Massive detection of cryptic recessive genetic defects in livestock mining millions of life trajectories

Ana Guintard, Florian Besnard, Margarita Cano, Chris Hoze, Hélène Leclerc, Mekki Boussaha, Cécile Grohs, Anne Barbat, Sebastien Fritz, Clémentine Escouflaire, et al.

Massive detection of cryptic recessive genetic defects in livestock mining millions of life trajectories

A. Guintard[1,2], F. Besnard[1,3], M. Cano[1,2], C. Hozé[1,2], H. Leclerc[1,2], M. Boussaha[1], C. Grohs[1], A. Barbat[1], S. Fritz[1,2], C. Escouflaire[1,2], J. Rivière[1,4], C. Péchoux[1], C. Danchin-Burge[3], G. Foucras[5], A. Relun[6], V. Plassard[7], M.-A. Arcangioli[8], S. Mattalia[1,3], D. Boichard[1] and A. Capitan[1,2*]

[1] Université Paris-Saclay, INRAE, AgroParisTech, GABI, 4 av. Jean Jaurès, 78350 Jouy-en-Josas, France; [2] ALLICE, 149 rue de Bercy, 75012 Paris, France; [3] IDELE, 149 rue de Bercy, 75012 Paris, France; [4] Université Paris-Saclay, INRAE, AgroParisTech, MICALIS, 4 av. Jean Jaurès, 78350 Jouy-en-Josas, France; [5] Université de Toulouse, ENVT, IHAP, 23 ch. des Capelles, 31300 Toulouse, France; [6] INRAE, Oniris, BIOEPAR, La Chantrerie, 44307 Nantes, France; [7] ENVA, 7 av. du Général de Gaulle, 94700 Maisons-Alfort, France; [8] VetAgro Sup, University of Lyon1, 1 av. Bourgelat, 69280 Marcy-l'Etoile, France; * aurelien.capitan@inrae.fr

Abstract

A data mining method applied to large-scale genotyping data is proposed to detect recessive loci responsible for increased mortality in cattle and that have remained undetected by previous approaches. It is based on a screen for homozygous haplotype enrichment/depletion in groups of females with different life trajectories. After validation of the results in at risk and control mating, 34 deleterious haplotypes (13 in Holstein, 11 in Montbéliarde, and 10 in Normande) were identified, with frequencies ranging from 1.5 to 7.6%. Profiles of survival curves and causes of mortality differed greatly between loci, with early juvenile, late juvenile and evenly distributed death events. Candidate causal variants were found for fifteen haplotypes. A frameshift mutation of *NOA1* and a disruptive inframe deletion of *RFC5*, affecting two genes with no previous record of live homozygous mutants in mammals, were subject to phenotypical characterization.

Introduction

Due to their limited effective size and intense selection, cattle breeds experience a gradual increase in inbreeding, favouring the emergence of recessive defects. While a few of them are highly visible, many remain overlooked because they induce mortality during gestation or after birth but with non-specific symptoms. With the tremendous development of genomic selection and genotyping, new methods have been proposed to detect these defects such as the deficit in homozygotes (VanRaden et al., 2011), reverse genetics (e.g. Michot et al., 2016), or GWAS accounting for dominance (Reynolds et al., 2021). In this paper, we propose another approach combining searches for enrichment/depletion in homozygotes and analyses of life trajectories. We demonstrate its detection power with the identification of several dozens of new defects responsible for increased mortality rate.

Material & Methods

Mapping of recessive loci and analysis of life trajectories. Information on pedigree and life trajectory (date of birth; lifespan; cause of death; and ages at first insemination, calving and

lactation) was extracted from the French National database for 5.96 million Holstein, 1.63 million Montbéliarde, and 1.24 million Normande females with Illumina Bovine SNP50 array genotypes available for their sire and maternal grand sire (MGS). Among them, 8,203, 6,198 and 2,254 cases (heifers that died of natural cause before 3 years and were never inseminated) and 291,529, 141,343 and 56,095 controls (females that calved and started a first lactation) from the three breeds, respectively, were also genotyped with various Illumina arrays and their genotypes were imputed and phased to the Bovine SNP50 as described in Mesbah-Uddin *et al.* (2019). For sliding haplotypes of 20 markers we counted the number of homozygotes (hmz) observed (Nobs) within each group of genotyped individuals and in parallel we estimated the expected number of hmz (Nexp) using within-family transmission probability. We filtered haplotypes satisfying the following criteria: Nobs ≥10 in cases, increase in hmz (i.e. (Nobs-Nexp)/Nexp)) ≥ 25% in cases and ≤ -25% in controls. Among stretches of consecutive haplotypes, we selected the one showing the highest increase in hmz in cases as the "peak haplotype". For validation we compared the proportions of animals belonging to three categories (case, control and others) using a χ^2 test with Benjamini-Hochberg correction; p-value ≤ 5%) among the descendants of at-risk mating ("1"; carrier sire and carrier MGS) or control mating ("0"; noncarrier sire and MGS) born between 2000 and 2015 (and that were mostly not genotyped themselves). Then, for 34 validated haplotypes, we calculated the daily proportion of animals that died of a natural death (D), were slaughtered (S) or were still alive (A) over a period of 6 years for mating types 0 and 1. We also calculated the D0-D1, S0-S1 and A0-A1 differences in proportions on a daily basis and scored the days for which 25, 50, 75 and 100 % of the maximum deviation between each difference in proportions was reached. Subsequently, we used these 12 parameters to perform a Principal Component Analysis (PCA) and a Hierarchical Clustering (HC) using RStudio package Factoshiny, v 1.2.5033.

Estimation of the effect of the haplotypes on various traits. For each of the 34 selected haplotypes we estimated the effects of the three genotypes on 11 traits (see results). We used yield deviation data corrected for environmental effects for each animal in a model including the haplotype as a fixed effect and a polygenic effect through a pedigree relationship matrix. Analyses were performed with BLUPF90 (Misztal et al., 2014).The size of genotype groups ranged from 4 (minimum set value) to 346,210 animals.

Analysis of whole genome sequences. The genomes of 611 cattle from 22 breeds (including 190 Holstein, 72 Montbéliarde and 97 Normande individuals) were sequenced with Illumina technology and processed as described in Boussaha et al. (2016). Breed specific SNPs or short Indels carried by at least 50% of the haplotype carriers, predicted to be deleterious and located within $\pm$ 10 Mb from the haplotype were considered as candidate causative mutations.

Clinical examination. Five *NOA1* and five *RFC5* homozygous mutants were subject to clinical examination ante and post mortem including blood dosage and histological analyses.

Results and discussion

Our screen for homozygous haplotype enrichment or depletion (HHED) in groups of females with different life trajectories identified numerous regions, which led us to consider the top 20 peak haplotypes per breed (figure 1a). In total, 34 haplotypes (13 in Holstein, 11 in Montbéliarde, and 10 in Normande) with frequencies comprised between 1.5 and 7.6% were confirmed to increase juvenile mortality in validation populations. Among them only one

haplotype colocalized with a deleterious locus previously reported, namely cholesterol deficiency in Holstein (CDH; e.g. Menzi et al., 2016). The proportion of mating at risk between sire and MGS for one or more validated haplotypes reached 4.7, 7.1 and 8.4 % for the three breeds, respectively. The subsequent analysis of the life trajectories of animals from at risk vs control mating using PCA and HC (figure 1 b-f) distinguished three categories of haplotypes depending on the age and causes of death: early juvenile mortality (d), progressive mortality in juveniles and subadults (e) and increased mortality and premature culling throughout life (f).

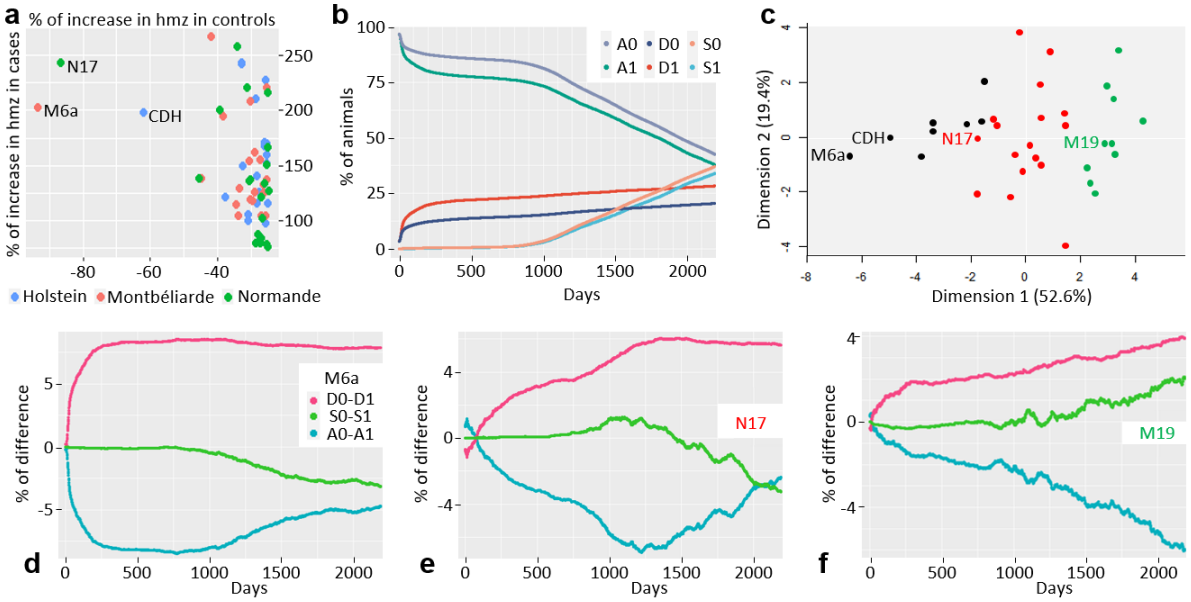

Figure 1. Mapping and characterization of life trajectories. a) increase in hmz in the case and control groups for the 60 selected haplotypes; c) daily proportion of animals from at risk (1) or control (0) mating for haplotype M6a that died (D), were slaughtered (S) or are alive (A); c) Results of PCA and HC for 34 haplotypes x 12 parameters. d-f) daily differences between D,S and E categories for mating types 0 and 1 for haplotypes M6a, N17 and M19, respectively.

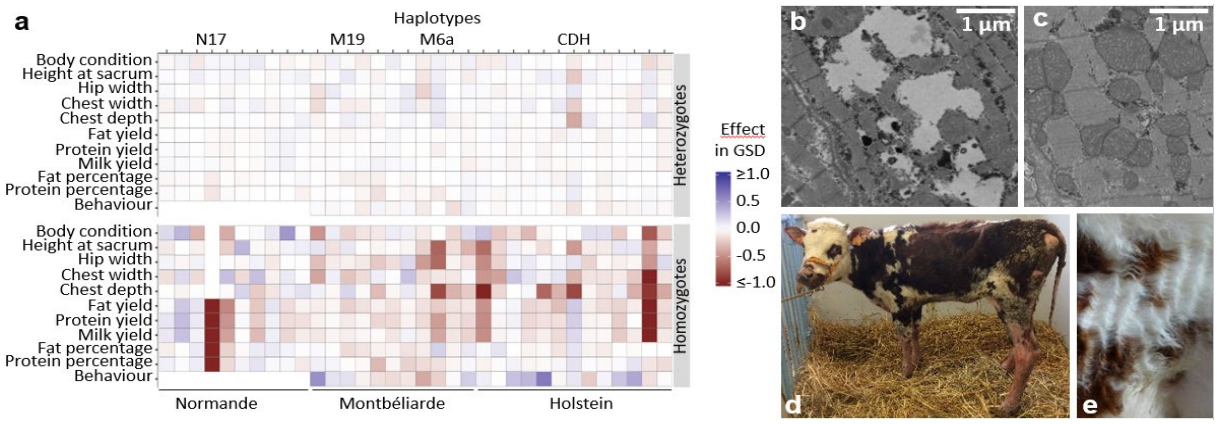

Figure 2. Effects of the 34 haplotypes on 11 traits (a) and results of clinical examinations (b-g). Electron microscopy of myocardium from NOA1-/- (b) and control (c) animals; General view (d) and detail of the fur of a RFC5-/- animal (e).

For 11 traits routinely collected for evaluation purpose, high and negative effects were observed for the hmz carrier genotype while the heterozygous one showed low or no effect as compared

with the noncarrier genotype (figure 2a). This suggest that these deleterious haplotypes owe their presence in the populations mostly to genetic drift rather than hitchhiking or balancing selection.

From whole genome sequences of 611 animals, we retained 22 promising candidates for 15 haplotypes, with predicted consequences including recessive metabolic, immune and neurological defects. Among them, candidate variants for haplotypes M6a in Montbéliarde and N17 in Normande affected genes with no previous record of live homozygous mutants in mammals, and were subject to phenotypical characterization. M6a candidate was a frameshift mutation of NOA1 (p.D400Rfs9) which is involved in the regulation of mitochondrial protein translation and respiration. Based on information of at risk mating data 50% of NOA1-/- individuals died before 41 days of life. The calves we examined were cachetic, had diarrhea and were unable to stand. Biochemical analyses (glycemia > 1g/L, creatine kinase > 200 UI/L and K < 2,5mmol/L of serum) suggested that the organism responded to metabolic demand by lytic activity of muscles, or even autophagy. After euthanasia, electron miscrocopy of myocardium revealed a nearly complete mitochondrial apoptosis and increased storage of glycogen (figure 2b,c). Finally, N17 candidate was a disruptive inframe deletion of the RFC5 subunit (p.E369del) of a protein complex involved in DNA replication. RFC5-/- animals showed retarded growth, thin and curly hair, and alopecia of body extremities (figure 2d,e). They died after recurrent episodes of diarrhea between six months and three years of age.

Conclusion

In conclusion, we propose an innovative strategy to detect deleterious recessive loci that remained overlooked by previous approaches. Taken as a whole, the dozen of loci validated within each breed is responsible for the death of 0.5 to 1% of the dairy females born each year in France. The management of all these new defects in selection will contribute to improve both animal welfare and the income of breeders.

Acknowledgments

This work was funded by APIS-GENE (grant Effitness). We thank A. Remot, M. Vilotte, S. Barbey (INRAE), T. Buronfosse, M. Bouchier, and E. Contat (VetAgroSup) for their help.

References

Boussaha M., Michot P., Letaief R., Hozé C., Fritz S., et al. (2016) Genet Sel Evol. 48(1):87. https://doi.org/10.1186/s12711-016-0268-z

Menzi F., Besuchet-Schmutz N., Fragnière M., Hofstetter S., Jagannathan V., et al. (2016) Anim Genet. 47(2):253-7. https://doi.org/10.1111/age.12410.

Mesbah-Uddin M., Hoze C., Michot P., Barbat A., Lefebvre R., et al. (2019) J Dairy Sci. 102(7):6340–6356. https://doi.org/10.3168/jds.2018-16100

Michot P., Chahory S., Marete A., Grohs C., Dagios D., et al. (2016) Genet Sel Evol 48, 56. https://doi.org/10.1186/s12711-016-0232-y

Misztal, I., Tsuruta S., Lourenco D.A.L., Aguilar I., Legarra A., et al. (2014). http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf

Reynolds E.G.M., Neeley C., Lopdell T.J., Keehan M., Dittmer K. et al (2021) Nat Genet 53(7), 949-+. https://doi.org/10.1038/s41588-021-00872-5

VanRaden P.M., Olson K.M., Null D.J., Hutchison J.L. (2011) J Dairy Sci 94: 6153–6161. https://doi.org/10.3168/jds.2011-4624