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

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

HAL Id: hal-04714371

https://hal.inrae.fr/hal-04714371v1

Submitted on 30 Sep 2024

Check for updates

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

Florian Besnard[1,2]* , Ana Guintard[1,3], Cécile Grohs[1] , Laurence Guzylack-Piriou[4] , Margarita Cano[1] , Clémentine Escouflaire[1,3] , Chris Hozé[1,3] , Hélène Leclerc[1,3] , Thierry Buronfosse[5] , Lucie Dutheil[4] , Jeanlin Jourdain[1,3] , Anne Barbat[1] , Sébastien Fritz[1,3] , Marie-Christine Deloche[1,3] , Aude Remot[6] , Blandine Gaussères[4] , Adèle Clément[4], Marion Bouchier[5], Elise Contat[5], Anne Relun[7] , Vincent Plassard[8] , Julie Rivière[1,9] , Christine Péchoux[1] , Marthe Vilotte[1], Camille Eche[10] , Claire Kuchly[10] , Mathieu Charles[1] , Arnaud Boulling[1] , Guillaume Viard[3,11], Stéphanie Minéry[2] , Sarah Barbey[12], Clément Birbes[13] , Coralie Danchin-Burge[2] , Frédéric Launay[12] , Sophie Mattalia[2] , Aurélie Allais-Bonnet[3,11] , Bérangère Ravary[8] , Yves Millemann[8] , Raphaël Guatteo[7] , Christophe Klopp[13] , Christine Gaspin[13] , Carole Iampietro[10] , Cécile Donnadieu[10] , Denis Milan[14] , Marie-Anne Arcangioli[5], Mekki Boussaha[1] , Gilles Foucras[4] , Didier Boichard[1] and Aurélien Capitan[1,3]*

*Correspondence:
florian.besnard@inrae.fr; aurelien.
capitan@inrae.fr

[1] Université Paris-Saclay,
INRAE, AgroParisTech, GABI,
78350 Jouy-en-Josas, France
Full list of author information is
available at the end of the article

Abstract

Background: Dairy cattle breeds are populations of limited effective size, subject to recurrent outbreaks of recessive defects that are commonly studied using positional cloning. However, this strategy, based on the observation of animals with characteristic features, may overlook a number of conditions, such as immune or metabolic genetic disorders, which may be confused with pathologies of environmental etiology.

Results: We present a data mining framework specifically 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.

Conclusions: 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

Besnard *et al. Genome Biology*   (2024) 25:248

Page 2 of 30

Background

Specialized dairy cattle breeds are genetic isolates that have experienced two recent bottlenecks with their creation from a limited number of founders about 150 years ago, and the overuse of a few elite artificial insemination sires in the last 50 years (e.g., [1, 2]). To illustrate this, a recent pedigree analysis of nine dairy breeds reared in France reported effective population sizes (Ne) ranging from 14 to 44 individuals, genetic contributions of the main ancestor to the breed's gene pool ranging from 7.9 to 21.3%, and average population inbreeding coefficients ranging from 2.1 to 6.6% (https://idele.fr/detail-dossier/varume-resultats-2023; accessed 2024–06-30). As a result, dairy cattle breeds are subject to recurrent outbreaks of recessive defects, which has historically led to the establishment of surveillance networks to allow early detection of affected animals with characteristic clinical features and homozygosity mapping of the locus [3].

With the tremendous development of high-throughput genotyping and sequencing technologies, alternative strategies such as searching for homozygous haplotype deficiency (e.g., [4, 5]), reverse genetics (e.g., [6, 7]), and non-additive association analysis using proxy phenotypes [8] have been developed over the years to detect genetic conditions that may easily go unnoticed due to a lack of specific signs. While these methods have proven effective in identifying new recessive defects, each has its own limitations.

The first strategy involves searching for haplotypes of markers with no observed homozygous carriers among the animals genotyped for genomic evaluation (Nobs = 0), whereas ten or more would be expected (Nexp ≥ 10) based on the genotypes of their sire and maternal grandsire sire or of their sire and dam, if the latter is available. In fact, dairy cattle are mainly bred by artificial insemination (AI) and all AI bulls are genotyped. To filter out false positives, the initial mapping step is followed by a statistical analysis of conception rates and juvenile mortality rates in independent validation populations consisting of at-risk mating (between a carrier sire and the daughters of a carrier sire) and control mating (non-carrier sire and maternal grandsire), under the assumption that the complete depletion in homozygotes is caused by embryonic lethality or juvenile mortality before the age of genotyping (which is usually about 3 months). Therefore, this method is only effective in detecting recessive loci that are responsible for early deaths and that show complete penetrance and complete linkage disequilibrium (LD) with the haplotypes of interest.

Reverse genetics consists of (i) screening whole-genome sequences of influential ancestors for heterozygous variants that are predicted to be deleterious in the homozygous state based on various sources of information, (ii) genotyping them in large populations to find homozygous individuals (or to demonstrate their absence in the case of embryonic lethality), and (iii) phenotyping case and control individuals to verify the initial assumption. After an initial craze that led to several publications in the mid-2010s (e.g., [6, 7, 9, 10]), this type of approach has been discontinued due to its high false positive rate (~ 90%; [7]), which is hardly compatible with the significant economic and logistical resources required for functional validation.

Finally, the principle of non-additive association analysis using proxy phenotypes [8] is to (i) impute the whole-genome sequences of tens of thousands of individuals initially genotyped with medium-density array as part of genomic evaluations, (ii) consider

phenotypes routinely recorded for selection purposes, such as stature or body condition score, which could be strongly affected by recessive defects, and (iii) perform genome-wide association studies using a model that accounts for recessive/dominant effects. As this approach is based on the analysis of phenotypes recorded during the first lactation, it does not allow the identification of defects that lead to the death of animals before this stage. In addition, the low imputation accuracy demonstrated for variants segregating at low frequencies ($< 5\%$) or in incomplete linkage disequilibrium (LD) with surrounding markers makes it not always suitable for mapping emerging defects due to recent mutations.

Considering the strengths and limitations of each of these methods, we developed a data mining framework specifically designed to detect recessive defects in livestock that were previously missed due to lack of specific signs, incomplete penetrance, or incomplete linkage disequilibrium using life history information from animals routinely genotyped for selection purposes. In this article, we present the different steps of this integrative approach (as summarized in Fig. 1) and provide a proof-of-concept with the identification of 33 new loci responsible for increased risk of juvenile mortality in three dairy cattle breeds, as well as the functional validation of three candidate variants affecting the *NOA1*, *RFC5*, and *ITGB7* genes.

Results

Mapping and validation of recessive haplotypes affecting juvenile mortality

The first step of our data mining framework is based on the principle of comparing the observed and expected number of homozygotes for given haplotypes in genotyped animals based on the genotypes of their paternal and maternal ancestors. However, instead of just looking for a depletion of homozygotes among all available animals, we propose to define groups of individuals with different life histories by mining databases for information on events that shaped their lives before looking for a simultaneous depletion of homozygotes among controls and enrichment among cases (Fig. 1a, b).

As a proof of concept, we decided to focus on recessive loci responsible for an increase in juvenile mortality during a period that has been poorly addressed in previous studies, namely between age at genotyping and the beginning of the productive career. To do this, we analyzed phased Illumina BovineSNP50 genotypes from 8203, 6198, and 2254 heifers that died of natural causes during the rearing period and 291,529, 141,343, and 56,095 adult cows of the Holstein, Montbeliarde, and Normande breeds, respectively (Additional file 1: Table S1). In each breed, we searched for sliding haplotypes of 20 markers (i.e., ~ 1 Mb; incremented by one marker) satisfying the following arbitrarily defined filters: ten or more observed dead heifers and at least 25% enrichment in dead heifers (($N_{obs} - N_{exp})/N_{exp} \geq 25\%$) and 25% depletion in cows (($N_{obs} - N_{exp})/N_{exp} \leq -25\%$) in terms of observed versus expected homozygous animals. The 25% depletion in cows was chosen to allow mapping of either (i) recessive variants that show complete LD with the haplotype and a penetrance of juvenile mortality as low as 25%, or (ii) loci with complete penetrance but a proportion of haplotype carriers that also carry the mutation as low as 50%, as well as (iii) various intermediate situations.
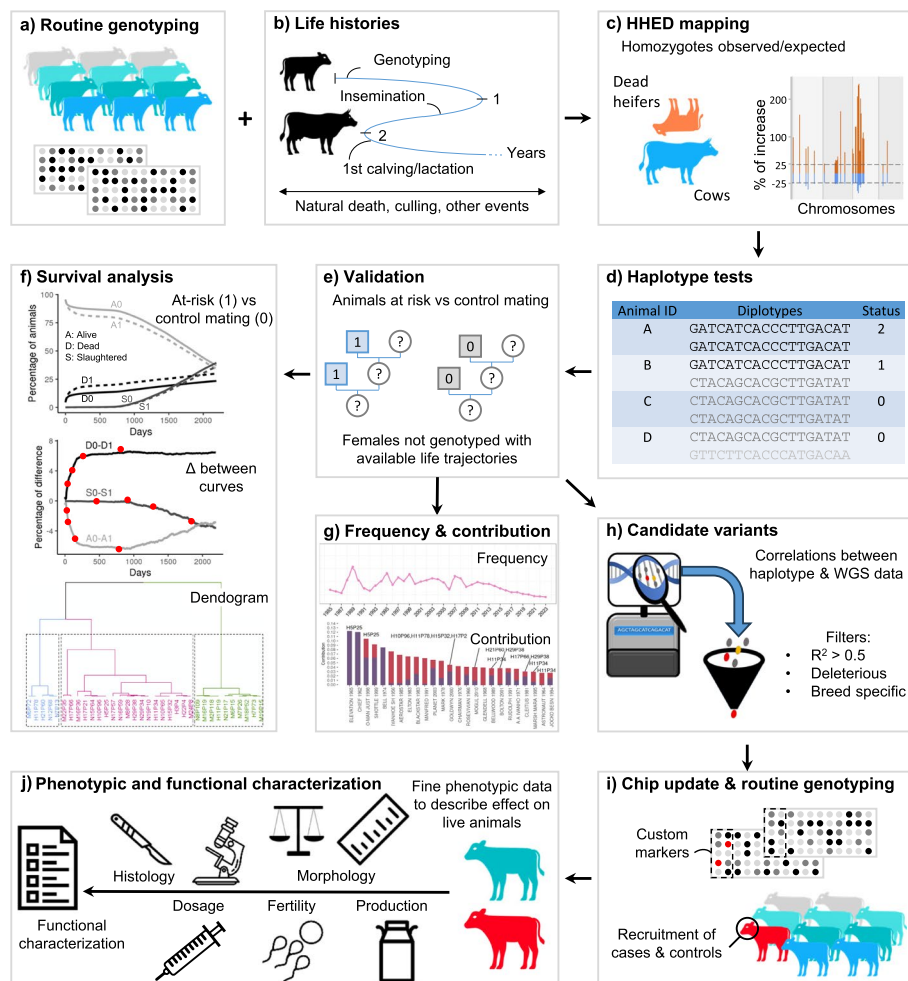
Fig. 1 Overview of the analysis plan developed to detect and characterize cryptic recessive genetic defects in cattle. Graphs are shown for illustrative purposes only. Detailed information is provided throughout the manuscript. The first two steps consist of listing animals genotyped for genomic evaluations 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

Besnard *et al. Genome Biology*   (2024) 25:248

Page 5 of 30

This approach, which we named "homozygous haplotype enrichment/depletion" (HHED) mapping, yielded numerous candidate regions (Figs. 1c and 2; Additional file 1: Tables S2 and S3). For further analysis, we therefore decided to select the top 20 peak haplotypes per breed and examine the life histories of independent validation populations of 6.0 million Holstein, 1.6 million Montbeliarde, and 1.2 million Normande ungenotyped 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; Additional file 1: Tables S4 and S5) and (Nobs − Nexp)/Nexp ratios in the discovery data set ranging from − 25 to − 94% in cows and + 78 to + 267% in dead heifers (Additional file 1: Table S4). Among them, a single haplotype (*H11P78* in Holstein cattle) was in linkage disequilibrium with a recessive 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; Additional file 1: Table S6), which causes death from cholesterol deficiency (CD) within the first weeks or months of life [11–13].
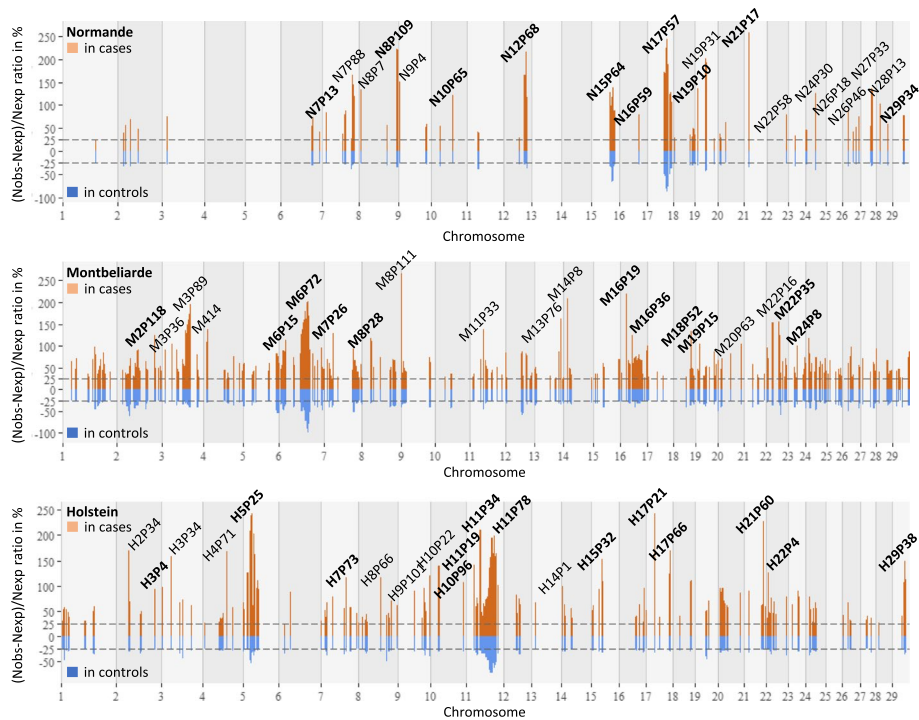
Fig. 2 Results of HHED mapping. Nobs and Nexp: Number of observed and expected homozygous animals, respectively, for sliding haplotypes of 20 markers (incremented by one marker; see "Methods"). Positive (Nobs − Nexp)/Nexp values correspond to enrichment in homozygotes, whereas negative values indicate depletion. Data supporting the graphs and details of the haplotypes are presented in Additional file 1: Table S3. The top 20 loci per breed are highlighted (also see Additional file 1: Table S4), 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; Additional file 1: Table S5). Note that the haplotype names are based on the breed (first character), chromosome (first number), and position (P) in Mb (last number)

**Prevalence of at-risk mating and survival analyses**

In a first attempt to measure the impact of the 34 validated haplotypes on the studied populations, we analyzed 1.3 million insemination records in Holstein, 0.5 million in Montbeliarde, and 0.2 million in Normande from the period 2013–2022 with both parents genotyped. We found 3.4, 3.6, and 8.0% of at-risk mating in the three breeds, respectively, and by extrapolation to 55.5 million inseminations, we 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 decade (Additional file 1: Table S7).

This substantial prevalence suggests that these previously undetected haplotypes collectively have a significant impact on dairy cattle breeding, although their magnitude is likely to vary between breeds and loci. Therefore, to better characterize the specific effects of each of these recessive haplotypes, we performed detailed survival analyses by calculating the daily difference in proportions between at-risk (1) and control (0) mating for female cattle that died of natural causes (D), were slaughtered (S), or were still alive (A) over 6 years (Additional file 1: Tables S8 and S9). In an attempt to define metrics to compare loci, we then scored the days on which 25, 50, 75, and 100% of the maximum deviation between the proportion differences were reached (Additional file 1: Table S10). 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. These are (i) early juvenile mortality—to which *H11P78* belongs in accordance with its association with the causative mutation for CD, (ii) late juvenile mortality, and (iii) increased mortality and premature culling throughout life (Fig. 3c, d; e, f; and g, h, respectively).

Evolution of haplotype frequencies and estimation of their effects on selected traits

To determine the forces driving the accumulation of these numerous deleterious loci in dairy cattle populations, we examined the evolution of haplotype frequencies per year from 1985 to the present, based on the genotypes of 180 thousand Normande, 591 thousand Montbeliarde, and 1.2 million Holstein individuals (Fig. 4a; Additional file 1: Tables S11 and S12). We observed sudden increases in frequency associated with the overuse of influential ancestors, and then slight and regular decreases due to the massive dissemination of semen from new elite bulls, diluting the genetic contribution of previous ones to the population. Notably, a single sire named GOLDWYN was the main source of seven of the 13 haplotypes validated in Holstein cattle, including *H11P78*/CD (Fig. 4b).

Since any large increase in frequency may not only be due to genetic drift, but may also be amplified by positive selection of heterozygous carriers, we sought to determine the extent to which the 34 haplotypes could influence 14 of the most important production and morphological traits for the dairy industry. To do this, we considered phenotypes adjusted for environmental effects (the so-called yield deviations) as part of the routine French genetic evaluation procedure and used 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 (see "Methods"; Additional file 1: Table S13; Fig. 4c). Due to the high statistical power provided by the large size of our cohorts ($1930 \leq n \leq 509{,}258$ cows), we found significant effects in 30% of the analyses
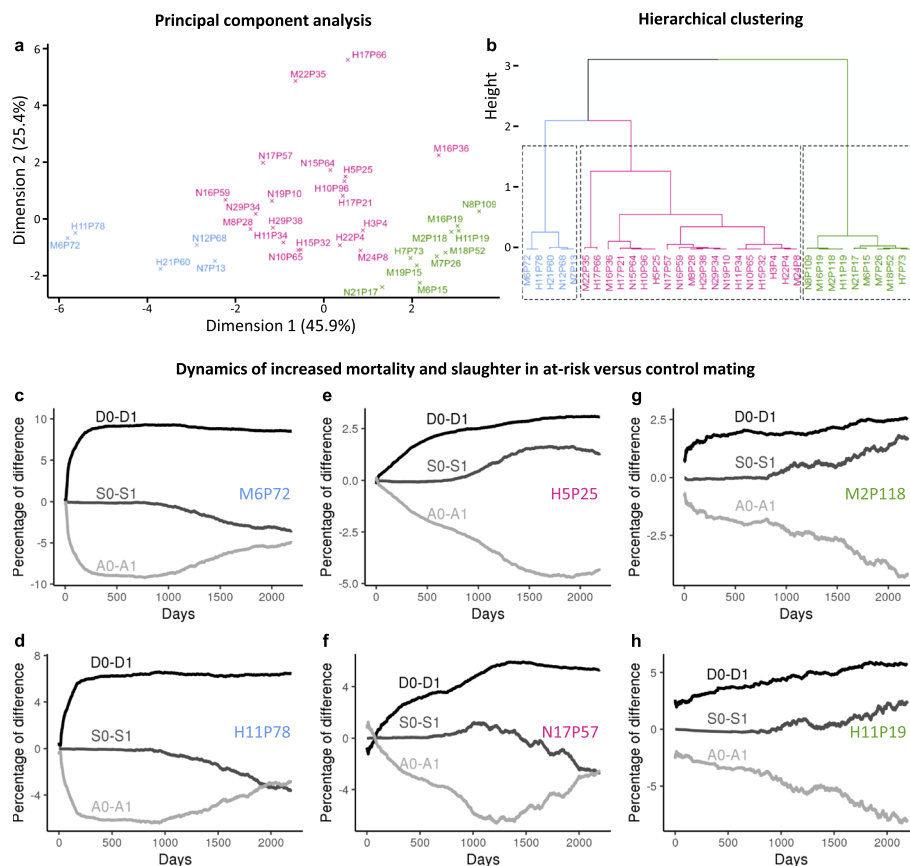
Fig. 3 Effects of the 34 validated haplotypes on female survival. **a** Results of principal component analysis (PCA) for 34 haplotypes × 12 parameters (Additional file 1: Table S10), selected to reflect the dynamics of increased mortality and slaughter in at-risk versus control mating. These parameters are the days on which 25, 50, 75, and 100% of the maximum deviations were reached after calculating the daily difference in proportions between at-risk (1) and control (0) mating for female cattle that died of natural causes (D), were slaughtered (S), or were still alive (A) over 6 years (Additional file 1: Tables S8, S9). The proportions of variance explained by the principal components are given in parentheses for the first two dimensions. **b** Results of hierarchical clustering (HC) on principal components. **c**–**h** Dynamics of increased mortality and slaughter in at-risk versus control mating expressed as 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 6 years (Additional file 1: Tables S8, S9). 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*)

(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 [14], and of alterations in lipid metabolism in healthy heterozygotes versus controls [15]. Therefore, our population-level findings shed new light on CD and support the idea that it is a co-dominantly inherited metabolic disorder.

Besnard *et al. Genome Biology*   (2024) 25:248

Page 8 of 30

Fig. 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 (Additional file 1: Table S11). **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"; Additional file 1: Table S12). **c** Effects of the heterozygosity for each Holstein, Montbeliarde, and Normande haplotype on 14 traits expressed in genetic standard deviation (GSD; Additional file 1: Table S13). Only significant effects are shown (Benjamini–Hochberg adjusted Student's *t*-test *p* value ≤ 0.01). Note the significant negative effects observed for *H11P78/CD* on several traits, including chest width and stature (− 0.30 and − 0.23 GSD, respectively), which support reduced growth in heterozygous carriers

Identification of candidate variants from whole-genome sequences

To gain insight into the molecular mechanisms associated with the 33 novel genetic determinants of increased juvenile mortality, we then analyzed the whole-genome sequences of 247 Holstein, 160 Montbeliarde, and 118 Normande cattle with known haplotype status, as well as 1344 controls from more than 70 breeds (Additional file 1: Tables S14 and S15; [16–66]). Except for *H17P21* in Holstein, which was represented by only one heterozygous carrier, the number of sequenced carriers for each haplotype ranged from 4 to 22 in this data set. By focusing on breed-specific variants that were predicted to be detrimental to protein function and had a haplotype-genotype square correlation (R^2) of 0.50 or greater, we prioritized 8 candidates (Additional file 1: Table S16). Of

note, six of these candidate variants had been included in the design of the SNP arrays used for genomic evaluation in France following reverse genetics studies (e.g., [6, 9]), and genotypes were available for 0.4 to 1.8 million animals from 15 breeds (Additional file 1: Table S17).

Validation of three causative variants

After verifying that these large-scale genotyping data yielded results consistent with those of whole-genome sequence analysis in terms of haplotype-genotype correlations (Additional file 1: Table S18), 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 β7 integrin (ITGB7 p.G375S) affecting a residue conserved among 128 vertebrate orthologs located at the contact interface with the α4 integrin (Fig. 5a; Additional file 1: Table S19). The α4β7 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; [67]). 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 α4β7^{pos} CD45RO^{pos} memory CD4 T cells in the lamina propria of the jejunum (Fig. 5c, Additional file 1: Table S20). 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; Additional file 1: Tables S21 and S22). 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).

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; Additional file 1: Table S23). Taking advantage of this incomplete LD, we examined the 2-year survival of the different haplotype × 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; Additional file 1: Table S24). Finally, population-level analyses revealed delayed age at first insemination (Fig. 5g; Additional file 1: Table S25), consistent with reduced growth (e.g., [68]), adverse effects on production traits (Fig. 5h; Additional file 1: Table S26), and premature culling in homozygous mutants compared to other genotypes (Fig. 5i; Additional file 1: Table S27).

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
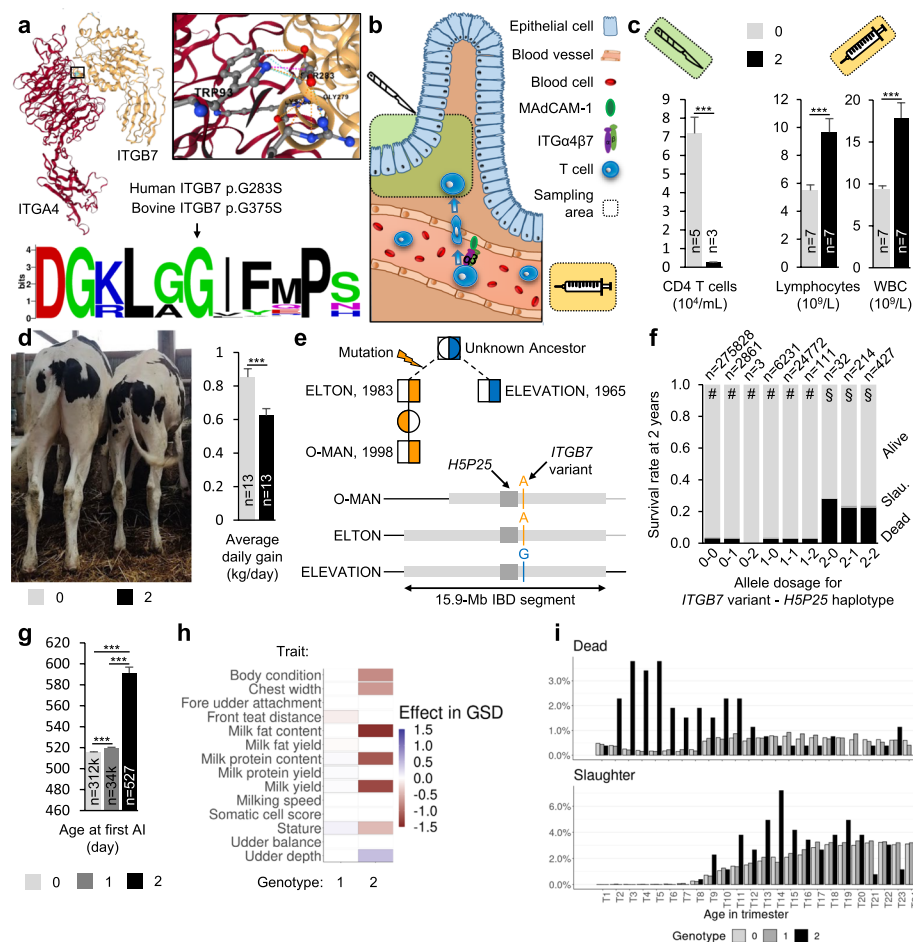
Fig. 5 Characterization of the ITGB7 substitution. **a** (top) Modeling of the 3D structure of the ITGα4β7 heterodimer using crystallography in humans, and evaluation of the consequences of glycine to serine substitution at amino acid position 283 in humans (375 in bovine). The purple dashed lines illustrate 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 (Additional file 1: Table S19). **b** Schematic representation of the ITGα4β7-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 α4β7pos 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 (Additional file 1: Tables S20 and S21). ***: 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 6 months to 2 years (Additional file 1: Table S22). **e** Origin of *H5P25* haplotype and *ITGB7* mutation (see Additional file 1: Table S23 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 (Additional file 1: Table S24). **g** Analysis of age at first AI for the three genotypes at the *ITGB7* variant (Additional file 1: Table S25). ***: Student *t*-test *p* value ≤ 0.001. **h** Effects of heterozygosity and homozygosity at the *ITGB7* variant on 14 recorded traits (Additional file 1: Table S26). Only significant effects are shown (Benjamini–Hochberg adjusted Student's *t*-test *p* value ≤ 0.01). GSD: genetic standard deviation. **i** Trimester repartition of the proportion of females that died or were slaughtered before reaching 6 years of age by genotype group at the *ITGB7* substitution (Additional file 1: Table S27)

p.E369del; $R^2 = 1$ with haplotype *N17P57*; Fig. 6a; Additional file 1: Tables S16 and S28), and a frameshift insertion in the nitric oxide-associated protein 1 (NOA1 p.D400RfsX9; $R^2 = 0.55$ with *M6P72*; Fig. 6i; Additional file 1: Table S16).

Besnard *et al. Genome Biology* (2024) 25:248

Page 11 of 30
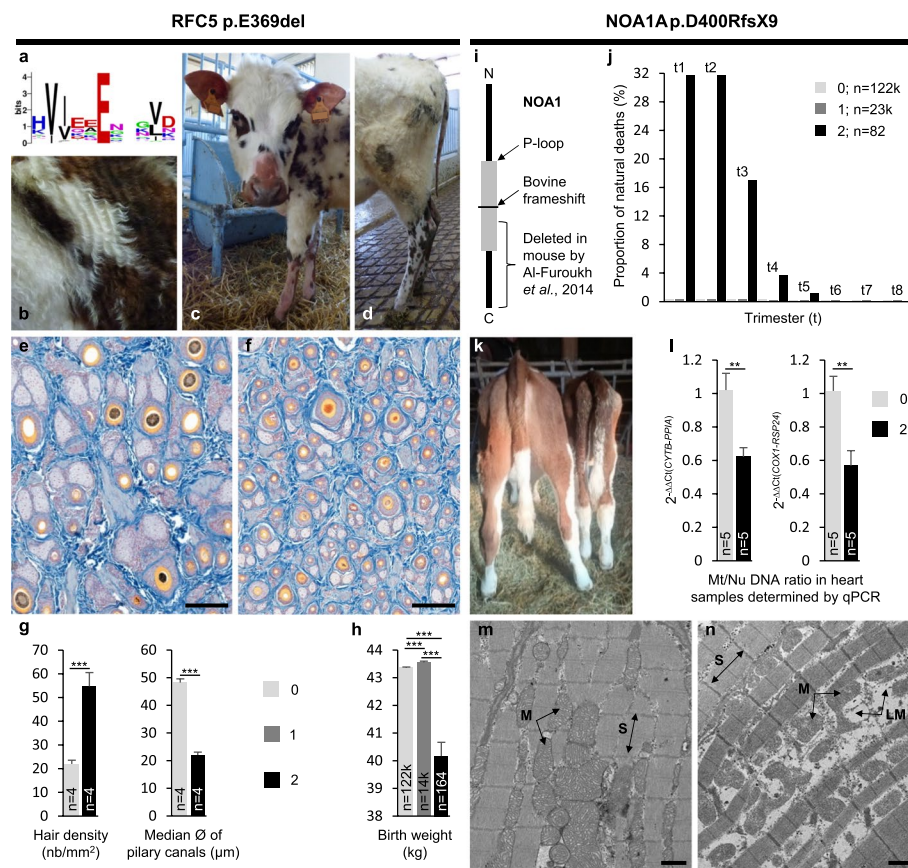
Fig. 6 Pathophysiological characterization of RFC5 and NOA1 mutants in Normande and Montbeliarde cattle, respectively. **a** WebLogo representation of multiple protein alignments showing perfect conservation of the glutamic acid residue at position 369 amino acids in bovine among 633 eukaryotic RFC5 orthologs (Additional file 1: Table S28). **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. 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 µM. **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 (Additional file 1: Table S29). **h** Analysis of birth weight for the three genotypes at the *RFC5* deletion (Additional file 1: Table S30). **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 [69]. **j** Percentage of natural deaths by trimester for 145 thousand females genotyped for the *NOA1* frameshift variant. Note that only the first 2 years of a 6-year study are shown (Additional file 1: Table S35). **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 (Additional file 1: Tables S37 and S38). Transmission electron microscopy images of myocardial samples from wild-type (**m**) and *NOA1* homozygous mutant (**n**) 4-month-old heifers. Scale bars = 1 µM. 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 [70–72]. 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

Besnard *et al. Genome Biology*   (2024) 25:248

Page 12 of 30

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 [73]. 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, Additional file 1: Table S29). 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; Additional file 1: Tables S30 and S31; Additional file 2: Fig. S1).

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 ([69, 74]; Fig. 6i). While *NOA1* knockout in mice causes midgestation lethality [74], we estimated that only 24.4% of the homozygous mutants died during pregnancy (Additional file 1: Table S32). An additional 50.7% died before reaching the age of genotyping (Additional file 1: Table S33), and the 24.9% that were genotyped died in most cases before 1 year of age (Additional file 1: Tables S34 and S35; Fig. 6j). Thus, the eight 2- to 4-month-old genotyped homozygotes (one male, seven females) that we followed clinically were among the longest-lived. The females developed ill-thrift between 3 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 (Additional file 1: Table S36, Additional file 2: Note S1). 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 (Additional file 1: Tables S36, S37, and S38; Fig. 6l–n). Finally, the only homozygous male never showed clinical signs until the end of its follow-up at 1 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.

Discussion

About the power of HHED mapping

As mentioned earlier in this article, the first steps of our data mining framework were inspired by the strategy originally developed by VanRaden and colleagues [4] to map embryonic lethal loci in cattle using only SNP array data from live animals, and which has since been successfully applied in dozens of studies in different species (e.g., pig [75], chicken [76], turkey [77], horse [78], and sheep [79]). In contrast to the latter, which consider all available genotypes, the originality of our adaptation is to create groups of cases and controls based on life history information in order to map recessive loci that cause depletion in homozygotes later in life than the usual age at which animals are genotyped.

In addition, the joint analysis of the number of observed (Nobs) and expected (Nexp) homozygotes for a given haplotype in both groups allowed us to apply a looser filter on the (Nobs − Nexp)/Nexp ratio in controls.

In the original version of the approach, this ratio is usually set to − 90% or even − 100% (i.e., corresponding to complete or near-complete depletion) to reduce the risk of false positives before performing validation analyses on at-risk and control mating, thus preventing the mapping of recessive loci showing incomplete LD or penetrance. Here, to simplify the calculation and demonstration, we decided to (i) apply the same arbitrary filters in each of the three breeds (ten or more observed dead heifers, (Nobs − Nexp)/Nexp ≥ 25% in dead heifers and (Nobs − Nexp)/Nexp ≤ − 25% in cows), (ii) select the top 20 haplotypes, and (iii) perform survival analyses in large validation populations to both tentatively confirm the results obtained in the initial mapping step and refine the effect of the haplotypes. We would like to emphasize that our approach includes a confirmation step with an independent validation population, which makes the choice of the initial threshold less critical. Also, since we selected only the top 20 peak haplotypes for each breed, they had a higher enrichment in the case group than our arbitrary filters. Actually, among the 60 selected haplotypes, the minimum values concerned the *N26P48* haplotype in Normande with a depletion of 25% in cows and an enrichment of 76% in dead heifers (Additional file 1: Table S4).

This strategy proved successful, as we validated 34 of the 60 haplotypes studied (with (Nobs − Nexp/Nexp) ratios in the discovery data set ranging from − 25 to − 94% in cows and + 78 to + 267% in dead heifers; Additional file 1: Table S4), including only one previously known locus, namely CD in Holstein, which we blindly retrieved. In addition, it is worth noting that ten of these haplotypes did not cause systematic death of homozygotes at a young age, but rather increased mortality and premature culling rates throughout life (Fig. 3). The fact that we were still able to detect them proves the sensitivity of our approach. Finally, a textbook example to demonstrate the power of our data mining framework is the description of BLIRD in Holstein, a 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 the world's most numerous and studied cattle breed.

Instead of choosing arbitrary preliminary filters, another option would have been to apply Fisher's exact test to the Nobs and Nexp values in the case and control groups. However, as shown in the simulations presented in Additional file 2: Note S2, the threshold corresponding to a chromosome-wide type I error risk of 0.05 after Bonferroni correction for multiple testing depends on the size of the case and control groups, the frequency of the haplotype, and the combination of (Nobs − Nexp)/Nexp ratios in each group (Additional file 1: Table S39). This implies adjusting the threshold for each of the hundreds of thousands of haplotypes examined within each analysis. Despite some computational complications, this option is necessary when a validation population is lacking, as in the case of terminal crosses between purebred strains. In this situation, HHED mapping can be performed in the purebred strains, but the commercial population cannot be used for validation because it is outbred.

Finally, another point of improvement in the methods is to adjust the size of the haplotype window to maximize the LD with the searched variant and to mitigate the effects

of local phasing and imputation errors that may be caused by large structural variations. Although several studies have been able to successfully (re)detect large deletions causing recessive embryonic lethality in cattle using the original approach developed by Van-Raden and colleagues [4] (e.g., a 1.4-Mb deletion encompassing *TFB1M* and a 660-kb deletion encompassing *RNASEH2B*; see [80, 81]), experience shows that detection power and mapping precision are somewhat compromised when using haplotypes that are too large. Here we examined sliding windows of 20 markers (incremented by one marker) from the Illumina BovineSNP50 covering an average of 1 Mb. This size is commonly used because the average r^2 value of LD between two markers separated by 1 Mb is 0.1 [82]. While this haplotype size offers simplicity, consistency, and computational advantages, it may sacrifice accuracy in recombination rates and resolution in genetic analyses compared to methods using genetic maps. One solution proposed by [81] is a two-step procedure: after identifying haplotypes based on fixed windows, they select the lowest *p* value haplotype of various sizes that is within the identified haplotype. Another option is to use the genetic map directly and select haplotypes with different marker sizes based on a fixed LD-distance between markers.

To conclude this section, HHED mapping can be summarized as a more powerful version of homozygosity mapping [3, 83] in that it (i) is based on haplotypes rather than simple markers, (ii) better corrects for population structure by taking into account the genotypes of the ancestors of case and control individuals, (iii) considers not only the observed affected homozygotes but also the missing ones, and, most importantly, (iv) does not assume that all cases are homozygous for the same Mendelian locus. To further demonstrate this, we applied the ASSHOM method described in the reference article for homozygosity mapping in animals by Charlier and colleagues [3] and did not obtain significant results in any of the three breeds studied after contrasting the genotypes of the dead heifers and cows (not shown).

Advantages and limitations of using haplotype information and population-level data to specify the effects of recessive loci

After the mapping and discovery steps, we leveraged the wealth of data at our disposal to gain preliminary insights into the effects of the deleterious haplotypes. While information on death is limited to a date and the mention of "slaughtered" or "dead" (which includes accidents, infections, euthanasia, and any other cause), its low accuracy was mitigated by the large sample sizes. Hence, by (i) comparing the daily difference in proportions between at-risk and control mating for female cattle that died of natural causes, were slaughtered, or were still alive over 6 years, (ii) scoring 12 parameters reflecting the dynamics of increased mortality and slaughter in at-risk versus control mating, and (iii) performing a principal component analysis followed by hierarchical clustering for 34 haplotypes, we distinguished three categories of haplotypes according to age and cause of death (Fig. 3). Of note, the reliability of these categories is supported by previous and follow-up studies that found that *H11P78*/*APOB* and *M6P72*/*NOA1* homozygous mutant calves die in the first weeks or months of life while *H5P25*/*ITGB7* and *N17P57*/*RFC5* homozygous mutants die at a later age or are culled during the first lactation for the long-lived ones.

Besnard *et al. Genome Biology*   (2024) 25:248

Page 15 of 30

We also attempted to estimate the effects of the 34 haplotypes on 14 traits routinely collected for selection purposes, but we were forced to limit our analyses to the comparison between heterozygous and wild-type carriers for two reasons. The first is that the size of the homozygous group was too small for haplotypes segregating at low frequencies. The second, and most critical, is that we do not know a priori what the level of LD is between the detected haplotype and the causative variant. In the case of a recessive mutation that causes early or late juvenile mortality with complete penetrance and that is in incomplete LD with the haplotype, all homozygous animals that survived and have production records are homozygous for the haplotype but not for the mutation, creating a bias. This is a major limitation of haplotype-based studies and it is of paramount importance to identify the causative variants for further characterization.

Exploration of WGS data and filtering strategies to uncover candidate variants

To do this, we relied on whole-genome sequences from 247 Holstein, 160 Montbeliarde, and 118 Normande cattle, as well as 1344 controls from more than 70 breeds that we generated or downloaded from public databases (Additional file 1: Table S14). Given the low number of effective ancestors of the dairy cattle breeds (e.g., Ae = 19, 17, and 21 in a recent pedigree-based estimate in the French Holstein, Montbeliarde, and Normande populations, respectively; https://idele.fr/detail-dossier/varume-resultats-2023; accessed 2024/07/07) and the fact that the sequenced individuals are mainly the most influential ancestors of their breeds, this data set allowed us to capture most of the genetic variability of the three breeds studied and, more generally, of the bovine species. As a result, we had between one and 22 carriers for each of the 33 new deleterious haplotypes, and we were able to search for candidate causative variants for all of them (Additional file 1: Table S15). We decided to apply a combination of one loose and two stringent filters, consisting of (i) selecting variants with a haplotype-genotype square correlation (R^2) of 0.50 or greater, to account for possible incomplete LD, (ii) filtering out variants that were observed in two or more breeds, since recessive defects are generally breed-specific, and (iii) focusing on coding variants predicted to be deleterious to protein function. In doing so, we identified only 8 candidates out of 33 haplotypes which may seem a somewhat low rate (Additional file 1: Table S16). We recognize that in some cases the same recessive defect can segregate in closely related breeds (e.g., for COL7A1-associated dystrophic epidermolysis bullosa in German Vorderwalder and Rotes Höhenvieh cattle; [84, 85]) and that variants located in non-coding regions can severely affect the splicing and expression regulation of neighboring genes (e.g., [86]), although their effects are more difficult to predict than coding variants. Further investigation is therefore required to prioritize the variants associated with haplotypes for which no candidate has yet been found, but here our goal was to identify the best candidates and minimize false positives before embarking on functional studies.

Interest of large-scale genotyping of candidate variants as part of genomic evaluations

The routine use of custom SNP arrays in genomic evaluation that are regularly updated and contain thousands or putative deleterious variants identified by reverse genetic studies (e.g., [6, 9]) offers several advantages. In our particular case, six of the eight candidates we identified had been added to the chip several years ago, giving us access to

genotype data for 0.4 to 1.8 million animals across 15 breeds. These genotype data were of primary importance in demonstrating the causality of the *ITGB7* mutation by analyzing the survival of animals with different genotype × haplotype combinations (see Additional file 1: Table S24). In this specific case, 3% of the 111 females homozygous for the haplotype and heterozygous for the variant died before 2 years of age, compared to 22% of the 214 females homozygous for the variant and heterozygous for the haplotype. These large-scale genotyping data also allow the identification of case and control animals in the field that can be recruited for clinical and functional studies aimed at describing the molecular mechanism underlying these recessive defects. It should be noted, however, that the recruitment of animals scattered over large areas requires a significant logistical and economic investment. To conclude this section, a final advantage is that once the causality of the mutation has been validated, a genetic test for counterselection of these defects can be proposed without delay.

Origin of outbreaks of previously undetected recessive defects

To understand the causes of the emergence of the new recessive defects identified in this study, we examined the evolution of haplotype frequencies over time. We observed patterns consistent with founder effects and genetic drift, characterized by a sudden increase in allele frequencies following massive use of the semen from elite AI bull carriers and from their descendants in later generations. Among them, 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. 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 addition, we sought to determine whether these haplotypes could influence productive traits and be positively selected. Analyzing 14 traits in cohorts of thousands of heterozygous carriers and hundreds of thousands of non-carriers for each of the 34 deleterious haplotypes, we found significant effects in 30% of the analyses (Benjamini–Hochberg adjusted Student's *t*-test *p* value ≤ 0.01), but these were mostly of small magnitude or negative. These results suggest that although there are iconic examples of recessive defects associated with balancing selection in the literature (e.g., see [87] for a review in livestock and [88, 89] for examples in cattle), this phenomenon is the exception rather than the rule.

Impact of previously undetected recessive loci on the industry and challenges of managing them in breeding programs

Although it was not the purpose of this study to quantify the impact of the newly discovered recessive defects on the industry, we can state that it is colossal. The costs associated with each locus depend on several factors (proportion of animals that died or were culled prematurely, age of premature death or culling, veterinary treatments that may have been administered,…) but we can estimate that they vary between 100 and 1000 euros per homozygous calf. Considering that we estimated that more than 290 thousand calves were born homozygous for at least one of the 34 validated haplotypes during the last decade, the economic loss during this period in France alone was approximately

29 to 290 million euros. In addition, these defects raise important environmental and animal welfare concerns which, taken together, strongly argue for their management in selection without any further delay.

This is a major challenge because the historical approach of simply removing carrier bulls from AI catalogs is no longer viable, as nearly all bulls carry at least one of the new deleterious recessive loci. This approach would create a bottleneck, severely limiting genetic progress and potentially leading to the emergence of new defects.

One option that accompanies the widespread use of advanced selection techniques is to avoid at-risk mating by incorporating the causative variants into the custom chips used for genomic evaluation and implementing these genotypes in mating plan software (e.g., [90]). In this scenario, by taking into account the genotypes of the AI bulls and the cows (or of the sires of the cows when they are not genotyped), the prevalence of homozygous births is significantly reduced and genetic progress is not affected. However, while managing mating presents advantages, it does not affect the frequency of the deleterious alleles at the population level. A solution to gradually reduce these frequencies, while limiting the impact on genetic progress consist in applying a selective pressure at the time of recruitment of young male calves destined to become AI bulls. For example, to eradicate the BLIRD anomaly in French Holstein cattle with ~10% carrier frequency, genotyping 4400 male calves instead of 4000 to select the necessary 200 candidates annually can halve the carrier frequency each generation. This approach is economically and genetically efficient and has been applied to manage embryonic loss haplotypes in Holstein, Montbeliarde, and Normande breeds over the past decade.

Prospects for the use of cattle as a model species

Beyond their positive impact on the livestock industry, our results also have interesting implications for basic research. The description of the pathological consequences associated with the inactivation of two genes (*NOA1* and *RFC5*), for which no live homozygous mutants had previously been reported in mammals, illustrates the prospects offered by the use of farm animals as model species. Historically, research in biology has focused on a very limited number of model species selected for their high reproductive rate, short generation interval, small size, and low maintenance costs (e.g., [91]). The availability of massive pedigree, phenotype, genotype, and whole-genome sequence data in cattle, combined with their unique population structure, invites a reconsideration of this paradigm. It is even more true that cattle offer a number of advantages, such as greater similarity to humans in terms of physiology, body size, and longevity than mice, and the ability to study paternal half-sib families of thousands of individuals in diverse environments thanks to the use of artificial insemination [9].

Conclusions

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 Montbeliarde, and 1.24 million Normande females whose sires and maternal grandsires (MGS) were genotyped. Of these, 593,445 Holstein, 333,520 Montbeliarde, and 103,975 Normande females were also genotyped, themselves.

The discovery population was restricted to genotyped females and included 8203, 6198, and 2254 "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 (Additional file 1: Table S1). 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, ~777 K). Raw genotypes were imputed and phased for 44,596 autosomal SNPs by FImpute [92] as part of the French routine genomic evaluation of cattle, as described by [93] (Additional file 1: Table S2). We considered sliding haplotypes of 20 markers (incremented by one marker) 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 ≥ 10 in cases, increase in homozygotes (i.e., (Nobs$-$Nexp)/Nexp) $\geq 25\%$ in cases and $\leq -25\%$ in controls (Additional file 1: Table S3). Among stretches of consecutive haplotypes, we selected the one showing the highest increase in homozygotes in cases as the "peak haplotype" (Additional file 1: Table S4). 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 ≤ 0.05) among the descendants of at-risk mating ("1"; carrier sire and carrier MGS) or control mating ("0"; non-carrier 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). This analysis was performed on a validation population of 5.96 million Holstein, 1.63 million Montbeliarde, and 1.24 million Normande females born between 2000 and 2015, most of whom were not genotyped themselves (Additional file 1: Table S5). Then, for 34 validated haplotypes, we calculated the daily proportion of females 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 (Additional file 1: Table S8). 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 (Additional file 1: Table S9). 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 (Additional file 1: Table S10).

Male calves were not included in this study, due to the fact that they are usually sent to feedlots farms at a young age (~1 month), making them difficult to monitor, especially when exported to other countries.

Estimation of the number of calves born homozygous for one or more validated haplotypes in the last 10 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): $NH = NAI*PR*CR*0.25/\%AI$ (Additional file 1: Table S7; [2]).

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 (Additional file 1: Table S11).

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 (Additional file 1: Table S12) using the PEDIG software [94].

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 [95]. The sample size ranged from 1930 to 509,258 cows per group (Additional file 1: Table S13). 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 haplotypes are in complete linkage disequilibrium with the causal variants. 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 1869 cattle from more than 70 different breeds generated by Illumina technology (Additional file 1: Table S14; [16–66]). Of these, 1137 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; [96]). SNPs and small InDels were identified using the GATK-HaplotypeCaller software [97] as previously described [98, 99], while structural variations (SVs) were detected using the Pindel [100], Delly [101], and Lumpy [102] software.

Variant filtering and annotation

R^2 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 (Additional file 1: Table S15). Variants with an R^2

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 patho-physiological 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 (Additional file 1: Table S16). 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 ([103]; https://www.genome.jp/tools-bin/clustalw). Finally, a sequence logo was generated using WebLogo ([104]; 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 (*Saccharomyces 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 Additional file 1: Tables S19 and S28.

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/; [105]). 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 Additional file 1: Tables S6, S16, S17 and S18. 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* (Additional file 1: Table S23) and *M6P72*/*NOA1* (Additional file 2: Fig. S2).

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 (Fig. 5; Additional file 1: Table S23). 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 Additional file 1: Table S2 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 × haplotype status (Additional file 1: Tables S24 and S34). 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 (Additional file 1: Tables S27, S31 and S35).

Effects of the *ITGB7*, *RFC5* and *NOA1* variants 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; Additional file 1: Table S26), for age at first insemination and age at first calving (Additional file 1: Table S25), and for birth weight (Additional file 1: Table S30). 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; Additional file 1: Table S32). 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 (Centered_CR_0*0 − 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 Additional file 1: Table S33 for more information).

Pathophysiological examination

ITGB7, *RFC5*, and *NOA1* homozygous mutant cattle aged 2 months to 3 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 Additional file 1: Table S36) 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 Additional file 1: Tables S36 and S37.

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) [106]. 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 (1000 ×). 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 milliliter 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 $HClO_4$ extracts were neutralized with 20% (w/v) KOH before analysis according to [107]. Biological parameters were compared with reference samples prepared from healthy animals according to ASVCP recommendations [108].

Quantification of CD4 + mem a4 + 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, $1-2 \times 10^6$ mononuclear cells were resuspended in HBSS, 0.5% BSA, and 10 mM Hepes. Cell viability was assessed using Viobility 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, USA), CD45-FITC (clone 1.11.32,

Biorad), CD45R0-A647 (clone IL-A116, Biorad), Beta7-RPE (clone FIB27, Biolegend, San Diego, USA), and 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 Sects. (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 Pannoramic Scan 150 and analyzed with CaseViewer 2.4 software (3D Histech). The number of hair follicles in a randomly selected 1 mm^2 square and the diameter of the pilary canal of 50 adjacent hair follicles were measured for each animal (Additional file 1: Table S29).

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 h 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 Sects. (70 nm) were collected on 200 mesh copper grids and counterstained with lead citrate. The grids were examined on a Hitachi HT7700 electron microscope operated at 80 kV (Milexia, France) and images were acquired using 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; Additional file 1: Tables S3 and S38). Quantitative PCR was performed in triplicate on a QuantStudio 12 K 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 µL 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 µL. For each animal, the ΔCt(*CYTB-PPIA*) and ΔCt(*COX1-RSP24*) values were calculated based on the mean cycle threshold (Ct) of the triplicates. Finally, the $2^{-\Delta\Delta Ct(CYTB-PPIA)}$ and $2^{-\Delta\Delta Ct(COX1-RSP24)}$ 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.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-024-03384-7.

Additional file 1: Supplementary tables. Table S1 Information about the animals involved in homozygous haplotype enrichment/depletion (HHED) mapping. Table S2 Information on the 44,596 markers from the Illumina BovineSNP50 array used in this study. Table S3 Results of HHED mapping. Table S4 Information on the top 20 loci per breed selected after HHED mapping. Table S5 Validation of the top 20 HHED loci per breed using at-risk and control mating. Table S6 Contingency table between *H11P78* haplotype status and genotypes at the *CD* locus in 721,006 Holstein cattle. Table S7 Analysis of artificial insemination records for at-risk mating [2]. Table S8 Survival analysis in at-risk and control mating for given haplotypes. Table S9 Daily difference in proportions between at-risk and control mating for animals that died of natural causes (D), were slaughtered (S), or were still alive (A) over a 6-year period. Table S10 Details of parameters recorded on survival curves for principal component analysis and hierarchical clustering. Table S11 Evolution of haplotype frequencies over time. Table S12 Details of the most important ancestors of the Holstein, Montbeliarde, and Normande female populations born within the period 2019–2022. Table S13 Effects of the 34 haplotypes in the heterozygous state on traits routinely recorded for evaluation purposes. Table S14 Details of the whole-genome sequences considered in this study [16–66]. Table S15 Number of carrier alleles for each of the validated haplotypes among the 247, 160, and 118 whole-genome sequenced Holstein, Montbeliarde, and Normande individuals. Table S16 Information on the candidate variants and their putative phenotypic consequences. Table S17 Large-scale genotyping of candidate variants. Table S18 Statistics of associations between haplotype status and genotypes for six candidate variants. Table S19 Extract from the multiple alignment of 128 vertebrate orthologs around the bovine ITGB7 candidate substitution plus or minus 10 amino acids. Table S20 Results of the quantification of CD4 + mem a4 + b7 + T lymphocytes in the intestinal lamina propria of *ITGB7* homozygous mutant and control heifers. Table S21 Blood analysis results from seven homozygous mutant heifers for the *ITGB7* substitution and seven homozygous wild-type matched controls. Table S22 Analysis of the average daily gain in 13 homozygous mutant heifers for the *ITGB7* substitution and 13 homozygous wild-type matched controls. Table S23 Details of the IBD segments shared by the bulls ELEVATION, ELTON, and O-MAN around haplotype H5P25 and of their genotype for *ITGB7* substitution. Table S24 Analysis of survival at 2 years of age for 310,479 females according to their combination of genotype for the *ITGB7* causative variant and haplotype status at H5P25. Table S25 Comparison of the mean age at first insemination and at the beginning of first lactation between groups defined on the basis of genotypes at the *ITGB7* causative variant. Table S26 Effects of the *ITGB7* candidate variant in the heterozygous and homozygous states on traits recorded for evaluation purposes. Table S27 Six-year survival analysis of Holstein females for each genotype group at the *ITGB7* substitution. Table S28 Extract from 633 eukaryotic orthologs around the bovine RFC5 inframe deletion plus or minus 30 amino acids. Table S29 Analysis of the hair density and median diameter of pilary canals on the shoulder skin of 4 homozygous mutant heifers for the *RFC5* inframe deletion and 4 homozygous wild-type matched controls. Table S30 Comparison of the mean birth weight between groups defined on the basis of imputed genotypes at the *RFC5*, *NOA1*, and *ITGB7* causative variants. Table S31 Six-year survival analysis of Normande females for each genotype group at the *RFC5* inframe deletion. Table S32 Analysis of conception rate in mating between males and females genotyped for the *NOA1* frameshift variant. Table S33 Estimation of the proportion of *NOA1* homozygous mutants that died between birth and genotyping. Table S34 Analysis of survival at 2 years of age of 402,729 females according to their combination of genotype for the *NOA1* frameshift variant and haplotype status at *M6P72*. Table S35 Six-year survival analysis of Montbeliarde females for each genotype group at the *NOA1* frameshift insertion. Table S36 Blood analysis results for *NOA1* homozygous mutants and controls. Table S37 Analysis of relative changes in the ratio of mitochondrial to nuclear DNA between *NOA1* homozygous mutant and control heifers by quantitative PCR. Table S38 Information on primers used for qPCR analyses. Table S39 Simulated data for HHED mapping based on the number of dead heifers and cows used in the 3 breeds.

Additional file 2: Supplementary figures and notes. Fig. S1 Trimester repartition of the proportion of females that died or were slaughtered before reaching 6 years of age by genotype group at the *RFC5* substitution. Fig. S2 Analyzing the origin of *M6P7*2 haplotype and *NOA1* frameshift variant. Supplementary Note S1 Neutrophils from *NOA1* homozygous mutant calves are functional [109–114]. Supplementary Note S2 Simulations to define significant thresholds for HHED mapping.

Additional file 3. Review history.

Peer review information

Andrew Cosgrove was the primary editor of this article at Genome Biology and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 3.

Authors' 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.

Availability of data and materials
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 [35] and PRJEB64023 [36]. Sequences from 1137 cattle from previous studies are available in the NCBI BioProject and ENA databases under the accession numbers listed in Additional file 1: Table S14 [16–66]. Light and transmission electron microscopy data are available in a public figshare repository accessible under the doi: https://doi.org/10.6084/m9.figshare.26819779.v2. All other datasets supporting the conclusions of this article are included in the article and its additional files, which consists of 39 supplementary tables for the first additional file, as well as two supplementary figures and two supplementary notes for the second additional file.

Declarations

Ethics approval and consent to participate
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.

Competing interests
The authors declare no competing interests.

Author details
[1]Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78350 Jouy-en-Josas, France. [2]IDELE, 149 Rue de Bercy, 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, INRAE, Ecole Nationale Vétérinaire d'Alfort, BREED, 78350 Jouy-en-Josas, France. [12]UE326, 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.

References
1. Danchin-Burge C, Leroy G, Brochard M, Moureaux S, Verrier E. Evolution of the genetic variability of eight French dairy cattle breeds assessed by pedigree analysis. J Anim Breed Genet. 2012;129:206–17.
2. Escouflaire C, Capitan A. Analysis of pedigree data and whole-genome sequences in 12 cattle breeds reveals extremely low within-breed Y-chromosome diversity. Anim Genet. 2021;52:725–9.
3. Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, et al. Highly effective SNP-based association mapping and management of recessive defects in livestock. Nat Genet. 2008;40:449–54.
4. VanRaden PM, Olson KM, Null DJ, Hutchison JL. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. J Dairy Sci. 2011;94:6153–61.
5. Fritz S, Capitan A, Djari A, Rodriguez SC, Barbat A, Baur A, et al. Detection of haplotypes associated with prenatal death in dairy cattle and identification of deleterious mutations in GART, SHBG and SLC37A2. PLoS One. 2013;8:e65550.

6.   Michot P, Chahory S, Marete A, Grohs C, Dagios D, Donzel E, et al. A reverse genetic approach identifies an ancestral frameshift mutation in RP1 causing recessive progressive retinal degeneration in European cattle breeds. Genet Sel Evol. 2016;48:56.
7.   Charlier C, Li W, Harland C, Littlejohn M, Coppieters W, Creagh F, et al. NGS-based reverse genetic screen for common embryonic lethal mutations compromising fertility in livestock. Genome Res. 2016;26:1333–41.
8.   Reynolds EGM, Neeley C, Lopdell TJ, Keehan M, Dittmer K, Harland CS, et al. Non-additive association analysis using proxy phenotypes identifies novel cattle syndromes. Nat Genet. 2021;53:949–54.
9.   Bourneuf E, Otz P, Pausch H, Jagannathan V, Michot P, Grohs C, et al. Rapid discovery of de novo deleterious mutations in cattle enhances the value of livestock as model species. Sci Rep. 2017;7:11466.
10. Li W, Sartelet A, Tamma N, Coppieters W, Georges M, Charlier C. 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. Anim Genet. 2016;47:110–3.
11. Kipp S, Segelke D, Schierenbeck S, Reinhardt F, Reents R, Wurmser C, et al. Identification of a haplotype associated with cholesterol deficiency and increased juvenile mortality in Holstein cattle. J Dairy Sci. 2016;99:8915–31.
12. Menzi F, Besuchet-Schmutz N, Fragnière M, Hofstetter S, Jagannathan V, Mock T, et al. A transposable element insertion in APOB causes cholesterol deficiency in Holstein cattle. Anim Genet. 2016;47:253–7.
13. Mock T, Mehinagic K, Menzi F, Studer E, Oevermann A, Stoffel MH, et al. Clinicopathological phenotype of autosomal recessive cholesterol deficiency in Holstein cattle. J Vet Intern Med. 2016;30:1369–75.
14. Häfliger IM, Hofstetter S, Mock T, Stettler MH, Meylan M, Mehinagic K, et al. APOB-associated cholesterol deficiency in Holstein cattle is not a simple recessive disease. Anim Genet. 2019;50:372–5.
15. Gross JJ, Schwinn A-C, Schmitz-Hsu F, Menzi F, Drögemüller C, Albrecht C, et al. Rapid communication: cholesterol deficiency–associated APOB mutation impacts lipid metabolism in Holstein calves and breeding bulls1. J Anim Sci. 2016;94:1761–6.
16. Tokyo University of Agriculture. PRJDA48395; 2010. https://www.ncbi.nlm.nih.gov/bioproject/PRJDA48395.
17. Department of Bioscience, Faculty of Applied Bioscience, Tokyo University of Agriculture. PRJDB2660; 2014. https://www.ncbi.nlm.nih.gov/bioproject/PRJDB2660.
18. Institute of Genetics, University of Bern, Switzerland. PRJEB5435; 2014. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB5435.
19. Institute of Genetics, University of Bern, Switzerland. PRJEB5965; 2014. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB5965.
20. Institute of Genetics, University of Bern, Switzerland. PRJEB7527; 2014. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB7527.
21. Institute of Genetics, University of Bern, Switzerland. PRJEB7528; 2014. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB7528.
22. Institute of Genetics, University of Bern, Switzerland. PRJEB8226; 2015. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB8226.
23. INRA JOUY-EN-JOSAS. PRJEB9343; 2015. https://www.ncbi.nlm.nih.gov/bioproject/292988.
24. Institute of Genetics, University of Bern, Switzerland. PRJEB11962; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB11962.
25. Institute of Genetics, University of Bern, Switzerland. PRJEB12093; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB12093.
26. Institute of Genetics, University of Bern, Switzerland. PRJEB12094; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB12094.
27. INRA JOUY-EN-JOSAS. PRJEB12703; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB12703.
28. Institute of Genetics, University of Bern, Switzerland. PRJEB14604; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB14604.
29. Institute of Genetics, University of Bern, Switzerland. PRJEB18113; 2016. https://www.ncbi.nlm.nih.gov/bioproject/356238.
30. INRA, UMR1313 GENETIQUE ANIMALE ET BIOLOGIE INTEGRATIVE. PRJEB27309; 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB27309.
31. Natural Resources Institute Finland (Luke). PRJEB28185; 2019. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB28185.
32. ETH ZURICH. PRJEB28191; 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB28191.
33. ETH ZURICH. PRJEB29487; 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB29487.
34. Trinity College Dublin, Republic of Ireland. PRJEB31621; 2019. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB31621.
35. inrae jouy-en-josas. PRJEB64022; 2024. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB64022.
36. INRAE - GeT-PlaGe. PRJEB64023; 2023. https://www.ncbi.nlm.nih.gov/bioproject/PRJEB64023.
37. Texas A&M University. PRJNA174819; 2012. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA174819.
38. Livestock Gentec, University of Alberta. PRJNA176557; 2012. https://www.ncbi.nlm.nih.gov/bioproject/176557.
39. Seoul National University. PRJNA210519; 2013. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA210519.
40. Seoul National University. PRJNA210523; 2013. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA210523.
41. The 1000 Bull Genomes Consortium. PRJNA238491; 2014. https://www.ncbi.nlm.nih.gov/bioproject/238491.
42. Livestock Gentec, University of Alberta. PRJNA256210; 2014. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA256210.
43. USDA. PRJNA277147; 2015. https://www.ncbi.nlm.nih.gov/bioproject/277147.
44. Ludwig-Maximilians-University. PRJNA279385; 2015. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA279385.
45. University College Dublin. PRJNA294709; 2015. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA294709.
46. Seoul National University. PRJNA318087; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA318087.
47. Seoul National University. PRJNA318089; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA318089.
48. USDA-ARS-USMARC. PRJNA324270; 2016. https://www.ncbi.nlm.nih.gov/bioproject/324270.

49. USDA, ARS, USMARC and Intrepid Bioinformatics. PRJNA324822; 2016. https://www.ncbi.nlm.nih.gov/bioproject/324822.
50. USDA, ARS, USMARC. PRJNA325058; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA325058.
51. University of Missouri. PRJNA343262; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA343262.
52. University of Veterinary Medicine Hannover, Foundation. PRJNA350739; 2016. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA350739.
53. Northwest A&F University. PRJNA379859; 2017. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA379859.
54. Ludwig-Maximilians-University. PRJNA411962; 2017. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA411962.
55. USDA-ARS. PRJNA422135; 2017. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA422135.
56. Agriculture Victoria Department of Economic Development, Jobs, Transport and Resources. PRJNA431934. 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA431934.
57. University of Veterinary Medicine Hannover, Foundation. PRJNA438601; 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA438601.
58. Cardiff University. PRJNA471656; 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA471656.
59. University of Alberta and Teagasc, Animal & Grassland Research and Innovation Centre. PRJNA474946; 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA474946.
60. Teagasc. PRJNA477833; 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA477833.
61. University of California, Davis. PRJNA494431; 2018. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA494431.
62. University of Nebraska Lincoln. PRJNA513064; 2019. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA513064.
63. Wageningen University and Research. PRJNA514237; 2019. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA514237.
64. University of Nebraska - Lincoln. PRJNA551500; 2019. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA551500.
65. University of Adelaide. PRJNA603764; 2020. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA603764.
66. The Royal Veterinary College. PRJNA642008; 2020. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA642008.
67. Gorfu G, Rivera-Nieves J, Ley K. Role of β7 integrins in intestinal lymphocyte homing and retention. Curr Mol Med. 2009;9:836–50.
68. Jourdain J, Barasc H, Faraut T, Calgaro A, Bonnet N, Marcuzzo C, et al. Large-scale detection and characterization of interchromosomal rearrangements in normozoospermic bulls using massive genotype and phenotype data sets. Genome Res. 2023;33:957–71.
69. Al-Furoukh N, Kardon JR, Krüger M, Szibor M, Baker TA, Braun T. NOA1, a novel ClpXP substrate, takes an unexpected nuclear detour prior to mitochondrial import. PLoS One. 2014;9:e103141.
70. Cullmann G, Fien K, Kobayashi R, Stillman B. Characterization of the five replication factor C genes of Saccharomyces cerevisiae. Mol Cell Biol. 1995;15:4661–71.
71. Furukawa T, Ishibashi T, Kimura S, Tanaka H, Hashimoto J, Sakaguchi K. Characterization of all the subunits of replication factor C from a higher plant, rice (Oryza sativa L.), and their relation to development. Plant Mol Biol. 2003;53:15–25.
72. Li Y, Gan S, Ren L, Yuan L, Liu J, Wang W, et al. Multifaceted regulation and functions of replication factor C family in human cancers. Am J Cancer Res. 2018;8:1343–55.
73. Reynolds N, Fantes PA, MacNeill SA. A key role for replication factor C in DNA replication checkpoint function in fission yeast. Nucleic Acids Res. 1999;27:462–9.
74. Kolanczyk M, Pech M, Zemojtel T, Yamamoto H, Mikula I, Calvaruso M-A, et al. NOA1 is an essential GTPase required for mitochondrial protein synthesis. Mol Biol Cell. 2011;22:1–11.
75. Derks MFL, Megens H-J, Bosse M, Lopes MS, Harlizius B, Groenen MAM. A systematic survey to identify lethal recessive variation in highly managed pig populations. BMC Genomics. 2017;18:858.
76. Derks MFL, Megens H-J, Bosse M, Visscher J, Peeters K, Bink MCAM, et al. A survey of functional genomic variation in domesticated chickens. Genet Sel Evol. 2018;50:17.
77. Abdalla EA, Id-Lahoucine S, Cánovas A, Casellas J, Schenkel FS, Wood BJ, et al. Discovering lethal alleles across the turkey genome using a transmission ratio distortion approach. Anim Genet. 2020;51:876–89.
78. Todd ET, Thomson PC, Hamilton NA, Ang RA, Lindgren G, Viklund Å, et al. A genome-wide scan for candidate lethal variants in Thoroughbred horses. Sci Rep. 2020;10:13153.
79. Ben Braiek M, Fabre S, Hozé C, Astruc J-M, Moreno-Romieux C. Identification of homozygous haplotypes carrying putative recessive lethal mutations that compromise fertility traits in French Lacaune dairy sheep. Genet Sel Evol. 2021;53:41.
80. Schütz E, Wehrhahn C, Wanjek M, Bortfeld R, Wemheuer WE, Beck J, et al. The Holstein Friesian lethal haplotype 5 (HH5) results from a complete deletion of TBF1M and cholesterol deficiency (CDH) from an ERV-(LTR) insertion into the coding region of APOB. PLoS ONE. 2016;11:e0154602.
81. Wu X, Mesbah-Uddin M, Guldbrandtsen B, Lund MS, Sahana G. Novel haplotypes responsible for prenatal death in Nordic Red and Danish Jersey cattle. J Dairy Sci. 2020;103:4570–8.
82. Hoze C, Fouilloux MN, Venot E, Guillaume F, Dassonneville R, Fritz S, et al. High density marker imputation efficiency in 16 French cattle breeds. Genet Sel Evol. 2013;45:33.
83. Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. Science. 1987;236:1567–70.
84. Menoud A, Welle M, Tetens J, Lichtner P, Drögemüller C. A COL7A1 mutation causes dystrophic epidermolysis bullosa in Rotes Höhenvieh cattle. PLoS One. 2012;7:e38823.
85. Pausch H, Ammermüller S, Wurmser C, Hamann H, Tetens J, Drögemüller C, et al. A nonsense mutation in the COL7A1 gene causes epidermolysis bullosa in Vorderwald cattle. BMC Genet. 2016;17:149.
86. Boulling A, Corbeau J, Grohs C, Barbat A, Mortier J, Taussat S, et al. A bovine model of rhizomelic chondrodysplasia punctata caused by a deep intronic splicing mutation in the GNPAT gene; 2024. p. 2024.06.13.598642. bioRxiv; 2024. Available from: https://www.biorxiv.org/content/10.1101/2024.06.13.598642v1. Cited 2024 Jul 8.
87. Derks MFL, Steensma M. Review: balancing selection for deleterious alleles in livestock. Front Genet. 2021;12:761728.

88. Fasquelle C, Sartelet A, Li W, Dive M, Tamma N, Michaux C, et al. Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the crooked tail syndrome in Belgian Blue cattle. PLoS Genet. 2009;5:e1000666.

89. Kadri NK, Sahana G, Charlier C, Iso-Touru T, Guldbrandtsen B, Karim L, et al. A 660-Kb deletion with antagonistic effects on fertility and milk production segregates at high frequency in Nordic Red cattle: additional evidence for the common occurrence of balancing selection in livestock. PLoS Genet. 2014;10:e1004049.

90. Brochard M, Boichard D, Capitan A, Guillaume G, Fritz S, Nicod L, et al. pANO, le risque d'anomalie létale pour les produits d'accouplements: principe et utilisation en race Montbéliarde sur la zone Gen'IATest. Proc 25th Rencontres Rercherches Ruminants, Paris, December 5-6 2018. https://journees3r.fr/spip.php?article4562.

91. Simmons D. The use of animal models in studying genetic disease: transgenesis and induced mutation. Nat Educ. 2008;1:70.

92. Sargolzaei M, Chesnais JP, Schenkel FS. A new approach for efficient genotype imputation using information from relatives. BMC Genomics. 2014;15:478.

93. Mesbah-Uddin M, Hoze C, Michot P, Barbat A, Lefebvre R, Boussaha M, et al. A missense mutation (p.Tyr452Cys) in the CAD gene compromises reproductive success in French Normande cattle. J Dairy Sci. 2019;102:6340–56.

94. Boichard D. Pedig: a fortran package for pedigree analysis suited for large populations. 2002; Available from: https://www6.jouy.inra.fr/gabi_eng/Our-resources/Tool-development/Pedig.

95. Misztal I, Lourenco D, Aguilar I, Legarra A, Vitezica Z. Manual for BLUPF90 family of programs. 2014. Available from: https://nce.ads.uga.edu/wiki/doku.php?id=application_programs.

96. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

97. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303.

98. Daetwyler HD, Capitan A, Pausch H, Stothard P, Van Binsbergen R, Brøndum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nat Genet. 2014;46:858–65.

99. Boussaha M, Michot P, Letaief R, Hozé C, Fritz S, Grohs C, et al. Construction of a large collection of small genome variations in French dairy and beef breeds using whole-genome sequences. Genet Sel Evol. 2016;48:87.

100. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009;25:2865–71.

101. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28:i333–9.

102. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15:R84.

103. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res. 1994;22:4673–80.

104. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.

105. Rodrigues CHM, Myung Y, Pires DEV, Ascher DB. mCSM-PPI2: predicting the effects of mutations on protein–protein interactions. Nucleic Acids Res. 2019;47:W338–44.

106. Herman N, Trumel C, Geffré A, Braun J-P, Thibault M, Schelcher F, et al. Hematology reference intervals for adult cows in France using the Sysmex XT-2000iV analyzer. J VET Diagn Invest. 2018;30:678–87.

107. Baird GD, Heitzman RJ. Gluconeogenesis in the cow. The effects of a glucocorticoid on hepatic intermediary metabolism. Biochem J. 1970;116:865–74.

108. Friedrichs KR, Harr KE, Freeman KP, Szladovits B, Walton RM, Barnhart KF, et al. ASVCP reference interval guidelines: determination of de novo reference intervals in veterinary species and other related topics. Vet Clin Pathol. 2012;41:441–53.

109. Cobessi D, Dumas R, Pautre V, Meinguet C, Ferrer J-L, Alban C. Biochemical and structural characterization of the Arabidopsis bifunctional enzyme dethiobiotin synthetase–diaminopelargonic acid aminotransferase: evidence for substrate channeling in biotin synthesis. Plant Cell. 2012;24:1608–25.

110. Frederickson Matika DE, Loake GJ. Redox regulation in plant immune function. Antioxid Redox Signal. 2014;21:1373–88.

111. MacMicking J, Xie Q, Nathan C. Nitric oxide and macrophage function. Annu Rev Immunol. 1997;15:323–50.

112. Lundberg JO, Weitzberg E. Nitric oxide signaling in health and disease. Cell. 2022;185:2853–78.

113. Guzik TJ, Korbut R, Adamek-Guzik T. Nitric oxide and superoxide in inflammation and immune regulation. J Physiol Pharmacol. 2003;54:469–87.

114. Rambault M, Borkute R, Doz-Deblauwe E, Le-Vern Y, Winter N, Dorhoi A, et al. Isolation of bovine neutrophils by fluorescence- and magnetic-activated cell sorting. In: Brandau S, Dorhoi A, editors., et al., Myeloid-derived suppressor cells. New York: Springer US; 2021. p. 203–17. https://doi.org/10.1007/978-1-0716-1060-2_16.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.