont estimées par les moyennes par année de naissance des femelles de la population (qui représente le niveau génétique de la race). Les moyennes d'index sont exprimées en écart-type génétique. On note une augmentation des valeurs génétiques après 2008 pour les trois races, coïncidant avec la mise en place de l'indexation VIN et VIV à cette date (cf Chapitre I). A noter que les index sont exprimés en solutions du modèle et ont un signe inversé pour qu'un index positif soit favorable, c'est-à-dire corresponde à moins de mortalité. On attend donc une relation négative avec le phénotype.

On note une augmentation des valeurs génétiques après 2008 pour les trois races, coïncidant avec la mise en place de l'indexation VIN et VIV à cette date (cf Chapitre I). On peut penser que le fait d'écarter les taureaux dont la descendance présente des risques de surmortalité périnatale, grâce à cette nouvelle évaluation, a permis de sélectionner indirectement des animaux présentant moins de risques de surmortalité durant les périodes suivantes, et ce surtout dans les races Montbéliarde et Holstein pour lesquelles les corrélations génétiques entre 0-2 jours et les périodes suivantes sont les plus élevées (voir tableau VII-1, corrélations génétiques avec la période 3-365 jours de 18,1 et 19,6% respectivement en Montbéliarde et en Holstein). Pour

autant, le niveau phénotypique, lui, se dégrade, ce qui laisse supposer une dégradation des effets du milieu. Puis après 2016, les niveaux génétiques moyens fluctuent selon les races. Le niveau génétique moyen se détériore en race Holstein, laissant supposer que la génétique explique partiellement l'augmentation du taux de mortalité juvénile observé dans cette race. Au contraire, les dernières années, la tendance est favorable en race Montbéliarde, tant au niveau phénotypique que génétique. Cela peut s'expliquer au moins en partie par l'éradication de l'anomalie MTCP chez les taureaux Montbéliards. En race Normande, on note une stabilisation sur les dernières années à la fois d'un point de vue génétique et phénotypique, cependant, la race possède les plus hauts taux de mortalité.

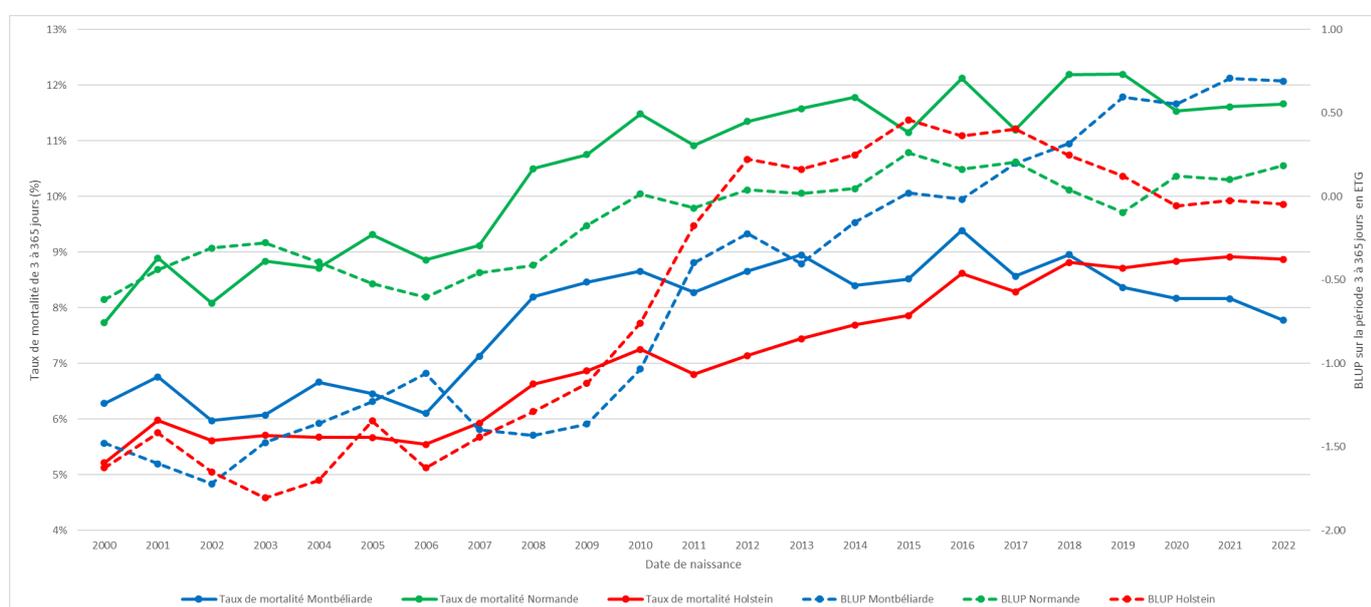


Figure VII-5 : Courbes de tendance des taux de mortalité et des BLUP sur la période 3 à 365 jours.

Les courbes de tendance des taux de mortalité sont représentées en trait plein et se réfèrent à l'axe vertical de gauche en pourcentage. Les niveaux moyens des mêmes animaux sur la période 3 à 365 jours sont exprimés en écart-type génétique (ETG) et représentés en pointillés ; ils se réfèrent à l'axe vertical de droite. Ces courbes sont colorées en bleu pour la Montbéliarde, vert pour la Normande et rouge pour la Holstein.

3.2. Validation de l'évaluation Single-step

Une méthode d'évaluation fournit des prédictions qu'il convient de valider pour vérifier l'absence de biais et mesurer la précision. Plusieurs approches sont possibles. Nous avons choisi ici la comparaison des prédictions obtenues dans une population d'apprentissage avec les mortalités observées dans une population de validation plus jeune. Deux évaluations ont été réalisées :

l'une partielle, avec les données de 2015 à 2020, permettant de prédire les valeurs des reproducteurs (apprentissage); l'autre sur la période complète de 2015 à 2022 pour le jeu de données de validation. L'évaluation complète permet d'ajuster les phénotypes de mortalité pour les effets de milieu et de calculer des indicateurs sur les données 2021-2022 non analysées dans l'étape d'apprentissage. Toutefois, le phénotype de mortalité est très peu héritable, donc l'utilisation de données individuelles n'est pas envisageable pour mesurer la précision des prédictions. En effet, la corrélation attendue serait de $\sqrt{h^2 CD}$ et serait donc inférieure à $\sqrt{h^2} \approx 0,10$, une valeur trop basse pour en tirer des conclusions fiables. La validation est donc réalisée à deux niveaux :

- En comparant les prédictions génomiques de taureaux sans filles dans la population d'apprentissage avec leurs taux de mortalité de leurs filles (au moins 50) dans la population de validation (nées après 2020). Ce test permet d'estimer le CD moyen des index génomiques par R^2 / h^2 avec ici h^2 l'héritabilité du phénotype « taux de mortalité moyen de la descendance ». Les résultats sont illustrés dans la Figure VII-6 et le Tableau VII-2
- En estimant la valeur génomique des femelles typées et nées après 2020 grâce aux effets SNP estimés dans la population d'apprentissage, puis en analysant le lien entre les classes d'index de ces femelles (déciles) et le taux moyen de mortalité observé (voir Figure VII-7).

Dans les trois races, les corrélations sont négatives et dans le sens attendu, en effet on attend à ce que plus un index est élevé moins le taureau aura de mortalité. Elles sont cependant sensiblement plus faibles en Montbéliarde que dans les deux autres races. Les taureaux avec peu de filles peuvent avoir une plus forte dispersion de la surface polynomiale (zone rouge de la Figure VII.6), ce qui traduit le manque de précision des index de ces taureaux.

La corrélation entre index génomiques et taux de mortalité des filles en population de validation permet d'obtenir une estimation empirique de la précision des index purement génomiques : « l'héritabilité » de la moyenne de mortalité de N filles peut être estimée par $N / (N + \lambda)$ avec $\lambda = (4 - h^2) / h^2$. En considérant que N est la médiane de la distribution du nombre de filles, on en déduit une estimation du CD des index génomiques des candidats $CD = R^2 / \ll h^2 \gg$. Les valeurs de CD, qui sont sans doute des sous-estimations compte tenu de la valeur de N considérée, sont rapportées dans le Tableau VII-2.

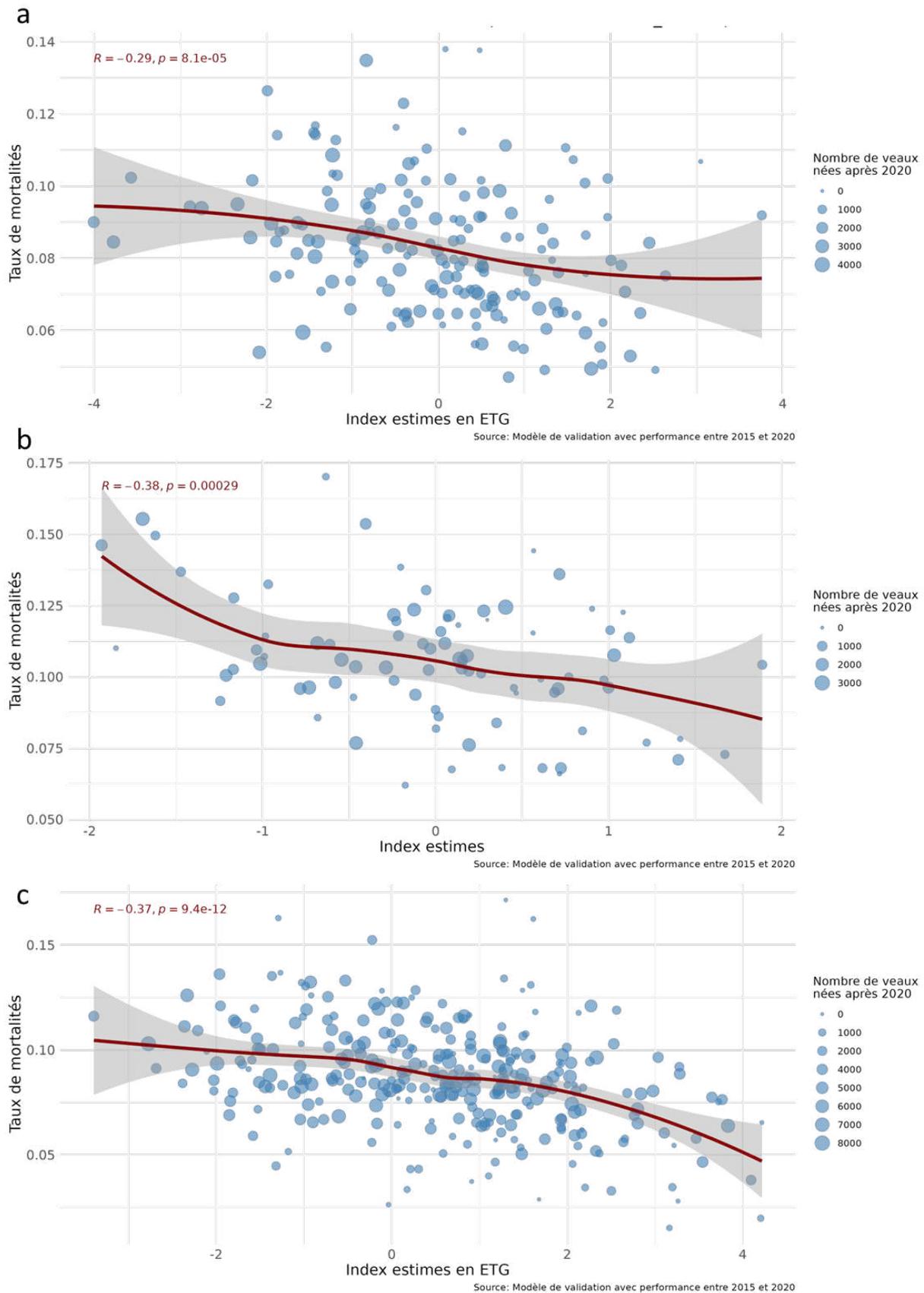


Figure VII-6 : Comparaison des taux de mortalité par descendance de taureaux candidats durant la période de validation (2021-2022) avec les prédictions génomiques issues de la population d'apprentissage (2015-2020) en races Montbéliarde (a), Normande (b) et Holstein (c).

Tableau VII-2: Estimations des CD à partir d'information purement génomique dans une population de validation.

Le CD génomique est estimé par la formule R^2/CD_{poly} avec différentes cohortes de taureaux selon leur nombre de filles N. Le CD_{poly} est estimé par $N / (N + (4-h^2)/h^2)$. R^2 est le carré de la corrélation observée entre index génomique et taux de mortalité observé dans la population de validation.

Race	Sélection	Taureaux candidats	Médiane (nb filles)	h^2 (%)	CD poly	R^2	CD génomique(%)
Montbéliarde	Au moins 100 filles	176	472	1,42	0.63	0,086	14
Montbéliarde	Au moins 500 filles	83	873	1,42	0.76	0,164	22
Normande	Au moins 100 filles	85	349	2,59	0.69	0,147	21
Normande	Au moins 500 filles	29	966	2,59	0.86	0,112	13
Holstein	Au moins 100 filles	311	563	1,07	0.60	0,140	23
Holstein	Au moins 500 filles	170	1199	1,07	0.76	0,201	26

Pour la Montbéliarde, les CD sont très faibles sur les taureaux avec peu de descendants, et difficilement utilisables. Cela est sans doute dû à la mortalité nettement plus faible dans cette race, ce qui joue sur l'héritabilité et la difficulté de mettre en évidence et donc prédire des différences génétiques entre descendance. Cela est illustré sur la Figure VII.6.a, pour laquelle la plupart des index présentent une faible dispersion et la tendance (traduite par la ligne rouge) est assez faible, même si elle va dans le bon sens.

Dans les deux autres races, les Figures VII-6.b VII-6.c montrent des prédictions raisonnablement précises compte tenu du caractère. Toutefois, les CD génomiques estimés pour les candidats sont de l'ordre de 0,20, ce qui est sensiblement plus bas que pour les autres index dans ces races. Ces index sont déjà largement utilisables en l'état mais ils gagneront en précision lorsque la population de référence sera plus large, en particulier si le nombre de veaux morts génotypés augmente.

Une autre façon de valider la capacité prédictive des index génomiques est de comparer le taux de mortalité observé par décile d'index génomiques prédits sur les génisses nées durant la période de validation. Ces index étant calculés grâce aux effets estimés des SNP à partir d'une population d'apprentissage, donc sans les performances de ces femelles. Ces résultats ne traduisent pas la précision des index, mais ils montrent si les différences prédites entre index génomiques se retrouvent en moyenne dans les taux de mortalité. Les résultats sont présentés dans la Figure VII.7 pour les trois races.

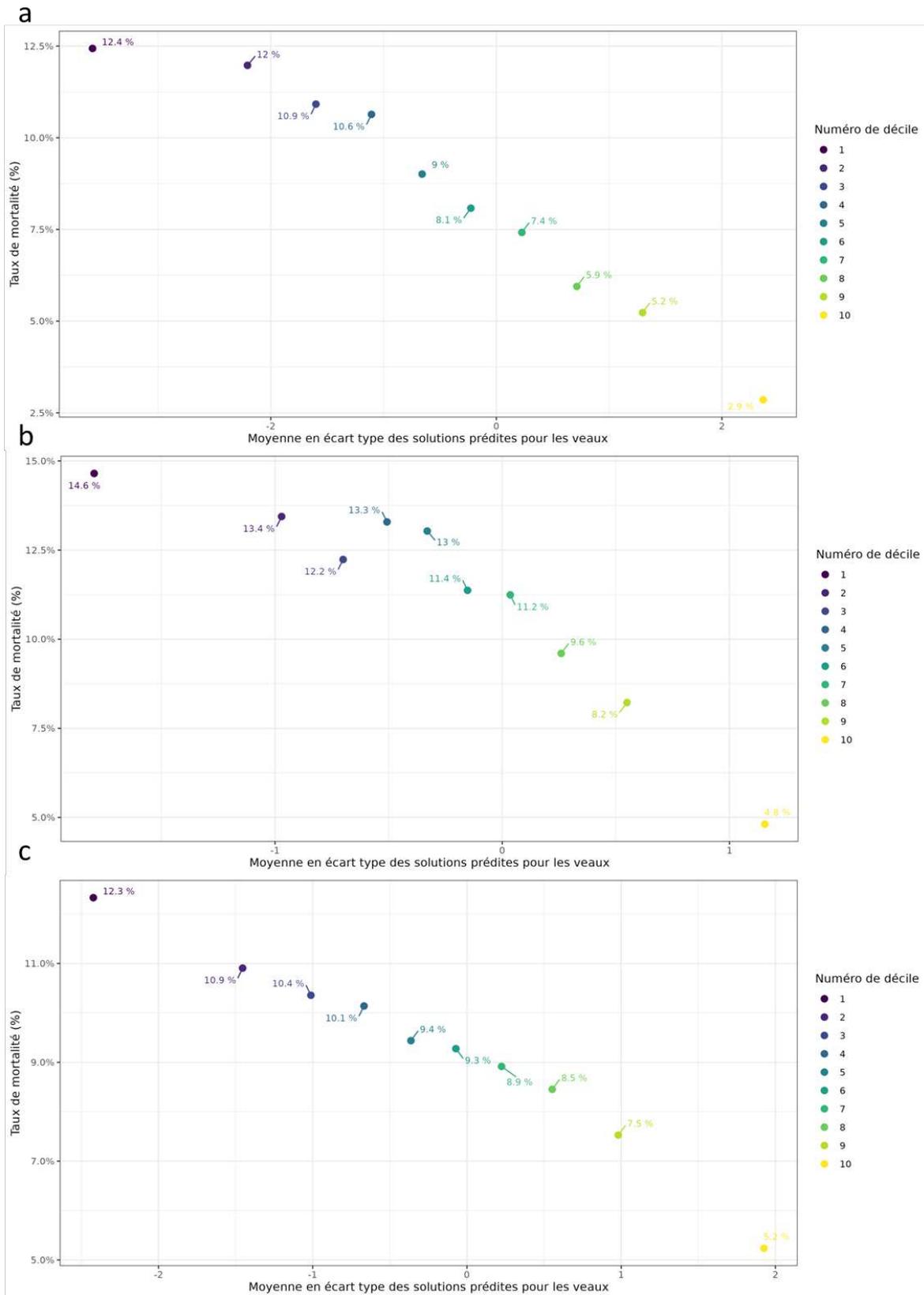


Figure VII-7 : Comparaison des taux de mortalité des veaux nés durant la période de validation (2021-2022) avec les classes d'index génomiques prédites sans les performances propres (découpés en déciles) en races Montbéliarde (a.), Normande (b.) et Holstein (c.).

Pour les trois races et même pour la race Montbéliarde pour laquelle les résultats précédents étaient moins prometteurs, on observe une bonne concordance entre les déciles et le niveau phénotypique de la descendance, avec une relation quasi linéaire. Cette relation linéaire montre que le potentiel génétique des reproducteurs prédit bien la mortalité attendue des génisses, avec une variabilité importante entre les moins bons et les meilleurs reproducteurs. Ainsi, 1 ETG correspond à une baisse de la mortalité de 1,75, 3,30 et 1,53 % dans les races Montbéliarde, Normande et Holstein. En race Montbéliarde, les CD très faibles lorsque nous considérons tous les taureaux dans cette race (1%), sont probablement causés par une mauvaise prédiction avec les index des taux de mortalités réels des taureaux porteurs de MTCP. Des études complémentaires sont nécessaires pour comprendre cette situation.

4. Exploration de pistes d'amélioration du modèle d'évaluation génomique

Le modèle précédent est un modèle standard de type GBLUP sans prise en compte d'information biologique. Il pourrait être amélioré de différentes manières pour gagner en précision. Tout d'abord, plusieurs mutations causales ont été identifiées. Elles pourraient être ajoutées dans le modèle qui estimerait alors l'effet de ces variants causaux et des valeurs génomiques résiduelles. La valeur génomique globale serait alors la somme des effets causaux et de la valeur génomique résiduelle. Selon le déterminisme des variants causaux (codominants ou récessifs), la valeur génomique (qui représente ce que l'individu transmet en moyenne à ses descendants) prend différentes formes, telles que présentées dans Falconer (1996), et rappelées par (Boichard et al. 2016). En cas d'effets additifs, la contribution du *locus* causal est la somme des effets des deux allèles portés. En cas d'effets non additifs, la contribution du *locus* causal dépend du génotype de l'individu, mais aussi de la fréquence des allèles dans la population. Ce type de modèle est conceptuellement simple et inclut les génotypes pour les variants causaux comme effets fixés, ou comme effets aléatoires avec leur variance a priori appropriée (assez forte). Cependant, dans une application de type Single Step, une difficulté est de prédire le génotype aux variants causaux pour les animaux non typés. Ceci est possible mais nécessite une étape préalable d'estimation des génotypes des animaux non typés, et de disposer d'un nombre suffisant d'animaux typés.

Un modèle approché du précédent consiste à remplacer les variants causaux par des haplotypes en déséquilibre de liaison et de ne cibler que des variants récessifs. C'est ce qui est présenté dans les paragraphes suivants.

Dans un premier temps, nous réalisons une phase de détection d'haplotypes associés à la mortalité à l'état homozygotes, ce qui nous permet de limiter le nombre d'haplotypes à prendre en compte dans le modèle d'évaluation. Dans un second temps, pour tous les haplotypes sélectionnés, nous calculons la probabilité des produits non typés d'être homozygotes à partir des typages parentaux. Enfin, dans un troisième temps, nous incluons ces haplotypes en covariables dans le modèle d'évaluation.

3.1. Estimation de la probabilité d'être homozygote pour un haplotype paternel

Dans cette approche, on se limite aux descendants de pères connus, typés et dont les génotypes sont phasés. Les haplotypes paternels sont donc connus. Quatre situations sont possibles : l'animal est typé (1.), l'animal est non typé mais la mère l'est (2.), la mère est non typée mais le grand-père maternel l'est (3.), ni la mère ni le grand-père maternel ne sont typés(4.). Soit $\{h_k, k=1, n_k\}$ l'ensemble des n_k haplotypes présents dans la population à une position du génome donnée, de fréquence f_k ; le père est porteur de (h_{pi}, h_{pj}) ; si la mère est typée, elle est porteuse de (h_{mi}, h_{mj}) ; alors la probabilité que le produit soit homozygote pour l'haplotype h_{pi} est de $\sum_{q=1}^2 0,25 \text{prob}(h_{pi} \equiv h_{mq})$. Lorsque la mère n'est pas typée mais que le grand-père maternel l'est (h_{gl}, h_{gm}) , la formule devient $\sum_{q=1}^2 0,125 \text{prob}(h_{pi} \equiv h_{gq}) + \sum_{k=1}^{n_k} 0,25 f_k \text{prob}(h_{pi} \equiv h_k)$. Lorsque ni la mère ni le grand-père maternel ne sont typés, la formule devient $0,5 f_{pi}$. Si le père est homozygote $(h_{pi} \equiv h_{pj})$, alors ces probabilités sont doublées (les calculs sont résumés dans la Figure VII-8).

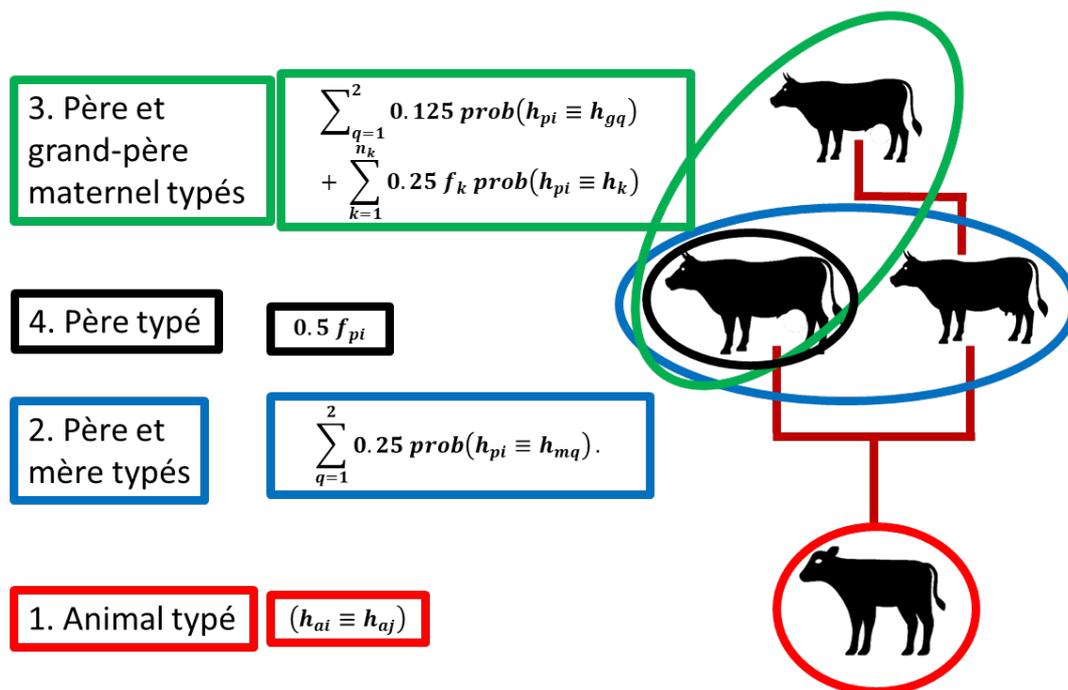


Figure VII-8 : Schéma du calcul de ROH pour un haplotype donné selon les typages disponibles.

3.2. Détection des haplotypes candidats

Afin de faciliter la détection des haplotypes influençant la mortalité juvénile et de ne pas calculer de probabilité de ROH sur des millions d'individus et ce pour des centaines de milliers d'haplotypes, nous avons procédé à une première étape de sélection des ROH. Pour ce faire, nous avons fait le choix de ne pas considérer les veaux qui n'avaient pas d'information génétique maternelle (animaux avec seulement le père génotypé ou sans aucune information). De plus, pour réduire les temps de calculs, la formule de probabilité lorsque le grand père est typé (3.) est approximé par $\sum_{q=1}^2 0,125 \text{prob}(h_{pp} \equiv h_{gq})$ en ne considérant que les haplotypes grand-paternels et non les fréquence haplotypiques dans la population.

Avec cette méthode, nous élargissons considérablement notre échantillon d'individus (cf. Tableau VII-3) pour inclure non seulement les veaux morts typés, mais également ceux dont l'information génomique n'est pas disponible. Ainsi, alors que seulement 9234, 3150 et 13255 veaux morts étaient typés, une probabilité d'homozgotie à l'haplotype étudié (appelée ensuite par simplicité ROH) peut être estimée pour 215 998, 175 226 et 755 741 veaux morts entre 3 et 365 jours dans chacune des trois races.

Tableau VII-3 : Tableau récapitulatif du nombre de veaux morts et vivants disponibles selon l'information génétique disponible pour ces animaux.

Type d'accouplement selon typage disponible (Mort/Total disponible)				
Génotypage		Montbéliarde	Normande	Holstein
Animal (1.)	Mort	9 234	3 150	13 255
	Total	531 113	154 324	1 033 649
Père mère (2.)	Mort	18 429	9 786	43 176
	Total	97 344	57 620	239 690
Père GPM (3.)	Mort	197 569	165 440	712 565
	Total	2 212 208	1 630 378	8 859 430
Père (4.)	Mort	32 748	23 137	219 174
	Total	375 628	214 384	2 808 074
Aucun typage	Mort	24 119	7 861	164 892
	Total	276 236	72 059	2 029 834
Total	Mort	282 099	209 374	1 153 062
	Mort_mère_gpm	215 998	175 226	755 741
	Total	3 492 529	2 128 765	14 970 677
	Ratio	23	56	57

Pour chaque haplotype considéré, nous calculons l'espérance du nombre de veaux homozygotes de chacune des deux classes « mort » et « vivant » et déduisons le nombre de veaux non homozygote à cet haplotype, résultant en un tableau de contingence 2x2. L'association est ensuite testée par χ^2 . L'haplotype est sélectionné s'il a une fréquence d'homozygotes de 10^{-5} au minimum chez les vivants ou chez les morts et si le test est significatif.

Le nombre de morts attendus est calculé en multipliant la fréquence d'homozygotes chez les vivants et les morts par le nombre total de morts. Ensuite, le nombre de veaux manquants (car non génotypés) est estimé par la différence entre le nombre d'homozygotes morts attendus et le nombre estimé de morts homozygotes.

À titre de preuve de concept dans ce manuscrit, nous illustrons l'exemple des anomalies CDH en race Holstein et MTCP en race Montbéliarde. Ces deux anomalies, localisées sur le chromosome 11 pour CDH et sur le chromosome 6 pour MTCP, ont un impact fort sur la mortalité juvénile. La mutation CDH manifeste son effet à l'état homozygote mais aussi, à un degré moindre, à l'état

hétérozygote (Gross *et al.* 2016; Menzi *et al.* 2016; Häfliger *et al.* 2019; Besnard *et al.* 2023), tandis que MTCP est strictement récessif.

Les résultats pour le chromosome 11 en race Holstein sont présentés dans la Figure VII.9. Entre la position 76 et 79 MB, plusieurs haplotypes montrent un fort déséquilibre entre le nombre de veaux attendus et l'espérance d'observés (sur la base des probabilités d'homozygotie). La mutation est une insertion de 1,38 kpb (77885995-77927965, <https://omia.org/OMIA001965/9913/>) d'un élément transposable dans la région codante du gène APOB (Kipp *et al.* 2016; Menzi *et al.* 2016; Mock *et al.* 2016). La région touchée par la mutation est très bien localisée par la méthode. Néanmoins la sélection de l'haplotype est plus complexe. En effet, dans cet exemple, l'haplotype avec la plus forte différence en nombre de veaux est un haplotype très fréquent dans la population. La p-value déterminée par le test de χ^2 sur le tableau de contingence offre un compromis acceptable entre un haplotype très fréquent et très partiellement porteur de la mutation et un haplotype plus rare avec un déséquilibre de liaison complet avec la mutation. Ce point sera abordé dans la Discussion générale.

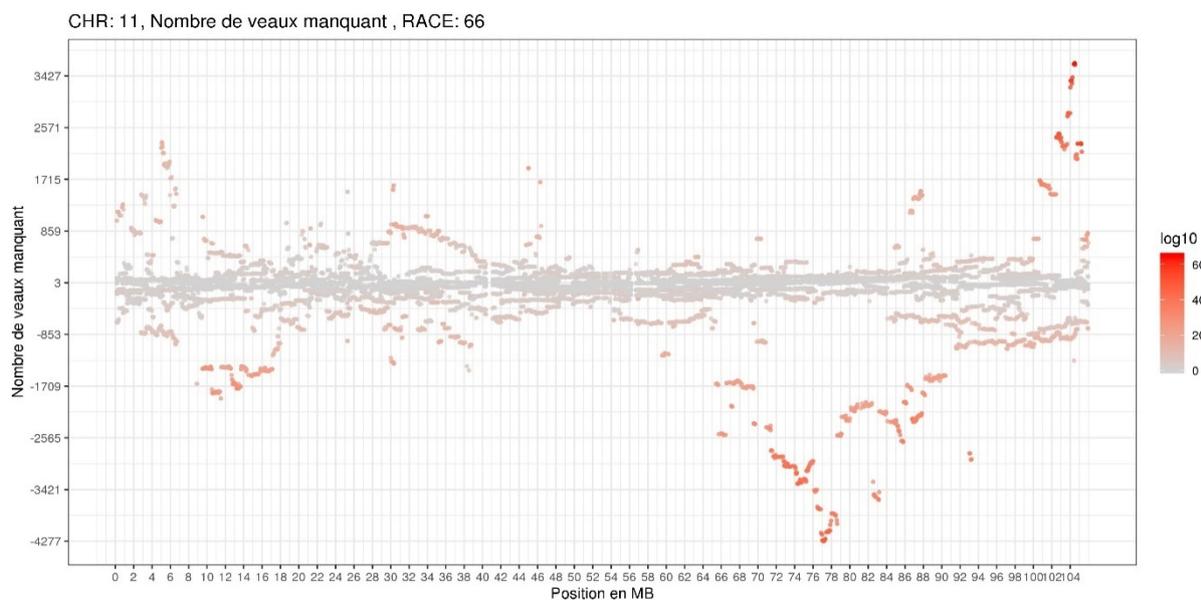


Figure VII-9 : Déficit de veaux pour chaque haplotype sur le chromosome 11 en race Holstein.

L'axe des abscisses montre la position en mégabases (MB) de chaque haplotype et la couleur montre la significativité [$-\log_{10}(p\text{-value})$] du test de χ^2 avec un gradient vers le rouge foncé.

L'exemple de la mitochondropathie avec la Figure VII.10 sur le chromosome 6 en Montbéliarde permet d'observer ce phénomène : la mutation est apparue sur un haplotype fréquent dans la population avec 23% de la population homozygote à l'haplotype et ayant une différence estimée

de -1617 veaux femelles et un $-\log_{10}(\text{p-value})$ égal à 11. Un autre haplotype adjacent (en rouge foncé sur le graphique) est beaucoup moins fréquent avec 0,3% d'homozygotes estimés dans la population vivante mais un $-\log_{10}(\text{p-value})$ de 94. En effet, cet haplotype ayant recombiné chez un taureau porteur, la majorité des porteurs de l'haplotype sont porteurs de la mutation.

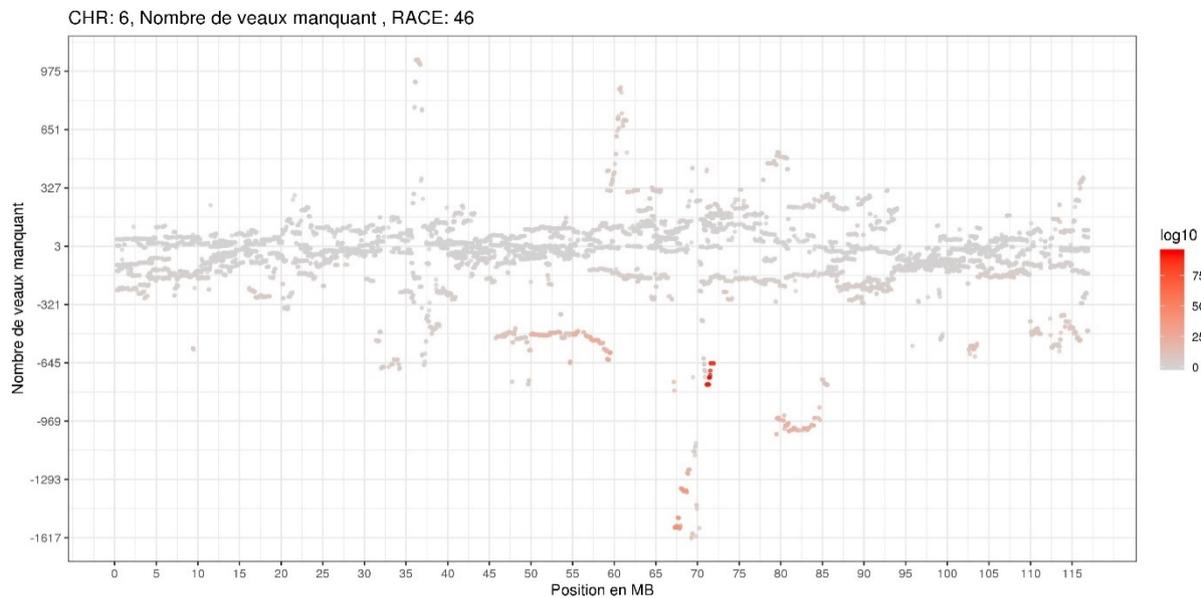


Figure VII-10 : Déficit de veaux pour chaque haplotype du chromosome 6 en race Montbéliarde.

L'axe des abscisses montre la position en mégabases (MB) de chaque haplotype. La couleur montre la significativité [$-\log_{10}(\text{p-value})$] du test χ^2 avec un gradient vers le rouge foncé.

Ainsi, les filtres suivants ont été mis en place :

- $\log_{10}(\text{p-value})$ supérieur à 6,3
- Sélection du plus haut $-\log_{10}(\text{p-value})$ dans un intervalle de 5 MB
- Sélection des 40 haplotypes les plus significatifs

Ces différents filtres nous ont permis de détecter de multiples haplotypes, dont certains déjà détectés avec les méthodes des chapitres précédents, les haplotypes les plus significatifs sont résumés dans le Tableau VII-4, avec NOA1 présenté en tant que deuxième exemple dans la Figure VII-10.

Tableau VII-4 : Liste des haplotypes les plus significatifs détectés dans les 3 races. La référence OMIA est indiquée si l'anomalie est déjà connue.

Race	Chr	Position en MB	Fréquence des homozygote morts	Fréquence homozygote	Nb attendu de morts	Différence entre attendus et observés	Taux de chute	$\log_{10}(p)$	Causal gene
Montbéliarde	6	71,393	0,60%	0,30%	749	-727	1,97	94	NOA1
Normande	17	56,479267	0,13%	0,08%	152	-111	1,73	12	RFC5
Holstein	5	32,334842	0,36%	0,25%	2164	-933	1,43	61	ITGB7
Montbéliarde	7	49,526867	0,35%	0,28%	683	-193	1,28	9	
Montbéliarde	16	58,944371	0,41%	0,34%	842	-178	1,21	7	
Holstein	22	56,192254	0,63%	0,54%	4620	-805	1,17	23	
Holstein	9	99,773758	0,40%	0,35%	2990	-500	1,17	14	
Normande	6	34,23983	1,61%	1,41%	2795	-394	1,14	10	
Normande	26	17,360622	16,01%	15,07%	29802	-1854	1,06	23	

3.3. Prise en compte des haplotypes dans une évaluation Single Step

La probabilité de porter un haplotype délétère à l'état homozygote est connue pour les animaux typés et estimée pour les animaux non typés à partir des typages des parents, comme nous l'avons vu précédemment (cf partie 4.1 de ce Chapitre). Ces probabilités sont ajoutées dans le modèle d'évaluation sous forme de covariables. L'analyse est donc limitée aux animaux de père typé. Ce modèle présente l'avantage potentiel de prendre en compte les anomalies récessives et une certaine partie du fardeau génétique, tandis que le modèle standard qui ne considère que des effets additifs tend à sous-estimer les variants récessifs létaux.

Quelques résultats sont présentés sur la Figure VII.11, qui résume les impacts d'être porteur d'un haplotype.

Globalement, les résultats ne sont pas complètement convaincants. La plupart des effets mis en évidence sont réduits. Alors que la différence attendue pour un variant létaux à l'état homozygote est proche de 1 (mort vs vivant), la plupart des effets sont de l'ordre de quelques pourcents. Cela reflète vraisemblablement un déséquilibre de liaison très partiel entre l'haplotype et le variant causal. En effet, la méthode tend à considérer des haplotypes fréquents, donc de fréquence sans doute bien plus élevée que la mutation causale. De sorte que les haplotypes porteurs de la mutation existent vraisemblablement sous deux formes : une majoritaire en DL très incomplet avec la mutation et différents rare haplotypes en fort DL avec l'allèle muté si des recombinaisons ont eu lieu.

L'effet estimé sur la mortalité est donc très fortement dilué. Par ailleurs, même si cela est moins important pour l'estimation de l'effet, il existe aussi d'autres haplotypes en déséquilibre partiel avec l'allèle létaux mais qui sont considérés ici comme non porteurs. L'efficacité de la méthode dépend donc de conditions assez restrictives : une fréquence de l'allèle létaux suffisamment élevée, un haplotype porteur en complet déséquilibre de liaison avec l'allèle létaux. Ces conditions sont partiellement remplies pour ROH6P71 qui colocalise avec MTCP. Son effet est de -16%, un effet très fort mais bien inférieur aux 100% qu'on pourrait attendre. La fréquence de l'haplotype est de 39%, bien supérieure à la fréquence de l'allèle MTCP, ce qui explique le déséquilibre incomplet.

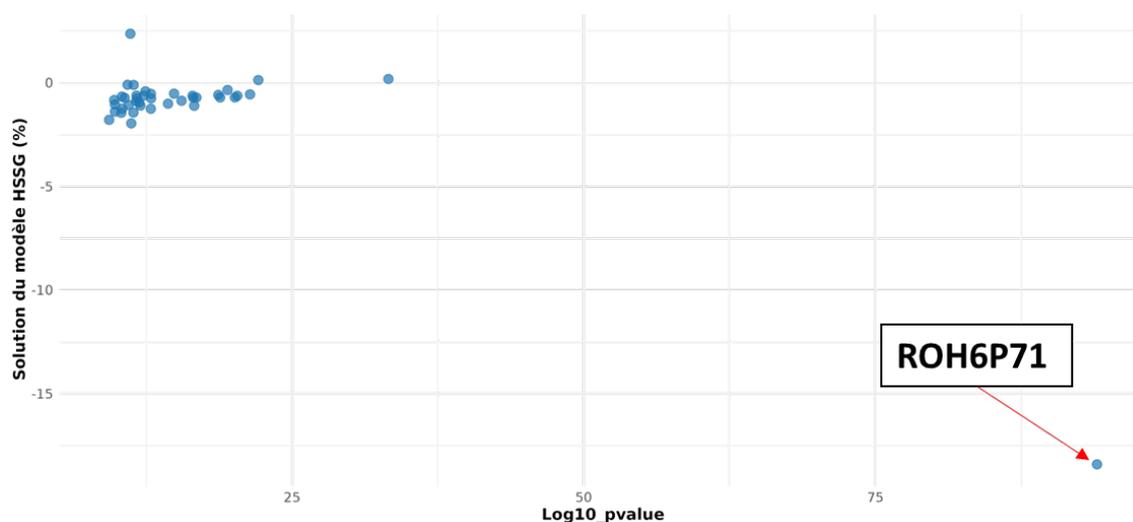


Figure VII-11 : Relation entre les résultats des tests de Chi2 en log 10 de p-value en abscisse et les solutions en % du RoH dans l'évaluation en ordonné pour la race Montbéliarde.

Un RoH, ROH6P71, est encadré et pointé par une flèche rouge, il est très significatif et semble être le seul RoH à avoir un effet sur la mortalité dans le modèle HSSG.

3.4. Limites et perspectives :

Cette approche visait à prendre en compte le fardeau génétique, c'est-à-dire les variants récessifs délétères. En l'absence de connaissance sur les variants causaux, les haplotypes associés à des effets délétères ont été recherchés comme proxy. Malheureusement, l'approche n'a pas été couronnée de succès car les haplotypes détectés, assez fréquents, semblent être en déséquilibre trop partiels pour être de bons proxys des variants causaux dans le modèle génétique.

Faute de temps, des modèles complémentaires n'ont pas pu être élaborés et mis en œuvre.

Toutefois, le modèle standard semble être une avancée très intéressante. Même si la précision est encore faible, son pouvoir prédictif est bien démontré et il est relativement facile à mettre en place dans le cadre d'un Single Step permettant d'utiliser toute l'information disponible. Sa mise en œuvre devrait être efficace pour réduire le taux de mortalité dans les races où il est relativement élevé. Elle est donc hautement souhaitable, à la fois pour des raisons économiques et éthiques. En race Montbéliarde, la précision semble très faible et il conviendra d'en comprendre les raisons, puisque l'héritabilité est assez similaire entre races. Une hypothèse est que la population d'apprentissage est fortement marquée par MTCP tandis que dans la population de validation, la mutation causale identifiée entretemps est fortement contre-

sélectionnée chez les taureaux et prise en compte dans les accouplements. Cette hypothèse pourrait être testée en supprimant les taureaux porteurs de MTCP et leurs descendants et en refaisant l'étude de validation.

Même si l'évaluation génomique est déjà satisfaisante, sa précision pourrait sans doute être améliorée de différentes façons :

- Augmenter le génotypage des morts. Ce travail difficile à valoriser économiquement par l'éleveur pourrait être pris en charge par les organismes ou entreprises de sélection. En effet, les morts constituent (heureusement) la classe minoritaire mais aussi la plus informative, et leur faible taux de génotypage ne permet pas de les caractériser finement, le génotypage des parents ne permettant pas compenser ce manque.
- L'héritabilité du caractère est faible. Elle pourrait sans doute être légèrement augmentée en focalisant l'étude dans les élevages avec la mortalité la plus élevée. Toutefois, elle conduit à une perte de données par troncation. Cette hypothèse pourrait être testée pour mesurer le changement de précision.
- La prise en compte de la consanguinité locale n'a pas été abordée. Elle pourrait mettre en évidence les régions du génome portant le fardeau génétique de façon plus systématique. Des travaux de ce type sont en cours dans le cadre du projet européen Rumigen.
- La prise en compte explicite des variants causaux serait probablement un plus, mais elle repose sur la disponibilité de l'information sur le maximum d'individus et une imputation efficace chez les animaux non typés. Après les travaux du projet Effitness (Besnard et *al.* 2023), ce travail est une priorité, mais n'a pas pu être réalisé dans cette thèse faute de temps. Il sera poursuivi après avec les projets WELCOW et Sévital, deux projets soutenus par APIS GENE et axés sur la mortalité en élevage.
- La méthode de détection des ROH en utilisant l'information des parents pourrait être appliquée au phénotype des inséminations artificielles en comparant les IA non fécondantes aux IA suivies de naissance.

Chapitre VIII. Discussion générale et perspectives

Dans ce chapitre de discussion générale, je récapitulerai les résultats principaux des chapitres précédents en les interprétant et en les comparant à la lumière des sources bibliographiques existantes. Puis je détaillerai leurs apports et les implications des différents travaux en termes de recherche fondamentale et appliquée tout en identifiant les limites des différentes études. Enfin, je ferai des recommandations pour gérer les anomalies en sélection et proposerai différentes stratégies pour repérer au plus tôt, voire anticiper l'émergence de nouvelles anomalies génétiques.

1. Résultats principaux des études sur la mortalité, avantages et limites

1.1. Un phénotype simple à observer, avec un déterminisme complexe

Du fait de sa faible héritabilité et de sa nature binaire, la génétique de la mortalité est un sujet difficile à étudier à l'échelle individuelle. C'est pourtant un phénotype facile à déterminer au niveau de l'élevage, et fondée sur des observations à déclaration obligatoire dans le cadre de l'IPG. Nous avons donc une vision exhaustive de la mortalité dans les races bovines laitières. Pour éviter des biais liés à un changement d'environnement, nous avons choisi d'écarter les veaux qui quittent leur élevage de naissance. C'est le cas pour la plupart des veaux mâles qui sont vendus à des engraisseurs au cours de leurs 3 premières semaines de vie alors que cela ne concerne que quelques pourcents de veaux femelles. Pour cette raison les analyses n'ont porté que sur la mortalité des jeunes femelles.

Du fait de sa nature binaire, le caractère n'est pas distribué normalement au niveau individuel dans la population, comme le sont par exemple la production laitière ou des mesures morphologiques. Néanmoins, du fait de la structure de la population bovine, il existe une distribution normale des moyennes des taux de mortalité observés au sein de la descendance des taureaux d'IA, comme nous l'avons vu dans le Chapitre II. Par ailleurs, la moyenne des performances des descendants présente une variance résiduelle réduite et reflète assez précisément la valeur du père. Cela nous a permis de mettre en évidence deux anomalies génétiques dues à des mutations *de novo* en ciblant des taureaux extrêmes dans le Chapitre III, de réaliser des analyses d'associations dans le Chapitre IV et d'estimer des paramètres

génétiques à l'aide de modèles père dans le Chapitre VII. De même dans le cadre de la recherche de déficit/enrichissement en homozygotes (Chapitre V), la faible précision du phénotype des animaux du groupe cas et du groupe contrôle à l'échelle individuelle a été compensée par les effectifs importants d'animaux étudiés, qui au final garantit une puissance de détection suffisante et permet la cartographie de *loci* passés inaperçus jusque-là.

Néanmoins, cette précision faible des phénotypes a certains désavantages lorsqu'il s'agit de caractériser les effets biologiques d'un *locus* en particulier et les causes de mort associées qui peuvent être multiples (dystocie, diarrhée, maladies respiratoires, etc...), comme nous l'avons vu dans le Chapitre I. Les grands jeux de données à notre disposition nous ont permis de préciser un peu ces phénotypes, en étudiant des phénotypes approchés ou proxys, comme l'âge à la première IA (pour contrôler la croissance), les différences de surmortalité (cause de sortie « mort ») ou de réformes anticipée (cause de sortie « boucherie ») entre animaux issus d'accouplements à risque ou contrôle (Besnard, en prep). Toutefois, cela ne remplace pas un phénotypage fin pour identifier les mécanismes moléculaires sous-jacents.

Cette absence de phénotype fin a priori complique grandement la recherche de résistance pour des pathogènes spécifiques. Cela est plus simple lorsque nous travaillons avec une maladie spécifique et que nous devons différencier les pathogènes responsables de la mortalité. Ces dernières années, différents projets ont été menés au sein l'UMT visant à exploiter les carnets sanitaires des élevages, et à nouer des partenariats avec les groupements de défense sanitaire et différents partenaires pour cartographier des *loci* de résistance/sensibilité à certaines maladies (paratuberculose, myosite eosinophile).

Toutefois, lorsque l'on travaille sur une maladie spécifique, différents biais existent qu'il faut prendre en compte, notamment lorsqu'on utilise uniquement les animaux génotypés pour les évaluations génomiques comme population d'étude. Le plus évident est le biais d'échantillonnage, dû à la sous-représentation de certains groupes d'animaux pour cause de mort prématurée ou de faible état corporel qui conduit les éleveurs à ne pas les génotyper. Un autre biais consiste à avoir une sous-représentation des cas due à un manque de remontées phénotypiques. Cela amène à un surnombre d'animaux notés sains et peut conduire à considérer à tort des animaux sains en réalité atteints, en l'absence de signalement phénotypique. Par ailleurs un animal sensible peut ne pas développer une pathologie s'il est élevé dans un milieu indemne et qu'il n'est pas exposé à ce pathogène. Enfin, l'infection par un premier pathogène est un facteur de risques pour d'autres maladies et certains animaux qui sont déclarés comme étant des cas peuvent être issus de diagnostic erroné. Afin de contourner ces erreurs de diagnostics,

une solution peut être de travailler avec un protocole unifié au niveau national reposant sur un diagnostic différentiel faisant appel à différentes biotechnologies (test Elisa, PCR quantitative, cultures bactériennes). Une autre solution plus simple à mettre en place est de cibler le recrutement d'animaux cas et contrôles dans les mêmes élevages et dans des cohortes de naissance identiques pour limiter les biais d'échantillonnages et d'environnement, comme pour la paratuberculose par exemple (voir « Des indicateurs génomiques de résistance à la paratuberculose disponibles en race Normande 2024 »).

Au global, le phénotype de la mortalité permet d'obtenir un grand jeu de données fiables sur les populations bovines. L'héritabilité est très faible mais ce handicap est compensé par la disponibilité de grandes familles de pères. L'écart type génétique est important par rapport à la moyenne de mortalité (pour la mortalité entre 0 à 365 jours sur les femelles, les écarts types génétiques associés sont de 3,3%, 3,9% et 3,4% pour les races Montbéliarde, Normande et Holstein respectivement). Une moyenne de descendance présentant un écart très élevé à la moyenne et hors distribution attendue, peut révéler une anomalie génétique.

1.2. Des extrêmes résultant d'apparition d'anomalies dominantes

Le calcul des taux de mortalité par descendance de taureaux peut révéler des écarts très forts à la moyenne. Ainsi, nous avons observé plusieurs taureaux avec des taux de mortalité très élevés porteurs d'anomalies génétiques à déterminisme dominant. Des anomalies allant du simple SNP (Besnard et al. 2022) à la translocation interchromosomique (Jourdain et al. 2023) ont été publiées. Puis nous avons testé une approche d'association en regardant si certaines parties du génome avaient un fort effet sur la mortalité. Ces études d'association à l'échelle du génome ont permis de détecter 25 QTL de mortalité juvénile, associés à différentes périodes de mortalité. Parmi ces QTL, nous avons pu proposer des variants candidats pour 5 d'entre eux. Néanmoins, la majorité des taureaux extrêmes dans la distribution de la mortalité restaient inexpliqués, ce qui nous a conduit à utiliser une nouvelle approche.

1.3. Des mutations récessives expliquant une partie de la mortalité

Nous avons ensuite recherché les anomalies récessives à l'aide d'une méthodologie de génétique inverse. Ces résultats ont permis de mettre en évidence 34 anomalies récessives dans

les trois races laitières et de proposer 8 variants candidats. Les mutations ayant le plus d'effet sur les populations ont été étudiées de façon fine en collaboration avec les écoles vétérinaires ; Chr17 g.57188407_57188409del, Chr6 g.72359797_72359798insG et Chr5 g.26807079G>A pour la Normande, Montbéliarde et Holstein respectivement. L'anomalie en Holstein a été trouvée dans les différentes populations mondiales comme aux Etats-Unis et différents pays Européens. Cette anomalie sur la race laitière la plus utilisée dans le monde a donc un très gros impact sur la mortalité et sa découverte permettra de réduire fortement la mortalité.

Ces trois mutations nous ont permis de mieux comprendre quels biais pouvaient intervenir avec les méthodes de cartographie utilisées. Notamment dans le cas de la mutation en Holstein sur *ITGB7* où l'haplotype porteur de la mutation est un haplotype très fréquent dans la population car le premier contributeur de la race ELEVATION est porteur de l'haplotype mais sans la mutation. Par conséquent, seulement une partie des plus jeunes porteurs de l'haplotype sont aussi porteurs de l'anomalie. A l'inverse, la mutation du gène *NOA1* responsable de la mitochondropathie, est présente sur des haplotypes différents (l'un dérivant de l'autre par recombinaison) provenant du deuxième et du huitième contributeur de la race (DEBOUT et GARDIAN). Cela amène des difficultés différentes ; dans le premier cas, les haplotypes en déséquilibre de liaison incomplet avec la mutation seront plus difficile à détecter car la majorité des homozygotes à l'haplotype restent en vie. Dans le cas de *ITGB7*, la détection a été permise par la recombinaison de certains haplotypes chez des taureaux reproducteurs plus jeunes, donnant ainsi des signaux significatifs sur des haplotypes bornant la mutation à 23MB et à 33MB. Dans le deuxième cas, pour la mutation sur *NOA1*, les effets calculés à partir des haplotypes peuvent être réduits. Aussi, dans le cas de plusieurs haplotypes recombinants et porteurs de la mutation, les méthodes de détection peuvent ne pas détecter des *loci* dont l'effet est réparti entre haplotypes. Finalement, pour détecter une mutation candidate récessive, la meilleure méthode reste la méthode HHED, mais elle pourrait être améliorée, ce que nous verrons en 4.4. Après avoir étudié les déficits d'homozygotie d'haplotypes par rapport à l'attendu, nous avons étudié les haplotypes pouvant être en déséquilibre mais à l'état hétérozygote.

1.4. Une détection difficile des mutations co-dominantes, par l'étude d'haplotypes en distorsion de ségrégation

Après avoir trouvé des anomalies récessives et dominantes dans les populations, je me suis demandé si des mutations pouvaient avoir un effet de dominance partielle ou de co-dominance.

Un exemple de mutation co-dominante est l'anomalie CDH (Gross et *al.* 2016; Häfliger et *al.* 2019), ayant été décrit comme ayant des effets à l'état homozygote mais aussi à l'état hétérozygote. Pour cela, nous avons étudié dans des familles de père les déséquilibres de transmission mendélienne des haplotypes à l'état hétérozygote chez les taureaux. Néanmoins, les résultats issus de cette méthodologie sont assez décevants car nous n'avons trouvé qu'un variant candidat pour un total de 3179 familles étudiées dans les trois races laitières.

De multiples variants avec différents déterminismes ont été trouvés lors des précédentes études, néanmoins les anomalies trouvées les plus importantes n'expliquent que quelques pourcents de la mortalité, nous allons donc examiner quelle part globale de la variabilité phénotypique est expliquée par la génétique.

1.5. Un déterminisme complexe avec une faible héritabilité

Le déterminisme génétique d'un caractère est caractérisé par l'héritabilité. L'héritabilité est la part de la variabilité phénotypique expliquée par la génétique. L'héritabilité est estimée à partir d'un modèle statistique dans lequel le phénotypique est décomposé en des effets de milieu, un effet génétique, et une résiduelle non expliquée. Dans le chapitre II et VII, nous avons étudié les facteurs environnementaux ayant le plus d'effet sur la mortalité, par exemple le sexe de l'animal, les conditions de naissance, la parité de la mère et la taille de l'élevage ainsi que les variations liées à la conduite d'élevage. Certains facteurs comme la taille et le poids du veau ont aussi un effet connu sur la mortalité mais du fait de leur déterminisme génétique, nous ne les avons pas inclus dans les modèles. Néanmoins, beaucoup d'effets de milieu restent non identifiés et constituent la variabilité résiduelle du modèle. Le modèle considéré décompose la variabilité phénotypique en différents effets de milieu et en une variabilité génétique. L'héritabilité correspond au ratio entre la variance génétique et la variance phénotypique (=variance génétique + variance résiduelle).

Nous avons montré que les coefficients d'héritabilité sont faibles, tant pour la mortalité globale qu'aux différentes périodes considérées (de 0,81 à 2,64%). Ces faibles valeurs ne signifient pas que la variabilité génétique soit faible, puisque l'écart type génétique est de l'ordre de 1,5 à 3,3% pour une moyenne de mortalité de 10%. La faible héritabilité signifie plutôt que la variance phénotypique est grande, soulignant le fait que la majorité des effets environnementaux ne sont pas expliqués par les effets du modèle. La situation est similaire à celle de la sensibilité aux

mammites cliniques, pour laquelle l'héritabilité est de l'ordre de 2%. La sélection pourrait donc prendre en compte ces caractères, en complément de la recherche d'anomalies génétiques.

1.6. Les index génomiques offrent une bonne prédiction mais l'estimation de la valeur des taureaux candidats reste imprécise

Même si l'héritabilité du caractère est assez faible, les résultats obtenus à l'aide des modèles génomiques de la mortalité juvénile utilisant toute l'information disponible sont intéressants. La caractérisation en déciles des veaux selon leurs index prédits montre de bonnes prédictions pour les animaux issus du pire et du meilleur décile, entre 12,4%- 2,9%, 14,6% - 4,8% et 12,3%- 5,2% pour les races Montbéliarde, Normande et Holstein respectivement. Pour les animaux des déciles intermédiaires, les prédictions sont très bonnes pour la Montbéliarde et la Holstein mais de qualité moyenne pour la Normande, vraisemblablement à cause d'un effectif plus réduit dans les données de validation, par rapport aux deux autres races. Malgré une capacité prédictive à l'échelle de la population, les estimations des CD restent modestes. Les valeurs de CD varient selon les échantillons mais sont majoritairement entre 15 et 26%. En conclusion ces résultats montrent des perspectives encourageantes pour l'application de notre modèle en sélection dans un avenir proche. Ces CD restent inférieurs aux attendus des sélectionneurs, mais il existe des marges de progression avec l'augmentation naturelle des populations de référence et, si possible, le génotypage des veaux morts.

Une des améliorations dans le modèle serait de prendre en compte les mutations candidates. Notre modèle, en effet, ne permet d'estimer que la composante génétique additive de la mortalité, et l'estimation de cette composante peut être ainsi améliorée dans le cas d'animaux porteurs d'anomalies génétiques récessives ou dominantes à pénétration incomplète. Par exemple dans le cas de la mitochondropathie et de l'alopecie des extrémités, qui sont des anomalies récessives, le modèle additif estime pour le SNP en DL un effet proportionnel à la fréquence de l'anomalie dans la population. Le modèle est efficace dans les cas graves où l'anomalie a une fréquence élevée, mais elle est quasi inopérante lorsque la fréquence allélique est faible, ce qui est le cas général. Le fait d'intégrer cette anomalie dans le modèle permettrait aux ES de pouvoir repérer dans la descendance des taureaux porteurs les candidats non porteurs et ayant une bonne composante génétique additive. Cela permettrait aussi, au global, d'améliorer la précision des résultats. La prise en compte de ces anomalies en effet fixes dans le modèle permettrait d'améliorer les prédictions des taureaux candidats.

2. Gestion des anomalies génétiques et sélection pour une amélioration de la survie des jeunes bovins

Les travaux menés dans le cadre de ma thèse ont permis d'identifier des dizaines de *loci* avec des déterminismes génétiques variés et un impact plus ou moins fort sur la capacité des jeunes bovins à survivre jusqu'à l'âge adulte. La prise en compte de ces résultats en sélection offre des perspectives intéressantes pour réduire la mortalité juvénile mais pose aussi quelques défis. En effet, historiquement, à chaque fois qu'une nouvelle anomalie génétique était découverte la politique de gestion à l'échelle raciale consistait à retirer les taureaux porteurs de la reproduction. Face la multiplication des découvertes de nouvelles anomalies, il n'est désormais plus possible d'agir ainsi de façon systématique. Il y aurait très peu de reproducteurs indemnes de tous les variants délétères connus et la réduction de la base de sélection impacterait sévèrement la variabilité génétique de la population et le progrès génétique sur les autres caractères. Pire, elle pourrait favoriser l'émergence de nouvelles anomalies génétiques par dérive génétique en créant un nouveau goulot d'étranglement. Dans cette partie je discuterai des différentes stratégies pouvant être adoptées en fonction des différents types de *loci* identifiés.

2.1. Retrait de certains reproducteurs du schéma de sélection

La stratégie de retrait de certains reproducteurs du schéma de sélection reste pertinente dans les cas de mutations *de novo* dominantes et de remaniements inter-chromosomiques. Le rapport avantage /inconvénients est en effet très favorable puisque la prévalence d'animaux atteints est généralement élevée dans la descendance des taureaux impactés (P.ex. 19,9% pour le taureau Pa et 30,6 % pour Mo; Chapitre III) et le fait de retirer une seule famille n'impacte que marginalement la diversité génétique à l'échelle populationnelle. Les risques de transmission de ces affections héréditaires à la génération suivante est généralement nulle pour les mutations *de novo* dominantes. En effet, le plus souvent la pénétrance est forte ou complète et les animaux ayant reçu la mutation *de novo* dominante expriment le phénotype délétère. Cette élimination de la diffusion large n'empêche pas d'utiliser le taureau comme père à taureau si son niveau le justifie : on l'utilise alors de façon ciblée pour procréer des fils et on ne sélectionne les candidats que parmi les non porteurs. Le potentiel génétique du reste du génome n'est ainsi pas perdu.

Dans le cas des anomalies de type remaniement chromosomique, dans la majorité des cas les veaux aneuploïdes (1/2 des veaux) n'arrivent pas jusqu'à la reproduction. Cependant, et cela a un fort impact sur la filière, 1/4 des génisses seront elles aussi transloquées-équilibrées, et auront les mêmes proportions d'animaux atteints dans leur descendance que leurs pères, comme nous l'avons montré dans le cadre de l'étude Jourdain et *al.* 2023 dont je suis co-auteur. Dans cette situation, le conseil est de génotyper tous les descendants de ces taureaux pour savoir, au moyen d'un test sur haplotype, lesquels ont reçu les chromosomes réarrangés. Si les trisomies et monosomies partielles n'entraînent que des morts embryonnaires précoces alors, au lieu d'être purement réformées, les femelles laitières pourront être inséminées en croisement de façon à leur permettre d'avoir une carrière productive et d'amortir le coût de leur élevage tout en stoppant la diffusion de cette anomalie. En effet, les coûts liés à une baisse de fertilité sont bien inférieurs aux coûts d'alimentation et de soin d'une femelle de la naissance à l'entrée en production. Si en revanche l'anomalie cause des morts après la naissance la mise à la reproduction des individus transloqués équilibrés est proscrite pour des raisons éthiques et économiques.

2.2. La prise en compte dans l'index de synthèse (ISU)

La sélection contre les anomalies est fréquemment mal dosée : en cas d'éradication, la sélection est trop forte et coûte cher en termes économiques et génétiques ; au contraire, en ne faisant rien, on risque une augmentation de fréquence dans les générations futures par dérive, rendant le problème plus difficile à régler. Une solution est d'inclure l'anomalie dans l'index de sélection. Elle convient pour tous les déterminismes mais elle est d'autant plus recommandée pour les anomalies récessives.

Cette stratégie a été proposée par (Cole 2015; Segelke et *al.* 2016). Elle est explicitée dans Boichard et *al.* (2016) et développée par Hoze et *al.* (2018). Boichard et *al.* formulent la composante de l'anomalie dans l'ISU comme suit :

Si les 3 génotypes A1A1, A1A2 et A2A2 ont des phénotypes de valeurs a , d , et $-a$, respectivement, et les fréquences des allèles A1 et A2 sont p et $q=1-p$, la valeur transmissible de l'allèle A1 est $\alpha_1 = q [a + d(q-p)]$ tandis que la valeur transmissible de l'allèle A2 est égale à $\alpha_2 = -p [a + d(q-p)]$. L'effet de substitution de l'allèle A2 par A1 est égal à $\alpha = \alpha_1 - \alpha_2 = (a + d(q-p))$. Les valeurs génétiques des 3 génotypes s'écrivent :

$$g(A1A1) = 2 \alpha_1 = 2q [a + d (q-p)] = 2q \alpha$$

$$g(A1A2) = \alpha_1 + \alpha_2 = (q-p) [a + d(q-p)] = (q-p) \alpha$$

$$g(A2A2) = 2 \alpha_2 = -2p [a + d(q-p)] = -2p\alpha$$

Dans le cas d'une anomalie récessive, $d=a$. En remplaçant q par f (fréquence de l'allèle délétère) et p par $(1-f)$ dans les expressions précédentes, on obtient :

$$\alpha = 2.f.a,$$

$$g(A1A1) = 4 f^2 a$$

$$g(A1A2) = -2 f (1-2f) a$$

$$g(A2A2) = -4 f (1-f) a$$

La différence de valeur entre les deux homozygotes étant égale à $2a$, on peut écrire $a=c/2$, où c représente le coût d'un animal atteint. Les 3 valeurs génétiques s'expriment donc comme suit :

$$g(A1A1) = 2 f^2 c$$

$$g(A1A2) = -f (1-2f) c$$

$$g(A2A2) = -2 f (1-f) c$$

Les valeurs génétiques dépendent donc de la fréquence de l'allèle délétère et du coût associé au génotype homozygote récessif. Plus la fréquence f est élevée et plus le coût c est élevé, plus le poids de l'anomalie dans l'index de synthèse est fort. Le coût est d'autant plus élevé que la mortalité est tardive, de quelques dizaines d'euros pour une mortalité embryonnaire à plus de 1000 euros pour un animal de plus d'un an. Cette valeur génétique est convertie en points d'ISU à raison de 5€ pour 1 point d'ISU. Ainsi un taureau porteur hétérozygote d'un variant délétère entraînant un coût c lié à la mort des homozygotes aura une perte moyenne par veau de $0,5*f*c$ et une valeur génétique de $-f(1-2f)c$. Par exemple un taureau porteur d'un variant responsable de perte embryonnaire (p.ex. MH1, MH2) dont le coût moyen serait de 70 € par *conceptus* mort et dont la fréquence serait de 10% dans la population aurait une valeur génétique pour cette anomalie de $-0,1*0,8*70/5=-1,1$ points d'ISU. Une perte tardive de 1000€ induirait dans les mêmes conditions une perte de 16 points d'ISU pour un reproducteur hétérozygote.

Hozé *et al.* (2018) présentent les fréquences et les coûts estimés pour différentes anomalies. Pour le poids économique, Segelke *et al.* (2016) et Cole *et al.* (2015) ont différencié les pertes embryonnaires précoces (coût de 70 €), les pertes embryonnaires tardives (97€), les mortalités à la naissance (130€), les mortalités juvéniles (219 €) et enfin les mortalités à l'âge adulte (supérieur à 400€). Finalement le fait d'être porteur d'un variant récessif délétère induisait des pertes d'ISU allant de -0,64 à -5,48 points pour les anomalies connues en races laitières en fonction de leur fréquence et du poids économique. Pour l'instant cette approche n'a pas encore été mise en place sur le terrain mais il est prévu qu'elle soit appliquée prochainement dans le

cadre de la réforme de l'ISU en race Normande et Holstein. Les porteurs de l'alopécie des extrémités que nous avons caractérisée dans le cadre de cette thèse et de l'allèle A de la Kappa-caséine (réduisant le rendement fromager du lait) ayant un effet négatif seront respectivement fortement et modérément pénalisés.

2.3. La gestion des accouplements

La réduction de fréquence prend du temps, plusieurs générations. En attendant, il faut gérer les anomalies dans les populations et éviter la naissance d'individus atteints. Une stratégie consiste à empêcher la naissance d'individus homozygotes en interdisant les accouplements à risque. Ce n'est qu'une solution transitoire, du moins pour les anomalies fréquentes, car elle ne réduit pas la fréquence allélique et reporte donc le problème à plus tard, surtout quand une grande partie de la population n'est pas génotypée. Par contre, le principal avantage de cette méthode est qu'elle n'impacte pas ou très marginalement le progrès génétique et la variabilité génétique raciale. Elle nécessite toutefois que les taureaux et les vaches/ou leurs pères soient génotypés et n'est donc applicable que dans les races ayant majoritairement recours à l'insémination et aux évaluations génomiques (en pratique les 3 principales races laitières nationales : Holstein, Montbéliarde et Normande et les 3 laitières régionales : Abondance, Tarentaise et Vosgienne). Ces données, au même titre que les index pour différents caractères sélectionnés, sont ensuite intégrées aux logiciels d'accouplements utilisés sur le terrain par les entreprises de sélection pour choisir le taureau le plus adapté pour chaque vache.

La méthode pANO proposée par Brochard et *al.* (2018), propose un indicateur unique pour toutes les mutations récessives détectées. Si les femelles sont génotypées alors on se base sur directement sur leurs informations propres. Si les femelles ne sont pas génotypées, on calcule leur probabilité d'être porteuse de chaque variant en fonction du génotype de leurs ancêtres génotypés. Ensuite on effectue des prédictions d'accouplement et on calcule le risque pour le produit à naître d'être homozygote à chacune des anomalies. Les probabilités sont ensuite combinées pour obtenir un indicateur unique facile à utiliser. L'accouplement est dit à risque entre 1 et 5% et à proscrire lorsque le risque dépasse 5% de naître homozygote pour une anomalie.

Cette méthode permet de réduire le nombre d'animaux homozygotes aux mutations et a déjà été mis en place pour la Montbéliarde sur les anomalies SHGC (Floriot et *al.* 2015), MH1, MH2 (Fritz et *al.* 2013; Michot et *al.* 2017) et MTCP (Fritz et *al.* 2013; Besnard et *al.* 2023). Elle a également

été mise en place et largement adoptée en races Holstein et Normande pour d'autres anomalies génétiques même si ces travaux n'ont pas fait l'objet de publications scientifiques.

Dans l'approche proposée par Bérodière et al. (2020), le plan d'accouplement est optimisé pour maximiser le revenu associé au produit à naître, incluant l'index de synthèse, la dépression de consanguinité et le risque d'homozygotie pour des anomalies, sous contrainte d'utilisation des taureaux.

2.4. Augmentation de l'effort de génotypage et de sélection des futurs taureaux d'insémination

Bien que la gestion des accouplements présente de nombreux avantages, cette approche ne modifie pas les fréquences des allèles délétères à l'échelle de la population. Ces fréquences peuvent même continuer à croître, rendant les choix d'accouplements potentiellement plus compliqués dans le futur. Il est donc conseillé d'avoir une approche proactive et de chercher à réduire progressivement les fréquences des allèles délétères en parallèles. L'une des façons les plus efficaces de le faire est d'appliquer une pression de sélection au moment du recrutement des jeunes veaux mâles qui seront amenés à devenir de futurs taureaux d'insémination. Prenons l'exemple de l'anomalie BLIRD en race Holstein dont la fréquence de porteurs (fp) est d'environ 10% (la fréquence allélique étant elle d'environ 5%). Si demain on veut éradiquer cette anomalie en impactant peu le progrès génétique annuel, il suffit de génotyper non pas 4000 veaux mâles mais 4400 ($=4000 \cdot (1+fp)$) veaux pour retenir les 200 candidats nécessaires chaque année pour devenir des taureaux d'insémination. Cette stratégie par la voie mâle permet de supprimer toute naissance d'homozygote et de diviser par deux la fréquence de porteurs à chaque génération, rendant progressivement l'effort nécessaire de plus en plus réduit. En définitive cette approche est similaire au retrait pur et simple des reproducteurs des catalogues d'insémination présentée au début de cette partie mais la pression de sélection est appliquée une génération après, sur la descendance des taureaux qui auraient été retirés, ce qui change tout en termes d'efficacité économique et génétique. La mise en service très rapide des taureaux évalués par génomique évite la difficile période de transition qu'on avait précédemment avec le testage. De plus elle peut être modulée et les entreprises optent le plus souvent pour une réduction progressive des fréquences de porteurs dans les cohortes de taureaux, plutôt qu'une éradication subite. En pratique elle est appliquée avec plus ou moins d'assiduité depuis une dizaine d'années pour

gérer certaines anomalies responsables de mort embryonnaire en races Holstein, Montbéliarde et Normande (HH1 à HH8, MH1 et MH2, NH2).

2.5. Mise en place d'une évaluation génomique sur la mortalité juvénile et perspective d'amélioration des modèles.

Enfin, une dernière approche pour réduire la mortalité juvénile des trois races laitières nationales consiste à mettre en place des évaluations génomiques de routine pour la mortalité juvénile. Même si les modèles peuvent être améliorés (nous y reviendrons juste après), ils pourraient déjà être utilisés pour effectuer une présélection des veaux mâles pour éviter l'augmentation de la mortalité à long terme, et même de la diminuer. Comme précisé précédemment, ces modèles pourraient être améliorés par la prise en compte des anomalies génétiques récessives, afin de mieux appréhender la composante génétique de ce caractère.

Dans les modèles mixtes linéaires, différents choix sont possibles quant à la prise en compte des anomalies dans les modèles génétiques. La première méthode est de mettre les différentes anomalies en effets fixés dans les modèles. L'index est ensuite déduit des solutions polygéniques des anomalies, comme présenté pour l'ISU en 2.2. Lorsque les animaux ne sont pas génotypés, la probabilité de chaque génotype peut être estimée. Nous avons ainsi proposé de calculer la probabilité d'être homozygote pour des anomalies récessives et avons ajouté ces probabilités en covariables dans le modèle. En l'absence de génotypage de la mutation causale, il est possible de réaliser le même processus avec des haplotypes en déséquilibre de liaison. Toutefois, cette approche a montré des résultats plutôt décevants, probablement du fait d'un déséquilibre trop incomplet entre haplotype et mutation causale (voir Figure VII.11).

En conclusion, la mise en place d'un index sur la mortalité juvénile est un outil très fortement recommandé, à la fois comme observatoire, outil d'analyse de la situation, et également outil de sélection. Il pourrait être intégré dans l'ISU, de même que les caractères de vitalité au vêlage (de la mère) et à la naissance (du veau). Il ne remplacera pas la détection d'anomalies génétiques mais est un outil à mettre en place conjointement avec la recherche de nouvelles mutations.

La surveillance doit être continue, car de nouvelles anomalies émergeront à l'avenir, différentes de celles connues actuellement. Elles peuvent déjà exister à une fréquence faible sans être identifiées mais voir leur fréquence augmenter par dérive. Elles peuvent aussi apparaître comme neo-mutation avant, éventuellement, de se disséminer.

3. Anticipation des futures émergences

Comme nous avons pu le voir, l'émergence régulière de nouvelles anomalies génétiques est un phénomène naturel et imprévisible mais plus on agit tôt, mieux on arrive à limiter leur impact. Ainsi je propose trois stratégies pour repérer au plus tôt ces émergences, voire les anticiper.

3.1. Mise en place d'un observatoire des extrêmes.

Les résultats obtenus dans le cadre de mon premier article m'ont conduit à réfléchir à la mise en place d'une détection en routine des taureaux qui seraient extrêmes du point de vue de leurs performances propres (fertilité) ou de la moyenne de leur descendance. L'hypothèse principale est qu'un écart à la normalité de ces caractères est associé à une anomalie génétique influençant le caractère en question. Je propose donc d'établir un observatoire des extrêmes pour différents caractères, avec des seuils de détection spécifiques (pire percentile ou pire 5% par exemple), afin d'identifier les taureaux porteurs potentiels de réarrangements chromosomiques ou de mutations *de novo* dominantes avec un taux de mosaïcisme élevé. Différents caractères peuvent être accessibles grâce aux données remontées pour les évaluations génomiques de routine, notamment les caractères de production, de morphologie, de fertilité ou encore la mortalité comme nous l'avons vu dans le chapitre III. Dans ce cas précis, des statistiques de routine menées par exemple tous les 3 ou 6 mois pourraient donner l'alerte et renforcer les remontées des acteurs de terrain faites à l'ONAB.

3.2. Détection des translocations équilibrées en routine

Nous avons vu avec l'exemple du taureau Mo (Besnard *et al.* 2023) et les travaux de la thèse de Jeanlin Jourdain auxquels j'ai participé qu'environ 1 taureau d'IA sur 400 était transloqué équilibré, soit environ un taureau d'IA par an. Dans le but de détecter en amont ces anomalies, la recommandation principale de l'article Jourdain *et al.* (2023) est d'effectuer un examen caryotypique systématique des veaux mâles avant leur recrutement par les entreprises de sélection. En complément, il serait souhaitable de mettre en place des analyses périodiques des déséquilibres de liaison interchromosomiques dans la descendance des taureaux afin de détecter les translocations passées inaperçues lors du contrôle de caryotype.

3.3. Anticipation de l'émergence de futures anomalies récessives par génétique inverse

Lorsque le génome des ancêtres fondateurs est séquencé, il devient possible d'accéder à la quasi-totalité de la variabilité génétique des races. Néanmoins, le matériel biologique n'est souvent plus disponible pour les fondateurs. Ainsi, parmi les 56, 61 et 61 taureaux avec plus de 1% de contribution génétique respectivement dans les races Montbéliarde, Normande et Holstein, seuls 9, 13 et 27 sont séquencés. (cf Figure VIII.1).

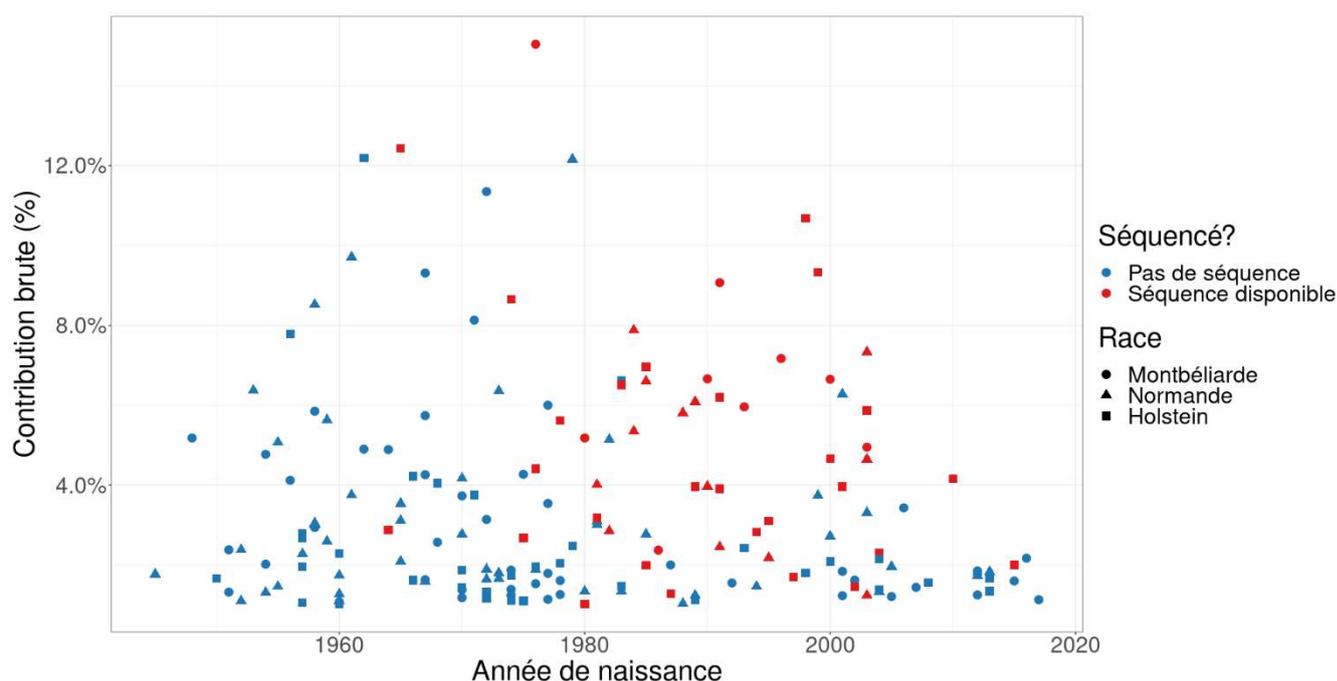


Figure VIII-1 : Graphique présentant la contribution brute au patrimoine génétique de leur race, l'année de naissance et la disponibilité d'une séquence de génome complet ou non pour les principaux taureaux Holstein, Montbéliards et Normands.

Seuls les taureaux avec une contribution brute supérieure à 1% ont été représentés. La forme du point informe sur la race du taureau et la couleur s'il a été séquencé.

Cette difficulté peut être contournée par le séquençage d'animaux plus actuels, porteurs de segments chromosomiques des fondateurs. A l'aide des informations récoltées durant les dernières années avec la mise en place de la sélection génomique et de la montée en puissance et de la baisse du coût du séquençage, des méthodes d'anticipation d'anomalies peuvent être mises en place. Dans le cadre du projet SEQOCCIN qui a permis de séquencer le génome de 571 taureaux d'IA, plusieurs taureaux ont été retirés du marché car porteurs d'anomalies considérées comme trop délétères. Aujourd'hui la qualité des prédictions est très perfectible et de nombreux

variants détectés sont des faux positifs. Par ailleurs, de nombreux variants causaux sont sans doute non détectés, par manque d'annotation, en particulier hors de la séquence codante. Avec la poussée de l'intelligence artificielle et des méthodes de caractérisation des variations du génome, les mutations délétères seront de mieux en mieux caractérisées, ce qui permettra de se doter d'annotations plus précises, et d'améliorer les prédictions faites par génétique inverse.

Si le projet SeqOccln pouvait faire figure de pionnier, il est quasi-acquis que l'ensemble des taureaux d'IA mis sur le marché seront séquencés. Cette évolution peut être mise en œuvre progressivement si nécessaire, même si les budgets impliqués sont maintenant très raisonnables grâce à la diminution du coût de séquençage. A minima, une première recommandation est donc de séquencer les pères à taureaux pour surveiller les anomalies émergentes. Au-delà de cet objectif, ces séquences contribueront à imputer précisément les génotypes de puce à une échelle plus dense, favoriseront la découverte des mutations causales, prépareront le futur génotypage par séquençage. Au-delà des pères à taureaux, tous les taureaux ayant atteint un certain seuil de nombre d'inséminations, variable selon les races, devraient être séquencés.

En effet, une anomalie récessive n'est exprimée qu'à l'état homozygote, c'est-à-dire 4 à 5 générations après la diffusion par un ancêtre porteur. Cette expression est très tardive et à ce moment la fréquence d'hétérozygotes est déjà importante. Même si l'annotation du génome et des variants reste à améliorer, le séquençage systématique est la méthode de choix d'identifier les mutations *de novo* ou récentes et prédire leur impact. Ces mutations peuvent être ajoutées sur la puce de génotypage, ce qui permet d'estimer la fréquence allélique et la suivre dans le temps, repérer rapidement des homozygotes pour les caractériser. Dans le cadre d'une politique volontariste, on peut même réaliser des accouplements consanguins pour valider ou réfuter le potentiel délétère des mutations.

4. Travaux à approfondir dans un futur proche

4.1. Analyses par familles de père (HSD)

Comme nous l'avons vu dans le Chapitre VI, la méthode de détection des haplotypes par HSD pourrait être adaptée pour tester les grandes familles génétiquement proches (par exemple un père et ses fils). Par exemple, le taureau Holstein LOUXOR possède à lui seul 14 000 veaux

génotypés et a eu de multiples fils eux-aussi taureaux d'insémination notamment NIVERNAIS et NEWSTAR avec plus de 4000 et 8000 veaux génotypés. Ces situations ne sont pas recommandées en termes de gestion de la diversité, mais elles constituent des dispositifs originaux pour tester l'effet de mutations *de novo* qui, rappelons-le, sont de l'ordre de plusieurs dizaines à apparaître à chaque génération et dont l'effet est peu caractérisé en dehors des quelques mutations responsables d'anomalies dominantes. Ces mutations sont le moteur de l'Evolution et certaines pourraient avoir des effets favorables sur des caractères de production par exemple.

4.2. Anomalies récessives

Un des problèmes majeurs dans la gestion des anomalies récessives dans une population est que les anomalies qui sont éliminées sont peu à peu remplacées par de nouvelles anomalies émergentes. Cela demande donc de régulièrement rechercher les nouvelles anomalies qui ségrégent dans la population. Comme nous l'avons vu précédemment, la surutilisation de taureaux élites aura à moyen terme un impact fort sur la fréquence d'une anomalie émergente dans la population. La majorité des anomalies récessives connues ont un ancêtre avec une forte contribution au patrimoine génétique de la race. Il y a cependant une difficulté de détection lorsque la mutation apparaît *de novo* sur les taureaux ayant une forte contribution, les haplotypes en DL avec la mutation sont presque indétectables car ils sont pour la plupart présents avec et sans la mutation, comme nous l'avons vu pour *ITGB7*. Ainsi, les anomalies récessives sont détectables par recherche de déficit en homozygote chez les contrôles et d'enrichissement chez les morts, et ces détections doivent être reconduites régulièrement. Nous avons proposé plusieurs approches dans ce manuscrit permettant de les détecter.

Dans la même idée, GOLDWYN, un taureau connu pour avoir apporté de la variabilité génétique dans la population Holstein Française est l'ancêtre commun pour sept haplotypes que nous avons détecté en HHED (cf Chapitre V). Le fait qu'il apporte certains haplotypes peu fréquents dans la population permet une facilité de détection par nos méthodes. Cependant, de multiples mutations récessives peuvent être apparues dans des régions très fréquentes dans la population et sont difficiles à trouver.

4.3. HHED des homozygotes

La méthode discutée dans le Chapitre V est très puissante et a permis la détection de la majorité des anomalies récessives. Elle devra être réutilisée périodiquement dans le but de détecter des potentielles nouvelles anomalies émergentes. Pour rappel, cette méthode se base sur l'étude des veaux génotypés dans la population et recherche des déficits qui ne respecte pas la distribution attendu sous l'équilibre de Hardy-Weinberg. C'est une méthode très puissante pour détecter des haplotypes à fort DL avec la mutation.

De multiples régions en déficit complets d'homozygotes n'ont pas encore été associées à leur variant causal, une explication est que nous basons nos recherches sur les SNP ou les petit InDels et nous négligeons les variants structuraux, principalement car leur détection n'est pas réalisée en routine. De plus, toutes nos études ont été basées sur des variants ayant un effet sur la traduction en protéine du gène, car mieux annotées, et non dans des régions non-codantes, qui peuvent pourtant avoir aussi des effets délétères. La proposition est d'abord d'intégrer dans la recherche les variants structuraux mais aussi d'intégrer d'autres variants dans des régions régulatrices des gènes, potentiellement mieux annotés dans les prochaines années avec les avancées des connaissances du fonctionnement du génome. Il est à noter que des travaux sur les variants structuraux, et l'inclusion de certaines de ces variants prédits comme délétères sur la puce utilisée en routine, sont en cours au sein de l'équipe.

4.4. HHED des homozygotes probables

Une autre approche se base sur la méthode proposée dans le point 4 du Chapitre VII Développement d'une évaluation génomique sur la mortalité juvénile. Celle-ci se basant sur les probabilités d'homozygotie à chaque haplotype, elle est possible même si les veaux ne sont pas typés dès lors que leurs ascendants sont typés. Cette méthode repose sur le fait que la majorité des pères reproducteurs sont génotypés mais que la majorité des veaux morts ne le sont pas. Cela est encore plus vrai dans les populations allaitantes où peu de génotypages sur les animaux commerciaux sont réalisés, alors que les pères sont souvent génotypés. Cette approche permettrait donc de compenser en partie le manque de couverture de ces populations en termes de génotypage. La méthode repose sur le même principe que HHED mais utilise les probabilités d'être homozygote sachant les informations génétiques des ascendants. Cela permettra de

détecter de nouvelles anomalies qui n'ont pas pu être détectées par HHED classique, grâce à une forte augmentation des effectifs informatiques.

4.5. Améliorations des modèles génétiques

4.5.1. Amélioration du phénotypage

Au niveau national, certaines données sont inexploitable dans les modèles et cela réduit donc la précision des différents modèles. Si les mortalités sont bien connues, leur cause en général ne l'est pas. Pour pouvoir sélectionner une résistance à certains pathogènes, il faudrait des phénotypages plus précis et en grande quantité. Cela demande la mise en place de protocoles au niveau national pour une collecte de données uniformisée. Ce travail n'est pour l'instant réalisé que de façon très ponctuelle pour suivre des animaux considérés comme à risque, avec l'aide de vétérinaires ou des écoles vétérinaires et des budgets forcément limités. L'utilisation des carnets sanitaires est une alternative en cours d'étude mais montre justement les défis d'uniformisation au niveau national et de précision de l'information. Une autre possibilité est de sélectionner les élevages à haute pression infectieuse, notamment les plus gros élevages avec un regroupement des veaux pour mieux mettre en évidence les taureaux transmettant une résistance aux pathogènes étudiés.

4.5.2. Détection des zones du génome responsables du fardeau génétique

Par manque de temps, cette approche n'a pas été étudiée dans cette thèse. Elle permettrait d'estimer les régions porteuses de la dépression de consanguinité (c'est-à-dire responsable d'une baisse de performance ou de fitness associée à l'homozygotie). La méthodologie serait la suivante : identifier, pour chaque animal et chaque région du génome, son état homozygote (ROH) ou non, réel ou en probabilité ; inclure ces probabilités de ROH en régression dans le modèle d'évaluation. Des régions de longueur constante, par exemple de 1 Mb, peuvent être considérées, ce qui impliquerait d'estimer environ 2 600 effets ROH. Cette approche permettrait de quantifier les régions du génome impactent la mortalité à l'état homozygotes. Cette cartographie du fardeau génétique pourra ensuite conduire à des méthodes d'optimisation en vue de réduire l'homozygotie dans ces régions du génome.

4.6. Le déploiement en routine et l'évaluation des races allaitantes et projet Sévital

Les études présentées dans ce manuscrit ont montré que la mortalité juvénile en élevage bovin était un problème grandissant et que la sélection peut contribuer à le gérer. Afin de permettre le déploiement des évaluations génomiques pour les professionnels, de tester et d'adapter les modèles aux allaitants, et de tester la prise en compte de gènes d'intérêts dans ces modèles, l'UMT a monté un projet nommé Sévital. Son livrable principal sera une mise en œuvre pratique d'une évaluation génomique de la mortalité, étape nécessaire à la prise en compte de ce caractère en sélection en vue de réduire la mortalité juvénile en élevage bovin.

4.7. La limite de la génétique

Il est pertinent d'aborder la diversité des taux de mortalité observée d'un élevage à l'autre comme une démonstration que la génétique ne peut pas expliquer entièrement ce phénomène. Il est nécessaire d'accompagner les efforts de sélection génétique par des démarches techniques spécifiques visant à identifier les causes de mortalité propres à chaque élevage. Cette approche permettrait de lutter efficacement contre la mortalité juvénile en tenant compte des spécificités de chaque exploitation. En effet, bien que les efforts de sélection génétique contribuent à améliorer la situation, leur efficacité peut être renforcée par des interventions ciblées dans les élevages présentant les résultats les moins favorables.

L'idée sous-jacente est que la gestion de la mortalité juvénile ne peut pas reposer uniquement sur des stratégies de sélection génétique, mais doit être complétée par des mesures adaptées aux particularités de chaque élevage. Cette approche, combinant sélection génétique et gestion technique, pourrait maximiser les chances de réduire efficacement la mortalité juvénile et d'améliorer la santé des troupeaux bovins laitiers.

Conclusion générale

Dans le but d'approfondir les connaissances sur le déterminisme génétique de la mortalité juvénile en élevage, nous avons développé plusieurs approches exploitant les vastes ensembles de données pour identifier des *loci* impliqués dans la mortalité avec des déterminismes génétiques différents.

Durant ce projet, j'ai pu participer à la découverte et la description d'une dizaine d'anomalies récessives et dominantes qui affectent la survie des jeunes animaux en élevage laitier dans les races Holstein, Montbéliarde et Normande. Grâce à l'utilisation des données nationales (généalogies, génotypes et identification) et des données de séquence, cette thèse a permis d'améliorer la connaissance des déterminismes génétiques de la mortalité en élevage laitier.

Les nouvelles connaissances sur ces mutations causales ainsi que l'utilisation des modèles quantitatifs par les professionnels de la filière en parallèle d'une meilleure gestion technique auront un effet immédiat et durable sur la réduction de la mortalité en élevage. Aussi, de multiples anomalies n'ont pas été encore reliées à des mécanismes génétiques sous-jacent précis et demanderont des analyses plus poussées dans l'avenir, notamment dans les races allaitantes.

J'aimerais remercier toutes les personnes qui auront aidé de près ou de loin à la réalisation de ce manuscrit, j'ai beaucoup appris lors de ces trois années de travail. J'espère que ces travaux auront un impact positif pour l'agriculture française, la filière bovine et en particulier les éleveurs.

Références bibliographiques

- Adams H. A., Sonstegard T.S., VanRaden P.M., Null D.J., Van Tassell C.P., Larkin D.M., and Lewin H.A. 2016. 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* 99 (8): 6693–6701. <https://doi.org/10.3168/jds.2015-10517>.
- Agreste. 2023. Graph'Agri 2023. <https://agreste.agriculture.gouv.fr/agreste-web/disaron/GraFra2023Integral/detail/>.
- Aguilar I., Misztal I., Johnson D.L., Legarra A., Tsuruta S., and Lawlor T.J. 2010. Hot Topic: A Unified Approach to Utilize Phenotypic, Full Pedigree, and Genomic Information for Genetic Evaluation of Holstein Final Score. *Journal of Dairy Science* 93 (2): 743–52. <https://doi.org/10.3168/jds.2009-2730>.
- Al-Khudhair A., VanRaden P.M., Null D.J., Neupane M., McClure M.C., and Dechow C.D. 2024. New Mutation within a Common Haplotype Is Associated with Calf Muscle Weakness in Holsteins. *Journal of Dairy Science*, 107(6):3768-3779. <https://doi.org/10.3168/jds.2023-24121>.
- Amberger J.S., Bocchini C.A., Scott A.F., and Hamosh A. 2019. OMIM.Org: Leveraging Knowledge across Phenotype–Gene Relationships. *Nucleic Acids Research* 47 (D1): D1038–43. <https://doi.org/10.1093/nar/gky1151>.
- Barbat A., Leclerc H., Philippe M., Fritz S., Daviere J.B., Guillaume F., De Bretagne T., Boichard D. 2016. GénoSanté: améliorer la santé productive des vaches laitières par la sélection génomique et la conduite d'élevage: une première étape avec l'acétonémie. *23èmes Rencontres Recherches Ruminants, Décembre 2016, Paris*. <https://www.journees3r.fr/spip.php?article4229>
- Becker D., Weikard R., Schulze C., Wohlsein P., and Kühn C. 2020. A 50-Kb Deletion Disrupting the RSPO2 Gene Is Associated with Tetradysmelia in Holstein Friesian Cattle. *Genetics Selection Evolution* 52 (1): 68. <https://doi.org/10.1186/s12711-020-00586-y>.
- Bérodier M., Brochard M., Boichard D., Minery S., and Ducrocq V. 2020. Plans d'accouplements sur la base d'informations génomiques: mieux conseiller les éleveurs en utilisant mieux toute l'information disponible. *25èmes Rencontre Recherche Ruminants, 2020, Paris*. <https://hal.inrae.fr/hal-03058008>
- Besnard F., Leclerc H., Boussaha M., Grohs C., Jewell N., Pinton A., Barasc H., et al. 2022. Detailed Analysis of Mortality Rates in the Female Progeny of 1,001 Holstein Bulls Allows the Discovery of New Dominant Genetic Defects. *Journal of Dairy Science*, 106(1):439-451. <https://doi.org/10.3168/jds.2022-22365>.
- Besnard F., Leclerc H., Cesbron N., Boussaha M., Grohs C., Jewell N., Pinton A., et al. 2022. Detailed Analysis of Mortality Rates in the Female Progeny of Holstein Bulls Allows the Discovery of New Dominant Genetic Defects. In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*, article 525, 2178–81. Wageningen Academic Publishers. https://doi.org/10.3920/978-90-8686-940-4_525.
- Besnard F., Guintard A., Grohs C., Guzylack-Piriou L., Cano M., Escouflaire C., Hozé C., et al. 2023. Massive Detection of Cryptic Recessive Genetic Defects in Cattle Mining Millions of Life Histories. *bioRxiv*. <https://doi.org/10.1101/2023.09.22.558782>.
- Binsbergen R. van, Bink M.C., Calus M.P., van Eeuwijk F.A., Hayes B.J., Hulsegge I., and Veerkamp R.F. 2014. Accuracy of Imputation to Whole-Genome Sequence Data in Holstein Friesian Cattle. *Genetics Selection Evolution* 46 (1): 41. <https://doi.org/10.1186/1297-9686-46-41>.
- Boichard D., Maignel L., and Verrier E. 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution* 29 (1): 5. <https://doi.org/10.1186/1297-9686-29-1-5>.
- Boichard D. 2020. Nouvelles Modalités de Diffusion de La Génétique. In: *Génétique Des Animaux d'élevage: Diversité et Adaptation Dans Un Monde Changeant*, 193–203. Quae.

- <https://www.quae.com/produit/1635/9782759231003/genetique-des-animaux-d-elevage>.
- Boichard D., Chung H., Dassonneville R., David X., Eggen A., Fritz S., Gietzen K.J., et al. 2012. Design of a Bovine Low-Density SNP Array Optimized for Imputation. *PLoS ONE* 7 (3): e34130. <https://doi.org/10.1371/journal.pone.0034130>.
- Boichard D., Grohs C., Michot P., Danchin-Burge C., Capitan A., Genestout L., Barbier S., Fritz S. 2016. Prise en compte des anomalies génétiques en sélection : le cas des bovins. *INRA Productions Animales*, 29(5) : 351-358.
- Boichard D., and Manfredi E. 1994. Genetic Analysis of Conception Rate in French Holstein Cattle. *Acta Agriculturae Scandinavica, Section A — Animal Science* 44 (3): 138–45. <https://doi.org/10.1080/09064709409410890>.
- Bonaiti B., Boichard D., Verrier E., Ducrocq V., Barbat A., and Briend M. 1990. La méthode française d'évaluation génétique des reproducteurs laitiers. *INRA Productions Animales* 3 (2): 83–92. <https://doi.org/10.20870/productions-animales.1990.3.2.4363>.
- Boubet B. 2021. Dossier-Pathologies-Respiratoires. <https://www.Gdscreuse.Fr/Wp-Content/Uploads/2017/10/2021-01-DOSSIER-PATHOLOGIES-RESPIRATOIRES.Pdf>.
- Bourneuf E., Otz P., Pausch H., Jagannathan V., Michot P., Grohs C., Piton G., et al. 2017. Rapid discovery of de novo deleterious mutations in cattle enhances the value of livestock as model species. *Scientific Reports* 7 (1): 11466. <https://doi.org/10.1038/s41598-017-11523-3>.
- Boussaha M., Boulling A., Wolgust V., Bourgeois-Brunel L., Michot P., Grohs C., Gaiani N., et al. 2023a. Integrin Alpha 6 Homozygous splice-site mutation causes a new form of Junctional Epidermolysis Bullosa in Charolais Cattle. *Genetics Selection Evolution* 55: 40. <https://doi.org/10.1186/s12711-023-00814-1>.
- Boussaha M., Eché C., Escouflaire C., Grohs C., Iampietro C., Capitan A., Milan D., et al. 2023b. Search for new mutations in cattle by systemac whole genome resequencing. *39th International Society of Animal Genetics Conference*, Jul 2023, Cape Town, South Africa. <https://hal.inrae.fr/hal-04170956>.
- Boussaha M., Michot P., Letaief R., Hozé C., Fritz S., Grohs C., Esquerré D., et al. 2016. Construction of a large collection of small genome variations in french dairy and beef breeds using whole-genome sequences. *Genetics Selection Evolution* 48 : 87. <https://doi.org/10.1186/s12711-016-0268-z>.
- Brochard M., Boichard D., Capitan A., Fayolle G., Fritz S., Nicod L., and Vacelet J.M. 2018. pANO, le risque d'anomalie létale pour les produits d'accouplements: principe et utilisation en race Montbéliarde sur la zone Gen'IAtest. *24èmes Rencontres Recherches Ruminants*, 2018, Paris. https://journées3r.fr/IMG/pdf/texte_7_genetique_m-brochard-2.pdf
- Bursted B., Zamariolli M., Teixeira Bellucco F., and Melaragno M.I. 2022. Mechanisms of structural chromosomal rearrangement formation. *Molecular Cytogenetics* 15 (1): 23. <https://doi.org/10.1186/s13039-022-00600-6>.
- Capitan A., Allais-Bonnet A., Pinton A., Marquant-Le Guienne B., Le Bourhis D., Grohs C., Bouet S., et al. 2012. A 3.7 Mb deletion encompassing zeb2 causes a novel polled and multisystemic syndrome in the progeny of a somatic mosaic bull. *PloS One* 7: e49084. <https://doi.org/10.1371/journal.pone.0049084>.
- Capitan A., Grohs C., Gautier M., and Eggen A. 2009. The Scurs Inheritance: New insights from the french Charolais breed. *BMC Genetics* 10: 33. <https://doi.org/10.1186/1471-2156-10-33>.
- Cavanagh J.A.L., Tammen I., Windsor P.A., Bateman J.F., Savarirayan R., Nicholas F.W., and Raadsma H.W. 2007. Bulldog dwarfism in Dexter cattle Is caused by mutations in ACAN. *Mammalian Genome* 18: 808–14. <https://doi.org/10.1007/s00335-007-9066-9>.
- Charlier C., Agerholm J.S., Coppieters W., Karlskov-Mortensen P., Li W., de Jong G., Fasquelle C., et al. 2012. A Deletion in the bovine FANCI gene compromises fertility by causing fetal

- death and brachyspina. *PLOS ONE* 7: e43085. <https://doi.org/10.1371/journal.pone.0043085>.
- Charlier C., Coppieters W., Rollin F., Desmecht D., Agerholm J.S., Cambisano N., Carta E., et al. 2008. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics* 40: 449–54. <https://doi.org/10.1038/ng.96>.
- Charlier C., Li W., Harland C., Littlejohn M., Coppieters W., Creagh F., Davis S., et al. 2016. NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. *Genome Research* 26: 1333–41. <https://doi.org/10.1101/gr.207076.116>.
- Christensen O.F., and Lund M.S. 2010. Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42: 2. <https://doi.org/10.1186/1297-9686-42-2>.
- Cingolani P., Platts A., Wang L.L., Coon M., Nguyen T., Wang L., Land S.J., Lu X., and Ruden D.M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the Genome of *Drosophila Melanogaster* Strain W1118; Iso-2; Iso-3. *Fly* 6: 80–92. <https://doi.org/10.4161/fly.19695>.
- Cole J. 2015. A simple strategy for managing many recessive disorders in a dairy cattle breeding program. *Genetics Selection Evolution* 47: 94. <https://doi.org/10.1186/s12711-015-0174-9>.
- Colleau J.J., Fritz S., Guillaume F., Baur A., Dupassieux D., Journaux L., Eggen A., and Boichard D. 2015. Simulation des potentialités de la sélection génomique chez les bovins laitiers. *INRA Productions Animales* 28: 251–58. <https://doi.org/10.20870/productions-animales.2015.28.3.3030>.
- Corbeau J., Grohs C., Jourdain J., Boussaha M., Besnard F., Barbat A., Plassard V., et al. 2024. A recurrent de novo missense mutation in COL1A1 causes Osteogenesis Imperfecta Type II and preterm delivery in Normande cattle. *Genetics Selection Evolution* 56: 39. <https://doi.org/10.1186/s12711-024-00909-3>.
- Crooks G.E., Hon G., Chandonia J.M., and Brenner S.E. 2004. WebLogo: A sequence logo generator. *Genome Research* 14: 1188–90. <https://doi.org/10.1101/gr.849004>.
- Cuttance E., and Laven R. 2019. Perinatal mortality risk factors in dairy calves. *Veterinary Journal* 253: 105394. <https://doi.org/10.1016/j.tvjl.2019.105394>.
- Daetwyler H.D., Capitan A., Pausch H., Stothard P., Van Binsbergen R., Brøndum R.F., Liao X., et al. 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nature Genetics* 46: 858–65. <https://doi.org/10.1038/ng.3034>.
- Danchin-Burge C., Danvy S., Laloë D., and Verrier E. 2017. Création d'un observatoire de la VARIabilité génétique des RUMinants et des Equidés (VARUME). *Innovations Agronomiques* 55, 235–245. <https://doi.org/10.15454/1.5137783633347627E12>.
- Dempster E.R., and Lerner I.M. 1950. Heritability of threshold characters. *Genetics* 35: 212–36.
- Duchesne A., Gautier M., Chadi S., Grohs C., Floriot S., Gallard Y., Caste G., Ducos A., and Eggen A. 2006. Identification of a doublet missense substitution in the bovine LRP4 gene as a candidate causal mutation for Syndactyly in Holstein cattle. *Genomics* 88: 610–21. <https://doi.org/10.1016/j.ygeno.2006.05.007>.
- Duchesne A., Vaiman A., Castille J., Beauvallet C., Gaignard P., Floriot S., Rodriguez S., et al. 2017. Bovine and Murine models highlight novel roles for SLC25A46 in mitochondrial dynamics and metabolism, with implications for human and animal health. *PLOS Genetics* 13: e1006597. <https://doi.org/10.1371/journal.pgen.1006597>.
- Duchesne A., Vaiman A., Frah M., Floriot S., Legoueix-Rodriguez S., Desmazières A., Fritz S., et al. 2018. Progressive Ataxia of Charolais cattle highlights a role of KIF1C in sustainable myelination. *PLOS Genetics* 14: e1007550. <https://doi.org/10.1371/journal.pgen.1007550>.
- Escouflaire C., Rebours E., Charles M., Orellana S., Cano M., Rivière J., Grohs C., Hayes H., and Capitan A. 2019. A de Novo 3.8-Mb Inversion affecting the EDA and XIST genes in a

- heterozygous female calf with generalized hypohidrotic ectodermal dysplasia. *BMC Genomics* 20: 715. <https://doi.org/10.1186/s12864-019-6087-1>.
- Falconer D.S. 1996. *Introduction to Quantitative Genetics*. Pearson Education India.
- Fasquelle C., Sartelet A., Li W., Dive M., Tamma N., Michaux C., Druet T., et al. 2009. 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 Genetics* 5: e1000666. <https://doi.org/10.1371/journal.pgen.1000666>.
- Fernando R.L., Cheng H., and Garrick D.J. 2016. An efficient exact method to obtain GBLUP and Single-Step GBLUP when the genomic relationship matrix is singular. *Genetics Selection Evolution* 48: 80. <https://doi.org/10.1186/s12711-016-0260-7>.
- Fingerlin T.E., Murphy E., Zhang W., Peljto A.L, Brown K.K., Steele M.P., Loyd J.E., et al. 2013. Genome-wide association study identifies multiple susceptibility loci for Pulmonary Fibrosis. *Nature Genetics* 45: 613–20. <https://doi.org/10.1038/ng.2609>.
- Fishman L., and McIntosh M. 2019. Standard deviations: the biological bases of transmission ratio distortion. *Annual Review of Genetics* 53: 347–72. <https://doi.org/10.1146/annurev-genet-112618-043905>.
- Floriot S., Vesque C., Rodriguez S., Bourgain-Guglielmetti F., Karaiskou A., Gautier M., Duchesne A., et al. 2015. C-Nap1 mutation affects centriole cohesion and is associated with a Seckel-like syndrome in cattle. *Nature Communications* 6: 6894. <https://doi.org/10.1038/ncomms7894>.
- Foulley J.L., Gianola D., and Thompson R. 1983. Prediction of genetic merit from data on binary and quantitative variates with an application to calving difficulty, birth weight and pelvic opening. *Génétique Sélection Évolution* 15: 401–24. <https://doi.org/10.1051/gse:19830305>.
- Frappeau V. 2023. Chiffres clés du dispositif génétique français 2022. https://idele.fr/?eID=cmis_download&oID=workspace%3A%2F%2FSpacesStore%2F5147239d-fcb0-443e-aa88-9923f7dcdd60&cHash=931fca920c222b56261bd72385043302.
- Fresco S., Boichard D., Fritz S., Lefebvre R., Barbey S., Gaborit M., and Martin P. 2023. Comparison of methane production, intensity, and yield throughout lactation in Holstein cows. *Journal of Dairy Science* 106: 4147–57. <https://doi.org/10.3168/jds.2022-22855>.
- Fritz S., Hoze C., Rebours E., Barbat A., Bizard M., Chamberlain A., Escouflaire C., Vander Jagt C., Boussaha M., Grohs C., Allais-Bonnet A., Philippe M., Vallee A., Amigues Y., Hayes B.J., Boichard D., Capitan A. 2018. An initiator codon mutation in SDE2 causes recessive embryonic lethality in Holstein cattle. *J Dairy Sci*, 101, 6220-6231. <https://doi.org/10.3168/jds.2017-14119>.
- Fritz S., Capitan A., Djari A., Rodriguez S.C., Barbat A., Baur A., Grohs C., et al. 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>.
- Fuerst-Waltl B., and Sørensen M.K. 2010. Genetic analysis of calf and heifer losses in Danish Holstein. *Journal of Dairy Science* 93: 5436–42. <https://doi.org/10.3168/jds.2010-3227>.
- Gonzalez-Peña D., Vukasinovic N., Brooker J.J., Przybyla C.A., and DeNise S.K. 2019. Genomic evaluation for calf wellness traits in Holstein cattle. *Journal of Dairy Science* 102: 2319–29. <https://doi.org/10.3168/jds.2018-15540>.
- Gourreau J.M. 2011. *Guide pratique des maladies des bovins*. Agriproduction. Paris: Éd. France agricole Institut de l'élevage.
- Grobet L., Martin L.J., Poncelet D., Pirottin D., Brouwers B., Riquet J., Schoeberlein A., et al. 1997. A deletion in the bovine Myostatin gene causes the double-muscléd phenotype in cattle. *Nature Genetics* 17: 71–74. <https://doi.org/10.1038/ng0997-71>.
- Grohs C., Corbeau J., Jourdain J., Boussaha M., Plassard V., Barbat A., Hamelin C., et al. 2022. A *de novo* missense mutation of COL1A1 causes Osteogenesis Imperfecta Type 2 and

- premature delivery in Normande Cattle. In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*, paper 521, 2162–65. Rotterdam, the Netherlands. https://doi.org/10.3920/978-90-8686-940-4_521.
- Grohs C., Duchesne A., Floriot S., Deloche M.C., Boichard D., Ducos A., and Danchin-Burge C. 2016. L'Observatoire National des Anomalies Bovines, son action et ses résultats pour une aide efficace à la gestion des anomalies génétiques, *INRA Productions Animales*, 29, 307-31811. <https://doi.org/10.20870/productions-animales.2016.29.5.2998>
- Gross J. J., Schwinn A.-C., Schmitz-Hsu F., Menzi F., Drögemüller C., Albrecht C., and Bruckmaier R.M. 2016. Rapid communication: Cholesterol Deficiency–Associated APOB mutation impacts lipid metabolism in Holstein calves and breeding bulls. *Journal of Animal Science* 94: 1761–66. <https://doi.org/10.2527/jas.2016-0439>.
- Häfliger I.M., Spengeler M., Seefried F.R., and Drögemüller C. 2022. Four novel candidate causal variants for deficient homozygous haplotypes in Holstein cattle. *Scientific Reports* 12: 5435. <https://doi.org/10.1038/s41598-022-09403-6>.
- Häfliger I.M., Hofstetter S., Mock T., Stettler M.H., Meylan M., Mehinagic K., Stokar-Regenscheit N., and Drögemüller C. 2019. APOB-associated Cholesterol Deficiency in Holstein cattle is not a simple recessive disease. *Animal Genetics* 50: 372–75. <https://doi.org/10.1111/age.12801>.
- Hazel L.N. 1943. The genetic basis for constructing selection indexes. *Genetics* 28: 476–90.
- Howie B., Fuchsberger C., Stephens M., Marchini J., and Abecasis G.R. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics* 44: 955–59. <https://doi.org/10.1038/ng.2354>.
- Hozé C., Escoufflaire C., Mesbah-Uddin M., Barbat A., Boussaha M., Deloche M.C., Boichard D., Fritz S., and Capitan A. 2020. Short Communication: A splice site mutation in CENPU is associated with recessive embryonic lethality in Holstein cattle. *Journal of Dairy Science* 103: 607–12. <https://doi.org/10.3168/jds.2019-17056>.
- Hozé C., Fritz S., Baur A., Grohs C., Danchin-Burge C., and Boichard D. 2018. Prise en compte des gènes d'intérêt dans les objectifs de sélection en bovins laitiers. In *24èmes Rencontres Recherches Ruminants*, Dec 2018, Paris, France. <https://hal.inrae.fr/hal-02735939>.
- Huang L.O., Labbe A., and Infante-Rivard C. 2013. Transmission ratio distortion: review of concept and implications for genetic association studies. *Human Genetics* 132: 245–63. <https://doi.org/10.1007/s00439-012-1257-0>.
- Id-Lahoucine S., Cánovas A., Jatón C., Miglior F., Fonseca P.A.S., Sargolzaei M., Miller S., Schenkel F.S., Medrano J.F., and Casellas J. 2019. Implementation of Bayesian methods to identify SNP and haplotype regions with transmission ratio distortion across the whole genome: TRDscan v.1.0. *Journal of Dairy Science* 102: 3175–88. <https://doi.org/10.3168/jds.2018-15296>.
- Id-Lahoucine S., Casellas J., Lu D., Sargolzaei M., Miller S., and Cánovas A. 2023. Distortion of Mendelian segregation across the Angus cattle genome uncovering regions affecting reproduction. *Scientific Reports* 13: 13393. <https://doi.org/10.1038/s41598-023-37710-z>.
- INRAE. 2024. Des indicateurs génomiques de résistance à la paratuberculose disponibles en race Normande. April 2024. <https://www.inrae.fr/actualites/indicateurs-genomiques-resistance-paratuberculose-disponibles-race-normande>.
- Jamrozik J., Fatehi J., Kistemaker G.J., and Schaeffer L.R. 2005. Estimates of genetic parameters for Canadian Holstein female reproduction traits. *Journal of Dairy Science* 88: 2199–2208. [https://doi.org/10.3168/jds.S0022-0302\(05\)72895-2](https://doi.org/10.3168/jds.S0022-0302(05)72895-2).
- Jourdain J., Barasc H., Faraut T., Calgaro A., Bonnet N., Marcuzzo C., Suin A., et al. 2023. Large-scale detection and characterization of interchromosomal rearrangements in normozoospermic bulls using massive genotype and phenotype data sets. *Genome Research* 33: 957–71. <https://doi.org/10.1101/gr.277787.123>.

- Kipp S., Segelke D., Schierenbeck S., Reinhardt F., Reents R., Wurmser C., Pausch H., et al. 2016. Identification of a haplotype associated with Cholesterol Deficiency and increased juvenile mortality in holstein cattle. *Journal of Dairy Science* 99: 8915–31. <https://doi.org/10.3168/jds.2016-11118>.
- Kishimoto K., Kato A., Osada S., Nishizuka M., and Imagawa M. 2010. FAD104, a positive regulator of adipogenesis, negatively regulates osteoblast differentiation. *Biochemical and Biophysical Research Communications* 397: 187–91. <https://doi.org/10.1016/j.bbrc.2010.05.077>.
- Kishimoto K., Nishizuka M., Kato D., Kato A., Osada S., and Imagawa M. 2013. FAD104, a regulatory factor of adipogenesis, acts as a novel regulator of calvarial bone formation. *The Journal of Biological Chemistry* 288: 31772–83. <https://doi.org/10.1074/jbc.M113.452961>.
- Kolishovski G., Lamoureux A., Hale P., Richardson J.E., Recla J.M., Adesanya O., Simons A., Kunde-Ramamoorthy G., and Bult C.J. 2019. The JAX synteny browser for mouse-human comparative genomics. *Mammalian Genome* 30: 353–61. <https://doi.org/10.1007/s00335-019-09821-4>.
- Leclerc H., Lefebvre R., Douguet M., Phocas F., and Mattalia S. 2016. Mortalité des veaux : analyse phénotypique et étude de la composante génétique. 23. *Rencontres autour des Recherches sur les Ruminants*, Dec 2016, Paris. <https://hal.archives-ouvertes.fr/hal-01608259/>.
- Legarra A., and Misztal I. 2008. Technical Note: Computing strategies in genome-wide selection. *Journal of Dairy Science* 91: 360–66. <https://doi.org/10.3168/jds.2007-0403>.
- Lund, Mogens S., Adrianus PW de Roos, Alfred G. de Vries, Tom Druet, Vincent Ducrocq, Sébastien Fritz, François Guillaume, et al. 2011. « A common reference population from four European Holstein populations increases reliability of genomic predictions ». *Genetics Selection Evolution* 43 (1): 43. <https://doi.org/10.1186/1297-9686-43-43>.
- Li W., Sartelet A., Tamma N., Coppieters W., Georges M., and Charlier C. 2016. Reverse genetic screen for loss-of-function mutations uncovers a frameshifting deletion in the Melanophilin gene accountable for a distinctive coat color in Belgian Blue Cattle. *Animal Genetics* 47: 110–13. <https://doi.org/10.1111/age.12383>.
- Li Y., Willer C., Sanna S., and Abecasis G. 2009. Genotype imputation. *Annual Review of Genomics and Human Genetics* 10:387–406. <https://doi.org/10.1146/annurev.genom.9.081307.164242>.
- Lynch C., Leishman E. M., Miglior F., Kelton D., Schenkel F., and Baes C.F. 2024. Review: Opportunities and challenges for the genetic selection of dairy calf disease traits. *Animal*, 18, 101141. <https://doi.org/10.1016/j.animal.2024.101141>.
- Maigret C. 2024. Rapport Bovins 2023 Productions Lait et Viande. https://idele.fr/inosys-reseaux-elevage/?eID=cmis_download&oID=workspace%3A%2F%2FSpacesStore%2F5173106d-ab95-464e-8ce4-073ec3c92720&cHash=6da68eec8bbb3ed5cd604af16949809e.
- Martin F. J., Amode M.R., Aneja A., Austine-Orimoloye O., Azov A.G., Barnes I., Becker A., et al. 2023. Ensembl 2023. *Nucleic Acids Research* 51 (D1): D933–41. <https://doi.org/10.1093/nar/gkac958>.
- Mattalia S., and Laloë D. 2009. UMT Evaluation génétique des bovins : bilan des travaux menés en 2008; projets pour 2009. Technical Report. Institut de l’Elevage. <https://hal.science/hal-01193622>.
- Matukumalli L.K., Lawley C.T., Schnabel R.D., Taylor J.F., Allan M.F., Heaton M.P., O’Connell J., et al. 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4: e5350. <https://doi.org/10.1371/journal.pone.0005350>.
- McClure M.C., Bickhart D., Null D., VanRaden P.M., Xu L., Wiggans G.R., Liu G., et al. 2014. 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: e92769. <https://doi.org/10.1371/journal.pone.0092769>.

- McLaren W., Gil L., Hunt S.E., Riat H.S., Ritchie G.R.S., Thormann A., Flicek P., and Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biology* 17: 122. <https://doi.org/10.1186/s13059-016-0974-4>.
- McVean G.A., Altshuler D.M., Durbin R.M., Abecasis G.R., Bentley D.R., Chakravarti A., Clark A.G., et al. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491 (7422): 56–65. <https://doi.org/10.1038/nature11632>.
- Menzi F., Besuchet-Schmutz N., Fragnière M., Hofstetter S., Jagannathan V., Mock T., Raemy A., et al. 2016. A transposable element insertion in *APOB* causes cholesterol deficiency in Holstein cattle. *Animal Genetics* 47: 253–57. <https://doi.org/10.1111/age.12410>.
- Mesbah-Uddin M., Hoze C., Michot P., Barbat A., Lefebvre R., Boussaha M., Sahana G., Fritz S., Boichard D., and Capitan A. 2019. A missense mutation (p.Tyr452Cys) in the CAD gene compromises reproductive success in French Normande cattle. *Journal of Dairy Science* 102: 6340–56. <https://doi.org/10.3168/jds.2018-16100>.
- Meuwissen T.H.E., Hayes B.J., and Goddard M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–29. <https://doi.org/10.1093/genetics/157.4.1819>
- Michot P., Chahory S., Marete A., Grohs C., Dagios D., Donzel E., Aboukadiri A., et al. 2016. A reverse genetic approach identifies an ancestral frameshift mutation in RP1 causing recessive progressive retinal degeneration in European cattle breeds. *Genetics Selection Evolution* 48: 56. <https://doi.org/10.1186/s12711-016-0232-y>.
- Michot P., Fantini O., Braque R., Allais-Bonnet A., Saintilan R., Grohs C., Barbieri J., et al. 2015. 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. *Genetics Selection Evolution* 47: 37. <https://doi.org/10.1186/s12711-015-0110-z>.
- Michot P., Fritz S., Barbat A., Boussaha M., Deloche M.C., Grohs C., Hozé C., et al. 2017. 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* 100: 8176–87. <https://doi.org/10.3168/jds.2017-12579>.
- Ministère de l'agriculture de l'agroalimentaire et de la forêt. 2016. *Le règlement zootechnique européen*. <https://agriculture.gouv.fr/le-reglement-zootechnique-europeen-un-nouveau-cadre-reglementaire-europeen-pour-la-genetique>.
- Misztal I., Lourenco D., Aguilar I., Legarra A., and Vitezica Z. 2014. Manual for BLUPF90 Family of Programs. https://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all.pdf
- Mock T., Mehinagic K., Menzi F., Studer E., Oevermann A., Stoffel M.H., Drögemüller C., Meylan M., and Regenscheit N. 2016. Clinicopathological phenotype of autosomal recessive Cholesterol Deficiency in Holstein cattle. *Journal of Veterinary Internal Medicine* 30: 1369–75. <https://doi.org/10.1111/jvim.13976>.
- Neupane M., Hutchison J.L., Van Tassell C.P., and VanRaden P.M. 2021. Genomic Evaluation of Dairy Heifer Livability. *Journal of Dairy Science*, 104, 8959-8965 <https://doi.org/10.3168/jds.2020-19687>.
- Ng P.C., and Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research* 31: 3812–14. <https://doi.org/10.1093/nar/gkg509>.
- Nishizuka M., Kishimoto K., Kato A., Ikawa M., Okabe M., Sato R., Niida H., Nakanishi M., Osada S., and Imagawa M. 2009. Disruption of the novel gene FAD104 causes rapid postnatal death and attenuation of cell proliferation, adhesion, spreading and migration. *Experimental Cell Research* 315: 809–19. <https://doi.org/10.1016/j.yexcr.2008.12.013>.
- Pedersen J., Kargo M., Fogh A., Pösö J., Eriksson J.A., Sander Nielsen U., and Pedersen Aamand G. 2015. Note on Economic Value of Young Stock Survival, August, 11. <https://www.nordicebv.info/wp-content/uploads/2015/10/Economic-value-of-Young-Stock-Survival.pdf>

- Pelt M.L. van, Eding H., Vessies P., and de Jong G. 2012. Developing a genetic evaluation for calf survival during rearing in The Netherlands. *Interbull Bulletin* 46, 5. <https://journal.interbull.org/index.php/ib/article/view/1741>
- Quaas R.L. 1988. Additive genetic model with groups and relationships. *Journal of Dairy Science* 71:91–98. [https://doi.org/10.1016/S0022-0302\(88\)79986-5](https://doi.org/10.1016/S0022-0302(88)79986-5).
- Rendel J.M., and Robertson A. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. *Journal of Genetics* 50: 1–8. <https://doi.org/10.1007/BF02986789>.
- Reynolds E.G.M., Neeley C., Lopdell T.J., Keehan M., Dittmer K., Harland C.S., Couldrey C., et al. 2021. Non-additive association analysis using proxy phenotypes identifies novel cattle syndromes. *Nature Genetics* 53: 949–54. <https://doi.org/10.1038/s41588-021-00872-5>.
- Rosen B.D., Bickhart D.M., Schnabel R.D., Koren S., Elsik C.G., Tseng E., Rowan T.N., et al. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. *GigaScience* 9: giaa021. <https://doi.org/10.1093/gigascience/giaa021>.
- Santman-Berends I.M.G.A., Schukken Y.H., and van Schaik G. 2019. Quantifying calf mortality on dairy farms: challenges and solutions. *Journal of Dairy Science* 102: 6404–17. <https://doi.org/10.3168/jds.2019-16381>.
- Schütz E., Wehrhahn C., Wanjek M., Bortfeld R., Wemheuer W.E., Beck J., and Brenig B. 2016. 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* 11: e0154602. <https://doi.org/10.1371/journal.pone.0154602>.
- Schwenger B., Schöber S., and Simon D. 1993. DUMPS cattle carry a point mutation in the Uridine Monophosphate Synthase gene. *Genomics* 16 (1): 241–44. <https://doi.org/10.1006/geno.1993.1165>.
- Segelke D., Täubert H., Reinhardt F., and Thaller G. 2016. Considering genetic characteristics in German Holstein breeding programs. *Journal of Dairy Science* 99: 458–67. <https://doi.org/10.3168/jds.2015-9764>.
- Sellier P., Boichard D., and Verrier E. 2019. La génétique animale à l'INRA: soixante ans d'une histoire scientifique en prise avec le monde de la sélection et riche en rebondissements technologiques. *Histoire de la Recherche Contemporaine*, 8, 86–97. <https://doi.org/10.4000/hrc.3222>.
- Shuster D.E., Kehrli M.E., Ackermann M.R., and Gilbert R.O. 1992. Identification and prevalence of a genetic defect that causes Leukocyte Adhesion Deficiency in Holstein cattle. *Proceedings of the National Academy of Sciences* 89: 9225–29. <https://doi.org/10.1073/pnas.89.19.9225>.
- Stevenson J.S., and Britt J.H. 2017. A 100-Year Review: practical female reproductive management. *Journal of Dairy Science* 100: 10292–313. <https://doi.org/10.3168/jds.2017-12959>.
- Tajima M., Miyake S., Takehana K., Kobayashi A., Yamato O., and Maede Y. 1999. Gene defect of Dermatan Sulfate Proteoglycan of cattle affected with a variant form of Ehlers-Danlos Syndrome. *Journal of Veterinary Internal Medicine* 13: 202–5. [https://doi.org/10.1892/0891-6640\(1999\)013<0202:gdodsp>2.3.co;2](https://doi.org/10.1892/0891-6640(1999)013<0202:gdodsp>2.3.co;2)
- Thomsen B., Horn P., Panitz F., Bendixen E., Petersen A.H., Holm L.E., Nielsen V.H., Agerholm J.S., Arnbjerg J., and Bendixen C. 2006. A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-Acetylglucosamine Transporter, causes complex vertebral malformation. *Genome Research* 16: 97–105. <https://doi.org/10.1101/gr.3690506>.
- Vaiman A., Fritz S., Beauvallet C., Boussaha M., Grohs C., Daniel-Carliet N., Relun A., Boichard D., Vilotte J.L., and Duchesne A. 2022. Mutation of the MYH3 gene causes recessive Cleft Palate in Limousine cattle. *Genetics Selection Evolution* 54: 71. <https://doi.org/10.1186/s12711-022-00762-2>.
- VanRaden P.M. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* 91: 4414–23. <https://doi.org/10.3168/jds.2007-0980>.

- VanRaden P.M., Olson K.M., Null D.J., and Hutchison J.L. 2011. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *Journal of Dairy Science* 94: 6153–61. <https://doi.org/10.3168/jds.2011-4624>.
- Veit H.P., and Farrell R.L. 1978. The anatomy and physiology of the bovine respiratory system relating to pulmonary disease. *The Cornell Veterinarian* 68: 555–81.
- Weller J.I., Gershoni M., Ezra E. 2021. Genetic and environmental analysis of female calf survival in the Israel Holstein cattle population. *Journal of Dairy Science* 104:3278-3291. <https://doi.org/10.3168/jds.2020-19434>.
- Wright S. 1922. Coefficients of inbreeding and relationship. *The American Naturalist* 56: 330–38.
- Yang J., Lee S.H., Goddard M.E., and Visscher P.M. 2011. GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* 88: 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>.
- Yang, Y., Si J., Lv X., Dai D., Liu L., Tang S., Wang Y., Zhang S., Xiao W., and Zhang Y. 2022. Integrated analysis of whole genome and transcriptome sequencing reveals a frameshift mutation associated with recessive embryonic lethality in Holstein cattle. *Animal Genetics* 53: 137–41. <https://doi.org/10.1111/age.13160>.
- Yuzbasiyan-Gurkan V., and Bartlett E. 2006. Identification of a unique splice site variant in *SLC39A4* in bovine Hereditary Zinc Deficiency, Lethal Trait A46: an animal model of acrodermatitis enteropathica. *Genomics* 88: 521–26. <https://doi.org/10.1016/j.ygeno.2006.03.018>.

Annexe

Annexe 1: Tableau contenant les SNP mis sur la puce EuroGMD V12.

Annexe 2: Résultats de l'analyse génomique pour la distorsion de ségrégation des haplotypes dans 1726 familles de taureaux Holstein, 974 familles de Montbéliarde et 479 familles de Normande.

Annexe 1: Tableau contenant les SNP mis sur la puce EuroGMD V12.

Le tableau contient les informations sur le nom du variant dans le manifest de la puce, les chromosomes, les positions de début des SNP et des petits InDel (insertions/Délétions) avec les allèles référence (REF) et alternatives (ALT) ainsi que le TYPE de variant.

Nom du variant	Chr	Position	REF	ALT	TYPE
FR12_EFF_FB_R66_5_27121368	5	27121368	T	C	SNP
FR12_EFF_FB_R66_5_27137418	5	27137418	GCTGGTGGCTTTGGAGGA	G	INDEL
FR12_EFF_FB_R66_5_94125315	5	94125315	G	T	SNP
FR12_EFF_FB_R66_5_94876545	5	94876545	G	C	SNP
FR12_EFF_FB_R66_5_94876810	5	94876810	G	A	SNP
FR12_EFF_FB_R66_5_94889653	5	94889653	C	T	SNP
FR12_EFF_FB_R66_5_94892453	5	94892453	T	G	SNP
FR12_EFF_FB_R66_5_94893199	5	94893199	G	C	SNP
FR12_EFF_FB_R66_5_94898422	5	94898422	C	T	SNP
FR12_EFF_FB_R66_5_94899418	5	94899418	C	T	SNP
FR12_EFF_FB_R66_5_94912034	5	94912034	C	T	SNP
FR12_EFF_FB_R66_5_94922078	5	94922078	G	A	SNP
FR12_EFF_FB_R66_5_94925345	5	94925345	A	G	SNP
FR12_EFF_FB_R66_5_96811047	5	96811047	G	A	SNP
FR12_EFF_FB_R66_5_100060121	5	100060121	G	A	SNP
FR12_EFF_FB_R66_11_49638557	11	49638557	G	A	SNP
FR12_EFF_FB_R66_11_49640629	11	49640629	C	T	SNP
FR12_EFF_FB_R66_11_49641869	11	49641869	C	G	SNP
FR12_EFF_FB_R66_16_64182653	16	64182653	T	C	SNP
FR12_EFF_FB_R66_16_64182675	16	64182675	G	A	SNP
FR12_EFF_FB_R66_16_64182912	16	64182912	T	C	SNP
FR12_EFF_FB_R66_16_64182957	16	64182957	G	A	SNP
FR12_EFF_FB_R66_16_64182996	16	64182996	A	G	SNP
FR12_EFF_FB_R66_16_64183522	16	64183522	T	C	SNP
FR12_EFF_FB_R66_16_64188083	16	64188083	G	T	SNP
FR12_EFF_FB_R66_16_64188174	16	64188174	T	C	SNP
FR12_EFF_FB_R66_16_64188229	16	64188229	G	GTC	INSER

FR12_EFF_FB_R66_16_64188541	16	64188541	T	G	SNP
FR12_EFF_FB_R66_23_24920830	23	24920830	A	G	SNP
FR12_EFF_FB_R66_23_24922231	23	24922231	C	T	SNP
FR12_EFF_FB_R66_23_24922635	23	24922635	C	T	SNP
FR12_EFF_FB_R66_23_24923087	23	24923087	C	T	SNP
FR12_EFF_FB_R66_23_24923189	23	24923189	C	G	SNP
FR12_EFF_FB_R66_23_24923230	23	24923230	TA	T	INDEL
FR12_EFF_FB_R66_23_24923895	23	24923895	C	G	SNP
FR12_EFF_FB_R66_23_24923910	23	24923910	G	A	SNP
FR12_EFF_FB_R66_23_24924381	23	24924381	C	A	SNP
FR12_EFF_FB_R66_23_24924389	23	24924389	G	A	SNP
FR12_EFF_FB_R66_29_29336190	29	29336190	A	C	SNP
FR12_EFF_FB_R66_29_29367341	29	29367341	A	G	SNP
FR12_EFF_FB_R66_29_29714191	29	29714191	T	TTG	INSER
FR12_EFF_FB_R66_29_29729829	29	29729829	T	C	SNP
FR12_EFF_FB_R66_29_29730070	29	29730070	C	T	SNP
FR12_EFF_FB_R66_29_29730119	29	29730119	G	T	SNP
FR12_EFF_FB_R66_29_29730201	29	29730201	C	T	SNP
FR12_EFF_FB_R46_66_11_49637951	11	49637951	C	G	SNP
FR12_EFF_FB_R46_66_11_49638997	11	49638997	A	G	SNP
FR12_EFF_FB_R46_66_11_49639150	11	49639150	T	C	SNP
FR12_EFF_FB_R46_66_11_49639291	11	49639291	A	G	SNP
FR12_EFF_FB_R46_66_11_49639606	11	49639606	G	GC	INSER
FR12_EFF_FB_R46_66_11_49639760	11	49639760	CA	C	INDEL
FR12_EFF_FB_R46_66_11_49639822	11	49639822	C	T	SNP
FR12_EFF_FB_R46_66_11_49639983	11	49639983	T	C	SNP
FR12_EFF_FB_R46_66_11_49640162	11	49640162	C	T	SNP
FR12_EFF_FB_R46_66_11_49640187	11	49640187	A	C	SNP
FR12_EFF_FB_R46_66_11_49640191	11	49640191	T	C	SNP
FR12_EFF_FB_R46_66_11_49640274	11	49640274	T	C	SNP
FR12_EFF_FB_R46_66_11_49640295	11	49640295	G	A	SNP

FR12_EFF_FB_R46_66_11_49640550	11	49640550	T	C	SNP
FR12_EFF_FB_R46_66_11_49641561	11	49641561	T	C	SNP
FR12_EFF_FB_R56_66_23_24916501	23	24916501	C	T	SNP
FR12_EFF_FB_R56_66_23_24917092	23	24917092	T	C	SNP
FR12_EFF_FB_R56_66_23_24918115	23	24918115	C	T	SNP
FR12_EFF_FB_R56_66_23_24918124	23	24918124	A	G	SNP
FR12_EFF_FB_R56_66_23_24918497	23	24918497	C	T	SNP
FR12_EFF_FB_R56_66_23_24918685	23	24918685	G	A	SNP
FR12_EFF_FB_R56_66_23_24918765	23	24918765	G	A	SNP
FR12_EFF_FB_R56_66_23_24920830	23	24920830	A	G	SNP
FR12_EFF_FB_R56_66_23_24922231	23	24922231	C	T	SNP
FR12_EFF_FB_R56_66_23_24922319	23	24922319	C	T	SNP
FR12_EFF_FB_R56_66_23_24922635	23	24922635	C	T	SNP
FR12_EFF_FB_R56_66_23_24923087	23	24923087	C	T	SNP
FR12_EFF_FB_R56_66_23_24923189	23	24923189	C	G	SNP
FR12_EFF_FB_R56_66_23_24923230	23	24923230	TA	T	INDEL
FR12_EFF_FB_R56_66_23_24923895	23	24923895	C	G	SNP
FR12_EFF_FB_R56_66_23_24923910	23	24923910	G	A	SNP
FR12_EFF_FB_R56_66_23_24924381	23	24924381	C	A	SNP
FR12_EFF_FB_R56_66_23_24924389	23	24924389	G	A	SNP
FR12_EFF_FB_R56_66_23_24926353	23	24926353	C	T	SNP
FR12_EFF_FB_R56_66_23_24927798	23	24927798	A	G	SNP
FR12_EFF_FB_R46_6_60147279	6	60147279	T	C	SNP
FR12_EFF_FB_R46_6_64582998	6	64582998	T	C	SNP
FR12_EFF_FB_R46_6_64583112	6	64583112	G	A	SNP
FR12_EFF_FB_R46_6_64586545	6	64586545	TATTAGGCATTAG	T	INDEL
FR12_EFF_FB_R46_6_67519214	6	67519214	G	A	SNP
FR12_EFF_FB_R46_6_72359797	6	72359797	C	CG	INSER
FR12_EFF_FB_R46_6_72432531	6	72432531	T	C	SNP
FR12_EFF_FB_R46_6_72517517	6	72517517	C	T	SNP
FR12_EFF_FB_R46_6_72567458	6	72567458	C	T	SNP

FR12_EFF_FB_R46_6_72567590	6	72567590	C	G	SNP
FR12_EFF_FB_R46_6_72577752	6	72577752	C	T	SNP
FR12_EFF_FB_R46_6_72579223	6	72579223	A	G	SNP
FR12_EFF_FB_R46_6_72579536	6	72579536	T	C	SNP
FR12_EFF_FB_R46_6_72579537	6	72579537	T	C	SNP
FR12_EFF_FB_R46_6_72579538	6	72579538	G	A	SNP
FR12_EFF_FB_R46_6_72579540	6	72579540	G	A	SNP
FR12_EFF_FB_R46_6_72583040	6	72583040	A	T	SNP
FR12_EFF_FB_R46_6_72583729	6	72583729	T	C	SNP
FR12_EFF_FB_R46_6_72584891	6	72584891	T	C	SNP
FR12_EFF_FB_R46_6_72584999	6	72584999	G	A	SNP
FR12_EFF_FB_R46_6_72585012	6	72585012	G	C	SNP
FR12_EFF_FB_R46_6_72585340	6	72585340	C	T	SNP
FR12_EFF_FB_R46_6_72585365	6	72585365	C	G	SNP
FR12_EFF_FB_R46_6_72586195	6	72586195	A	G	SNP
FR12_EFF_FB_R46_6_72586393	6	72586393	T	A	SNP
FR12_EFF_FB_R46_6_72651516	6	72651516	G	A	SNP
FR12_EFF_FB_R46_6_72664964	6	72664964	T	C	SNP
FR12_EFF_FB_R46_7_88705815	7	88705815	A	C	SNP
FR12_EFF_FB_R46_7_88705816	7	88705816	G	C	SNP
FR12_EFF_FB_R46_7_88707238	7	88707238	T	G	SNP
FR12_EFF_FB_R46_8_96508486	8	96508486	C	G	SNP
FR12_EFF_FB_R46_8_96569715	8	96569715	C	T	SNP
FR12_EFF_FB_R46_11_49638434	11	49638434	G	A	SNP
FR12_EFF_FB_R46_23_47872269	23	47872269	A	C	SNP
FR12_EFF_FB_R46_23_47878021	23	47878021	A	G	SNP
FR12_EFF_FB_R46_26_15491247	26	15491247	G	C	SNP
FR12_EFF_FB_R46_26_15502036	26	15502036	C	T	SNP
FR12_EFF_FB_R46_26_15518110	26	15518110	A	T	SNP
FR12_EFF_FB_R46_26_15518315	26	15518315	G	A	SNP
FR12_EFF_FB_R46_26_15520633	26	15520633	G	A	SNP

FR12_EFF_FB_R46_26_15553725	26	15553725	G	A	SNP
FR12_EFF_FB_R46_26_15553906	26	15553906	G	A	SNP
FR12_EFF_FB_R46_27_32271077	27	32271077	C	T	SNP
FR12_EFF_FB_R46_27_32271550	27	32271550	C	T	SNP
FR12_EFF_FB_R46_27_32271551	27	32271551	A	G	SNP
FR12_EFF_FB_R46_27_32272067	27	32272067	CT	C	INDEL
FR12_EFF_FB_R46_27_32276254	27	32276254	A	G	SNP
FR12_EFF_FB_R46_27_32277018	27	32277018	A	G	SNP
FR12_EFF_FB_R46_27_32277281	27	32277281	A	G	SNP
FR12_EFF_FB_R56_6_36028197	6	36028197	A	G	SNP
FR12_EFF_FB_R56_6_36234189	6	36234189	T	C	SNP
FR12_EFF_FB_R56_6_36404086	6	36404086	A	G	SNP
FR12_EFF_FB_R56_6_37402193	6	37402193	C	T	SNP
FR12_EFF_FB_R56_6_37410605	6	37410605	T	C	SNP
FR12_EFF_FB_R56_9_75584533	9	75584533	C	T	SNP
FR12_EFF_FB_R56_9_75586734	9	75586734	ATGAT	A	INDEL
FR12_EFF_FB_R56_9_75589450	9	75589450	A	G	SNP
FR12_EFF_FB_R56_9_75591443	9	75591443	A	G	SNP
FR12_EFF_FB_R56_9_75592264	9	75592264	G	GA	INSER
FR12_EFF_FB_R56_9_75596148	9	75596148	AT	A	SNP
FR12_EFF_FB_R56_9_75596416	9	75596416	T	G	SNP
FR12_EFF_FB_R56_9_75596653	9	75596653	G	C	SNP
FR12_EFF_FB_R56_9_75601157	9	75601157	C	T	SNP
FR12_EFF_FB_R56_9_75602043	9	75602043	T	G	SNP
FR12_EFF_FB_R56_16_78956587	16	78956587	A	G	SNP
FR12_EFF_FB_R56_16_78957668	16	78957668	G	GAAC	INSER
FR12_EFF_FB_R56_16_78958773	16	78958773	G	A	SNP
FR12_EFF_FB_R56_16_78959286	16	78959286	T	C	SNP
FR12_EFF_FB_R56_16_78959292	16	78959292	G	A	SNP
FR12_EFF_FB_R56_16_78970453	16	78970453	C	G	SNP
FR12_EFF_FB_R56_16_78970984	16	78970984	C	T	SNP

FR12_EFF_FB_R56_16_78976775	16	78976775	GCA	G	INDEL
FR12_EFF_FB_R56_16_78977918	16	78977918	C	A	SNP
FR12_EFF_FB_R56_16_78977921	16	78977921	C	G	SNP
FR12_EFF_FB_R56_18_15466142	18	15466142	T	C	SNP
FR12_EFF_FB_R56_18_15473314	18	15473314	C	G	SNP
FR12_EFF_FB_R56_18_15745694	18	15745694	G	A	SNP
FR12_EFF_FB_R56_18_15784669	18	15784669	GA	G	INDEL
FR12_EFF_FB_R56_18_15843275	18	15843275	C	T	SNP
FR12_EFF_FB_R56_26_19053372	26	19053372	C	T	SNP
FR12_EFF_FB_R56_26_19054604	26	19054604	A	G	SNP
FR12_EFF_FB_R56_26_19055971	26	19055971	C	T	SNP
FR12_EFF_FB_R56_26_19056054	26	19056054	TC	T	INDEL
FR12_EFF_FB_R56_26_19057397	26	19057397	T	C	SNP

Annexe 2 Résultats de l'analyse génomique pour la distorsion de ségrégation des haplotypes dans 1 726 familles de taureaux Holstein, 974 familles de Montbéliarde et 479 familles de Normandie.

ID 1er mrq : Identifiant interne du premier marqueur de l'haplotype (Voir Tableau ST1 de Besnard *et al.* (2023) pour la corrélation avec l'ID Illumina). Pos 1er mrq : Position du premier marqueur de l'haplotype en Mb. Nb htz pere : Nombre de taureaux porteurs hétérozygotes. NObs et NAtt : Nombres observés et attendus de la progéniture porteuse sur les phases paternelles. P-value : Pour un haplotype donné, valeur p du test du Chi2 de conformité sur les dénombrements observés pour cet haplotype et pour les allèles alternatifs dans la progéniture de chaque taureau hétérozygote par rapport aux dénombrements attendus en supposant un ratio de distribution de 50-50. Adj_p-value : Valeur p ajustée selon la méthode de Bonferroni. Seuls les haplotypes observés à l'état hétérozygote chez au moins dix taureaux et avec une valeur p inférieure à 0,01 après correction de Bonferroni ont été conservés (en tenant compte de 2 488 067 haplotypes testés, comprenant 939 326 chez les Holstein, 890 745 chez les Montbéliarde, et 657 996 chez les Normandie).

Race	Identifiant de l'haplotype	Identifiant du 1 ^{er} marqueur	Chr	Position du 1 ^{er} marqueur	Nb de peres hétérozygotes	NObs	Natt	NObs/NAtt	P-value	Adj_p-value
HOL	HSDH1P3	F0100750	1	3,01091	21	5166	6060	0,85	1,90E-32	4,73E-26
HOL	HSDH1P20	F0103840	1	19,93872	14	6235	7284	0,86	2,34E-28	5,82E-22
HOL	HSDH1P51	F0110430	1	51,45324	26	5046	6586	0,77	1,57E-83	3,91E-77
HOL	HSDH1P69	F0114260	1	68,62321	14	2575	3645	0,71	9,16E-86	2,28E-79
HOL	HSDH1P98	F0120230	1	97,92606	135	28251	29634	0,95	3,50E-100	8,71E-94
HOL	HSDH1P157	F0133230	1	156,9104	12	3754	4067	0,92	2,82E-10	7,02E-04
HOL	HSDH2P5	F0201320	2	5,082951	12	2553	2930	0,87	3,25E-11	8,09E-05
HOL	HSDH2P18	F0203970	2	18,14261	68	15266	17433	0,88	4,12E-30	1,03E-23
HOL	HSDH2P27	F0206310	2	27,36423	76	10469	13037	0,80	1,46E-102	3,63E-96
HOL	HSDH2P35	F0208470	2	35,28259	26	3200	3713	0,86	2,77E-18	6,89E-12
HOL	HSDH2P42	F0209880	2	42,34209	48	4199	6389	0,66	5,55E-157	1,38E-150
HOL	HSDH2P60	F0213190	2	59,52265	48	6430	7354	0,87	3,10E-48	7,71E-42
HOL	HSDH2P72	F0215610	2	72,0312	171	21494	23426	0,92	2,44E-96	6,07E-90
HOL	HSDH2P128	F0225550	2	128,0404	135	24712	24213	1,02	1,26E-195	3,13E-189
HOL	HSDH3P64	F0311690	3	63,92325	206	23917	24905	0,96	9,69E-29	2,41E-22
HOL	HSDH3P70	F0313170	3	69,65091	243	26695	28867	0,92	3,91E-78	9,73E-72
HOL	HSDH4P9	F0401300	4	8,57928	11	859	1019	0,84	6,34E-10	1,58E-03
HOL	HSDH4P33	F0406670	4	33,37592	77	10431	13251	0,79	4,29E-167	1,07E-160
HOL	HSDH4P49	F0410160	4	48,95223	171	26246	30965	0,85	7,01E-68	1,74E-61
HOL	HSDH5P22	F0504700	5	22,12235	35	11380	12519	0,91	3,00E-13	7,46E-07
HOL	HSDH5P35	F0507100	5	35,1194	26	4308	5402	0,80	9,06E-37	2,25E-30
HOL	HSDH5P43	F0508430	5	43,02452	49	18283	19360	0,94	8,07E-41	2,01E-34
HOL	HSDH5P94	F0516880	5	93,95363	132	31483	33125	0,95	6,60E-15	1,64E-08
HOL	HSDH6P15	F0603640	6	15,48659	183	21179	22156	0,96	5,54E-74	1,38E-67
HOL	HSDH6P55	F0612030	6	55,27716	63	8979	13282	0,68	1,51E-291	3,76E-285
HOL	HSDH6P65	F0613830	6	64,90578	43	5685	7376	0,77	4,73E-68	1,18E-61
HOL	HSDH6P110	F0623370	6	109,5523	114	24225	25997	0,93	1,20E-62	2,99E-56
HOL	HSDH7P68	F0714210	7	67,69334	39	6034	7120	0,85	5,43E-31	1,35E-24
HOL	HSDH7P75	F0715770	7	75,09102	76	14416	20279	0,71	0,00E+00	0,00E+00
HOL	HSDH7P85	F0718150	7	85,45407	252	32434	33146	0,98	1,36E-63	3,38E-57
HOL	HSDH7P109	F0722700	7	109,1607	27	4293	5603	0,77	2,10E-41	5,22E-35
HOL	HSDH8P17	F0803230	8	16,98202	34	5034	8605	0,59	0,00E+00	0,00E+00

HOL	HSDH8P34	F0806810	8	33,88081	37	3840	4914	0,78	2,15E-55	5,35E-49
HOL	HSDH8P65	F0813590	8	65,14817	10	3205	3511	0,91	3,14E-15	7,81E-09
HOL	HSDH8P82	F0817050	8	81,95105	50	4780	5690	0,84	7,21E-49	1,79E-42
HOL	HSDH9P26	F0905090	9	26,31787	32	4788	6364	0,75	4,32E-66	1,07E-59
HOL	HSDH9P80	F0915140	9	79,63527	103	24218	23636	1,02	5,65E-22	1,41E-15
HOL	HSDH9P90	F0917290	9	89,52244	327	43825	45289	0,97	6,76E-126	1,68E-119
HOL	HSDH10P21	F1004610	10	21,36331	23	2880	3116	0,92	1,06E-09	2,64E-03
HOL	HSDH10P39	F1007610	10	38,90898	45	7058	8394	0,84	6,02E-24	1,50E-17
HOL	HSDH10P50	F1010080	10	49,85869	42	3504	4139	0,85	6,62E-10	1,65E-03
HOL	HSDH10P74	F1015290	10	73,75179	52	7124	11505	0,62	0,00E+00	0,00E+00
HOL	HSDH11P6	F0007490	11	6,390618	55	7119	10359	0,69	5,43E-206	1,35E-199
HOL	HSDH11P29	F1107480	11	28,76195	13	1044	1809	0,58	4,44E-74	1,10E-67
HOL	HSDH11P52	F1112060	11	51,87027	84	12284	14768	0,83	2,24E-50	5,57E-44
HOL	HSDH11P59	F1113040	11	59,03588	15	1708	2205	0,77	6,88E-15	1,71E-08
HOL	HSDH11P70	F1115270	11	70,21626	126	11285	12866	0,88	4,73E-43	1,18E-36
HOL	HSDH11P84	F1117940	11	84,44799	16	1937	2449	0,79	6,34E-37	1,58E-30
HOL	HSDH12P6	F1201130	12	6,211444	65	9238	10384	0,89	3,68E-09	9,16E-03
HOL	HSDH12P23	F1204330	12	23,45011	96	14967	20966	0,71	0,00E+00	0,00E+00
HOL	HSDH12P44	F1207740	12	44,1058	37	5734	9241	0,62	1,57E-288	3,91E-282
HOL	HSDH12P66	F1212620	12	66,12396	33	4744	6586	0,72	1,61E-98	4,01E-92
HOL	HSDH12P86	F1217660	12	85,79035	18	2321	2768	0,84	2,11E-10	5,25E-04
HOL	HSDH13P15	F1302630	13	15,09967	29	5478	8096	0,68	4,10E-159	1,02E-152
HOL	HSDH13P79	F1316970	13	79,39143	24	5186	5839	0,89	7,61E-14	1,89E-07
HOL	HSDH14P9	F1402800	14	9,452923	15	2507	3141	0,80	1,13E-41	2,81E-35
HOL	HSDH14P54	F1411710	14	54,29726	35	7834	8524	0,92	1,17E-23	2,91E-17
HOL	HSDH14P76	F1416100	14	76,05439	110	25252	27938	0,90	8,43E-12	2,10E-05
HOL	HSDH15P40	F1507710	15	39,97584	33	6750	8836	0,76	5,42E-101	1,35E-94
HOL	HSDH15P55	F1510740	15	55,35624	45	5565	6215	0,90	1,13E-16	2,81E-10
HOL	HSDH15P77	F1515670	15	77,34282	17	1956	3358	0,58	6,98E-126	1,74E-119
HOL	HSDH16P44	F1607900	16	44,30102	11	1284	1647	0,78	1,14E-17	2,84E-11
HOL	HSDH16P54	F1610110	16	53,7965	51	8456	9336	0,91	3,42E-15	8,51E-09
HOL	HSDH16P68	F1613180	16	67,75583	10	2414	2162	1,12	3,09E-37	7,69E-31
HOL	HSDH17P13	F1703270	17	13,25542	228	36260	44867	0,81	1,84E-294	4,58E-288
HOL	HSDH17P65	F1714060	17	65,40878	20	2514	3246	0,77	2,52E-37	6,27E-31
HOL	HSDH18P60	F1812400	18	59,95347	10	5253	6073	0,86	2,03E-16	5,05E-10
HOL	HSDH19P1	F1900100	19	1,493892	222	44996	46205	0,97	2,08E-29	5,18E-23
HOL	HSDH19P16	F1902720	19	16,10723	91	10551	15814	0,67	0,00E+00	0,00E+00
HOL	HSDH19P33	F1906670	19	32,77976	68	6921	8486	0,82	2,55E-85	6,34E-79
HOL	HSDH19P44	F1909250	19	44,18233	33	6155	6371	0,97	1,52E-40	3,78E-34
HOL	HSDH19P60	F1912950	19	59,70275	10	1895	2981	0,64	8,67E-77	2,16E-70
HOL	HSDH20P6	F2001510	20	6,395922	152	19566	18814	1,04	8,06E-50	2,01E-43
HOL	HSDH20P22	F2004960	20	22,41143	167	29707	34530	0,86	1,39E-57	3,46E-51
HOL	HSDH20P43	F2009040	20	43,4836	53	6827	8484	0,80	9,51E-52	2,37E-45
HOL	HSDH20P59	F2012510	20	59,37067	14	1299	2006	0,65	3,16E-57	7,86E-51
HOL	HSDH21P31	F2106560	21	30,92052	15	609	946	0,64	7,97E-19	1,98E-12

HOL	HSDH22P55	F2211670	22	54,91764	24	4116	4898	0,84	1,94E-15	4,83E-09
HOL	HSDH23P35	F2306780	23	34,61406	13	7541	8019	0,94	1,27E-09	3,16E-03
HOL	HSDH24P12	F2402850	24	12,07401	153	17812	18716	0,95	3,21E-10	7,99E-04
HOL	HSDH24P30	F2406320	24	30,311894	265	34587	34610	1,00	1,80E-49	4,48E-43
HOL	HSDH26P25	F0009190	26	24,529	372	71858	82363	0,87	7,33E-250	1,82E-243
HOL	HSDH26P43	F2608950	26	43,1762	106	14231	20587	0,69	0,00E+00	0,00E+00
HOL	HSDH28P5	F2800680	28	4,942362	35	2329	3406	0,68	8,40E-59	2,09E-52
HOL	HSDH28P24	F2804650	28	24,11742	14	1822	2626	0,69	2,07E-55	5,15E-49
HOL	HSDH28P41	F2807980	28	40,52356	22	1607	2663	0,60	6,68E-70	1,66E-63
HOL	HSDH29P6	F2901140	29	6,373421	119	18591	23857	0,78	1,31E-278	3,26E-272
HOL	HSDH29P14	F2902760	29	14,10388	61	9814	11058	0,89	6,21E-11	1,55E-04
HOL	HSDH29P34	F2906780	29	33,56077	12	1133	1446	0,78	1,81E-12	4,50E-06
HOL	HSDH29P50	F2910390	29	49,72219	32	15792	16376	0,96	1,42E-12	3,53E-06
MON	HSDM1P42	F0108320	1	41,81178	22	3756	4732	0,79	8,88E-57	2,21E-50
MON	HSDM1P54	F0111100	1	53,71381	28	3030	3757	0,81	1,00E-37	2,49E-31
MON	HSDM1P91	F0118710	1	90,54595	10	1412	1802	0,78	1,56E-21	3,88E-15
MON	HSDM1P98	F0120170	1	97,72904	125	29906	30690	0,97	1,48E-173	3,68E-167
MON	HSDM1P108	F0122330	1	108,3985	11	2173	2577	0,84	4,51E-27	1,12E-20
MON	HSDM2P31	F0207550	2	31,43174	29	4822	5257	0,92	6,88E-10	1,71E-03
MON	HSDM2P61	F0213350	2	60,53904	176	17882	19841	0,90	8,07E-73	2,01E-66
MON	HSDM2P118	F0223430	2	117,9628	19	2370	3282	0,72	2,26E-64	5,62E-58
MON	HSDM2P133	F0226780	2	133,2757	33	11232	12851	0,87	7,42E-46	1,85E-39
MON	HSDM3P49	F0309160	3	49,44959	11	1384	1715	0,81	1,44E-09	3,58E-03
MON	HSDM3P75	F0314160	3	75,02458	25	4159	5471	0,76	5,17E-85	1,29E-78
MON	HSDM3P101	F0320610	3	101,2704	31	3382	4264	0,79	4,91E-55	1,22E-48
MON	HSDM4P30	F0406120	4	29,97334	40	5476	6351	0,86	5,84E-10	1,45E-03
MON	HSDM4P44	F0409060	4	44,37473	10	685	836	0,82	1,24E-10	3,09E-04
MON	HSDM4P54	F0411070	4	53,58046	82	12196	14948	0,82	1,85E-103	4,60E-97
MON	HSDM5P24	F0504990	5	24,01187	12	1440	2050	0,70	5,92E-40	1,47E-33
MON	HSDM5P35	F0507070	5	34,8945	15	2963	2724	1,09	1,41E-22	3,51E-16
MON	HSDM5P48	F0509230	5	48,22938	67	14709	14334	1,03	5,48E-42	1,36E-35
MON	HSDM5P98	F0517500	5	97,84554	30	5165	8252	0,63	1,46E-300	3,63E-294
MON	HSDM6P32	F0606960	6	31,8024	31	5742	6331	0,91	1,92E-44	4,78E-38
MON	HSDM6P40	F0608400	6	39,95973	52	7243	10108	0,72	2,39E-160	5,95E-154
MON	HSDM6P70	F0614880	6	70,38797	28	10971	10739	1,02	2,20E-55	5,47E-49
MON	HSDM6P78	F0616200	6	77,92632	11	3326	3399	0,98	8,47E-11	2,11E-04
MON	HSDM6P88	F0617960	6	88,0248	117	18984	22157	0,86	4,81E-121	1,20E-114
MON	HSDM6P94	F0619560	6	94,49385	29	3237	3506	0,92	6,26E-17	1,56E-10
MON	HSDM6P111	F0623840	6	111,2224	30	6089	6566	0,93	1,20E-11	2,99E-05
MON	HSDM7P18	F0703360	7	17,50252	78	11835	13199	0,90	1,77E-27	4,40E-21
MON	HSDM7P39	F0707950	7	38,50824	43	9678	10104	0,96	3,97E-49	9,88E-43
MON	HSDM7P48	F0710290	7	48,32596	19	3455	3680	0,94	8,08E-13	2,01E-06
MON	HSDM7P100	F0720500	7	99,9802	21	1979	2574	0,77	4,09E-23	1,02E-16
MON	HSDM8P1	F0800230	8	0,568381	28	9717	9319	1,04	2,85E-51	7,09E-45
MON	HSDM8P24	F0804590	8	24,02787	54	7919	11952	0,66	0,00E+00	0,00E+00

MON	HSDM8P82	F0817090	8	82,09495	69	6493	7362	0,88	1,96E-36	4,88E-30
MON	HSDM9P9	F0901570	9	8,50954	20	2286	3795	0,60	7,02E-150	1,75E-143
MON	HSDM9P16	F0903090	9	15,97594	17	1385	1696	0,82	8,11E-13	2,02E-06
MON	HSDM9P54	F0910460	9	53,53693	97	8749	9721	0,90	5,63E-24	1,40E-17
MON	HSDM9P82	F0915830	9	82,43165	483	115714	104673	1,11	0,00E+00	0,00E+00
MON	HSDM9P90	F0917360	9	89,94259	146	16835	18034	0,93	6,39E-35	1,59E-28
MON	HSDM11P4	F0009630	11	3,935679	12	1039	1644	0,63	3,93E-48	9,78E-42
MON	HSDM11P26	F1106960	11	26,36249	31	6713	7591	0,88	1,17E-14	2,91E-08
MON	HSDM11P34	F1108840	11	34,12948	63	6103	6992	0,87	2,52E-35	6,27E-29
MON	HSDM11P61	F1113520	11	61,10609	62	13937	15504	0,90	4,39E-17	1,09E-10
MON	HSDM12P55	F1210620	12	55,28498	12	1150	1381	0,83	2,14E-09	5,32E-03
MON	HSDM13P1	F1300100	13	0,86624	22	1839	2311	0,80	2,59E-16	6,44E-10
MON	HSDM15P20	F1502900	15	19,7329	23	1906	2186	0,87	1,34E-10	3,33E-04
MON	HSDM17P2	F1700370	17	1,53275	16	2307	2863	0,81	5,00E-52	1,24E-45
MON	HSDM17P32	F1706760	17	32,31461	60	8390	9920	0,85	3,77E-70	9,38E-64
MON	HSDM17P48	F1710090	17	48,03894	14	1314	2051	0,64	2,93E-73	7,29E-67
MON	HSDM17P72	F1715580	17	72,13222	73	10513	12893	0,82	3,36E-87	8,36E-81
MON	HSDM19P0	F0005470	19	0,450023	18	1238	2542	0,49	3,32E-131	8,26E-125
MON	HSDM19P53	F1911230	19	52,82455	23	6958	7737	0,90	9,85E-12	2,45E-05
MON	HSDM20P24	F2005180	20	23,69503	24	2346	3619	0,65	1,15E-96	2,86E-90
MON	HSDM20P62	F2012960	20	61,61079	50	8776	9835	0,89	7,14E-32	1,78E-25
MON	HSDM21P21	F2104160	21	20,64279	64	9320	9943	0,94	2,12E-11	5,27E-05
MON	HSDM21P53	F2114630	21	52,88857	42	6837	10063	0,68	2,08E-210	5,18E-204
MON	HSDM22P32	F2206680	22	31,58384	15	2042	2255	0,91	1,99E-09	4,95E-03
MON	HSDM23P33	F2306530	23	33,31792	16	5859	6730	0,87	5,59E-35	1,39E-28
MON	HSDM24P12	F2402870	24	12,19564	88	7368	7987	0,92	1,62E-17	4,03E-11
MON	HSDM24P23	F2404740	24	22,7723	15	1578	1888	0,84	2,78E-13	6,92E-07
MON	HSDM24P33	F2406890	24	32,82697	242	40843	52847	0,77	0,00E+00	0,00E+00
MON	HSDM24P43	F2409200	24	42,6591	27	4667	5325	0,88	3,15E-37	7,84E-31
MON	HSDM25P33	F2507620	25	33,16547	25	4676	5422	0,86	5,17E-12	1,29E-05
MON	HSDM27P22	F2704420	27	22,21629	17	3068	5788	0,53	1,50E-274	3,73E-268
MON	HSDM29P13	F2902550	29	12,84943	21	3210	4065	0,79	2,03E-32	5,05E-26
MON	HSDM29P36	F2907330	29	35,99961	33	4720	5497	0,86	1,46E-14	3,63E-08
MON	HSDM29P50	F2910390	29	49,72219	57	12409	13360	0,93	9,51E-27	2,37E-20
NOR	HSDN1P19	F0103650	1	18,67909	11	740	1057	0,70	1,20E-24	2,99E-18
NOR	HSDN1P97	F0120090	1	97,3435	21	3090	3665	0,84	3,32E-43	8,26E-37
NOR	HSDN1P125	F0125560	1	125,0737	17	1062	1511	0,70	5,49E-28	1,37E-21
NOR	HSDN2P1	F0227660	2	1,436462	19	2201	2499	0,88	1,67E-10	4,16E-04
NOR	HSDN2P56	F0212530	2	56,04539	91	13631	13485	1,01	5,65E-21	1,41E-14
NOR	HSDN5P82	F0514790	5	82,22651	16	946	1290	0,73	1,16E-16	2,89E-10
NOR	HSDN6P15	F0603460	6	14,51555	14	1319	1683	0,78	2,05E-26	5,10E-20
NOR	HSDN6P111	F0623840	6	111,2224	47	5357	5488	0,98	7,12E-21	1,77E-14
NOR	HSDN7P43	F0709070	7	43,08561	18	3392	4159	0,82	7,85E-31	1,95E-24
NOR	HSDN9P90	F0917290	9	89,52244	53	6687	6340	1,05	1,74E-17	4,33E-11
NOR	HSDN12P58	F0001080	12	58,33069	31	2599	3313	0,78	1,76E-27	4,38E-21

NOR	HSDN13P79	F1316930	13	79,23672	35	2922	4434	0,66	8,16E-84	2,03E-77
NOR	HSDN14P5	F1401250	14	4,586387	54	5136	6372	0,81	3,04E-31	7,56E-25
NOR	HSDN14P68	F1414670	14	68,22894	11	633	830	0,76	1,57E-10	3,91E-04
NOR	HSDN14P75	F1415980	14	75,43926	19	2390	2457	0,97	1,29E-09	3,21E-03
NOR	HSDN16P72	F1614150	16	71,52721	67	7353	9323	0,79	7,54E-63	1,88E-56
NOR	HSDN21P17	F2103210	21	16,56176	10	474	663	0,71	8,99E-13	2,24E-06
NOR	HSDN21P42	F2108740	21	41,79044	11	1241	1852	0,67	3,03E-39	7,54E-33
NOR	HSDN23P29	F2305700	23	29,13745	13	1013	1963	0,52	3,10E-90	7,71E-84
NOR	HSDN24P12	F2402890	24	12,24306	39	4885	4504	1,08	3,00E-19	7,46E-13
NOR	HSDN26P19	F2603890	26	18,6128	103	17304	14970	1,16	3,32E-222	8,26E-216
NOR	HSDN26P32	F2606340	26	31,56206	13	1392	1314	1,06	2,41E-13	6,00E-07
NOR	HSDN29P28	F2905490	29	28,15407	10	538	1047	0,51	2,37E-46	5,90E-40

Titre : Exploration génomique de la mortalité juvénile des bovins

Mots clés : Mortalité - Génomique – Génétique – Veaux – Bovins – Anomalies

Résumé : Dans le secteur de l'élevage bovin laitier, la mortalité des jeunes animaux représente un problème important, avec environ 10% des veaux mourant de manière naturelle au cours de leur première année. Ce phénomène est complexe, influencé par divers facteurs environnementaux et génétiques, ce qui le rend difficile à mesurer à grande échelle.

L'Observatoire National des Anomalies Bovines (ONAB) en France a permis d'identifier de nombreuses mutations génétiques impliquées dans des anomalies, bien que ces découvertes ne représentent qu'une petite partie des morts observées.

Depuis la fin des années 2000, l'accumulation de vastes bases de données sur le pedigree, le phénotype et le génotype bovins offre de nouvelles opportunités. Ces données, bien que parfois peu précises dans leur description des historiques de vie des animaux, offrent un volume considérable pour mener des analyses approfondies.

Dans le cadre de ma thèse, j'ai développé plusieurs approches exploitant ces vastes ensembles de données pour identifier des *loci* impliqués dans la mortalité avec des déterminismes génétiques dominants, récessifs et parfois avec une pénétrance incomplète. Ces *loci*, ne correspondent pas aux cas remontés à l'ONAB en raison de l'absence de signes cliniques caractéristiques, et comprennent notamment les gènes *ITGB7*, *NOA1* et *RFC5*, associés respectivement aux anomalies génétiques BLIRD, MTCP et à l'alopécie des extrémités. Ces découvertes ont été confirmées par des observations phénotypiques fines. De plus, j'ai développé des évaluations génomiques pilotes sur la mortalité.

L'intégration de ces travaux dans les programmes de sélection promet de réduire la mortalité en élevage. Par ailleurs, la découverte de mutations délétères affectant des gènes peu connus ouvre des perspectives prometteuses pour la recherche fondamentale.

Title : Genomic exploration of juvenile mortality in dairy cattle farming

Keywords: Mortality - Genomics - Genetics - Calves - Cattle - Abnormalities

Abstract: Juvenile animal mortality is a major concern in the dairy industry, with approximately 10% of calves dying naturally in their first year. This is a complex phenomenon influenced by a variety of environmental and genetic factors, making it difficult to measure on a large scale.

The French "Observatoire National des Anomalies Bovines" (ONAB) has made it possible to identify numerous genetic mutations associated with abnormalities, although these discoveries represent only a small proportion of observed deaths.

Since the late 2000s, the accumulation of vast databases of bovine pedigrees, phenotypes and genotypes has opened up new possibilities. These data, although sometimes not very precise in their description of the animal's life history, offer a considerable volume for in-depth analysis.

As part of my Ph.D. thesis, I developed several approaches using these large datasets to identify *loci* involved in mortality with dominant, recessive and sometimes incomplete penetrance genetic determinism. These *loci*, which cannot be tracked with ONAB due to the absence of characteristic clinical signs, include mutations in the *ITGB7*, *NOA1*, and *RFC5* genes, which are associated with BLIRD, MTCP, and limb alopecia genetic abnormalities, respectively. These findings were supported by detailed phenotypic observations. In addition, I developed pilot genomic evaluations of mortality.

The integration of this work into selection programs promises to reduce mortality in breeding. In addition, the discovery of deleterious mutations in little known genes opens up promising prospects for basic research.

