GENETIC ANALYSIS OF NOVEL TRAITS IN HOLSTEIN DAIRY CATTLE

by

Larissa Carolina Novo

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Animal and Dairy Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 08/14/2023

The dissertation is approved by the following members of the Final Examination Committee:

Francisco Peñagaricano, Assistant Professor, Animal and Dairy Sciences
Kent A. Weigel, Professor, Animal and Dairy Sciences
Guilherme J. M. Rosa, Professor, Animal and Dairy Sciences
Brian W. Kirkpatrick, Professor, Animal and Dairy Sciences
Brian Parks, Assistant Professor, Nutritional Sciences

DEDICATION

This dissertation is dedicated to my mother, my aunt and uncle, and my husband. Always by my side, they were my rock that sustained me and the wave that carried me further, beyond the storms. My dream became theirs as has this achievement. For you, and to you.

ACKNOWLEDGMENTS

I would like to acknowledge my lab group and their uncountable scripts shared with me and thoughtful suggestions. Our coffee breaks on the fourth floor, conference travels and barbecues are certainly a large part of this amazing experience. You made it easier every day.

I would also like to acknowledge the valuable contributions of my committee members to this dissertation. It was a pleasure to share my research with you and to receive input from great researchers that are references in human and animal genetics.

I would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), along with Fulbright Brazil for believing in me and for providing the funding necessary to obtain a PhD in a renowned institution like the University of Wisconsin -Madison.

Finally, I would like to express my immense gratitude to my advisor, Dr Francisco Peñagaricano. You did more than instruct. You were a mentor, a friend, and a model. You provided the tools and guided me through all the unknowns that starting a PhD in a new country and in a new research area could have. Thank you for your patience and for pushing me to become a better student, a better researcher, and a better person.

ABSTRACT

The availability of new technologies, as well as a greater attention to the needs of consumers and society, has promoted the development of selection strategies for novel traits in dairy cattle. However, several steps must be followed from the proposal until the implementation of novel traits into breeding programs. This process requires efforts from the definition of the new trait, exploration of the genetic basis of the trait, the development of proper analysis methods, and finally the implementation of a genetic evaluation. This thesis addresses different aspects of the development and implementation of novel traits, including definition of the phenotype, better understanding of the biological processes affecting the trait, identification of putative causal variants and genes, and finally a feasibility study of a national genetic evaluation. Specifically, this thesis evaluated different body temperature phenotypes as potential indicators of feed efficiency traits, investigated visceral fat deposition and its relationship with metabolic disorders, uncovered putative causal variants for periparturient hypocalcemia, and explored the implementation of a national genetic evaluation for Johne's disease. The first study showed that body temperature can be used as an indicator of feed efficiency and energy-related traits. The second study revealed that visceral fat accumulation is associated with inflammation, insulin resistance, and immune response, and is genetically linked to displaced abomasum. The third study identified genetic variants and individual genes affecting postpartum blood calcium concentration. The last study showed that a national genetic evaluation for Johne's Disease in Holsteins is feasible. Overall, this work has performed a comprehensive investigation of the different steps required for the development and implementation of novel traits in dairy cattle.

TABLE OF CONTENTS

CHAPTER ONE: INTRODUCTION

The selection of dairy cattle has evolved from empirical and visual selection to the use of pedigree and more recently, genomic information. The productivity of dairy cows has increased dramatically in the last six decades. Animal welfare, animal health, feed efficiency, and environmental traits are receiving more and more attention and are expected to be under the spotlight of research and governmental policies for many years to come (Chesnais et al., 2016).

The exploration of novel traits starts with the need for a specific phenotype that has not been explored yet, or the adoption of a new technology for existing traits. From the identification to the implementation of the novel trait, several steps are involved, including exploratory phenotypic studies followed by validation of the results, genetic and genomic analysis of the new trait, establishment of reference populations, and finally the implementation of national genetic evaluations (Brito et al., 2020b). Selection for feed efficiency of dairy cows is an example of a successful implementation of a novel trait in the U.S. dairy industry. Both residual feed intake and feed saved were implemented in December 2020 by the Council on Dairy Cattle Breeding (VanRaden et al., 2021).

The emphasis on health traits changed with the advent of genomics, given that genomics specially benefits traits that are difficult and expensive to measure, as well as traits with low heritability (Martin et al., 2018). The large-scale genotyping of dairy cattle in the United States and more accessible research tools, such as next generation sequencing, have promoted further biological investigations of well-known health traits such as metritis and ketosis.

On the other hand, new computational technologies, image analysis, machine-learning algorithms, and the development of precision-farming devices have created a massive demand

for the study of recently proposed phenotypes (Bell and Tzimiropoulos, 2018). New monitoring systems have been guiding the development of new traits, such as automated body condition score, feeding behavior, activity, and enteric methane emissions, among others, that may be combined with genomic information to select cattle (Egger-Danner et al., 2014).

The aim of this work was to investigate the development and implementation of novel traits, from the development of new phenotypes and unravel the genetic and biological basis to the adoption of a national genetic evaluation. Chapter three evaluates three body temperature traits as indirect measures for feed efficiency traits, chapter four explores the biological basis of visceral fat in dairy cows using genomic information and both additive and non-additive models, chapter five investigates causal variants and individual genes for periparturient hypocalcemia, a metabolic disorder that affects dairy cows around parturition, and finally chapter six evaluates the feasibility of a national genetic evaluation for Johne's Disease for U.S. Holstein dairy cattle.

CHAPTER TWO: LITERATURE REVIEW

Novel traits are added to the genetic selection standards every few years, according to the industry needs. The criteria for the inclusion of novel traits into selection indexes depend on different factors including the ability of collection of the new phenotype, quality of the data collected, relationship with other important traits, heritability, and other aspects.

With the development of new technologies, the possibilities for the collection of information once unknown are considered and a new phenotype is added to the selection processes. The assessment of the quality of a phenotype itself, however, is not a guarantee of a good candidate for selection and a comprehensive exploratory approach is needed. The decision of a trait starts with the design to the data to be collected according to the information needed and technology available, for example, behavior information based on devices capable to read the movement of the head.

Definition of phenotype

Studies on economic traits in dairy cattle cover various subjects and the research has led to increased specificity and detailed research, yet a considerable part of biological responses underlying important traits are not fully understood. More technological alternatives are available, but the development of novel traits passes through intricate study design to determine the data to be collected. The hypothesis-based research process, in most cases, depend on descriptive studies for formulation that represent an essential part of selection for novel traits.

The decision of the data to be collected must meet not only the scientific standards of reproducibility by other research groups for example, but also meet realistic expectations for the

future. Traits such as methane emission for instance, have been limited for a long time to research settings, even though the studies have been developed as early as the decade of 1940 (Moe and Tyrrell, 1979). The necessity of respiration chambers for the data collection largely prevented the expansion of the research being developed for the trait, but easier-to-measure methods such as milk spectra or handheld devices are expected to be the key factor for the expansion for large scale data (Bell and Tzimiropoulos, 2018).

The amount of information needed for an accurate representation of the trait is also taken into consideration as it can impact considerably its use in research given limited funding sources. For instance, several traits follow a circadian pattern that can bias the data if measured daily but a data collection as frequent as hourly is a labor-intensive task that most of commercial farms cannot fulfill. Cases such as behavioral traits and body temperature are examples of monitoring demanding phenotypes that can be assisted by wearable devices that can automatically collect information throughout the day (Bell and Tzimiropoulos, 2018).

Descriptive studies also allow for determination of the adequate number of animals once the quality of measurements has been assessed. Traits with known low heritability or diseases with low incidence (e.g., lameness) demand greater study population to provide enough information for prediction (Mottram, 2016). For rectal temperature of Holstein cows for example, with heritability of 0.17, approximately 20,000 would be needed to achieve a genomic reliability of 0.4 (Cheruiyot et al., 2022). In such cases, descriptive studies describing the distribution, behavior and influential factors on the data were central to the development of algorithms and new devices for data collection. Once a good indicator is established, validation studies are performed to verify the behavior of the methods used on other populations.

Reference populations

Most of the phenotyping methods currently being developed under precision farming perspective rely on technological gadgets and infrastructure not available to many commercial farms, limiting the development of the novel trait. As alternative, effort has been placed by research stations to collect the data and information of a specific and reduced population to be used as reference for the formulation of predictive models for target populations (Egger-Danner et al., 2014). Examples of reference populations undergoing expansion are for residual feed intake, enteric methane emission, behavior, temperament and milkability, and heat tolerance.

Genomic studies

In a genetic selection standpoint, the relationship of the phenotype with other economically important traits is another essential part of novel trait investigation. While a new phenotype might have been stablished as a management tool in commercial farms for example, its use on selection is not advisable while the possible long-term effects have not been determined. A novel trait in this case may be limited to herd supervision instead of included in a selection index. A very high genetic correlation with a traditional trait can be used to rule out its inclusion or decrease its weight on a genetic evaluation as it would bring redundancy to the models.

The biology underlying the trait is essential for the understanding of the phenotype as a management tool, and the perspectives for future genetic selection. One of the main tools for basic biological investigation are genome-wide association studies (GWAS) for the scanning and identification of the candidate genes affecting the phenotype. With more than 6.5 million

genotyped dairy cattle reported in 2022, studies using genomic information are becoming more frequent (Wiggans and Carrillo, 2022).

GWAS can not only indicate the candidate genes for a specific phenotype, but their use can be extended to pathway analysis and link the trait under study to important traits not yet well understood or with difficult phenotyping such as metabolic diseases. Significant SNPs and genomic regions identified through GWAS may be included to SNP panels, but its application is population dependent. For instance, Luo et al. (2021) used GWAS to identify 5 candidate genes for heat stress measured as rectal temperature in dairy cows and after biological validation indicated its inclusion in SNP panels or prediction models. Although genomic analysis had been performed to investigate heat stress previously, the replication of these in a population of Chinese Holstein cows identified different candidate genes.

Genomic prediction of novel traits

For a novel trait, the accuracy of phenotype prediction may be lower due to insufficient phenotyped population (Chesnais et al., 2016). The incorporation of biological information, specifically candidate causal variants, can improve accuracies of genomic prediction based on low-density SNP panel (Cheruiyot et al., 2022). With respect to prediction methods, the genomic best linear unbiased prediction (GBLUP) assumes that all variants have small effects with the same distribution, which is unlikely for traits that are influenced by a major SNP, while methods such as weighted GBLUP assign different weights for SNPs, according to prior information (Wang et al., 2012). The inclusion of prior biological information in prediction models is particularly beneficial to small populations such as the reference populations for novel traits but can be irrelevant with larger populations or highly polygenic traits. In a scenario with 2,000

animals, an effective population size of 20 and 10 quantitative trait loci (QTL), Lourenco et al (2017) reported an increase of 0.15 in the accuracy of prediction when SNP weight were added but the improvement was almost nil when the population increased to 25,000 or with 500 QTL.

On the other hand, Van Binsbergen et al (2015) reported that biological information is decisive to the positive outcomes of including whole-genome sequence information in comparison to high-density SNP panels. Furthermore, by focusing on markers closer to causal variants or the variants themselves, it is possible to aggregate data from distantly related breeds to balance the small size of the studied population and increase the reliability of predictions (Lund et al., 2016). The emphasis on causal variants can also help predictions across populations as it bypasses the different linkage disequilibrium (LD) between the marker and actual variant seen in diverse populations, as long as the density of markers is sufficient to provide a consistent LD (Lund et al., 2016).

Another tool available to increase accuracy of prediction in smaller reference populations is the semi-supervised learning that combine methods that predict future phenotypes based on a prior and known one, and unsupervised models that does not rely on previous phenotypes for model training. Yao et al. (2016) reported that semi-supervised learning was able to improve the accuracy of genomic prediction and achieve values comparable to larger and fully-phenotyped populations.

Implementation of national genetic evaluations

Ideally, the outcome for the development of novel measurement tools is not only to be applied on management decisions, but also on genetic selection. Once the impact of the trait on other important selection criteria is determined and a desirable reliability is achieved the

feasibility of a national genetic evaluation can be assessed. In recent years, RFI and clinical mastitis have been incorporated to the net merit (NM$) index used by the Council on Dairy Cattle Breeding to rank dairy animals based on their genetic merit (VanRaden et al., 2021). Their proposal as selection trait, however, started several decades ago and has only advanced with the establishment of reference populations, indicator traits and automation of data collection. RFI for instance was proposed in 1963 by Koch et al, but it was not until the introduction of genomics that its use was truly considered in breeding programs due to limited phenotypic records, unstandardized measurements, lack of breeding goal and biological information that could help to increase accuracies (Brito et al., 2020c). The inclusion of both traits are examples of long, yet successful, implementation of novel traits.

The efforts in enteric methane emission tend to focus on phenotyping and the use of genomic selection as mitigation tool (Lassen and Difford, 2020). Biology investigation through GWAS on genes and variants, establishment of reference populations, association of microbiome, and definition and standardization of phenotypes are currently undergoing research and may turn methane into a candidate for inclusion in selection indexes in the near future (Garnsworthy et al., 2019; Pszczola et al., 2019). Fortunately, the correlations between different methods for methane emission determinations suggest that data can be combined to provide larger reference population for genetic selection which can accelerate its inclusion in breeding programs (Garnsworthy et al., 2019).

Dairy cattle breeding programs rely on predicted transmitting abilities (PTA) to provide the ranking of animals based on the selection index. The decision of inclusion or not of a novel trait into the program can consider factors such as impact on overall selection (e.g., reranking of

animals), the number of phenotyped individuals, reliability of PTA predictions, heritability and repeatability of the trait, and economic impact (Brito et al., 2020b; VanRaden et al., 2021).

Ongoing research on novel traits

There are several novel traits being currently considered for research in different fronts to meet industry's and government's requirements. Although some are still under the development of a phenotype, other are closer to implementation in breeding programs (Egger-Danner et al., 2014). Functional traits are an important point of focus of funding and research efforts. Phenotypes related to udder health, for example, aim to reduce the incidence of mastitis and other infectious diseases affecting mammary gland function (Egger-Danner et al., 2014).

Clinical mastitis has been included along with other direct health traits in the genetic evaluation of dairy cows in the United States, but the standardization of data recorded remains as one limiting factor to unifying data from multiple sources across US (Parker Gaddis et al., 2020). Martin et al. (2018) described indirect measures of mastitis such as somatic cell count (SCC) and its variants such as average SCC and standard deviation of SCC, and udder depth and indicated them as strong candidates for inclusion in evaluations. However, the authors also highlighted the opportunity for research on other proxy traits and a solid reference population for an accurate selection. Currently even though health traits are part of US national genetic evaluation, efforts are still being placed on the biological understanding of the disease for a prevention point, but the results can also contribute to better prediction in the future (McConnel et al., 2020; Bisutti et al., 2023).

Workability traits combine phenotypes such as milking speed, average flow rate, milking frequency, handling time and others and aim to describe the facility of handling and milking.

Workability has been included in the Norwegian dairy breeding program since 1970 but is yet to be implemented in the US. Although the difficulties fall mainly on the larger herd sizes, the use of automated milking systems is potentially a favorable direction into phenotyping while a task force assigned by the Council on Dairy Cattle Breeding studies the feasibility of national genetic evaluation for the traits (Miles et al., 2022). Genomic-based heritability for daily milkability traits in North America were estimated to be from 0.02 and 0.56 (values for milking failure and milking efficiency, respectively), and indicated a potential for genetic selection based on genomic information (Pedrosa et al., 2023). Additionally, a study using whole-genome sequence was performed and provided the discovery of eight candidate mutations for milking temperament that could be used as an improved tool for predictions for the trait in future evaluations (Chen et al., 2020a). While maximizing the use of current records for workability traits may assist with estimations, the combining of traditional and automated milking system data should be taken into consideration as they are highly genetically correlated, yet can yield different heritabilities (Wethal et al., 2020).

Sustainability is also currently under the spotlight as consumers and governments push for higher regulations of production systems. The enteric methane emission is a trait under consideration for genetic evaluation considering that installation of data recording devices in commercial farms is unlikely. However, the reference population is not yet sufficient to provide reliable predictions and more data is needed over the next years (de Haas et al., 2021). Different measures of methane emission are being used and combining methods is a possibility to increase the dataset available to established a good reference population (Garnsworthy et al., 2019). The heritabilities of methane emission vary from 0.05 to 0.27, with higher estimates for measurements using sniffers, and the genetic correlation with production and health traits

indicate that selection for reduced methane emission on its own would lead to lower milk production, lower dry matter intake, and smaller animals (Pszczola et al., 2019; Lassen and Difford, 2020). These results point out the need to investigate the genetic selection of methane emission in a selection index context.

Research efforts and large-scale phenotyping groups similar to the ones created for feed efficiency can promote standardized measurements for novel traits by sharing data and generate a reference population with phenotypic and genetic variability. Projects such as METHAGENE (www.methagene.eu), Ruminomics (https://cordis.europa.eu/project/rcn/101163/en), Efficient Dairy Genome Project (https://genomedairy.ualberta.ca), and The Resilient Dairy Genome Project (http://www.resilientdairy.ca/) will lead to phenotyping efforts for novel traits definition and implementation.

The introduction of automated phenotyping based on accelerometer, imaging analysis and computer vision, wheatear data as temperature and humidity and others under the umbrella of precision livestock generates a load of data for research use. However, such information originating from automation and sensor data still faces several challenges. Aside from large size and high-dimensionality of the data, using the so-called "big-data" requires substantial computer processing, well-trained and validated models and run into issues with large diversity, high variability, data consolidation and integration, and data ownership issues (Gengler, 2019). The use of precision livestock farming to provide large-scale phenotyping for genomic selection programs is possible but the variables measured over time depend on complex methods to deal with factors such as missing errors, redundancy of information (Brito et al., 2020a; Rosa, 2021). In this context, current breeding programs would need to change to accommodate automated phenotyping as the current genetic evaluations are continually revised and the weekly updates

cannot support the time needed for data processing and for inclusion, some calculations would need to be processed on-farm to be included in a timely manner (Gengler, 2019).

REFERENCES

Bell, M. J., and Tzimiropoulos, G. (2018). Novel Monitoring Systems to Obtain Dairy Cattle Phenotypes Associated With Sustainable Production. *Front. Sustain. Food Syst.* 2. doi:10.3389/fsufs.2018.00031.

Bisutti, V., Mach, N., Giannuzzi, D., Vanzin, A., Capra, E., Negrini, R., et al. (2023). Transcriptome-wide mapping of milk somatic cells upon subclinical mastitis infection in dairy cattle. *J. Anim. Sci. Biotechnol.* 14, 93. doi:10.1186/s40104-023-00890-9.

Brito, L. F., Oliveira, H. R., McConn, B. R., Schinckel, A. P., Arrazola, A., Marchant-Forde, J. N., et al. (2020a). Large-Scale Phenotyping of Livestock Welfare in Commercial Production Systems: A New Frontier in Animal Breeding. *Front. Genet.* 11, 1–32. doi:10.3389/fgene.2020.00793.

Brito, L. F., Oliveira, H. R., Silva, F. F., and Schenkel, F. S. (2020b). "Developments in genomic predictions in dairy cattle breeding: a historical overview of methods, technologies, and applications," in *Advances in Animal Breeding*, 357–382. doi:10.19103/AS.2019.0058.18.

Brito, L., Rojas de Oliveira, H., Houlahan, K., Fonseca, P. A. S., Lam, S., Butty, A. M., et al. (2020c). *INVITED REVIEW: Genetic mechanisms underlying feed utilization and implementation of genomic selection for improved feed efficiency in dairy cattle*. doi:10.1139/cjas-2019-0193.

Chen, S. Y., Oliveira, H. R., Schenkel, F. S., Pedrosa, V. B., Melka, M. G., and Brito, L. F. (2020). Using imputed whole-genome sequence variants to uncover candidate mutations and genes affecting milking speed and temperament in Holstein cattle. *J. Dairy Sci.* 103, 10383–10398. doi:10.3168/jds.2020-18897.

Cheruiyot, E. K., Haile-Mariam, M., Cocks, B. G., and Pryce, J. E. (2022). Improving Genomic Selection for Heat Tolerance in Dairy Cattle: Current Opportunities and Future Directions. *Front. Genet.* 13, 1–18. doi:10.3389/fgene.2022.894067.

Chesnais, J. P., Cooper, T. A., Wiggans, G. R., Sargolzaei, M., Pryce, J. E., and Miglior, F. (2016). Using genomics to enhance selection of novel traits in North American dairy cattle. *J. Dairy Sci.* 99, 2413–2427. doi:10.3168/jds.2015-9970.

de Haas, Y., Veerkamp, R. F., de Jong, G., and Aldridge, M. N. (2021). Selective breeding as a mitigation tool for methane emissions from dairy cattle. *Animal* 15, 100294. doi:10.1016/j.animal.2021.100294.

Egger-Danner, C., Cole, J. B., Pryce, J. E., Gengler, N., Heringstad, B., Bradley, A., et al. (2014). Invited review: Overview of new traits and phenotyping strategies in dairy cattle with a focus on functional traits. *Animal* 9, 191–207. doi:10.1017/S1751731114002614.

Garnsworthy, P. C., Difford, G. F., Bell, M. J., Bayat, A. R., Huhtanen, P., Kuhla, B., et al. (2019). Comparison of methods to measure methane for use in genetic evaluation of dairy cattle. *Animals* 9, 1–12. doi:10.3390/ani9100837.

Gengler, N. (2019). Symposium review: Challenges and opportunities for evaluating and using the genetic potential of dairy cattle in the new era of sensor data from automation. *J. Dairy Sci.* 102, 5756–5763. doi:10.3168/jds.2018-15711.

Lassen, J., and Difford, G. F. (2020). Review: Genetic and genomic selection as a methane mitigation strategy in dairy cattle. *Animal* 14, s473–s483. doi:10.1017/S1751731120001561.

Lourenco, D. A. L., Fragomeni, B. O., Bradford, H. L., Menezes, I. R., Ferraz, J. B. S., Aguilar, I., et al. (2017). Implications of SNP weighting on single-step genomic predictions for

different reference population sizes. *J. Anim. Breed. Genet.* 134, 463–471. doi:10.1111/jbg.12288.

Lund, M. S., Van Den Berg, I., Ma, P., Brøndum, R. F., and Su, G. (2016). Review: How to improve genomic predictions in small dairy cattle populations. *Animal* 10, 1042–1049. doi:10.1017/S1751731115003031.

Luo, H., Li, X., Hu, L., Xu, W., Chu, Q., Liu, A., et al. (2021). Genomic analyses and biological validation of candidate genes for rectal temperature as an indicator of heat stress in Holstein cattle. *J. Dairy Sci.* 104, 4441–4451. doi:10.3168/jds.2020-18725.

Martin, P., Barkema, H. W., Brito, L. F., Narayana, S. G., and Miglior, F. (2018). Symposium review: Novel strategies to genetically improve mastitis resistance in dairy cattle. *J. Dairy Sci.* 101, 2724–2736. doi:10.3168/jds.2017-13554.

McConnel, C. S., Crisp, S. A., Biggs, T. D., Ficklin, S. P., Parrish, L. M., Trombetta, S. C., et al. (2020). A Fixed Cohort Field Study of Gene Expression in Circulating Leukocytes From Dairy Cows With and Without Mastitis. *Front. Vet. Sci.* 7, 1–11. doi:10.3389/fvets.2020.559279.

Miles, A., Fourdraine, R., Sievert, S., Gaddis, K., Bewley, J., Eaglen, S., et al. (2022). Considerations in using quantitative measurements of milking speed for genetic evaluations for all dairy breeds in the USA.

Moe, P. W., and Tyrrell, H. F. (1979). Methane Production in Dairy Cows. *J. Dairy Sci.* 62, 1583–1586. doi:10.3168/jds.S0022-0302(79)83465-7.

Mottram, T. (2016). Animal board invited review: Precision livestock farming for dairy cows with a focus on oestrus detection. *Animal* 10, 1575–1584. doi:10.1017/S1751731115002517.

Parker Gaddis, K. L., VanRaden, P. M., Cole, J. B., Norman, H. D., Nicolazzi, E., and Dürr, J. W. (2020). Symposium review: Development, implementation, and perspectives of health evaluations in the United States. *J. Dairy Sci.* 103, 5354–5365. doi:10.3168/jds.2019-17687.

Pedrosa, V. B., Boerman, J. P., Gloria, L. S., Chen, S. Y., Montes, M. E., Doucette, J. S., et al. (2023). Genomic-based genetic parameters for milkability traits derived from automatic milking systems in North American Holstein cattle. *J. Dairy Sci.* 106, 2613–2629. doi:10.3168/jds.2022-22515.

Pszczola, M., Calus, M. P. L., and Strabel, T. (2019). Short communication: Genetic correlations between methane and milk production, conformation, and functional traits. *J. Dairy Sci.* 102, 5342–5346. doi:10.3168/jds.2018-16066.

Rosa, G. J. M. (2021). Grand Challenge in Precision Livestock Farming. *Front. Anim. Sci.* 2, 1–3. doi:10.3389/fanim.2021.650324.

Van Binsbergen, R., Calus, M. P. L., Bink, M. C. A. M., Van Eeuwijk, F. A., Schrooten, C., and Veerkamp, R. F. (2015). Genomic prediction using imputed whole-genome sequence data in Holstein Friesian cattle. *Genet. Sel. Evol.* 47, 1–13. doi:10.1186/s12711-015-0149-x.

VanRaden, P. M., Cole, J. B., and Parker Gaddis, K. L. (2021). Net merit as a measure of lifetime profit: 2021 revision. Available at: https://www.ars.usda.gov/research/publications/publication/?seqNo115=353707.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb).* 94, 73–83. doi:10.1017/S0016672312000274.

Wethal, K. B., Svendsen, M., and Heringstad, B. (2020). Are farmer assessed temperament, milking speed, and leakage genetically the same traits in automatic milking systems and

traditional milking systems? *J. Dairy Sci.* 103, 3325–3333. doi:10.3168/jds.2019-17503.

Wiggans, G. R., and Carrillo, J. A. (2022). Genomic selection in United States dairy cattle. *Front. Genet.* 13, 1–7. doi:10.3389/fgene.2022.994466.

Yao, C., Zhu, X., and Weigel, K. A. (2016). Semi-supervised learning for genomic prediction of novel traits with small reference populations: an application to residual feed intake in dairy cattle. *Genet. Sel. Evol.* 48, 1–9. doi:10.1186/s12711-016-0262-5.

CHAPTER THREE: ASSOCIATIONS BETWEEN BODY TEMPERATURE AND FEED EFFICIENCY TRAITS IN HOLSTEIN DAIRY COWS

Larissa C. Novo[1], Ligia Cavani[1], Heather M. White[1], Kent Weigel[1], and Francisco Peñagaricano[1]

[1]Department of Animal and Dairy Sciences, University of Wisconsin-Madison, Madison WI, 53706

Abstract

Heat load and body temperature can impact the performance of dairy cows and are influenced by several external and internal (individual) factors. The aim of this study was to investigate the associations between three body temperature traits, namely average body temperature, consistency of body temperature, and change in body temperature after the largest meal of the day, with four feed efficiency traits, namely dry matter intake **(DMI)**, metabolic body weight **(MBW),** milk energy **(MilkE)**, and residual feed intake **(RFI)** in lactating Holstein cows. Data were collected on 304 mid-lactation Holstein cows enrolled in 11 different feed efficiency trials from 2020 to 2023 at the University of Wisconsin-Madison. Temperature records were obtained using an automatic temperature logger placed vaginally during 2 weeks. Body temperature was recorded every five minutes with a resolution of 0.0625 ºC. Average body temperature was calculated as the individual mean during the trial, consistency of body temperature was calculated as the log transformed variance of the deviations of individual records from the cow's mean, and change in body temperature was calculated as the difference in the temperature before and after the largest meal of the day. Data for DMI, MBW, MilkE and RFI were collected from the same cows in the same feeding trials for 6-7 weeks. The

associations between body temperature traits and feed efficiency traits were assessed using linear regression models including body temperature and cohort (trial-treatment) as independent variables. Average body temperature was positively associated with RFI (slope = 2.14; $p < 0.05$) indicating that cows with higher body temperature are less feed efficient. Dairy cows with higher MilkE showed less consistent body temperature (slope = 1.42; $p < 0.05$) and had greater decrease of body temperature after the largest meal of the day (slope = -14.61; $p < 0.001$). Larger changes in body temperature after meal were also associated with greater DMI (slope = -10.26; $p < 0.001$). Overall, our results suggest that body temperature can be used as an indicator of feed efficiency in lactating dairy cows and that more variations in body temperature are associated with higher MilkE and DMI.

Introduction

Thermoregulation is an important aspect of the metabolism of a dairy cow and interferes with the prioritization of blood flow, metabolites requirements, and energy partitioning (Lees et al., 2019). In a challenging environment, dairy cows try to balance feed and water intake to reduce the heat increment of digestion, spend more time standing for better heat dissipation, increase the passage rate of digesta, and prefer carbohydrate over lipid utilization (Wheelock et al., 2010; Allen et al., 2015).

The change in energy partitioning promoted by heat load directly impacts milk yield, fat and protein content and exacerbates the importance of thermoregulation at expense of production (Aharoni et al., 2005). Variation in rumen and body temperature are not only associated with less efficiency in energy utilization but also with lower feed efficiency (DiGiacomo et al., 2014; Fischer et al., 2018).

Although the relationship between feed efficiency and body temperature is an active research topic, data are usually limited to a small number of animals, with one or a few records per animal, collected during one or a few days. Rectal temperature records depend on manual collection, restraining and moving the animal, and can interfere with an individual's temperature. Vaginal measurements provide similar results in comparison to rectal temperature and allow the use of data loggers for longer periods (Vickers et al., 2010). Note that the automatic measurement of body temperature provides more frequent records and consequently, more accurate daily average (Tresoldi et al., 2020). Despite its benefits, the relationship between high-frequency body temperature records with feed efficiency in lactating cows is not well stablished in the literature.

The objective of this study was to investigate the associations between body temperature measured as average body temperature, consistency of body temperature, and change in body temperature after meal, with dry matter intake, metabolic body weight, milk energy, and residual feed intake in mid-lactation Holstein cows. We hypothesized that (1) body temperature is associated with feed efficiency, (2) more consistent cows are more efficient, and (3) the variation in body temperature after meal is related with feed efficiency.

Material and methods

Animals

Animal handling and sampling procedures were approved by the University of Wisconsin-Madison College of Agricultural and Life Sciences Animal Care and Use Committee. Data were collected from mid-lactation Holstein cows (n = 304) in 11 feeding trials performed between 2020 and 2023 at the University of Wisconsin-Madison, as part of the national dairy cow feed efficiency project (Tempelman et al., 2015; Li et al., 2019). Cows were housed in a sand-bedded freestall facility at the Emmons Blaine Dairy Cattle Research Center (Arlington, WI) or at Marshfield Agricultural Research Station (Stratford, WI). Cows were provided ad libitum access to feed and water, with diets formulated to meet or exceed nutrient requirements.

Internal temperature records

Cows enrolled in feed efficiency trials were selected to monitor body temperature based on their reproductive status. Cows in early stages of pregnancy or cows enroll in insemination protocols were not included for temperature collection. Body temperature was collected with a Thermochron iButtons device (Embedded Data Systems, Lawrenceburg, KY), placed in an intravaginal device (CIDR, Zoetis US, New York, NY) lacking progesterone supplement. To avoid malfunction and record failure, the CIDR with the device was assembled following an internal protocol consisting of filling the gaps between device and CIDR and covering it with shrinking tube and electrical tape to ensure it was securely waterproof. Each eligible cow received one device and had temperature recorded every five minutes for 14 days, with a precision of 0.0625 °C. Animals with signs of severe discomfort or lesions were removed from the study. Cows with less than nine days of records were excluded. In addition, individual

temperature records outside three standard deviations from the cow's mean were not considered. The final dataset consisted of 304 primiparous and multiparous lactating cows with a total of 1,183,204 body temperature records, with an average of 6,899 records per cow.

Feed efficiency traits

Daily feed intakes records were measured for 6-7 weeks via a roughage intake control system (Hokofarm Group) or Calan Broadbent Feeding System (American Calan). Milk weights were obtained daily, milk samples were obtained 4 times a week for determination of milk composition, and BW were obtained on 3 consecutive days at the beginning, middle, and end of the experimental period. Milk energy was calculated weekly using the following equation:

$$MilkE = (0.0929 \times fat\,\% + 0.0563 \times protein\,\% + 0.0395 \times lactose\,\%) \times milk\ yield$$

The RFI for each cow was calculated according to the following equation:

$$DMI = DIM + partity + \beta_1 MilkE + \beta_2 MBW + \beta_3 \Delta BW + cohort + week + \varepsilon,$$

where DIM represents the effect of days in milk with 9 levels (grouped by 16 d), parity represents the effect of parity (lactation number) with 3 levels (1, 2, and 3), MilkE is secreted milk energy with partial regression coefficient β_1, MBW is metabolic body weight with partial regression coefficient β_2, ΔBW is the change in body weight with partial regression coefficient β_3, $cohort$ represents the random effect of trial-treatment, $week$ represents the random effect of week of experiment, and ε is the random residual of the model, representing RFI.

Body temperature traits

Three body temperature phenotypes were considered in this study: average body temperature, consistency of body temperature, and change in body temperature after the largest meal of the day. All these phenotypes were averaged for the trial period. Consistency of body

temperature was calculated as the log transformed variance of the deviations of individual records from the mean. Change in body temperature after the largest meal was calculated as the difference in body temperature between 15 minutes before the meal and 30 minutes after the end of the meal. The largest meal of the day consisted of the meal with the most feed consumed based on the work of Cavani et al., (2022).

Statistical Analysis

The associations between body temperature traits and feed efficiency traits were assessed using linear regression models containing the temperature phenotype and the cohort (trial-treatment) as independent variables. Pearson's correlations between body temperature traits and feed efficiency traits were also calculated.

Results

Table 1 shows the descriptive statistics for the three body temperature traits. As expected, no considerable differences in the mean and standard deviations of body temperature were observed across trials and seasons. Notably, an important between-animal variation was observed for the three body temperature phenotypes. The change in body temperature after meal was consistently negative, showing that during the period following food ingestion, the body temperature decreased on average -0.24 ºC ± 0.09. Interestingly, the high-frequency body temperature collected made it possible to observe daily temperature peak patterns between 1400 and 2000, and lowest temperatures between 0300 and 0500, across trials. Table 2 shows the Pearson's correlations calculated between body temperature traits and feed efficiency traits.

Figure 1 shows the associations between average body temperature and RFI, DMI, MBW, and MilkE. Interesting, body temperature was positively associated with RFI (P-value < 0.05) indicating that animals with higher body temperature also have higher RFI, and hence, are less feed efficient. The other traits evaluated, DMI, MBW and MilkE, did not show significant associations with average body temperature.

Figure 2 shows the associations between consistency of body temperature and RFI, DMI, MBW, and MilkE. Note that consistency of body temperature is presented as a log transformed variance, and thus the least variable cows, a.k.a. the most consistent cows, present the most negative values. Only MilkE showed a significant association with consistency of body temperature with animals with high variability in body temperature (values close to zero) tend to allocate more energy towards milk production. Interestingly, this association cannot be attributed

to higher DMI as DMI was not associated with consistency of body temperature (P-value > 0.05).

Figure 3 shows the associations between change in body temperature after meal and RFI, MBW, MilkE, and DMI. Most cows decreased their body temperature after the largest meal of the day. Larger variations in body temperature were positively associated with milk energy and also dry matter intake (P-value < 0.05). No significant associations were observed between change in body temperature and RFI nor MBW.

Discussion

Body temperature is a known factor involved in the efficiency of energy utilized by animals (Lees et al., 2019). Heat load and the capacity of an individual to adapt to environmental challenges interfere in organism's priorities and shift energy normally allocated to production. Adaptative processes in dairy cows may include decreased feed intake, lower feeding time, lower milk yield and quality, behavioral changes, among others (Tao et al., 2020). Negative relationships between body temperature and feed efficiency traits have been described for beef cattle (Martello et al., 2016), dairy calves (Leão et al., 2018), grower pigs (Cook et al., 2020), among others. In a scenario of global warming where thermotolerance should be prioritized, a better understanding of the relationship between body temperature and production efficiency is of paramount importance for the dairy industry.

Different measures of body temperature have been used, such as infrared, skin, rectal and vaginal temperature, with their respective advantages and weaknesses. While easier to obtain, external measurements like skin and infrared temperatures might be a better indicator of heat dissipation than increment *per se* (DiGiacomo et al., 2014). Internal measures, on the other hand, provide more frequent core body temperature but have implementation challenges (Suthar et al., 2013). To the best of our knowledge, this study describes the largest dataset of high-frequency body temperature records collected on lactating dairy cows. The data collected consisted of 20 records per hour for 14 days totaling 1,183,204 temperature records for 304 lactating Holstein cows.

The three phenotypes herein presented aimed to explore the individual variations in body temperature and the possible associations with DMI, MBW, MilkE, and RFI. It was

hypothesized that the ability of animals to maintain their temperature throughout the day and their ability to limit the heat increment after feeding is related to feed efficiency. The fact that RFI was found to be significantly associated only with average body temperature is an indicator that feed efficiency might be a product of overall body temperature and not necessarily its variability. The positive and significant association between RFI and average body temperature suggests that higher body temperature is an indicator of poor efficiency, and hence body temperature could be used as an indicator trait of feed efficiency. Note that feed efficiency is based on feed intake records, and these measurements are typically performed only in research farms.

Previous studies have suggested that high body temperature negatively affects feed efficiency. In fact, under challenging conditions, animals allocate energy towards metabolic adaptations and blood flow from gastrointestinal tract and digestion to peripheral tissues for heat dissipation (Ríus, 2019). Some authors argue that the reduction in milk yield is mainly caused by the decrease in DMI as a protective measure to avoid heat increment of feed, but data indicate that DMI accounts for only 51.4% of the variation in milk loss and that metabolic adaptations might explain the remaining (Wheelock et al., 2010; Tao et al., 2020). Note that cows used in this study were not exposed to heat stress conditions, and hence, the association between body temperature with RFI but not with DMI, MilkE and MBW indicates that individual differences are possibly due to metabolic adaptations not triggered by heat stress conditions.

The second phenotype of interest was the variability of body temperature and it was calculated as the log transformed variance of the deviations from the animal's average temperature for the whole trial period. Only MilkE was significantly associated with consistency of body temperature, indicating that animals capable of maintaining their temperature constant

over the trial have a slightly lower amount of energy allocated to milk production. This agrees with the theory that the energy devoted to thermoregulation is determined at the expense of production, even under a thermoneutral environment. Interestingly, most variable cows had lower consistency not only across days but also within a day.

Different factors can be source of temperature variation during a period of 24h such as diet, feeding behavior, feed delivery and milking time. Fischer et al. (2018) reported that more variable rumen temperatures were associated with fewer bunk visits and larger meal size and that the higher the rumen temperatures, the lower the time spent eating. The same authors also detected that dairy cows with higher residual energy intake have more variable rumen temperatures. The fact that approximately 65.5% of the net energy intake variability among animals is attributed to milk energy allocation and that milk fat and protein can likely explain heat production, the association of MilkE with consistency of temperature seen in this work is expected (Fischer et al., 2018; Morris et al., 2021).

Body temperature may rise from 1 to 3 hours after feeding, following an increase in metabolism and consequently a heat load due to digestion functions and rumination and is considered an additional source of body temperature variation during a day (Purwanto et al., 1990; de Melo Costa et al., 2018). In this study, except during milking time, cows had unrestricted access to feed throughout the day and therefore, a more distributed feeding behavior pattern was expected. To investigate change in body temperature after feeding, the largest meal of the day was used. Opposed to what was expected and to what has been shown in the literature, the body temperature of the animals decreased on average 0.24 ºC 30 minutes after the meal. This result could be explained by the redirection of blood flow from the extremities to the gastrointestinal tract expected under the thermoneutral conditions to which the cows were

exposed. The shift in the blood flow after food ingestion occurs to prioritize digestion and nutrient absorption and to supply metabolites necessary to complete metabolic pathways (Sangsritavong et al., 2002).

As previously reported, although a good indicator of core body temperature, the temperature loggers placed in the vagina might not be able to detect, for example, changes in rumen temperature (AlZahal et al., 2011). Interestingly, larger changes in body temperature were significantly associated with MilkE and DMI indicating that animals with greater temperature decrease after feeding ingested more food and allocated more energy towards milk production. These results agree with the hypothesis that, under thermoneutral conditions, the metabolism would prioritize digestion and nutrients absorption, differing from the response of individuals observed under heat stress that may lead to production decline.

Conclusion

Overall, the results presented in this work suggest that average body temperature is associated with feed efficiency in mid-lactation dairy cows. Our results revealed that cows with higher body temperature tend to be less feed efficient. We also observed that cows with less consistent temperature throughout the day tend to direct more energy towards milk production. Cows with larger changes in body temperature after feeding tend to eat more and produce more milk energy.

Table 1. Descriptive statistics for average body temperature, consistency of body temperature and change of temperature after meal per feeding trial

			Average temperature	Consistency of temperature	Change of temperature
Trial	Cows	Season	Mean ± SD	Mean ± SD	Mean ± SD
1	17	Fall	38.69 ± 0.08	-3.20 ± 0.50	-0.20 ± 0.05
2	20	Summer	38.69 ± 0.11	-2.74 ± 0.34	-0.35 ±0.10
3	24	Summer	38.73 ± 0.09	-2.72 ± 0.43	-0.29 ± 0.09
4	41	Winter	38.74 ± 0.10	-3.26 ± 0.49	-0.19 ± 0.07
5	27	Spring	38.71 ± 0.10	-3.12 ± 0.49	-0.26 ± 0.07
6	29	Summer	38.75 ± 0.12	-2.37 ± 0.64	-0.23 ± 0.09
7	37	Fall	38.73 ± 0.10	-3.03 ± 0.33	-0.25 ± 0.09
8	23	Spring	38.74 ± 0.12	-3.10 ± 0.30	-
9	39	Winter	38.74 ± 0.09	-3.38 ± 0.28	-0.21 ± 0.08
10	21	Winter	38.73 ± 0.12	-3.29 ± 0.32	-0.23 ± 0.05
11	26	Spring	38.75 ± 0.12	-3.13 ± 0.42	-
Total	304	-	38.74 ± 0.11	-2.89 ± 0.59	-0.24 ± 0.09

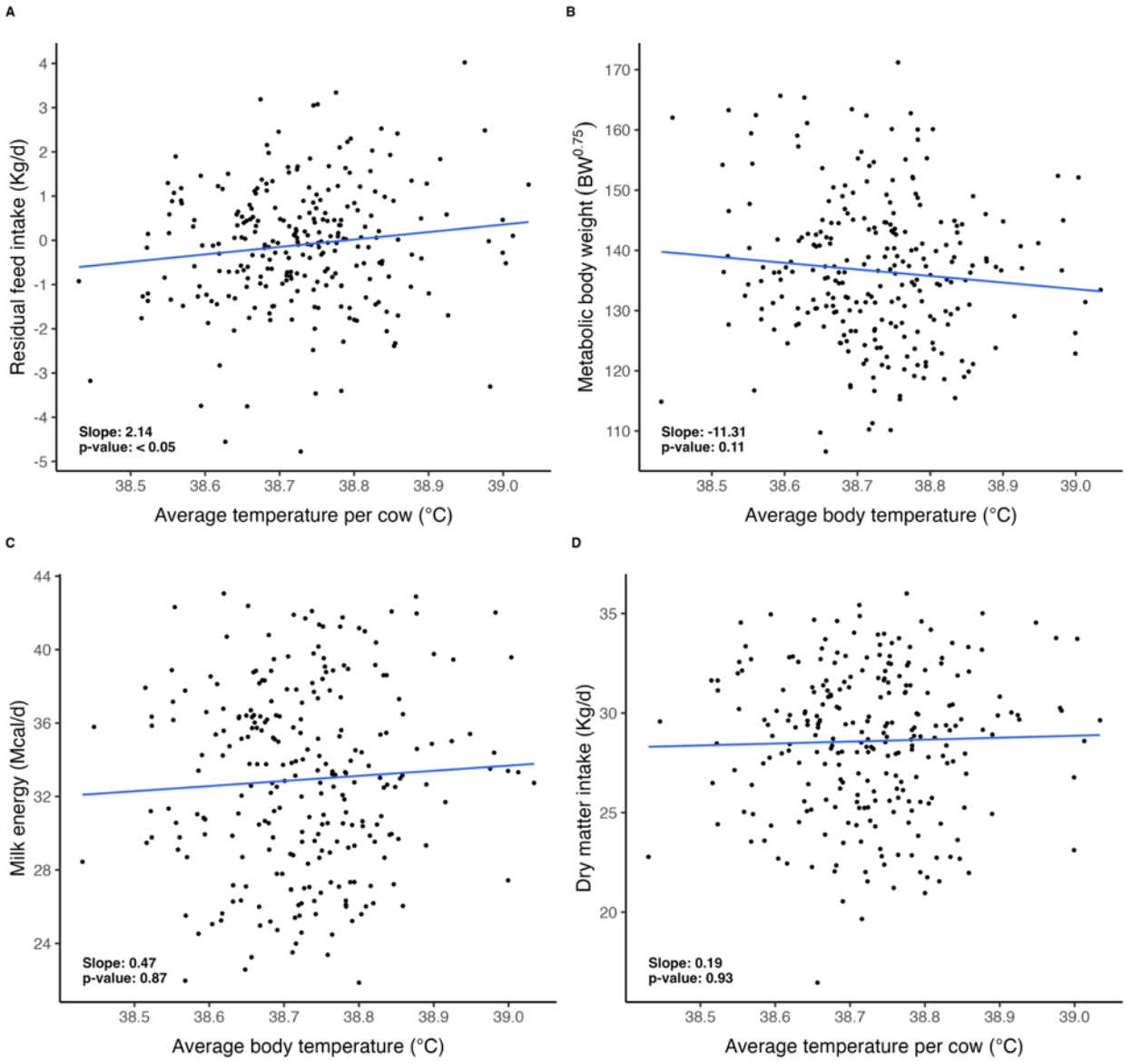

Figure 1. Associations between average body temperature with residual feed intake (RFI), metabolic body weight (MBW), milk energy (MilkE) and dry matter intake (DMI). The average body temperature was calculated as the average temperature during 2 weeks of the feeding trial. The slopes and P-values correspond to the effect of average body temperature adjusted by cohort.

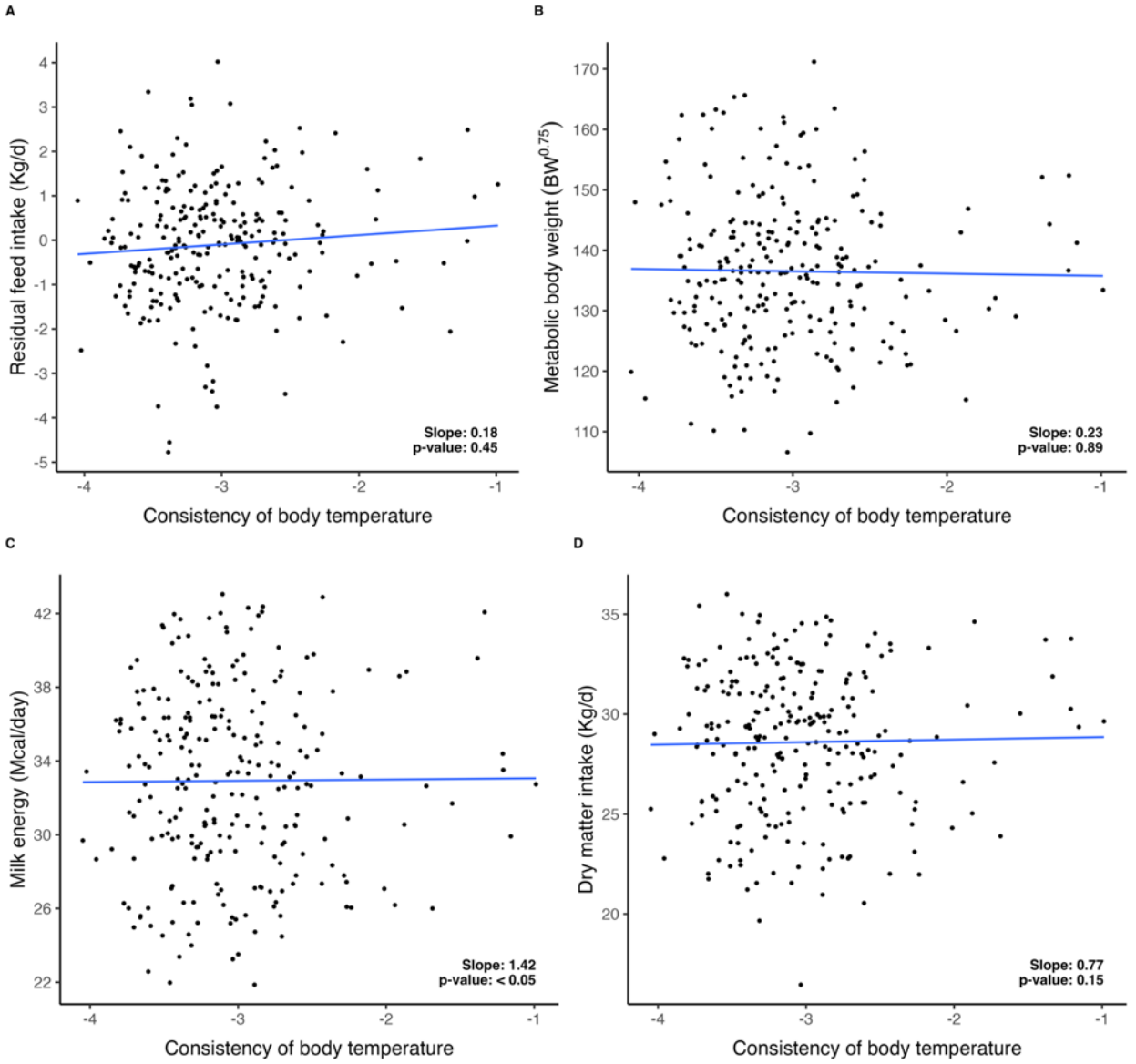

Figure 2. Associations between consistency of body temperature with residual feed intake (RFI), metabolic body weight (MBW), milk energy (MilkE) and dry matter intake (DMI). Consistency of body temperature was obtained as the log transformed variance of the deviations from the mean. The slopes and P-values correspond to the effect of average body temperature adjusted by cohort.

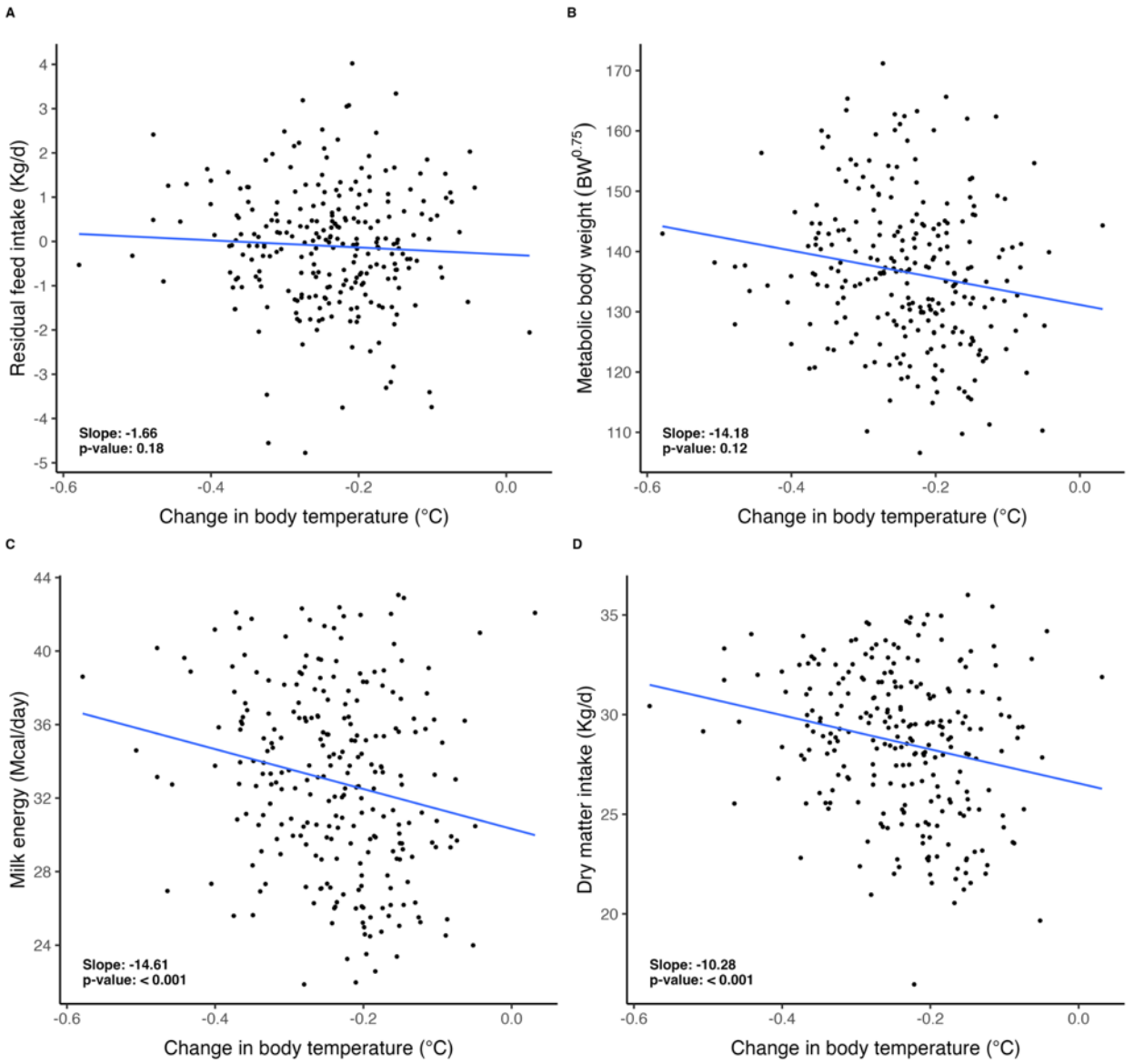

Figure 3. Associations between change in body temperature with residual feed intake (RFI), metabolic body weight (MBW), milk energy (MilkE) and dry matter intake (DMI). Change in body temperature was obtained as the diference in temperature before and after the largest meal of the day. The slopes and P-values correspond to the effect of average body temperature adjusted by cohort.

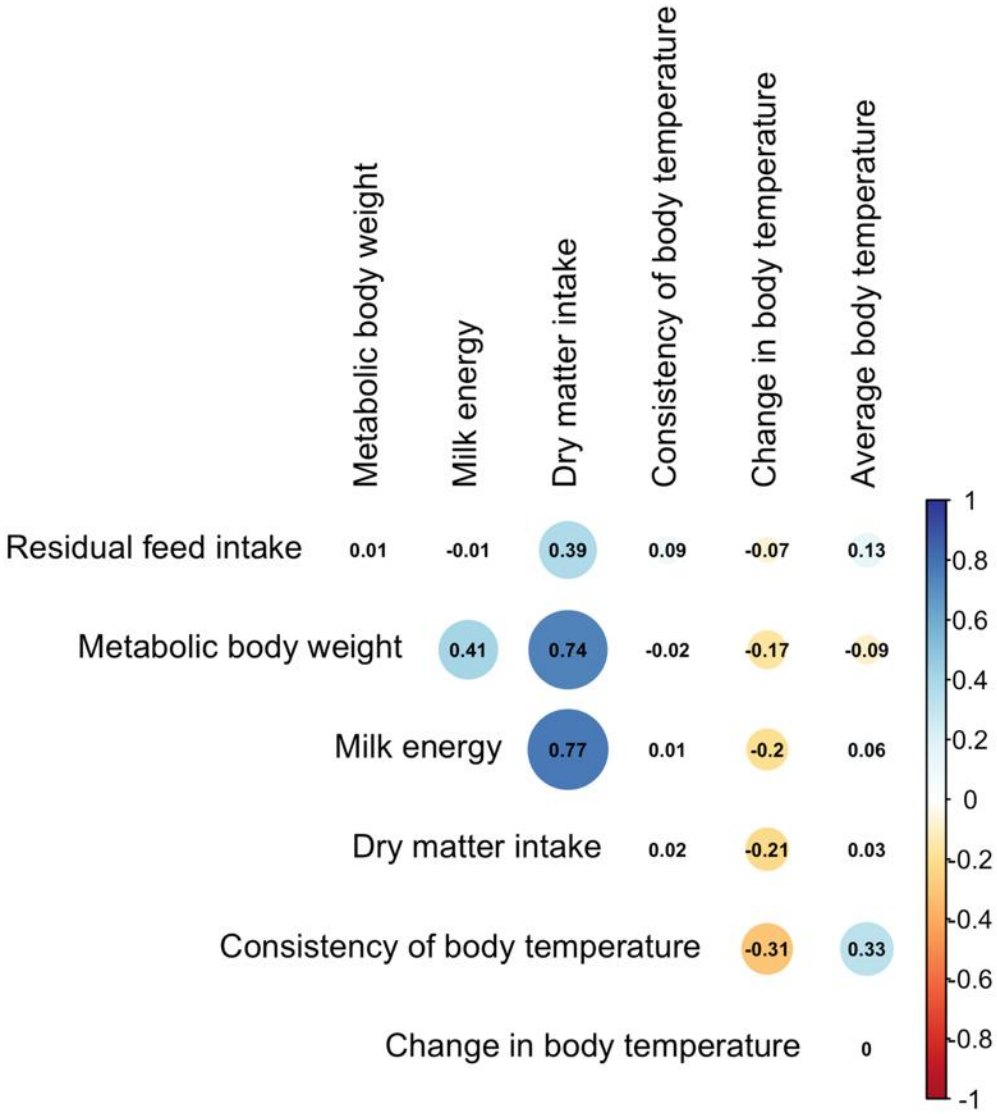

Figure 4. Pearson correlations between body temperature measures and residual feed intake, metabolic body weight, milk energy, and dry matter intake.

References

Aharoni, Y., Brosh, A., and Harari, Y. (2005). Night feeding for high-yielding dairy cows in hot weather: Effects on intake, milk yield and energy expenditure. *Livest. Prod. Sci.* 92, 207–219. doi:10.1016/j.livprodsci.2004.08.013.

Allen, J. D., Hall, L. W., Collier, R. J., and Smith, J. F. (2015). Effect of core body temperature, time of day, and climate conditions on behavioral patterns of lactating dairy cows experiencing mild to moderate heat stress. *J. Dairy Sci.* 98, 118–127. doi:10.3168/jds.2013-7704.

AlZahal, O., AlZahal, H., Steele, M. A., Van Schaik, Kyriazakis, I., Duffield, T. F., et al. (2011). The use of a radiotelemetric ruminal bolus to detect body temperature changes in lactating dairy cattle. *J. Dairy Sci.* 94, 3568–3574. doi:10.3168/jds.2010-3944.

Cavani, L., Brown, W. E., Parker Gaddis, K. L., Tempelman, R. J., VandeHaar, M. J., White, H. M., et al. (2022). Estimates of genetic parameters for feeding behavior traits and their associations with feed efficiency in Holstein cows. *J. Dairy Sci.* 105, 7564–7574. doi:10.3168/jds.2022-22066.

Cook, N., Chabot, B., Liu, T., Froehlich, D., Basarab, J., and Juarez, M. (2020). Radiated temperature from thermal imaging is related to feed consumption, growth rate and feed efficiency in grower pigs. *J. Therm. Biol.* 94, 102747. doi:10.1016/j.jtherbio.2020.102747.

de Melo Costa, C. C., Campos Maia, A. S., Brown-Brandl, T. M., Chiquitelli Neto, M., and de França Carvalho Fonsêca, V. (2018). Thermal equilibrium of Nellore cattle in tropical conditions: an investigation of circadian pattern. *J. Therm. Biol.* 74, 317–324.

doi:10.1016/j.jtherbio.2018.04.014.

DiGiacomo, K., Marett, L. C., Wales, W. J., Hayes, B. J., Dunshea, F. R., and Leury, B. J. (2014). Thermoregulatory differences in lactating dairy cattle classed as efficient or inefficient based on residual feed intake. *Anim. Prod. Sci.* 54, 1877–1881. doi:10.1071/AN14311.

Fischer, A., Delagarde, R., and Faverdin, P. (2018). Identification of biological traits associated with differences in residual energy intake among lactating Holstein cows. *J. Dairy Sci.* 101, 4193–4211. doi:10.3168/jds.2017-12636.

Leão, J. M., Coelho, S. G., Machado, F. S., Azevedo, R. A., Lima, J. A. M., Carneiro, J. C., et al. (2018). Phenotypically divergent classification of preweaned heifer calves for feed efficiency indexes and their correlations with heat production and thermography. *J. Dairy Sci.* 101, 5060–5068. doi:10.3168/jds.2017-14109.

Lees, A. M., Sejian, V., Wallage, A. L., Steel, C. C., Mader, T. L., Lees, J. C., et al. (2019). The impact of heat load on cattle. *Animals* 9, 1–20. doi:10.3390/ani9060322.

Li, B., VanRaden, P. M., Guduk, E., O'Connell, J. R., Null, D. J., Connor, E. E., et al. (2020). Genomic prediction of residual feed intake in US Holstein dairy cattle. *J. Dairy Sci.* 103, 2477–2486. doi:10.3168/jds.2019-17332.

Martello, L. S., da Luz e Silva, S., da Costa Gomes, R., da Silva Corte, R. R. P., and Leme, P. R. (2016). Infrared thermography as a tool to evaluate body surface temperature and its relationship with feed efficiency in Bos indicus cattle in tropical conditions. *Int. J. Biometeorol.* 60, 173–181. doi:10.1007/s00484-015-1015-9.

Morris, D. L., Brown-Brandl, T. M., Miller, P. S., Weiss, W. P., White, R. R., and Kononoff, P.

J. (2021). Factors that affect heat production in lactating Jersey cows. *J. Dairy Sci.* 104, 346–356. doi:10.3168/jds.2020-18976.

Purwanto, B. P., Abo, Y., Sakamoto, R., Yamamoto, S., and Furumoto, F. (1990). Diurnal patterns of heat production and heart rate under thermoneutral conditions in Holstein Friesian cows differing in milk production. *J. Agric. Sci.* 114, 139–142. doi:10.1017/S0021859600072117.

Ríus, A. G. (2019). INVITED REVIEW: Adaptations of protein and amino acid metabolism to heat stress in dairy cows and other livestock species. *Appl. Anim. Sci.* 35, 39–48. doi:10.15232/aas.2018-01805.

Sangsritavong, S., Combs, D. K., Sartori, R., Armentano, L. E., and Wiltbank, M. C. (2002). High feed intake increases liver blood flow and metabolism of progesterone and estradiol-17β in dairy cattle. *J. Dairy Sci.* 85, 2831–2842. doi:10.3168/jds.S0022-0302(02)74370-1.

Suthar, V., Burfeind, O., Maeder, B., and Heuwieser, W. (2013). Agreement between rectal and vaginal temperature measured with temperature loggers in dairy cows. *J. Dairy Res.* 80, 240–245. doi:10.1017/S0022029913000071.

Tao, S., Orellana Rivas, R. M., Marins, T. N., Chen, Y. C., Gao, J., and Bernard, J. K. (2020). Impact of heat stress on lactational performance of dairy cows. *Theriogenology* 150, 437–444. doi:10.1016/j.theriogenology.2020.02.048.

Tempelman, R. J., Spurlock, D. M., Coffey, M., Veerkamp, R. F., Armentano, L. E., Weigel, K. A., et al. (2015). Heterogeneity in genetic and nongenetic variation and energy sink relationships for residual feed intake across research stations and countries. *J. Dairy Sci.* 98, 2013–2026. doi:10.3168/jds.2014.8510.

Tresoldi, G., Schütz, K. E., and Tucker, C. B. (2020). Sampling strategy and measurement device affect vaginal temperature outcomes in lactating dairy cattle. *J. Dairy Sci.* 103, 5414–5421. doi:10.3168/jds.2019-16667.

Vickers, L. A., Burfeind, O., von Keyserlingk, M. A. G., Veira, D. M., Weary, D. M., and Heuwieser, W. (2010). Technical note: Comparison of rectal and vaginal temperatures in lactating dairy cows. *J. Dairy Sci.* 93, 5246–5251. doi:10.3168/jds.2010-3388.

Wheelock, J. B., Rhoads, R. P., VanBaale, M. J., Sanders, S. R., and Baumgard, L. H. (2010). Effects of heat stress on energetic metabolism in lactating Holstein cows. *J. Dairy Sci.* 93, 644–655. doi:10.3168/jds.2009-2295.

CHAPTER FOUR: GENOMIC ANALYSIS OF VISCERAL FAT ACCUMULATION IN HOLSTEIN COWS

Larissa C. Novo[1], Ligia Cavani[1], Pablo Pinedo[2], Pedro Melendez[3], and Francisco Peñagaricano[1]

[1] Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI, USA

[2] Department of Animal Sciences, Colorado State University, Fort Collins, CO, USA

[3] School of Veterinary Medicine, Texas Tech University, Amarillo, TX, USA

Abstract

Visceral fat is related to important metabolic processes, including insulin sensitivity and lipid mobilization. The goal of this study was to identify individual genes, pathways, and molecular processes implicated in visceral fat deposition in dairy cows. Data from 172 genotyped Holstein cows classified at slaughterhouse as having low (n = 77; omental fold < 5 mm in thickness and minimum fat deposition in omentum) or high (n = 95; omental fold ≥ 20 mm in thickness and marked fat deposition in omentum) omental fat were analyzed. The identification of regions with significant additive and non-additive genetic effects was performed using a two-step mixed model-based approach. Genomic scans were followed by gene-set analyses in order to reveal the genetic mechanisms controlling abdominal obesity. The association mapping revealed four regions located on BTA19, BTA20 and BTA24 with significant additive effects. These regions harbor genes, such as SMAD7, ANKRD55, and the HOXB family, that are implicated in lipolysis and insulin tolerance. Three regions located on BTA1, BTA13, and BTA24 showed

marked non-additive effects. These regions harbor genes MRAP, MIS18A, PRNP and TSHZ1, that are directly implicated in adipocyte differentiation, lipid metabolism, and insulin sensitivity. The gene-set analysis revealed functional terms related to cell arrangement, cell metabolism, cell proliferation, cell signaling, immune response, lipid metabolism, and membrane permeability, among other functions. We further evaluated the genetic link between visceral fat and two metabolic disorders, ketosis, and displaced abomasum. For this, we analyzed 28k records of incidence of metabolic disorders from 14k cows across lactations using a single-step genomic BLUP approach. Notably, the region on BTA20 significantly associated with visceral fat deposition was also associated with the incidence of displaced abomasum. Overall, our findings suggest that visceral fat deposition in dairy cows is controlled by both additive and non-additive effects. We detected at least one region with marked pleiotropic effects affecting both visceral fat accumulation and displaced abomasum.

Introduction

Fat deposition occurs essentially in three regions, namely intramuscular, subcutaneous, and around visceral organs. These fat depots differ in structural organization, cellular size, biological function, and metabolic characteristics. Historically, visceral fat was considered to protect and insulate the internal organs; however, its critical role as a form of energy storage and endocrinological signaling has been recently recognized (Booth et al., 2016). Indeed, visceral adipose tissue carries out relevant functions, including lipogenesis to store surplus energy as triglycerides during periods of overnutrition, lipolysis to release energy as free fatty acids during periods of undernutrition, and secretion of a broad spectrum of active molecules, such as proinflammatory cytokines (Item and Konrad, 2012). Visceral adiposity has been linked to several metabolic disorders, including impaired glucose and lipid metabolism, and insulin resistance (Konrad and Wueest, 2014).

Dairy cows typically experience a state of negative energy balance around parturition and early lactation when the energy demand for maintenance and milk production exceeds that of dietary energy intake (Pascottini et al., 2020). This negative energy balance leads to fat mobilization, and consequently, an increase in plasma concentrations of free fatty acids, which are used as a fuel source by peripheral tissues and the mammary gland for milk fat synthesis. Visceral adipocytes in dairy cows are more metabolically active and sensitive to lipolysis than subcutaneous adipocytes, and hence, visceral adipose tissue is a key player in metabolic health during the transition period (Ji et al., 2014b). Note that extreme lipid mobilization is associated with different metabolic disorders, including displacement of the abomasum, ketosis, and fatty liver (Hostens et al., 2012; Contreras et al., 2015).

There is growing evidence that body fat distribution is influenced by genetic factors. For instance, in humans, waist-hip ratio, a measure of body fat distribution independent of overall adiposity, is a heritable trait controlled by multiple significant loci (Heid et al., 2010). In dairy cows, Melendez et al. (2018) reported that visceral fat accumulation is also a heritable trait, and hence, genetic selection to reduce the prevalence of excessive visceral fat is possible.

The first objective of this study was to reanalyze the data from Melendez et al (2018) using alternative methods for gene mapping and the application of novel gene-set tools. Given that visceral fat accumulation is associated with metabolic health, the second objective of this study was to identify genomic regions with pleotropic effects on visceral adiposity and incidence of ketosis or displaced abomasum.

Material and Methods

Visceral Fat Accumulation

Data were collected at a slaughterhouse located in Green Bay, WI, USA (Melendez et al., 2018). Briefly, adult Holstein cows were evaluated from the processing line. After the carcass was eviscerated, the amount of omental fat at the level of the insertion of the lesser omentum over the pylorus area was evaluated. Low visceral fat accumulation was defined as an omental fold < 5 mm in thickness and minimum fat deposition observed throughout the entire omentum. High visceral fat accumulation was defined as an omental fold ≥ 20 mm in thickness and marked fat deposition observed throughout the entire omentum. Only cows with body condition score between 2.75 and 3.25 were considered. As mentioned by Melendez et al. (2018), the goal of this sampling protocol was to obtain two groups of Holstein cows with extreme differences in visceral fat accumulation but with very similar subcutaneous fat deposition. A total of 172 cows were finally selected for this study, 77 with low and 95 with high visceral fat accumulation. It is important to note that an important drawback of the sampling approach used in this work was the lack of management information of animals before arriving at the slaughterhouse. The absence of information such as nutritional strategy, pregnancy status and origin of the animals may influence on the results of the genomic scan and the significant signals observed can be partially explained by a factor not accounted in the model.

These 172 cows were genotyped using the Illumina BovineHD Beadchip with over 777k single nucleotide polymorphism (SNP) probes. After removing monomorphic markers and those located in the X chromosome, a total of 584,557 SNPs remained for the genomic analyses.

Genomic Scans

The importance of additive and non-additive effects on visceral fat accumulation was evaluated using a two-step mixed-model-based approach (Aulchenko et al., 2007).

In the first step, the following model was fitted:

$$\mathrm{y} = \mathrm{Xb} + \mathrm{Zu} + \mathrm{e}$$

where y is the vector of visceral fat accumulation scores (low = 0 or high = 1), b is the vector of fixed effects, u is the vector of random animal effects, and e is the vector of random residual effects. The incidence matrices X and Z relate visceral fat accumulation scores to fixed and animal effects, respectively. The random effects were assumed to follow a multivariate normal distribution with $\mathrm{u} \sim N(0, \mathrm{G}\sigma_u^2)$ and $\mathrm{e} \sim N(0, \mathrm{I}\sigma_e^2)$, where σ_u^2 and σ_e^2 are the animal additive genetic and residual variances respectively, G is the genomic relationship matrix, and I an identity matrix. The variance-covariance matrix for this first model was estimated as $\mathrm{V}_0 = \mathrm{ZGZ}'\sigma_u^2 + \mathrm{I}\sigma_e^2$.

In the second step, the following model was fitted for every SNP:

$$\mathrm{y} = \mathrm{X}\beta + X_{SNP}\beta_{SNP} + \epsilon$$

where X_{SNP} is the design matrix for the marker under consideration and β_{SNP} is the regression coefficient, also known as SNP effect. Every SNP genotype was coded using single numeric variables as (0, 1, 2), (0, 1, 1), (0, 0, 1) and (0, 1, 0) for testing additive, dominance, recessive and overdominance effects, respectively. This model assumes that $\epsilon \sim N(0, \mathrm{V}_0\sigma_e^2)$. The significance of each SNP effect was evaluated using the following test statistic:

$$\mathrm{z} = \frac{\mathrm{X}'_{SNP}\mathrm{V}_0^{-1}(\mathrm{y} - \mathrm{X}\hat{\beta})}{\sqrt{\mathrm{X}'_{SNP}V_0^{-1}\mathrm{X}_{SNP}}}$$

which approximates the Wald test, and hence, is asymptotically standard normal. These analyses were performed using the R package MixABEL (Aulchenko et al., 2007). The possible inflation of the test statistics was evaluated using quantile-quantile (Q-Q) plots. (Zare et al., 2014)

Overrepresentation Analysis

The overrepresentation analysis, also known as gene-set analysis, was performed in three steps as described by Han and Peñagaricano (2016). The first step was the assignment of SNP markers to annotated genes. The latest bovine reference annotation (ARS-UCD1.2) was used to retrieve the exact location of each annotated bovine gene in the genome. SNP markers were assigned to annotated genes if they were located within the genomic sequence of the gene or at most 10 kb upstream or downstream the gene. Significant genes were defined as those genes containing at least one SNP with a significant additive effect (P-value ≤ 0.01). The second step was the assignment of annotated genes to gene-sets. Six different gene-set databases were explored: GO, KEGG, MeSH, InterPro, MSigDB, and Reactome. Finally, in the third step, the enrichment or overrepresentation of significant genes in each gene-set was tested using a Fisher's exact test. All these analyses were performed using the *R* package EnrichKit (https://github.com/liulihe954/EnrichKit), developed by our group.

Genomic Analysis of Metabolic Disorders

Data consisted of 27.4k producer-recorded lactation incidence records of displaced abomasum and ketosis from 13.4k Holstein cows that calved between January 2010 and December 2015 in one large commercial dairy herd in the State of Florida, USA. Metabolic disorders were recorded as binary, i.e., Y = 1 if the cow had clinical symptoms, and Y = 0

otherwise. Genotype data for 60,671 single nucleotide polymorphism (SNP) markers were available for 5.9k cows with health records and 1.4k sires in the pedigree. Markers that mapped to the sex chromosomes, or were monomorphic, or had minor allele frequency less than 1% were removed from the SNP dataset. After data editing, a total of 54,043 SNPs were retained for subsequent analyses.

The incidence of metabolic disorders was analyzed using a threshold model (Gianola, 1982). This model, also known as probit model, describes the observable response variable (0 or 1) using an underlying linear model, $z = \eta + \varepsilon$, where η is a vector of linear predictors and ε is a vector of independent and identically distributed standard normal random variables. Here, the liner predictor η had the following form:

$$\eta = X\beta + Z_1 hys + Z_2 u + Wpe$$

where β is a vector of fixed effects in the model, hys is a vector of random heard-year-season effects, u is a vector of random additive genetic effects, and pe is a vector of random permanent environmental effects. The vector β includes the intercept and the lactation number as a class variable with 5 levels (1, 2, 3, 4, and 5+). The matrices X, Z_1, Z_2, and W are the incidence matrices relating health records to fixed, hear-year-season, animal, and permanent environmental effects, respectively. The random effects were assumed to follow a multivariate normal distribution with $hys \sim N(0, I\sigma^2_{hys})$, $u \sim N(0, H\sigma^2_u)$, and $pe \sim N(0, I\sigma^2_{pe})$, where σ^2_{hys}, σ^2_u and σ^2_{pe} are the hear-year-season, animal additive genetic, and permanent environmental variances respectively, H is a relationship matrix, and I an identity matrix. The matrix H combines pedigree and genotypic information (Aguilar et al., 2010).

Candidate genomic regions associated with metabolic disorders were identified based on the amount of genetic variance explained by 2.0 Mb window of adjacent SNPs. The SNP effects

were estimated as $\hat{s} = DM'[MDM']^{-1}\hat{a}_g$, where $\hat{s}$ is the vector of SNP marker effects, D is a diagonal matrix of weights of SNPs, M is a matrix relating genotypes of each SNP marker to observations, and $\hat{a}_g$ is the vector of genomic estimated breeding values for genotyped animals (Wang et al., 2012). The percentage of genetic variance explained by a 2.0 Mb region was calculated as,

$$\frac{Var(u_i)}{\sigma_u^2} \times 100 = \frac{Var\left(\sum_{j=1}^{N} M_j s_j\right)}{\sigma_u^2} \times 100$$

where u_i is the genetic value of the i^{th} genomic region under consideration, N is the total number of adjacent SNPs within the 2.0 Mb region, and s_j is the marker effect of the j^{th} SNP within the i^{th} region. These analyses were performed using the program POSTGSF90 (Aguilar et al., 2014).

Results

Genomic Scans for Visceral Fat Accumulation

Figure 1 shows the results of the whole-genome single marker scans for testing both additive and non-additive (recessive) genetic effects on visceral fat accumulation in Holstein cows. Six different genomic regions, distributed on chromosomes BTA13, BTA19, BTA20 and BTA24, showed the most significant additive effects. The two significant regions in BTA19 (10.32-10.62 Mb and 37.89-38.19Mb) harbor candidate genes *DHX40*, *YPEL2*, *CLTC*, *SKAP1*, and HOXB1-6 family. These genes are implicated in different functions, including cell proliferation, cell differentiation and immunity. Moreover, the significant region detected in BTA20 (22.88-23.18 Mb) harbors the gene *ANKRD55*, which is related to adipocyte proliferation and lipolysis. Finally, the significant region in BTA24 (48.46-48.76 Mb) harbors the gene *SMAD7*, which is associated with glucose uptake and obesity.

Of particular interest, three genomic regions located in BTA1, BTA13 and BTA24 showed purely non-additive (recessive) effects (Figure 1). The region in BTA1 (3.14-3.44 Mb) harbors the candidate genes *MRAP* and *MIS18A* that are implicated in insulin sensitivity and obesity. The region in BTA13 (46.92-47.22 Mb) harbors the gene *PRNP* which is directly involved in visceral fat adipose tissue deposition. The significant region in BTA24 (3.48-3.78 Mb) harbors the gene *TSHZ1* that is associated with adipocyte differentiation and lipid metabolism.

Overrepresentation Analysis

The overrepresentation analysis, namely the search for gene-sets or gene pathways that show an overrepresentation of significant genes, was performed using a Fisher's exact test, a test of proportions based on the cumulative hypergeometric distribution. Figure 2 shows a set of terms that were significantly enriched with genes associated with visceral fat accumulation. These functional terms are related to calcium signaling, cell arrangement, cell metabolism, cell proliferation, cell signaling, immune response, lipid metabolism, membrane permeability, and nervous signaling, among other functions.

Genomic Analysis of Metabolic Disorders

The identification of genomic regions affecting displaced abomasum or ketosis was performed using the single-step genomic BLUP. This method combines all the available phenotypic, genotypic, and pedigree information, and fits all the SNP simultaneously. Candidate regions were identified based on the amount of genetic variance explained by 2.0 Mb SNP-windows. Figure 3 shows the gene mapping results for displaced abomasum. Notably, the prominent peak in BTA20, which harbors gene *ANKRD55*, was also significantly associated with visceral fat accumulation, suggesting a pleiotropic action. On the other hand, there were not common regions between ketosis and visceral fat accumulation (data not shown).

Discussion

Visceral fat is a highly active tissue involved in complex metabolic processes, such as inflammation and insulin sensitivity. Dairy cows with excessive visceral fat are more susceptible to metabolic disorders (Drackley et al., 2014; Ji et al., 2014a; Singh et al., 2014). Although it has importance for health and production traits, very few studies have investigated the genetic basis of visceral fat accumulation in dairy cattle. As such, this study was specially conducted to identify individual genes, functional gene-sets and biological pathways associated with visceral fat accretion in Holstein cows. We also investigated the genetic link between visceral fat accumulation and the incidence of metabolic disorders, namely ketosis and displaced abomasum.

As a preparation process for the negative energy balance experienced by dairy cows during lactation, the organism undergoes several adaptations during late pregnancy, including energy storage accompanied by hypertrophy and active remodeling of internal adipose tissue (Kenéz et al., 2015). Several pathways reveled in this work coincide with cell rearrangement of adipose tissue identified in other species in obesity studies. In fact, our gene-set analysis detected biological pathways directly involved in the visceral fat expansion, such as cell arrangement, cell proliferation and cytoskeleton regulation. Additionally, some of the most significant genes detected in our whole-genome scans are directly implicated in lipid accumulation and tissue rearrangement. For instance, gene *PRNP* encodes the cellular prion protein known to regulate visceral fat volume, body fat weight, adipocyte cell size, and body weight gain in mice (Jeong et al., 2019). Similarly, the significant gene *ZADH2*, also known as *PTGR-3*, negatively modulates adipocyte differentiation through regulation of *PPARγ*, a major regulator of adipogenesis (Yu et al., 2013). Concerning rearrangement, the significant gene *YPEL2* is known to be involved in the

cell division (Hosono et al., 2004) and the *HOXB* family to encode for transcription factors for genes involved in the anatomical structure of omental fat (Ahn et al., 2019). Genes implicated in the cascade signaling due to the active release of free fatty acids on liver were also detected. Notably, gene *SMAD7*, one of the most significant genes revealed in the association mapping, is associated with higher levels of circulating free fatty acids, lower expression of lipolytic genes, and more proinflammatory proteins in obese mice (Seong et al., 2018). While the action of *SMAD7* occurs by downregulating the TGF-β pathway, the significant gene *CLTC* stimulates this pathway while downregulating *NADPH* oxidase to protect against the negative effects of highly active free fatty acids oxidation (Han, 2016; Caballero-Díaz et al., 2020). Additionally, our overrepresentation analysis detected pathways related to fibroblast growth factor receptors (FGFR), calcium signaling, protein kinases and glutamate signaling. The FGFR1 signaling pathway is related to lipid droplet dynamics, phospholipid homeostasis, protection against oxidative stress and to hypertrophy in obese individuals (Ye et al., 2016). The FGFR2 signaling pathway, inhibited by the significant gene *SMAD7*, indirectly promotes lipid biosynthesis by reducing cAMP pool and protein kinase A (PKA) activity (Ornitz and Itoh, 2015; Huang et al., 2018). Interestingly, both cAMP and PKA are affected by the gene *MRAP*, a significant gene detected in the non-additive scan. In an interesting way, this gene is associated with mitochondrial fatty acid oxidation and is indispensable for the lipolytic response to adrenocorticoid hormone, and consequently, insulin sensitivity (Zhang et al., 2018).

Research has shown that inflammation is directly impacted by higher availability of glucose and free fatty acids (Patel and Abate, 2013). Interestingly, our study revealed many genes and gene-sets associated with inflammation. For instance, the gene *NOX4*, the main source of reactive oxygen species (ROS) is highly expressed in adipocytes and is controlled by the gene

CLTC detected in this work (Den Hartigh et al., 2017). Curiously, the insulin resistance mentioned before as an effect of the gene *MRAP* has ROS as one of the major causes (Han, 2016). Our enrichment analysis detected calcium transport terms that are known to be affected by ROS in the form of impaired calcium homeostasis that can lead to cell death (Dejos et al., 2020). Other adipose-related inflammation response is also represented is this work. The significant gene *SLC23A2* codes for SVCT1, a transporter of vitamin C, a well-known antioxidant that is able to inhibit adipocyte differentiation and lipid accumulation (Rahman et al., 2014). Significant functional terms such as protein kinase C, dendritic spine, and metalloproteinases may indicate the response to local inflammation via proliferation, activation and communication of T-cells, respectively (Black and Black, 2012; Khokha et al., 2013; Sundara Rajan and Longhi, 2016). Interestingly, the gene *SKAP1* identified in the genomic scan is an immune cell adaptor responsible for regulating multiple functions of T-cells (Raab et al., 2019). Enriched terms such as metalloproteases and bacterial humoral defense can also be related to systemic inflammation as the threshold between the beneficial acute inflammation and damaging effect of chronic inflammation is controlled by metalloproteinases via macrophage activity (Khokha et al., 2013).

There is growing evidence that excessive visceral fat may lead to metabolic disorders. In fact, the significant genes and pathways identified in this study suggest a differential inflammatory response among cows with different levels of visceral fat. It is known that cows with displaced abomasum have preferable mobilization of visceral over subcutaneous fat and present higher macrophage infiltration in the omental adipose tissue compared to healthy individuals (Hostens et al., 2012; Contreras et al., 2015). Interestingly, our work revealed one region in chromosome 20 that has significant effects on both visceral fat accumulation and

susceptibility to displaced abomasum. Notably, this region harbors the gene *ANKRD55*, which encodes a scaffold protein related to proliferation of pre-adipocytes, insulin sensitivity, and even more important, higher visceral fat accumulation in human subjects (Harder et al., 2013; Ji et al., 2019; Chen et al., 2020b). Gene *ANKRD55* is highly active in immune diseases, which corresponds to the state of the clinically diagnosed displaced abomasum cows used in this study, as studies have shown that cows with displaced abomasum are under active lipolysis, negative energy balance, and under higher infiltration of macrophages in adipose tissues (Hostens et al., 2012; Contreras et al., 2015). The whole-body insulin resistance possibly promoted by *ANKRD55* would endorse the insulin resistance stimulated by cytokines release from visceral fat and the impaired glucose-stimulated insulin secretion in pancreatic β cells mediated by the significant gene *TSHZ1* (Raum et al., 2015). Changes in insulin concentration and blood calcium levels, two mechanisms identified in our gene-set analysis, are notably one of the causes for displaced abomasum in cows (Van Winden et al., 2003). This common peak in BTA20 for visceral fat accumulation and displaced abomasum suggests a genetic link between visceral fat levels and the incidence of metabolic diseases that deserves further investigation.

Conclusions

In this study, we performed an integrative genomic analysis to understand the genetic and biological basis of visceral fat accumulation in Holstein dairy cattle. Our analysis revealed significantly associated regions located on BTA19, BTA20 and BTA24 for additive model and unique regions on BTA1, BTA13 and BTA24 with purely non-additive (recessive) effects. These regions harbor genes such as *SMAD7*, *SKAP1*, *ANKRD55*, *MRAP* and *MIS18A* that are directly associated with adipocyte differentiation, immune response, lipid metabolism and insulin tolerance. The gen-set analysis also revealed pathways related to tyrosine-kinase receptors, cell signaling, calcium channels and several related to metabolism, including negative regulation of fibroblast growth factor receptor 1 and 2. We also performed an analysis to investigate the link between visceral fat and metabolic diseases. One common region was identified for displaced abomasum in BTA20 harboring the gene *ANKRD55*, an autoimmune risk protein linked to adipocyte differentiation and insulin resistance. Our study suggests that visceral fat deposition in dairy cows is controlled by both additive and non-additive effects, and a genetic link between visceral fat accumulation and metabolic diseases that may be better investigated in further studies.

Tables

Table 1. Genetic markers and candidate genes associated with visceral fat accumulation in Holstein cows

Chromosome	Position	Gene	Function	p-value
Additive effects				
BTA19	10.32-10.62	*YIPEL2*	Cell cycle	5.047331e-07
BTA19	37.89-38.19	*CLTC*	Embryonic development and cell apoptosis	4.928219e-06
		HOXB1-6	Characterization of omental adipose tissue	4.928219e-06
		SKAP1	Immune cell adaptor	4.928219e-06
BTA20	22.88-23.18	*ANKRD55*	Pre-adipocyte proliferation and/or differentiation	4.032977e-06
BTA24	48.46-48.76	*SMAD7*	Adipocyte differentiation and apoptosis	5.358331e-06
Recessive effects				
BTA1	3.14-3.39	*MRAP*	Adrenocorticotropic hormone-induced lipolysis and insulin sensitivity	2.418982e-05
		MIS18A	Cell cycle and cell proliferation	2.418982e-05
BTA13	46.92-47.22	*PRNP*	Adipocyte differentiation via autophagic flux	8.937391e-06
		SLC23A2	Vitamin C transport	8.937391e-06
BTA24	3.48-3.78	*TSHZ1*	Part of b-cell transcriptional network	6.281956e-07
		ZADH2 (PTGR3)	Adipocyte differentiation via PPARy	6.281956e-07

Figures

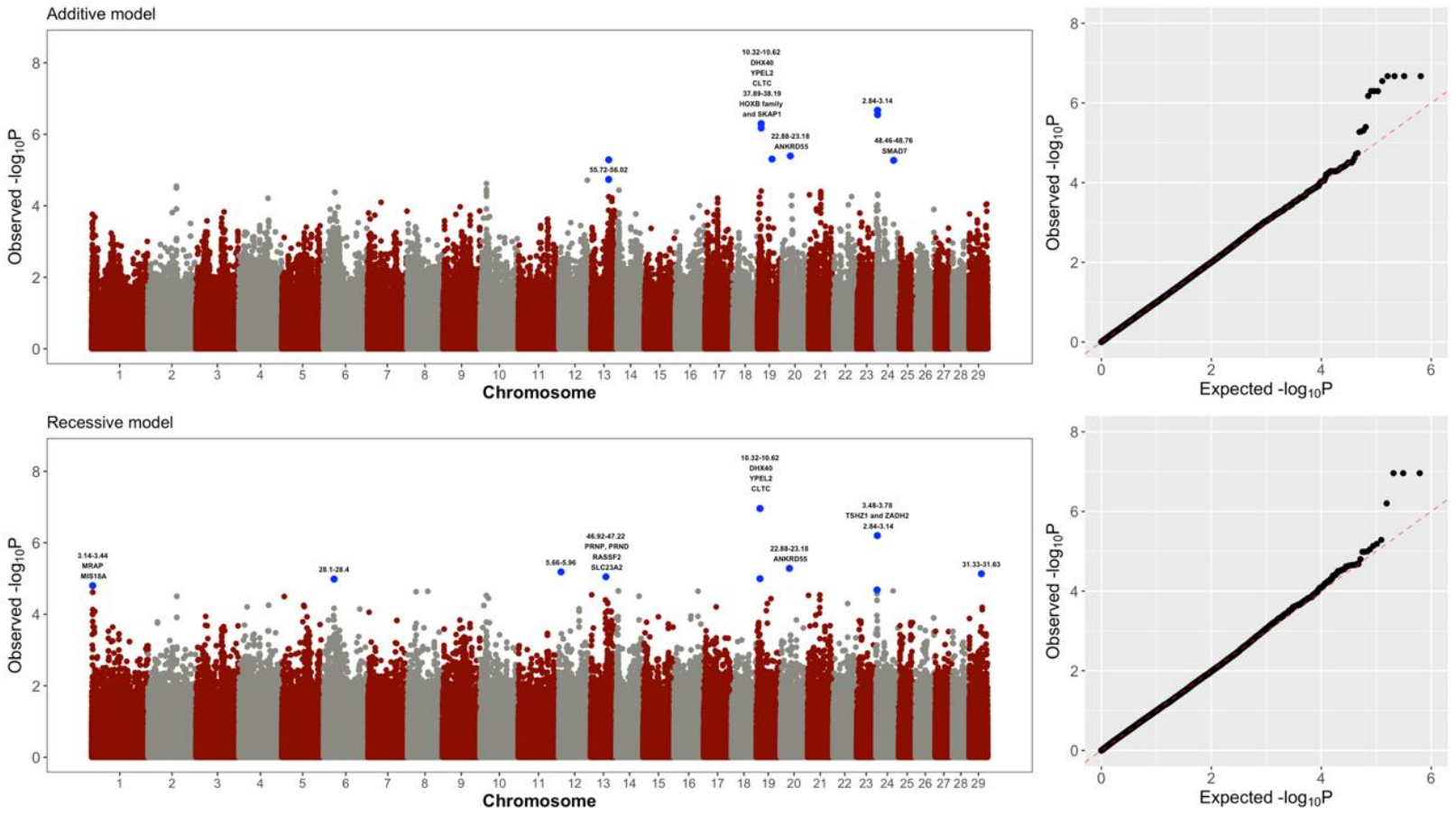

Figure 1. Manhattan plots and quantile-quantile plots showing the significance of additive and recessive effects on visceral fat accumulation across the entire bovine genome. Genes directly implicated in adipocyte differentiation, lipid metabolism, and insulin sensitivity are highlighted in the Manhattan plots.

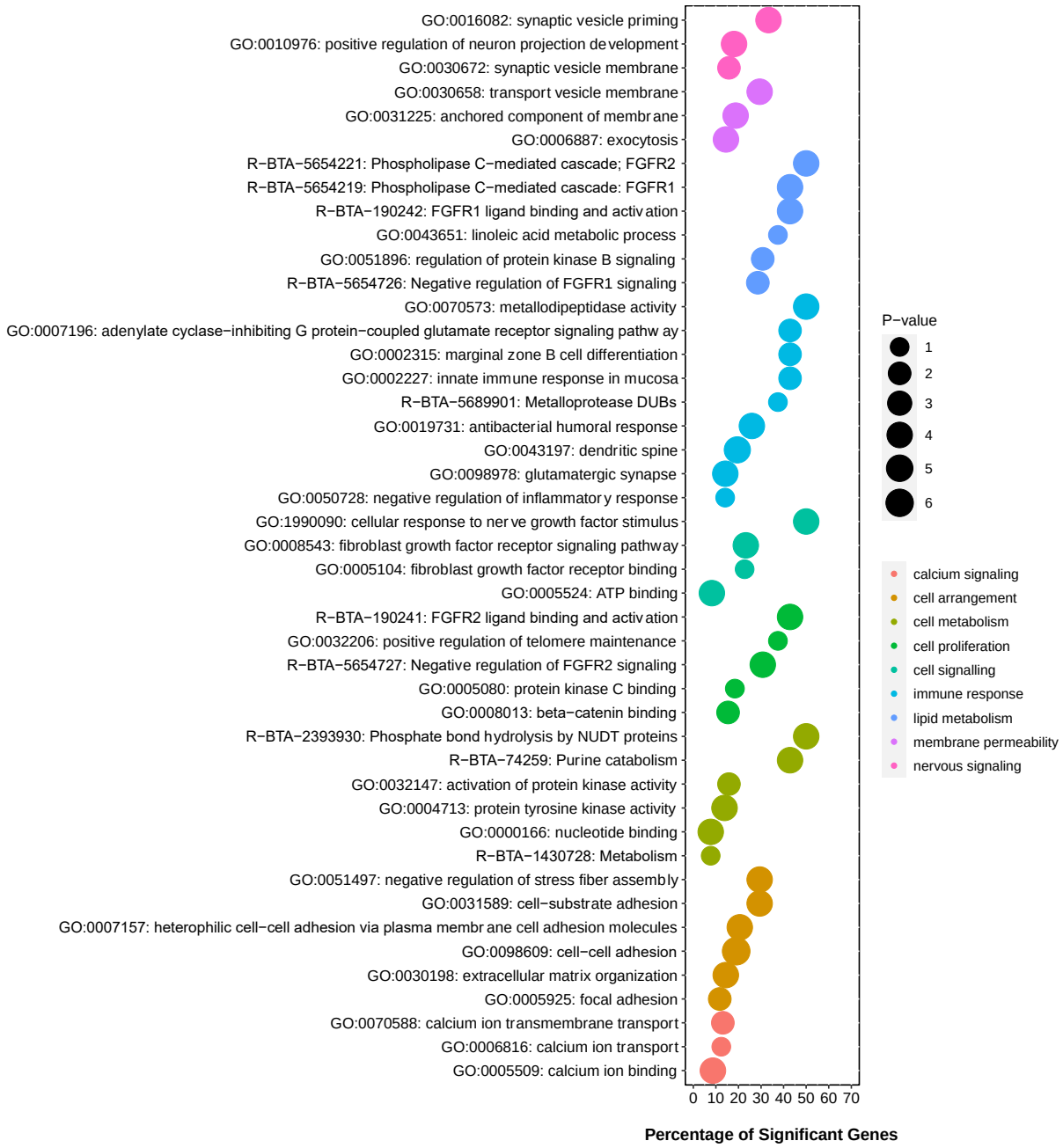

Figure 2. Functional terms significantly enriched with genes associated with visceral fat accumulation. Six gene-set databases were analyzed: GO, KEGG, MeSH, InterPro, MSigDB, and Reactome. The y-axis displays the name while the x-axis displays the percentage of significant genes in each functional enriched term.

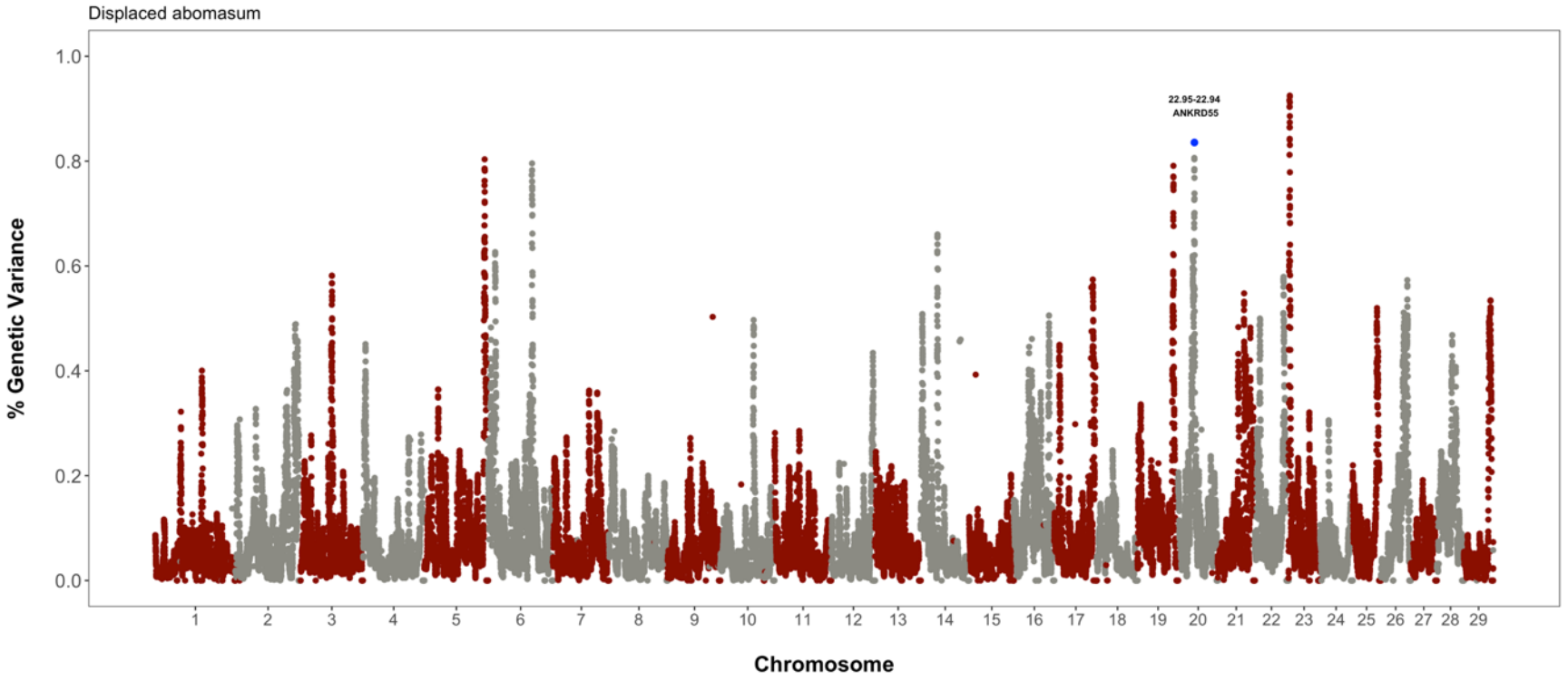

Figure 3. Whole-genome scan for incidence of displaced abomasum: percentage of additive genetic variance explained by 2.0 Mb SNP-windows across the entire bovine genome. Candidate gene *ANKRD55* is implicated in both visceral fact accumulation and displaced abomasum.

References

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S., and Lawlor, T. J. (2010). Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93, 743–752. doi:10.3168/jds.2009-2730.

Aguilar, I., Misztal, I., Tsuruta, S., and Legarra, A. (2014). PREGSF90 – POSTGSF90: Computational Tools for the Implementation of Single-step Genomic Selection and Genome-wide ... in *Proceedings, 10th World Congress of Genetics Applied to Livestock Production* (Vancouver, BC, Canada), 0–3. doi:10.13140/2.1.4801.5045.

Ahn, J., Wu, H., and Lee, K. (2019). Integrative Analysis Revealing Human Adipose-Specific Genes and Consolidating Obesity Loci. *Sci. Rep.* 9, 1–13. doi:10.1038/s41598-019-39582-8.

Aulchenko, Y. S., Ripke, S., Isaacs, A., and van Duijn, C. M. (2007). GenABEL: An R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296. doi:10.1093/bioinformatics/btm108.

Black, A. R., and Black, J. D. (2012). Protein kinase C signaling and cell cycle regulation. *Front. Immunol.* 3, 1–17. doi:10.3389/fimmu.2012.00423.

Booth, A., Magnuson, A., Fouts, J., and Foster, M. T. (2016). Adipose tissue: An endocrine organ playing a role in metabolic regulation. *Horm. Mol. Biol. Clin. Investig.* 26, 25–42. doi:10.1515/hmbci-2015-0073.

Caballero-Díaz, D., Bertran, E., Peñuelas-Haro, I., Moreno-Càceres, J., Malfettone, A., López-Luque, J., et al. (2020). Clathrin switches transforming growth factor-β role to pro-tumorigenic in liver cancer. *J. Hepatol.* 72, 125–134. doi:10.1016/j.jhep.2019.09.012.

Chen, Z., Yu, H., Shi, X., Warren, C. R., Lotta, L. A., Friesen, M., et al. (2020). Functional Screening of Candidate Causal Genes for Insulin Resistance in Human Preadipocytes and Adipocytes. *Circ. Res.*, 330–346. doi:10.1161/CIRCRESAHA.119.315246.

Contreras, G. A., Kabara, E., Brester, J., Neuder, L., and Kiupel, M. (2015). Macrophage infiltration in the omental and subcutaneous adipose tissues of dairy cows with displaced abomasum. *J. Dairy Sci.* 98, 6176–6187. doi:10.3168/jds.2015-9370.

Dejos, C., Gkika, D., and Cantelmo, A. R. (2020). The Two-Way Relationship Between Calcium and Metabolism in Cancer. *Front. Cell Dev. Biol.* 8, 1–14. doi:10.3389/fcell.2020.573747.

Den Hartigh, L. J., Omer, M., Goodspeed, L., Wang, S., Wietecha, T., O'Brien, K. D., et al. (2017). Adipocyte-Specific Deficiency of NADPH Oxidase 4 Delays the Onset of Insulin Resistance and Attenuates Adipose Tissue Inflammation in Obesity. *Arterioscler. Thromb. Vasc. Biol.* 37, 466–475. doi:10.1161/ATVBAHA.116.308749.

Drackley, J. K., Wallace, R. L., Graugnard, D., Vasquez, J., Richards, B. F., and Loor, J. J. (2014). Visceral adipose tissue mass in nonlactating dairy cows fed diets differing in energy density1. *J. Dairy Sci.* 97, 3420–3430. doi:10.3168/jds.2014-8014.

Han, C. Y. (2016). Roles of reactive oxygen species on insulin resistance in adipose tissue. *Diabetes Metab. J.* 40, 272–279. doi:10.4093/dmj.2016.40.4.272.

Han, Y., and Peñagaricano, F. (2016). Unravelling the genomic architecture of bull fertility in Holstein cattle. *BMC Genet.* 17, 1–11. doi:10.1186/s12863-016-0454-6.

Harder, M. N., Ribel-Madsen, R., Justesen, J. M., Sparsø, T., Andersson, E. A., Grarup, N., et al. (2013). Type 2 diabetes risk alleles near BCAR1 and in ANK1 associate with decreased β-cell function whereas risk alleles near ANKRD55 and GRB14 associate with decreased insulin sensitivity in the danish Inter99 cohort. *J. Clin. Endocrinol. Metab.* 98, 801–806.

doi:10.1210/jc.2012-4169.

Heid, I. M., Jackson, A. U., Randall, J. C., Winkler, T. W., Qi, L., Ssteinthorsdottir, V., et al. (2010). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* 42, 949–960. doi:10.1038/ng.685.

Hosono, K., Sasaki, T., Minoshima, S., and Shimizu, N. (2004). Identification and characterization of a novel gene family YPEL in a wide spectrum of eukaryotic species. *Gene* 340, 31–43. doi:10.1016/j.gene.2004.06.014.

Hostens, M., Fievez, V., Leroy, J. L. M. R., Van Ranst, J., Vlaeminck, B., and Opsomer, G. (2012). The fatty acid profile of subcutaneous and abdominal fat in dairy cows with left displacement of the abomasum. *J. Dairy Sci.* 95, 3756–3765. doi:10.3168/jds.2011-5092.

Huang, X., Liu, G., Guo, J., and Su, Z. Q. (2018). The PI3K/AKT pathway in obesity and type 2 diabetes. *Int. J. Biol. Sci.* 14, 1483–1496. doi:10.7150/ijbs.27173.

Item, F., and Konrad, D. (2012). Visceral fat and metabolic inflammation: The portal theory revisited. *Obes. Rev.* 13, 30–39. doi:10.1111/j.1467-789X.2012.01035.x.

Jeong, J. K., Lee, J. H., Kim, S. W., Hong, J. M., Seol, J. W., and Park, S. Y. (2019). Cellular prion protein regulates the differentiation and function of adipocytes through autophagy flux. *Mol. Cell. Endocrinol.* 481, 84–94. doi:10.1016/j.mce.2018.11.013.

Ji, P., Drackley, J. K., Khan, M. J., and Loor, J. J. (2014a). Inflammation- and lipid metabolism-related gene network expression in visceral and subcutaneous adipose depots of Holstein cows. *J. Dairy Sci.* 97, 3441–3448. doi:10.3168/jds.2013-7296.

Ji, P., Drackley, J. K., Khan, M. J., and Loor, J. J. (2014b). Overfeeding energy upregulates peroxisome proliferator-activated receptor (PPAR)γ-controlled adipogenic and lipolytic

gene networks but does not affect proinflammatory markers in visceral and subcutaneous adipose depots of Holstein cows. *J. Dairy Sci.* 97, 3431–3440. doi:10.3168/jds.2013-7295.

Ji, Y., Yiorkas, A. M., Frau, F., Mook-Kanamori, D., Staiger, H., Thomas, E. L., et al. (2019). Genome-wide and abdominal MRI data provide evidence that a genetically determined favorable adiposity phenotype is characterized by lower ectopic liver fat and lower risk of type 2 diabetes, heart disease, and hypertension. *Diabetes* 68, 207–219. doi:10.2337/db18-0708.

Kenéz, Á., Kulcsár, A., Kluge, F., Benbelkacem, I., Hansen, K., Locher, L., et al. (2015). Changes of adipose tissue morphology and composition during late pregnancy and early lactation in dairy cows. *PLoS One* 10, 1–11. doi:10.1371/journal.pone.0127208.

Khokha, R., Murthy, A., and Weiss, A. (2013). Metalloproteinases and their natural inhibitors in inflammation and immunity. *Nat. Rev. Immunol.* 13, 649–665. doi:10.1038/nri3499.

Konrad, D., and Wueest, S. (2014). The gut-adipose-liver axis in the metabolic syndrome. *Physiology* 29, 304–313. doi:10.1152/physiol.00014.2014.

Melendez, P., Poock, S. E., Pithua, P., Pinedo, P., Manriquez, D., Moore, S. G., et al. (2018). Genome-wide study to detect single nucleotide polymorphisms associated with visceral and subcutaneous fat deposition in Holstein dairy cows. *Animal* 13, 487–494. doi:10.1017/S1751731118001519.

Ornitz, D. M., and Itoh, N. (2015). The fibroblast growth factor signaling pathway. *Wiley Interdiscip. Rev. Dev. Biol.* 4, 215–266. doi:10.1002/wdev.176.

Pascottini, O. B., Leroy, J. L. M. R., and Opsomer, G. (2020). Metabolic stress in the transition period of dairy cows: Focusing on the prepartum period. *Animals* 10, 1–17. doi:10.3390/ani10081419.

Patel, P., and Abate, N. (2013). Body fat distribution and insulin resistance. *Nutrients* 5, 2019–2027. doi:10.3390/nu5062019.

Raab, M., Strebhardt, K., and Rudd, C. E. (2019). Immune adaptor SKAP1 acts a scaffold for Polo-like kinase 1 (PLK1) for the optimal cell cycling of T-cells. *Sci. Rep.* 9, 1–13. doi:10.1038/s41598-019-45627-9.

Rahman, F., Frouh, F. Al, Bordignon, B., Fraterno, M., Landrier, J. F., Peiretti, F., et al. (2014). Ascorbic acid is a dose-dependent inhibitor of adipocyte differentiation, probably by reducing cAMP pool. *Front. Cell Dev. Biol.* 2, 1–10. doi:10.3389/fcell.2014.00029.

Raum, J. C., Soleimanpour, S. A., Groff, D. N., Coré, N., Fasano, L., Garratt, A. N., et al. (2015). Tshz1 regulates pancreatic β-cell maturation. *Diabetes* 64, 2905–2914. doi:10.2337/db14-1443.

Seong, H. A., Manoharan, R., and Ha, H. (2018). Smad proteins differentially regulate obesity-induced glucose and lipid abnormalities and inflammation via class-specific control of AMPK-related kinase MPK38/MELK activity. *Cell Death Dis.* 9. doi:10.1038/s41419-018-0489-x.

Singh, S. P., Häussler, S., Heinz, J. F. L., Akter, S. H., Saremi, B., Müller, U., et al. (2014). Lactation driven dynamics of adiponectin supply from different fat depots to circulation in cows. *Domest. Anim. Endocrinol.* 47, 35–46. doi:10.1016/j.domaniend.2013.12.001.

Sundara Rajan, S., and Longhi, M. P. (2016). Dendritic cells and adipose tissue. *Immunology* 149, 353–361. doi:10.1111/imm.12653.

Van Winden, S. C. L., Jorritsma, R., Müller, K. E., and Noordhuizen, J. P. T. M. (2003). Feed intake, milk yield, and metabolic parameters prior to left displaced abomasum in dairy cows. *J. Dairy Sci.* 86, 1465–1471. doi:10.3168/jds.S0022-0302(03)73730-8.

Wang, H., Misztal, I., Aguilar, I., Legarra, A., and Muir, W. M. (2012). Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res. (Camb).* 94, 73–83. doi:10.1017/S0016672312000274.

Ye, M., Lu, W., Wang, X., Wang, C., Abbruzzese, J. L., Liang, G., et al. (2016). FGF21-FGFR1 coordinates phospholipid homeostasis, lipid droplet function, and ER stress in obesity. *Endocrinology* 157, 4754–4769. doi:10.1210/en.2016-1710.

Yu, Y. H., Chang, Y. C., Su, T. H., Nong, J. Y., Li, C. C., and Chuang, L. M. (2013). Prostaglandin reductase-3 negatively modulates adipogenesis through regulation of PPARγ activity. *J. Lipid Res.* 54, 2391–2399. doi:10.1194/jlr.M037556.

Zare, Y., Shook, G. E., Collins, M. T., and Kirkpatrick, B. W. (2014). Genome-wide association analysis and genomic prediction of Mycobacterium avium subspecies paratuberculosis infection in US Jersey cattle. *PLoS One* 9. doi:10.1371/journal.pone.0088380.

Zhang, X., Saarinen, A. M., Campbell, L. E., De Filippis, E. A., and Liu, J. (2018). Regulation of lipolytic response and energy balance by melanocortin 2 receptor accessory protein (MRAP) in adipocytes. *Diabetes* 67, 222–234. doi:10.2337/db17-0862.

CHAPTER FIVE: IDENTIFICATION OF GENETIC VARIANTS AND INDIVIDUAL GENES ASSOCIATED WITH POSTPARTUM HYPOCALCEMIA IN HOLSTEIN COWS

Larissa C. Novo,[1] Michael B. Poindexter,[2] Fernanda M. Rezende,[2] José E. P. Santos,[2] Corwin D. Nelson,[2] Laura L. Hernandez,[1] Brian W. Kirkpatrick,[1] and Francisco Peñagaricano[1]

[1] Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI 53706, USA

[2] Department of Animal Sciences, University of Florida, Gainesville, FL 32611, USA

Abstract

Periparturient hypocalcemia is a complex metabolic disorder that occurs at the onset of lactation because of a sudden irreversible loss of Ca incorporated into colostrum and milk. Some cows are unable to quickly adapt to this demand and succumb to clinical hypocalcemia, commonly known as milk fever, whereas a larger proportion of cows develop subclinical hypocalcemia. The main goal of this study was to identify putative causative mutations and candidate genes affecting postpartum blood calcium concentration in Holstein cows. Data consisted of blood calcium concentration measured in 2,513 Holstein cows on the first three days after parturition. All cows had genotypic information for 79k SNP markers. Two consecutive rounds of imputation were performed: first, the 2,513 Holstein cows were imputed from 79k to 312k SNP markers. This imputation was performed using a reference set of 17,131 proven Holstein bulls with 312k SNP markers. Then, the 2,513 Holstein cows were imputed from 312k markers to whole-genome sequence data. This second round of imputation used 179 Holstein

animals from the 1000 Bulls Genome Project as a reference set. Three alternative phenotypes were evaluated: (i) total calcium concentration in the first 24 h postpartum, (ii) total calcium concentration in the first 72 h postpartum calculated as the area under the curve; and (iii) the recovery of total calcium concentration calculated as the difference in total calcium concentration between 72 and 24 h. The identification of genetic variants associated with these traits was performed using a two-step mixed model-based approach implemented in the R package MixABEL. The most significant variants were located within or near genes involved in calcium homeostasis and vitamin D transport (*GC*), calcium and potassium channels (*JPH3* and *KCNK13*), energy and lipid metabolism (*CA5A*, *PRORP*, and *SREBP1*), and immune response (*IL12RB2* and *CXCL8*), among other functions. This work provides the foundation for the development of novel breeding and management tools for reducing the incidence of periparturient hypocalcemia in dairy cattle.

Keywords: calcium homeostasis, causal mutation, genomic scan

Introduction

Periparturient hypocalcemia is a complex metabolic disorder that occurs at the onset of lactation because of a sudden irreversible loss of calcium incorporated into colostrum and milk. Some cows are unable to quickly adapt to this demand and succumb to clinical hypocalcemia, commonly known as milk fever. A larger proportion of cows develop subclinical hypocalcemia which increases the risk of the cow to develop other peripartum diseases, such as retained placenta, uterine prolapse, endometritis, displaced abomasum, ketosis, and mastitis. The economic losses due to periparturient hypocalcemia are substantial due to reduced milk production, extended days open, and early culling, among others (Liang et al., 2017).

Periparturient hypocalcemia is a heritable trait with heritability estimates between 0.01 to 0.35 depending on the breed and the methodology (Pryce et al., 2016). The magnitude of these heritability estimates suggests that genetic selection can be effective to reduce the incidence of periparturient hypocalcemia in dairy cattle. Little is known, however, about the genes underlying cow's susceptibility to periparturient hypocalcemia. Previous genomic studies identified genes implicated in calcium and potassium homeostasis, such as *GC*, *LRRC38*, *KCNK9*, and *CYP27A*, as associated with milk fever and postpartum blood calcium concentration in Holstein cows (Pacheco et al., 2018; Cavani et al., 2022b). In addition, Sasaki et al. (Sasaki et al., 2014), evaluating gene expression of peripheral blood mononuclear cells, reported genes *PKIB*, *DDIT4*, *PER1*, and *NUAK1*, as potential biomarkers for milk fever predisposition.

The identification of causal mutations and individual genes affecting periparturient hypocalcemia could have multiple benefits, including better understanding of the molecular mechanisms underlying this complex metabolic disorder, promote the development of new

drugs, therapies, and prevention strategies, and contribute to the design of novel breeding strategies. As such, the aim of this study was to use whole-genome sequence data to identify genetic variants associated with three alternative periparturient hypocalcemia traits, namely (i) total calcium concentration in the first 24 h postpartum, (ii) total calcium concentration in the first 72 hrs. postpartum, and (iii) the difference in total calcium concentration between 72 and 24 h postpartum. We also characterized the genes located near the most significant variants.

Materials and Methods

Phenotypic and genotypic data

The data consisted of 2,513 Holstein cows (938 primiparous and 1,575 multiparous) with blood calcium concentrations measured on the first, second, and third day after parturition. The samples were collected in five different experiments performed in 2 different dairy herds in the US from December 2015 to June 2020. Records with blood calcium concentration ≥ 3.0 mmol/L were excluded from the analysis based on distribution of data and because values above this threshold cannot be considered biologically feasible. After data editing, the mean (± SD) for blood calcium concentration was 2.126 (± 0.275) mmol/L, with a minimum and maximum of 0.675 and 2.99, respectively. All cows with blood calcium records had genotypic information for 79,060 SNP markers.

Alternative postpartum hypocalcemia traits

Three alternative postpartum blood calcium concentration traits were analyzed (Figure 1): (i) total calcium concentration in the first 24 h postpartum (mmol/L), (ii) total calcium concentration in the first 72 h postpartum calculated as the area under the curve (mmol/L/day) using the trapezoid function; and (iii) the recovery of calcium concentration calculated as the difference in calcium concentration between 72 and 24 h (mmol/L). Table 1 shows the descriptive statistics for these three alternative postpartum hypocalcemia phenotypes.

Imputation process

The 2,513 Holstein cows with postpartum blood calcium records and 79,060 SNP genotypes were imputed to whole-genome sequence in order to identify causal mutations associated with periparturient hypocalcemia. The imputation was performed in two steps, first from 79k to 312k markers, and then from 312k markers to whole-genome sequence.

From 79k to 312k markers.

The first imputation step aimed to provide a bridge between a medium density SNP chip and whole-genome sequence. As such, the 79k SNP genotypes were imputed to 312,615 SNP genotypes using 17,131 Holstein bulls born between 1995 and 2008 as reference population. The imputation process was performed using FImpute3 (Sargolzaei et al., 2014) using the default parameters.

From 312k markers to whole-genome sequence.

After the first imputation step, the imputed 312k markers were used as a new target for the imputation to whole-genome sequence. A total of 179 US Holstein bulls born in United States and part of the 1000 Bulls Genome project were used as reference panel. The imputation was performed again using Fimpute3, and each chromosome was processed separately.

Genomic scan

Only autosomal markers with a call rate > 0.9 and a minor allele frequency ≥ 1% for SNP densities 79k and 312k, and ≥ 0.01% for whole-genome sequence were retained for this analysis. After quality control, a total of 76,389, 300,621 and 11.6 million markers were available for the alternative genomic scans. The three alternative periparturient hypocalcemia phenotypes were analyzed using a two-step mixed-model-based approach (Aulchenko et al., 2007).

In the first step, the following model was fitted:

$$\mathrm{y} = \mathrm{Xb} + \mathrm{Zu} + \mathrm{e}$$

where y is the vector of periparturient hypocalcemia records, b is the vector of fixed effects, u is the vector of random animal effects, and e is the vector of random residual effects. The incidence matrices X and Z relate phenotypic records to fixed and animal effects, respectively. The two random effects were assumed to follow a multivariate normal distribution with $\mathrm{u} \sim N(0, \mathrm{G}\sigma_u^2)$ and $\mathrm{e} \sim N(0, \mathrm{I}\sigma_e^2)$, where σ_u^2 and σ_e^2 are the animal additive genetic and residual variances respectively, G is the genomic relationship matrix, and I an identity matrix. For the genomic scan using whole-genome sequence, matrix G was created using 50,196 SNPs randomly selected across the entire genome. The variance-covariance matrix for this first model was estimated as $\mathrm{V}_0 = \mathrm{ZGZ}'\sigma_u^2 + \mathrm{I}\sigma_e^2$.

In the second step, the following model was fitted for every SNP:

$$\mathrm{y} = \mathrm{X}\beta + X_{SNP}\beta_{SNP} + \epsilon$$

where X_{SNP} is the design matrix for the marker under consideration and β_{SNP} is the regression coefficient, also known as SNP effect. This model assumes that $\epsilon \sim N(0, \mathrm{V}_0\sigma_e^2)$. The significance of each SNP effect was evaluated using the following test statistic:

$$\mathrm{z} = \frac{\mathrm{X}'_{SNP}\mathrm{V}_0^{-1}(\mathrm{y} - \mathrm{X}\hat{\beta})}{\sqrt{\mathrm{X}'_{SNP}V_0^{-1}\mathrm{X}_{SNP}}}$$

which approximates the Wald test, and hence, is asymptotically standard normal. These analyses were performed using the R package MixABEL(Aulchenko et al., 2007). The *P*-values were

adjusted for multiple comparisons using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Statistical significance was declared using an adjusted *P*-values smaller than 0.05.

The assignment of significant genetic variants to bovine genes was based on the latest bovine genome reference ARS-UCD 1.2 using Ensembl (Cunningham et al., 2022). Genetic variants located upstream, downstream, or within annotated genes were considered.

Results

Total calcium concentration in the first 24 h postpartum

Five different genomic regions located on BTA3, BTA6, BTA10, BTA12 and BTA21 showed significant associations with calcium concentration in the first 24 h postpartum (Figure 2, Table 2). The most significant variant on BTA3 is upstream to the gene *SREBP1*, a major gene implicated in lipid metabolism. The most significant variant on BTA6 is localized between the genes *GC* and *NPFFR2*, which are implicated in the vitamin D signaling pathway and regulation of MAPK cascade, respectively. This genomic region on BTA6 also harbors gene *CXCL8*, which encodes for a pro-inflammatory cytokine, and gene *ADAMTS3*, which encodes a protease involved in the biosynthesis of collagen. Interestingly, both peaks on BTA3 and BTA6 were detected even using 79k and 312k SNP chips. Moreover, the most significant variant on BTA10 is located in an intron of gene *KCNK13*, a two-pore domain potassium channel that is regulated by extracellular calcium concentration. The most significant variant on BTA12 is located upstream of gene *NDFIP2*, which is involved in metal ion transport and positive regulation of protein ubiquitination. Finally, the most significant variant on BTA21 is located in an intron of gene *PRORP*, a member of the mitochondrial ribonuclease P complex involved in mitochondrial tRNA 5'-end processing. The major peaks detected on BTA10, BTA12, and BTA21 were detected only using whole-genome sequence data.

Area under the curve of total calcium concentration in the first 72 h postpartum

Six genomic regions located on BTA3, BTA6, BTA11, BTA12, BTA18, and BTA26 showed significant associations with the area under the curve of total calcium concentration in

the first 72 h postpartum (Figure 3, Table 2). The major peaks on BTA3 and BTA6, which harbor genes *SREBP1* and *GC*, respectively, were also identified as significantly associated with calcium concentration in the first 24 h postpartum. The most significant variant on BTA11 is located in an intron of gene *DENND1A*, a member of the connecdenn family, which is involved in endocytosis and protein transport. Two significant regions were detected on BTA12, one region harbors gene *CAB39L*, which encodes for a calcium-binding protein involved in energy stress, and the other region harbors gene *FNDC3A*, a transmembrane protein with RNA binding activity. The most significant region on BTA18 harbors several genes, including *JPH3*, *IRX3*, *BANP* and *CA5A*. The genes are implicated in calcium ion transport into cytosol, energy homeostasis, regulation of cell cycle, and one-carbon metabolism.

Recovery of calcium concentration

Three genomic regions located on BTA1, BTA9 and BTA12 showed significant associations with recovery of blood calcium concentration from first to third day postpartum (Figure 4, Table 2). The most significant variant on BTA1 is located in a copy number variant segment previously associated with the immune process. The most significant variant on BTA9 is upstream genes *SYNCRIP* and *SNX14*. Gene *SYNCRIP* encodes a member of the heterogeneous nuclear ribonucleoprotein family, RNA binding proteins that regulate alternative splicing, polyadenylation, and other aspects of mRNA metabolism and transport. Gene *SNX14* encodes a member of the sorting nexin family which is implicated in intracellular trafficking. The significant region on BTA12 has no genes currently annotated.

Discussion

We investigated the genetic basis of three interrelated postpartum blood calcium traits, namely calcium concentration in the first 24 h postpartum, area under the curve for total calcium concentration in the first 72 h postpartum, and the recovery of calcium concentration calculated as the difference in calcium concentration between 72 and 24 h. Periparturient hypocalcemia is a complex metabolic disorder, and the use of alternative phenotypes enables to capture different nuances in changes in postpartum blood calcium. In fact, Neves and collaborators have shown that the association of blood calcium concentration with cow performance varies depending on the timing of assessment during the early postpartum period (Neves et al., 2018a, 2018b).

Some of the most significant variants are located near or within genes directly implicated in calcium homeostasis. Indeed, one of the regions highly associated with calcium concentration in the first 24 h and also the first 72 h harbors gene *GC*. This gene encodes the vitamin D binding protein, which is responsible for the transport of most vitamin D3 metabolites in the plasma (Bouillon et al., 2020). Among vitamin D3 metabolites, 25-hydroxyvitamin D3 is the main circulating form in the plasma, and is converted into 1,25-dihydroxyvitamin D3, the biologically active vitamin D metabolite, in the kidney. When there is a decrease of blood calcium concentration, the parathyroid gland responds with increased secretion of parathyroid hormone, which upregulates the production of 1,25-dihydroxyvitamin D3. The 1,25-dihydroxyvitamin D3 stimulates renal reabsorption and gastrointestinal absorption of calcium and its effect is dependent on the blood ionized Ca concentration (Horst et al., 1994; Wilkens et al., 2020).

The genomic scans identified genes that are involved in calcium and potassium channels. For instance, gene *JPH3*, associated with total calcium concentration in the first 72 h postpartum,

encodes a member of the junctophilin family which is implicated in the physical approximation of plasmalemmal and sarcoplasmic/endoplasmic reticulum membranes. Junctophilins, such as *JPH3*, facilitate signal transduction in excitable cells between plasmalemmal voltage-gated calcium channels and intracellular calcium release channels (Lehnart and Wehrens, 2022). Gene *KCNK13*, associated with calcium concentration in the first 24 h postpartum, encodes a potassium channel containing two pore-forming domains. This channel is regulated by different factors, including arachidonic acid, halothane, and high extracellular calcium concentration(Enyedi and Czirják, 2010).

Genes implicated in energy metabolism were found to be associated with postpartum blood calcium concentration. For instance, gene *CA5A*, associated with area under the curve for total calcium concentration in the first 72 h postpartum, encodes a member of carbonic anhydrases, a large family of zinc metalloenzymes that catalyze the reversible hydration of carbon dioxide and participate in a variety of biological processes, including respiration, calcification, acid-base balance, and bone resorption. *CA5A* is localized in the mitochondria, it is expressed primarily in the liver, and it is involved in one-carbon metabolic process(Mani Urmila et al., 2022). Gene *PRORP*, also known as *MRPP3*, associated with calcium concentration in the first 24 h postpartum, encodes a mitochondrial RNA processing enzyme within the rNase P complex that is implicated in mitochondrial energy metabolism. Recent studies have shown that mutations on *PRORP* can reduce mitochondrial calcium, which in turn reduces insulin release from the pancreatic islet β cells, and reduced insulin secretion results in decreased insulin concentrations which contributes to imbalanced metabolism and insulin resistance (Rossetti et al., 2021). The region on BTA3 that showed significant associations with calcium concentration in the first 24 h and also 72 h postpartum harbors gene *SREBP1*. This gene encodes a

transcription factor that binds to the sterol regulatory element 1, a DNA motif that is found in the promoter of the low-density lipoprotein receptor gene and other genes involved in sterol biosynthesis. Note that there is a close link between negative energy balance and hypocalcemia in periparturient dairy cows. Indeed, negative energy balance has been demonstrated to decrease circulating calcium while the lack of intracellular calcium impairs carbohydrate metabolism, exacerbating the negative energy balance state (Chamberlin et al., 2013).

Periparturient cows experience significant immune dysregulation, and because intracellular calcium signaling is important for immune cell activation, the development of periparturient hypocalcemia contributes to periparturient immune suppression (Kimura et al., 2006). Interestingly, we found several genes associated with postpartum blood calcium that are directly implicated in the immune response. For instance, gene *IL12RB2*, associated with area under the curved for total calcium concentration in the first 72 h postpartum, encodes a subunit of the interleukin 12 receptor complex and plays an important role in Th1 cell differentiation (Prigione et al., 2016). Gene *CXCL8*, associated with calcium concentration in the first 24 h postpartum, encodes interleukin 8, a member of the CXC chemokine family and a major mediator of the inflammatory response. Interleukin 8 is secreted by mononuclear macrophages, neutrophils, eosinophils, and T lymphocytes, among others, and it functions as a chemotactic factor by guiding immune cells to the site of infection (Qazi et al., 2011). Gene *CCDC186*, associated with area under the curve for total calcium concentration in the first 72 h postpartum, is not yet well characterized but it appears it enables small GTPase binding, and it is involved in insulin secretion and response to bacterium.

Of special interest, the most significant variants associated with postpartum blood calcium traits are all located in non-coding regions, either upstream or downstream of genes or

within intronic regions. Causative mutations in non-coding regions such as regulatory elements (enhancers, insulators, and promoters) and untranslated regions (5'UTR and 3'UTR) typically affect the level of gene expression, whereas mutations on introns could affect both level of gene expression but also the structure of the protein. Interestingly, The Encyclopedia of DNA Elements (ENCODE) project, launched as a follow-up to the Human Genome Project with the goal of identifying and annotating functional elements of the human genome, revealed that many DNA variants associated with diseases lie within non-coding elements (Ecker et al., 2012).

Conclusions

We performed a comprehensive genomic analysis of three alternative postpartum blood calcium concentration traits in dairy cattle. The most significant variants were located within or near genes involved in calcium homeostasis and vitamin D transport, calcium and potassium channels, energy and lipid metabolism, and immune response, among other functions. These findings can contribute to the development of novel breeding and management strategies for reducing periparturient hypocalcemia in dairy cattle.

Table 1. Descriptive statistics for three alternative postpartum blood calcium profiles.

Trait	N	Mean	SD
Total calcium concentration in the first 24 h postpartum, mmol/L	2,513	2.03	0.29
Total calcium concentration in the first 72 h postpartum, AUC[1]	2,513	4.56	1.08
Recovery of calcium concentration[2]	2,513	0.21	0.30

[1] AUC = area under the curve in mmol/L/day from 24 to 72 h postpartum.

[2] Difference in blood calcium concentration (mmol/L) between 72 and 24 h postpartum.

Table 2. Most significant variants and candidate genes associated with three alternative postpartum blood calcium concentration traits.

Chr	Pos	P-value	Gene	Location	Process
Total calcium concentration in the first 24 h postpartum					
3	77.8	6.6e-10	*SREBP1*	Upstream	lipogenesis
6	87.2	1.5e-09	*GC*	Upstream	vitamin D transport
6	87.2	1.5e-09	*NPFFR2*	Inside	regulation MAPK cascade
6	87.2	1.5e-09	*ADAMTS3*	Downstream	biosynthesis of collagen
6	88.8	6.6e-07	*CXCL8*	Upstream	pro inflammatory cytokine
10	101.7	8.2e-07	*KCNK13*	Inside	potassium channel
12	54.5	4.1e-11	*NDFIP2*	Upstream	protein ubiquitination
21	45.5	1.6e-08	*PRORP*	Inside	energy metabolism
Area under the curve of total calcium concentration in the first 72 h postpartum					
3	77.9	8.8e-08	*SREBP1*	Upstream	lipogenesis
3	78.0	4.8e-08	*IL12RB2*	Inside	interleukin 12 receptor
6	87.2	1.0e-06	*GC*	Upstream	vitamin D transport
11	94.8	1.0e-06	*DENND1A*	Inside	membrane trafficking
12	18.9	1.6e-07	*CAB39L*	Downstream	calcium-binding protein
12	18.8	1.3e-06	*FNDC3A*	Inside	RNA binding activity
18	13.2	4.9e-08	*JPH3*	Inside	calcium ion transport
18	13.4	1.4e-06	*CA5A*	Inside	one-carbon metabolism
18	13.4	1.4e-06	*BANP*	Upstream	regulation of cell cycle
18	22.7	2.6-e06	*IRX3*	Upstream	energy homeostasis
26	34.8	5.9e-07	*CCDC186*	Inside	response to bacterium
Recovery of calcium concentration[1]					
9	63.6	4.6e-18	*SYNCRIP*	Upstream	mRNA metabolism
9	63.6	4.6e-18	*SNX14*	Upstream	intracellular trafficking

[1] *Difference in blood calcium concentration between 72 and 24 h postpartum.*

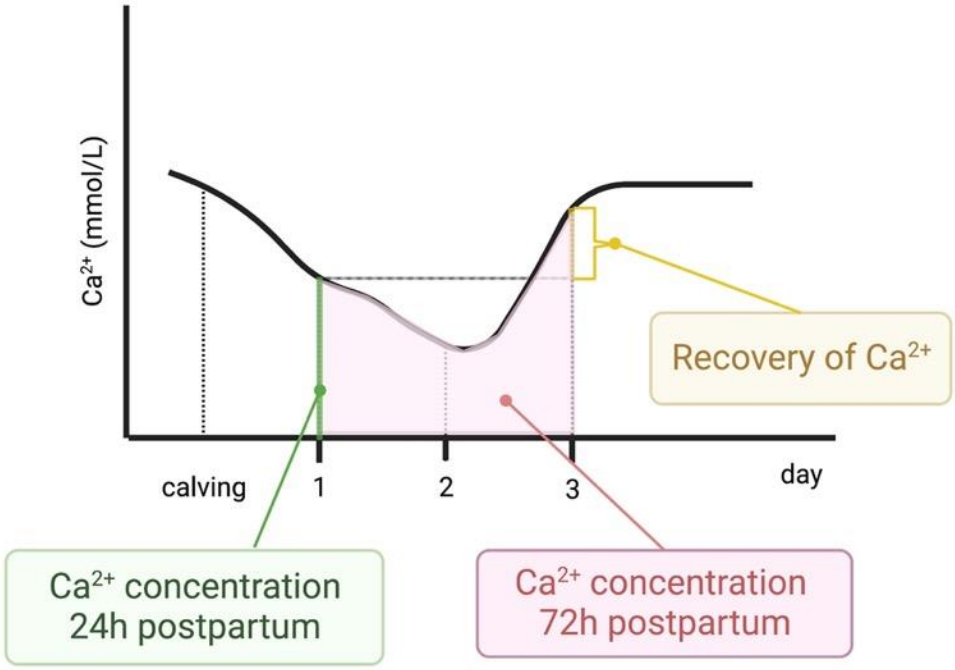

Figure 1. Three alternative traits phenotypes: total calcium concentration in the first 24 h postpartum, total calcium concentration in the first 72 h postpartum calculated as the area under the curve, and the recovery of calcium concentration calculated as the difference in calcium concentration between 72 and 24 hrs.

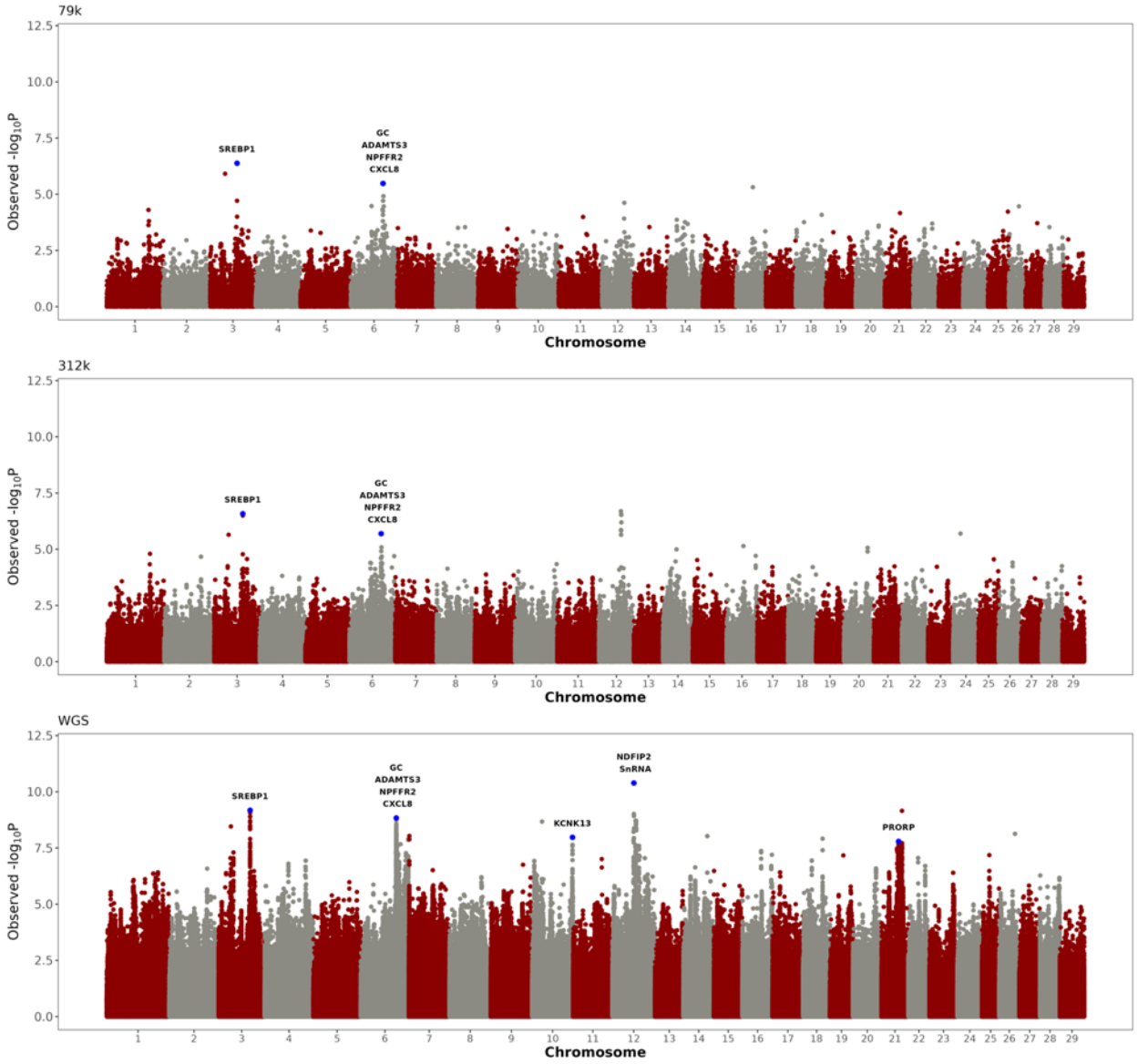

Figure 2. Genomic scans for total blood calcium concentration measured in the first 24 h postpartum. The scans were performed using three different SNP densities, namely 79k SNP, 312k SNP, and whole-genome sequence (WGS). Genes directly implicated in calcium homeostasis, calcium and potassium channels, energy metabolism, and immune response, are highlighted in the Manhattan plots.

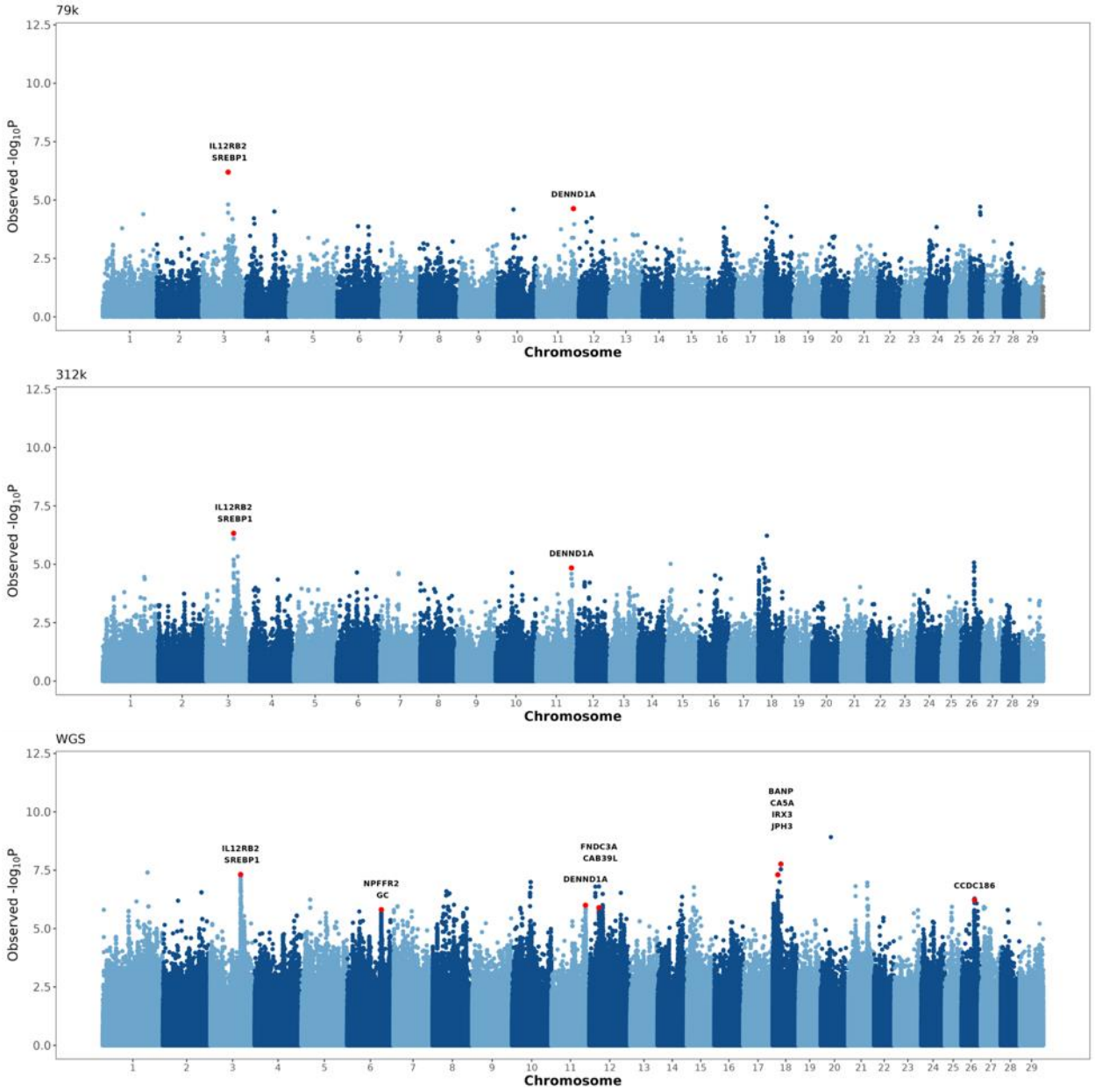

Figure 3. Genomic scans for total blood calcium concentration measure in the first 72 h postpartum. The scans were performed using three different SNP densities, namely 79k SNP, 312k SNP, and whole-genome sequence (WGS). Genes directly implicated in calcium homeostasis, calcium and potassium channels, energy metabolism, and immune response, are highlighted in the Manhattan plots.

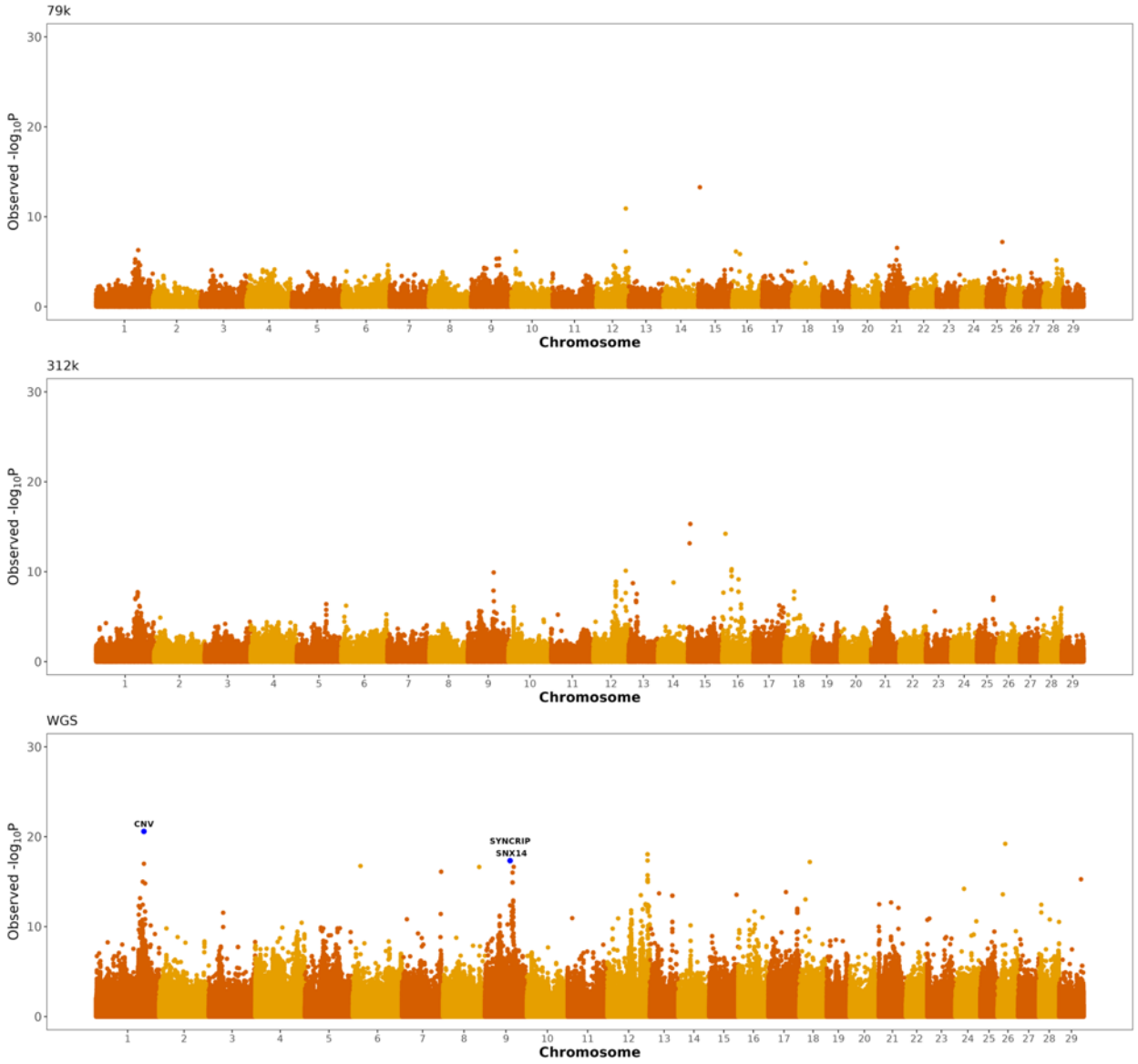

Figure 4. Genomic scans for the recovery of blood calcium concentration calculated as the difference in calcium concentration between 72 and 24 hrs. The scans were performed using three different SNP densities, namely 79k SNP, 312k SNP, and whole-genome sequence (WGS). Genes implicated in mRNA metabolism, transport, and intracellular trafficking are highlighted in the Manhattan plots.

References

Aulchenko, Y. S., Ripke, S., Isaacs, A., and van Duijn, C. M. (2007). GenABEL: An R library for genome-wide association analysis. *Bioinformatics* 23, 1294–1296. doi:10.1093/bioinformatics/btm108.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 (1995), Publi. *J. R. Stat. Soc.* 57, 289–300.

Bouillon, R., Schuit, F., Antonio, L., and Rastinejad, F. (2020). Vitamin D Binding Protein: A Historic Overview. *Front. Endocrinol. (Lausanne).* 10, 1–21. doi:10.3389/fendo.2019.00910.

Cavani, L., Poindexter, M. B., Nelson, C. D., Santos, J. E. P., and Peñagaricano, F. (2022). Gene mapping, gene-set analysis, and genomic prediction of postpartum blood calcium in Holstein cows. *J. Dairy Sci.* 105, 525–534. doi:10.3168/jds.2021-20872.

Chamberlin, W. G., Middleton, J. R., Spain, J. N., Johnson, G. C., Ellersieck, M. R., and Pithua, P. (2013). Subclinical hypocalcemia, plasma biochemical parameters, lipid metabolism, postpartum disease, and fertility in postparturient dairy cows. *J. Dairy Sci.* 96, 7001–7013. doi:10.3168/jds.2013-6901.

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., et al. (2022). Ensembl 2022. *Nucleic Acids Res.* 50, D988–D995. doi:10.1093/nar/gkab1049.

Ecker, J. R., Bickmore, W. A., Barroso, I., Pritchard, J. K., Gilad, Y., and Segal, E. (2012). ENCODE explained. *Nature* 489, 52–54. doi:10.1038/489052a.

Enyedi, P., and Czirják, G. (2010). Molecular background of leak K+ currents: Two-pore domain potassium channels. *Physiol. Rev.* 90, 559–605. doi:10.1152/physrev.00029.2009.

Horst, R. L., Goff, J. P., and Reinhardt, T. A. (1994). Calcium and Vitamin D Metabolism in the Dairy Cow. *J. Dairy Sci.* 77, 1936–1951. doi:10.3168/jds.S0022-0302(94)77140-X.

Kimura, K., Reinhardt, T. A., and Goff, J. P. (2006). Parturition and hypocalcemia blunts calcium signals in immune cells of dairy cattle. *J. Dairy Sci.* 89, 2588–2595. doi:10.3168/jds.S0022-0302(06)72335-9.

Lehnart, S. E., and Wehrens, X. H. T. (2022). the Role of Junctophilin Proteins in Cellular Function. *Physiol. Rev.* 102, 1211–1261. doi:10.1152/physrev.00024.2021.

Liang, D., Arnold, L. M., Stowe, C. J., Harmon, R. J., and Bewley, J. M. (2017). Estimating US dairy clinical disease costs with a stochastic simulation model. *J. Dairy Sci.* 100, 1472–1486. doi:10.3168/jds.2016-11565.

Mani Urmila, N., Kewalramani, D., Balakrishnan, U., and Manokaran, R. K. (2022). A case of carbonic anhydrase type VA deficiency presenting as West syndrome in an infant with a novel mutation in the CA-VA gene. *Epilepsy Behav. Reports* 20, 100573. doi:10.1016/j.ebr.2022.100573.

Neves, R. C., Leno, B. M., Bach, K. D., and McArt, J. A. A. (2018a). Epidemiology of subclinical hypocalcemia in early-lactation Holstein dairy cows: The temporal associations of plasma calcium concentration in the first 4 days in milk with disease and milk production. *J. Dairy Sci.* 101, 9321–9331. doi:10.3168/jds.2018-14587.

Neves, R. C., Leno, B. M., Curler, M. D., Thomas, M. J., Overton, T. R., and McArt, J. A. A. (2018b). Association of immediate postpartum plasma calcium concentration with early-lactation clinical diseases, culling, reproduction, and milk production in Holstein cows. *J.*

Dairy Sci. 101, 547–555. doi:10.3168/jds.2017-13313.

Pacheco, H. A., Da Silva, S., Sigdel, A., Mak, C. K., Galvão, K. N., Texeira, R. A., et al. (2018). Gene mapping and gene-set analysis for milk fever incidence in holstein dairy cattle. *Front. Genet.* 9, 1–8. doi:10.3389/fgene.2018.00465.

Prigione, I., Covone, A. E., Giacopelli, F., Bocca, P., Risso, M., Tripodi, G., et al. (2016). IL12RB2 Polymorphisms correlate with risk of lung adenocarcinoma. *Immunobiology* 221, 291–299. doi:10.1016/j.imbio.2015.10.006.

Pryce, J. E., Parker Gaddis, K. L., Koeck, A., Bastin, C., Abdelsayed, M., Gengler, N., et al. (2016). Invited review: Opportunities for genetic improvement of metabolic diseases. *J. Dairy Sci.* 99, 6855–6873. doi:10.3168/jds.2016-10854.

Qazi, B. S., Tang, K., and Qazi, A. (2011). Recent Advances in Underlying Pathologies Provide Insight into Interleukin-8 Expression-Mediated Inflammation and Angiogenesis. *Int. J. Inflam*. 2011, 1–13. doi:10.4061/2011/908468.

Rossetti, G., Ermer, J. A., Stentenbach, M., Siira, S. J., Richman, T. R., Milenkovic, D., et al. (2021). A common genetic variant of a mitochondrial RNA processing enzyme predisposes to insulin resistance. *Sci. Adv.* 7. doi:10.1126/sciadv.abi7514.

Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics* 15. doi:10.1186/1471-2164-15-478.

Sasaki, K., Yamagishi, N., Kizaki, K., Sasaki, K., Devkota, B., and Hashizume, K. (2014). Microarray-based gene expression profiling of peripheral blood mononuclear cells in dairy cows with experimental hypocalcemia and milk fever. *J. Dairy Sci.* 97, 247–258. doi:10.3168/jds.2013-7049.

Wilkens, M. R., Nelson, C. D., Hernandez, L. L., and McArt, J. A. A. (2020). Symposium review: Transition cow calcium homeostasis—Health effects of hypocalcemia and strategies for prevention. *J. Dairy Sci.* 103, 2909–2927. doi:10.3168/jds.2019-17268.

CHAPTER SIX: FEASIBILITY STUDY OF GENETIC EVALUATION FOR JOHNE'S DISEASE IN U.S. HOLSTEIN DAIRY CATTLE

Larissa C. Novo,[1,2*] Kristen L. Parker Gaddis,[2] Xiao-Lin Wu,[1,2] T. M. McWhorter,[2] Javier Burchard,[2] H. Duane Norman,[2] João Dürr, [2] Robert Fourdraine,[3] and Francisco Peñagaricano[1]

[1] Department of Animal and Dairy Sciences, University of Wisconsin, Madison, WI 53706, USA

[2] Council on Dairy Cattle Breeding, Bowie, MD 20716, USA

[3] Dairy Records Management Systems, Raleigh, NC 27603, USA

Abstract

Johne's disease (JD) is an infectious enteric disease in ruminants, causing substantial economic loss annually worldwide. This work aimed to investigate the feasibility of a national genetic evaluation for Johne's Disease susceptibility in Holstein cattle in the United States (U.S.). The data were extracted from Johne's disease data repository, maintained at the Council on Dairy Cattle Breeding (CDCB) and initially supplied by two dairy records processing centers. The data comprised 365,980 Holstein cows from 1,048 herds participating in the voluntary control program for JD. Two protocol kits, Idexx Paratuberculosis Screening Ab Test (IDX) and Parachek® 2 (PCK), were used to analyze milk samples with the Enzyme-Linked Immunosorbent Assay (ELISA) technique. Test results from the first five parities were considered. An animal was considered infected if it had at least one positive result; otherwise, it was uninfected. The overall average of JD incidence was 4.72% in these U.S. Hosltein cattle.

Genotypes of 78,964 SNP markers were extracted for 25,000 animals. Variance components and genetic parameters were estimated based on three models, namely, a pedigree-only threshold model (THR), a single-step threshold model (ssTHR), and a single-step linear model (ssLR). The posterior heritability estimates of JD susceptibility were low to moderate: 0.11 - 0.16 based on the two threshold models 0.05 - 0.09 based on the linear model. The average reliability of the estimated breeding values of JD susceptibility based on the three models varied from 0.18 (THR) to 0.22 (ssLR) for IDX and from 0.14 (THR) to 0.18 (ssTHR and ssLR) for PCK. Despite no prior direct genetic selection on JD, the estimated genetic trends of JD susceptibility were negative and highly significant (P-value < 0.01). Our results suggest that an official genetic evaluation of JD susceptibility is feasible in the U.S. Holstein cattle using ELISA-based tests, which can potentially reduce the JD incidence rate in the long term.

Key words: infectious disease, heritability, genetic trend, reliabilities

Introduction

Paratuberculosis, also known as Johne's disease (JD), is a chronic infectious disease caused by *Mycobacterium avium* ssp. *paratuberculosis* (MAP) that affects cattle and other ruminants. After infection, the MAP resides in macrophages in the intestines and triggers a chronic immune response resulting in inflammation of host intestine walls and potentially leading to death. Like tuberculosis in humans, infected animals silently carry the mycobacteria and spread the disease among herds (Whittington et al., 2017). Though the true herd prevalence of JD is difficult to assess, studies have reported up to 75% of herds with at least one positive case in the United States (Garcia and Shalloo, 2015). JD is commonly associated with increased premature culling, diminished production and reproduction, reduced feed efficiency, and increased veterinary costs and penalty payments due to reduced milk quality (Garcia and Shalloo, 2015). The economic losses caused by JD in the United States were estimated to be around $198 million per year (Rasmussen et al., 2021).

The epidemiology of JD is complex and depends on many factors such as herd management, age, infection dose, and disease prevalence. Clinically, this disease has a silent period lasting between 2 to 10 years, and test protocols have varying sensitivity levels, making JD-controlling programs arduous. Substantial research effort has been placed on vaccination for JD in the past decades, yet effects of vaccination are far from elucidated (e.g., Bannantine et al., 2014), Hence, the control for this disease currently relies on a voluntary program implemented by the US dairy producers (Davis and Park, 2018). Although JD is a heritable trait, the genetics

underlying host resistance remains largely unknown. As a result, indicator traits are often used to assist genetic programs in reducing JD incidence (Brito et al., 2018).

Throughout the last century, genetic evaluations in dairy cattle have heavily emphasized production over reproductive and health traits. Consequently, an antagonistic genetic relationship between production and health has become increasingly apparent (e.g., Pryce et al., 1998; Rauw et al., 1998). Meanwhile, the dairy industry has been shifting toward larger herd sizes and increasingly recognizing the importance of cows with strong resistance to diseases. Genetic improvement of disease resistance in dairy cattle has been an area of research for more than fifty years (reviewed by Parker Gaddis et al., 2020). Still, official genetic evaluations on large populations began considering health traits in the mid-1990s through indirect selection and indicator traits. In the U.S., for example, somatic cell score and productive life were combined with yield traits into a total net merit selection index (NM$; VanRaden and Wiggans, 1995); both representing indirect selection and indicator traits. Somatic cell score was a convenient indicator trait for clinical mastitis (VanRaden, 2017). Productive life provides an indirect indicator of overall health by estimating the length of time a cow remains in the milking herd, and hence a measurement related to the cow's health (VanRaden, 2017). However, direct recording and selecting health traits are more appealing because they can add value and make progress toward breeding healthier dairy cattle (Parker Gaddis et al., 2020). Over the past thirty years, there has been a shift in the selection index from yield traits to fertility, health, and fitness traits (Cole et al., 2020). Health traits such as hoof health (Cole et al., 2020) and stillbirth (e.g., Yao et al., 2014; Sigdel et al., 2022) are being considered. In the past decade, genetic evaluation and prediction on JD susceptibility have also been a topic of great interest (e.g., Zare et al., 2014; Kirkpatrick et al., 2022).

In the past few decades, the search for genes involved in the genetic determination of JD susceptibility has continued through candidate gene studies (e.g., Pinedo et al., 2009a, 2009b; Pant et al., 2011), QTL mapping (e.g., Gonda et al., 2007; Minozzi et al., 2010), and genome-wide association studies (GWAS; e.g., Settles et al., 2009; Kirkpatrick et al., 2011, 2022; van Hulzen et al., 2012; Zare et al., 2014; Kirkpatrick and Lett, 2018; Sanchez et al., 2022). These studies have linked JD with immune-mediated diseases, maturation of lymphocytes, mitochondrial health, and innate immune response that are also present in other infectious diseases in dairy cattle. However, genetic evaluations limited to significantly identified markers, linked to functional genes or QTL, are suboptimal in their prediction power. For example, Zare et al. (2014) reported an early study to predict JD in U.S. Jersey cattle using estimated SNP effects from genome-wide association studies in a dataset of approximately 5,000 mature cows. However, the accuracy measured by the receiver operating characteristic curve was low. Recently, Kirkpatrick et al. (2022) evaluated the efficacy of genomic prediction using a sire data set (897 sires with genotypes and daughters with phenotypes) in U.S. Holstein cattle. The average accuracy, assessed by the correlation of genomic-estimated breeding values (GEBV) for the testing tested sires and their daughter averages in 5-fold cross-validations, ranged between 0.43 and 0.53. Kirkpatrick et al. (2022) thus suggest that genomic predictions using genome-wide SNP markers for susceptibility to MAP infection could be used by producers in selecting AI service sires or replacement females as a means of producing a herd with lesser susceptibility to MAP infection. Still, the results were not representative of a full population genetic evaluation on JD resistance because genomic and proven sires tended to have higher prediction accuracies than other categories of animals (e.g., cows with/without genotypes and phenotypes or bulls without genotypes and daughters). Sanchez et al. (2022) reported genomic predictions with

relatively higher reliability for JD resistance (reliability = 0.55). These predictions were estimated from a reduced reference population that included about 75% of cows with phenotypes in a large Holstein population consisting of (161,253 animals which included 56,766 cows with Enzyme-Linked Immunosorbent Assay (ELISA) serological phenotypes and 12,431 animals with genotypes) raised in northwestern France.

Genetic evaluation of JD susceptibility (or resistance) depends on two factors. Firstly, collecting phenotypic and genotypic data can take many years. Secondly, given the available JD phenotypic and genetic data, it remains crucial to determine how accurate or reliable the genetic evaluations will be. In the past two decades, the JD records collected in U.S. Holstein cattle have accumulated substantially, thus providing a dedicated opportunity to assess a genetic evaluation on JD. As such, the primary objectives of this study were to characterize the available JD data repository in terms of incidence rate, variance compositions, and genetic parameters, and assess the reliability of estimated breeding values for JD susceptibilities in the U.S. Holstein cattle. The goal was to determine the feasibility of implementing an official genetic evaluation of JD susceptibility in Holstein cattle in the U.S.

Material And Methods

The data were extracted from the Johne's disease data repository maintained at the Council on Dairy Cattle Breeding, initially supplied by two dairy records processing centers. This data repository contained a total of 973,652 enzyme-linked immunosorbent assay (ELISA) scores. After removing redundant records, 713,395 ELISA scores were obtained representing 413,255 unique animals from 17 pure breeds and crossbred animals. The majority (92.38%) of these records were collected from Holstein cattle, and approximately 3.90% of the ELISA scores were collected from crossbred cattle. The remaining records were collected from Jersey (2.94%), Brown Swiss (0.36%), and other purebred cattle breeds (0.42%). The JD tests used two ELISA kits, namely, Idexx Paratuberculosis Screening Ab Test (Idexx, Montpellier, France) (IDX) and Parachek® 2 (Prionics AG, Schlieren, Switzerland (PCK). Individual JD records were collected in the states of Wisconsin (50.28%), Michigan (29.6%), Minnesota (18.48%), and the other 15 states (1.64%).

Considering the epidemiological profile of the disease, we kept only test results from the first five parities. An animal was classified as infected if it had at least one positive result, while animals without any positive test results were deemed uninfected. For animals with multiple positive (or negative) tests, only the first positive or last negative result was retained. The data was further cleaned by removing animals with inconclusive or suspect JD results and herds with less than 15 animals. In total, ELISA scores for 365,980 animals from 1,048 herds participating in the voluntary control program for JD were retained, covering the period from 1992 to 2022 with two protocol kits: IDX and PCK. The PCK milk analysis results represented 94,639 samples analyzed between 1992 and 2017, while IDX results represented 271,341 samples analyzed between 2002 and 2020. Genotypes of 78,964 SNP markers were extracted for 35,420 animals

with IDX and 170 animals with PCK test results. From the animals with IDX tests, only 25,000 genotyped animals, which were randomly selected, were included in the present study due to a limitation of the free software version (Lourenco et al., 2020) used for variance component estimation. Although the programs used to calculate breeding values could handle a larger number of genotyped animals, for consistency, the same genotyped animals were used for all analyses in this study. The pedigree consisted of 834,853 animals traced up to five generations.

The JD data were analyzed separately for the two JD diagnostic tests using the single-step approach implemented in the BLUPF90 suite of programs (Misztal et al., 2014) versions 3.16 and 2.47 for GIBBSF90+ and BLUPF90+, respectively. The PREGSF90 (version 1.4) program was used to obtain genomic and pedigree relationship matrices and to perform quality control of the genotype file. The quality control removed monomorphic markers, markers with minor allele frequency of less than 0.05, animals and markers with a call rate of less than 0.90, and animals with parent-progeny conflicts. After the data quality control, the dataset included 75,313 markers on 24,927 animals in the IDX-test category and 164 animals in the PCK-test category. The phenotypes of JD test results were coded as a binary (1 = infected, 0 = uninfected) trait.

Linear model

The linear model is described as the following:

$$y = \mathrm{X}\beta + \mathrm{Z}_1 h + \mathrm{Z}_2 u + e$$

where β is a vector of fixed effects for age within lactation (15 levels) and stage of lactation (3 levels), h and u are the vectors of random effects for herd-year-season of the tests (6,367 levels) and additive genetic, respectively, X, Z_1 and Z_2 are the corresponding incidence matrices, and e is a vector of residuals. To correct the age discrepancies of tests for each

lactation within herds, the age categories equally divided the animals into three groups (younger, average, and older; 33% of total animals, each) for each lactation. The groups of lactation stages (s) were defined as early (s < 150d), peak (150d ≤ s < 300d), and late (s ≥ 300d) lactation when the tests were conducted.

Threshold model

The threshold model postulates a continuous and normally distributed variable, namely liability (η_i), that delimits the observable binary trait (y_i) according to a threshold κ (Gianola, 1982). That is,

$$y_i|\eta_i, \kappa = \begin{cases} 1 & if\ \eta_i > \kappa \\ 0 & otherwise \end{cases}$$

The threshold κ is fixed arbitrarily to center the distribution, hence not an unknown parameter in a binary threshold model. The liability variable is then modeled by the linear model as is shown in (1) but with the observable vector $\boldsymbol{y}$ replaced by the unobservable liability vector $\boldsymbol{\eta}$. The conditional probability of observing a realization vector of $\boldsymbol{\eta}$, given a vector of observations (**y**) for a population of *n* animals, and the threshold κ, is the following (Wu et al., 2008):

$$Pr(y|\eta, \kappa) = \prod_{i=1}^{n}\{I(\eta_i \leq \kappa)Pr(y_i = 0) + I(\eta_i > \kappa)Pr(y_i = 1)\}$$

where I(A) is an indicator function, which takes the value one if condition *A* is true and a zero otherwise. The threshold model was analyzed by Bayesian analysis implemented via Markov chain Monte Carlo simulation (Korsgaard et al., 2003).

Estimations of variance components

Variance components and genetic parameters (e.g., heritabilities) were estimated using three models: a pedigree-only threshold model (THR), a single-step threshold model (ssTHR), and a single-step linear model (ssLR). In single-step methodology, the pedigree relationship matrix (A) is replaced by the realized relationship matrix (H; Legarra et al., 2009; Misztal et al., 2009). However, H is difficult to calculate. Therefore, Aguilar et al. (2010) proposed the direct construction of the inverse of the realized relationship matrix (H^{-1}) as:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix}$$

where G^{-1} is the inverse of the genomic relationship matrix for genotyped animals, and A_{22}^{-1} is the inverse of the pedigree relationship matrix for genotyped animals.

The GIBBSF90+ and BLUPF90+ programs (Misztal et al., 2014) were used to implement the Bayesian threshold and linear models, respectively. Convergence was determined by visual inspection of the trace plots and by the Geweke convergence diagnostic value (Geweke, 1992). The Geweke method assesses a convergence diagnostic for Markov chains based on a test for equality of the means of the first and last part of a Markov chain (by default, the first 10% and the last 50%). The test assumes that if the samples are drawn from the stationary distribution of the chain, then the two means are equal, and Geweke's statistic has an asymptotically standard normal distribution. The test statistic is a standard Z-score, the difference between the means of the two samples divided by its estimated standard error. The standard error is calculated from the spectral density at zero and thus considers any autocorrelation. The Markov chain Monte Carlo simulation consisted of 80,000 iterations after discarding a burn-in period of 15,000 iterations and thinned for every 100th sample for posterior inferences.

Breeding values and reliabilities

Breeding values, reliabilities, and genetic trends were obtained with univariate analysis for both JD tests and models using the BLUP90IOD2 and CBLUP90IOD2 programs (Tsuruta et al., 2001; Misztal et al., 2014). The THR model calculated traditionally estimated breeding values (EBV), whereas the ssTHR and ssLR calculated genomic estimated breeding values (GEBV) as these two methods include genomic information. The reliabilities were obtained from ACCF90GS as in:

$$REL = 1 - \left(\frac{PEV}{\sigma_u^2}\right)$$

Here, $PEV = var(\hat{u}_i - u_i) = c^{ii}\sigma_e^2$ is the prediction error variance of $\hat{u}_i$, where c^{ii} is the diagonal elements in the subblock of the inversed matrix of the left-hand side of the mixed model equation (MME) corresponding to the random animal effect for individual *i* and σ_e^2 is the residual variance, and $var(u_i) = \sigma_u^2$. Due to the large dataset, an additional analysis considering only the pedigree relationship was performed as a benchmark model for comparison.

Estimation of genetic trends

Genetic trends were estimated with the breeding values of the JD incidence by adjusting the two diagnostic test results to a common base, assuming that the differences in the JD incidence rate between the two JD tests were essentially non-genetic. Hence, we fitted a temporal linear regression on the estimated JD breeding values, assuming heterogenous intercepts and a common regression coefficient for the two tests. That is,

$$\bar{u}_{kj} = a_k + b\left(t_{kj} - \bar{t}_k\right) + \epsilon_{kj},$$

where $\bar{u}_{kj}$ is the mean of breeding values computed in year j for the kth test (i.e., $k = 1$ for PCK tests and $k = 2$ for IDX tests), t_{kj} is the year of the test (i.e., 1992 – 2017 for PCK tests and 2002 – 2020 for IDX) with $\bar{t}_k$ being the corresponding mean, a_k is the heterogenous intercept, b is the common coefficient pertaining to the two tests, and ϵ_{kj} is an error term. The model parameters were solved by least squares. In brief, the solutions of the model parameters are the following:

$$\begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{b} \end{pmatrix} = \begin{pmatrix} n_1 & 0 & \sum_{j=1}^{n_1}(t_{1j} - \bar{t}_1) \\ 0 & n_2 & \sum_{j=1}^{n_2}(t_{2j} - \bar{t}_2) \\ \sum_{j=1}^{n_1}(t_{1j} - \bar{t}_1) & \sum_{j=1}^{n_2}(t_{2j} - \bar{t}_2) & \sum_{j=1}^{n_1}(t_{1j} - \bar{t}_1)^2 + \sum_{j=1}^{n_2}(t_{2j} - \bar{t}_2)^2 \end{pmatrix}^{-1} \times \begin{pmatrix} \sum_{j=1}^{n_1} \bar{u}_{1j} \\ \sum_{j=1}^{n_2} \bar{u}_{2j} \\ \sum_{j=1}^{n_1} \bar{u}_{1j}(t_{1j} - \bar{t}_1) + \sum_{j=1}^{n_2} \bar{u}_{2j}(t_{2j} - \bar{t}_2) \end{pmatrix}$$

where n_1 and n_2 are the numbers of years having IDX and PCK test results, respectively.

Results And Discussion

Data summarization and characterization for Johne's disease diagnostic tests

A major limitation in the development of direct genetic measures for health traits over the past decades has been the lack of a centralized system to collect health data at a national level (Gaddis et al., 2020). Zwald et al. (2004) demonstrated that producer-recorded health event data were a viable source of phenotypes for genetic evaluation if collected into a national data set. In the U.S., the Animal Improvement Program Laboratory (now the Animal Genomics and Improvement Laboratory, AGIL; Beltsville, MD) has collaborated with industry partners and veterinary experts to develop a standardized health recording system since 2008 (Cole et al., 2008) and advance the development of health evaluations. On the other hand, the standardization of health traits may overlook the discrepancy between JD objective diagnosis and producer-reported occurrences of other traits such as ketosis, milk fever, and others. The CDCB JD data repository was established to collect phenotypes and genotypes that could be used for genetic evaluations on JD.

The CDCB JD data repository represented the world's largest one in Holstein cattle. After data quality control, ELISA records for 365,980 animals from 1,048 herds participating in the voluntary control program for JD tests were retained. Compared to the studies conducted a decade ago (e.g., Gonda et al., 2006), the JD data size has increased by over a hundred folds. The dataset used by Kirkpatrick et al. (2020) partially overlapped with the dataset used in this study, including only data collected up to 2016, which fell into three categories. The data in the first category represented the U.S. commercial dairy cattle population from 2010 to 2016 with farm records for MAP infections (i.e., milk ELISA test results for US Holstein cows) provided by the

Dairy Records Management System and AgSource Cooperative (Kirkpatrick and Lett, 2018). They retained JD records for 202,367 cows after the data editing. The other two categories included testing cows from commercial dairy herds developed as resource populations at the University of Minnesota and Michigan State University. In contrast, the dataset used in this study contained twice as many animals. Recently, Sanchez et al. (2022) merged two JD datasets to evaluate the genetic resistance to paratuberculosis in French Holstein cattle. The first dataset contained non-redundant MAP statuses (serum ELISA and fecal PCR results) for 4,100 cows, while the second dataset included non-redundant MAP statuses of 243,274 cows from 15,476 herds, deduced from serological tests routinely recorded since 2015. After the data cleaning, the merged dataset comprised 161,253 animals, including 56,766 cows with ELISA serological phenotypes and 12,431 animals with genotypes. Approximately three-fourths of the serum samples were diagnosed with IDX, whereas around one-fourth were diagnosed with Idvet ID Screen ® Paratuberculosis Indirect (Idvet, Montpellier, France).

The records in the CDCB JD repository are results from two diagnostic tests, IDX and PCK. Figure 1 shows the yearly distribution of the number of these two tests reported through the voluntary program. The number of IDX tests increased substantially after 2012 due to the rising use of IDX as the primary test and the USDA's subsidies provided for testing. The overall JD positive rate was 2.06% with IDX tests and 6.7% with PCK tests, reflecting the differences between the two tests. Still, such a difference might also be confounded by the non-zero yearly phenotypic or genetic trends due to the switch of JD tests over the years. This question was addressed in the later section with estimated phenotypic and genetic trends. The averages age of cows with positive results did not change considerably between the two tests (53.6 ± 16.24 months for IDX vs. 52.65 ± 16.31 months for PCK), excluding the age of cows at testing as an

influential factor for the different JD positive rates between the two tests. The JD positive rates varied substantially in lactations (Table 1). For the PCK tests, the average JD-positive rate increased steadily from 4.8% at lactation 1 to 9.1% at lactation 3 and then declined afterwards (6.3 – 7.9%). For the IDX tests, the average JD positive rate increased from 1.8% at lactation 1 to 3.8% at lactation 4 and declined afterward (3.4 – 4.3%). Overall, the JD positive rate per lactation was higher with PCK compared to IDX. The weighted average of JD-positive rates between the two tests increased drastically from lactation 1 (1.8%) to 4 (5.0%) and did not change substantially for lactation five and beyond. For either test, the overall positive rate at the fourth lactation was approximately twice as much as that at the first lactation. Also due to the temporal switch of JD tests, the number of genotyped animals with PCK tests was far less than that with IDX tests, because PCK kits were primarily used before 2010, when high-throughput genotyping was less popular.

Variance components and heritability estimates

Variance components and heritability estimates were obtained using three models: a pedigree-only threshold model (THR), a single-step threshold model (ssTHR), and a single-step linear model (ssLR), respectively (Table 2). With ssLR, we estimated the genetic variance to be between 0.001 and 0.005, the herd-year-season variance between 0.001 and 0.003, and the residual variance between 0.02 and 0.05, respectively, for both tests. The estimated heritability was 0.05 for IDX and 0.09 for PCK. The two threshold models postulated a liability variable underlying the binary JD test outcomes (positives vs. negatives) when fixing the residual variance to 1.0. These two threshold models estimated the genetic variances to be between 0.16 and 0.25 and the herd-year-season variance between 0.18 and 0.25. The heritability estimates

were 0.13 and 0.17 for IDX and PCK, respectively. While ssLR had lower variance components than the two threshold models, they had proportionally higher residual variances. Hence, the estimated JD heritability with ssLR was lower than those obtained with the two threshold models. Overall, our heritability estimates for JD susceptibility (or liability to JD susceptibility) are within the range (from 0.04 to 0.18) of previous heritability estimates for serum ELISA phenotypes measured in various Holstein populations in the literature (e.g., Gonda et al., 2006; Hinger et al., 2008; Berry et al., 2010; Shook et al., 2012). Note that the previous studies mainly reported within-herd heritabilities, whereas ours were cross-herd heritability estimates. If the herd-year-season variance component is removed from the denominator of the heritability formula, the within-herd heritability estimates would be slightly higher (i.e., 0.05 - 0.20).

The two threshold models gave similar heritability estimates, and adding genomic information did not significantly alter the posterior variance components. However, the estimated heritability for PCK test results was higher than that for IDX test results based on all three models. This difference could result from the intrinsic differences between the diagnostic tests that possibly captured genetic and environmental variations differently. Still, the impact of sampling sizes could also matter. In the present study, the number of animals with IDX tests was 2.87 times as many as those with PCK tests. Statistically, a larger sample size leads to more precise estimates. Hence, the standard errors of heritability estimates for the IDX tests were significantly smaller than those for the PCK tests (Table 2). The heritability estimates obtained from the two threshold models were higher than the linear model for the reason explained before. Somewhat differently, Gonda et al. (2006) reported higher heritability estimates with linear models than with threshold models. This difference could arise from varied JD incidence rates between studies. The JD incidence rates in this study were 2.8% for IDX tests, 6.6% for PCK

tests, and 3.8% as a weighted average, respectively, which were considerably lower than the JD incidence rates (7.7%, ranging from 4.7% to 12.4%) in the Holstein cows used by Gonda et al. (2006). For a ratio variable, the variance depends on its mean (i.e., average incidence rate) such that the variance of a linear binary trait decreases substantially as the incidence rate drops. For example, let the phenotypic variance for a 50:50 (JD positives vs. negatives) incidence rate be 100%. Then, a 20% positive incidence rate would retain 64% of the phenotypic variance. When the positive incidence rate dropped to 7.7%, it could retain 28.4% of the phenotypic variance. However, when the positive incidence rate fell further to 2.9%, it could only retain 11.3% of the phenotypic variance. The smaller the JD incidence rate, the more the JD phenotypes deviate from the normal distributions. Possibly, the linear model did not effectively capture the variation of the phenotypes. Instead, our results showed that a threshold model could perform better than a linear model to effectively capture the variance components particularly when the disease incidence rate was low. Hence, given the binary nature of these traits, a threshold model can be more appropriate when the incidence rate in the minor category is low (Lynch and Walsh, 1998).

Kirkpatrick and Lett (2018) also reported higher heritability estimates with threshold models (0.157 - 0.186) than linear models (0.041 - 0.062) using a dataset that partially overlapped with the dataset in this study, which is consistent with our findings. They showed that heritability estimates increased as data were restricted to herds with presumed higher MAP exposure for the linear model. In another study, van Hulzen et al. (2011) analyzed log-transformed ELISA test results (percentage S/P + 50) as the dependent variable in a linear mixed-effects model. They showed that the phenotypic variance increased concurrently with an increase in within-herd test prevalence. Additionally, they observed an increase in heritability due to increased genetic variance relative to phenotypic variance. Attalla et al. (2010) reported

higher heritability estimates for milk ELISA test results with the linear model ($h^2 = 0.08$) than the threshold model ($h^2 = 0.065$) when ignoring the maternal genetic effects. Still, they found lowered heritability estimates with the linear model ($h^2 = 0.075$) than with the threshold model ($h^2 = 0.095$) when the genetic variance included the maternal genetic variance. Based on these results, Attalla et al. (2010) stated that there was no solid evidence to recommend a linear model over a threshold model for genetic evaluation. Note that the linear trait in their analysis is not binary scores per se but the natural log of the optical density results obtained with a PCK diagnostic test.

The herd-year-season variances are similar to the additive genetic variance for all models, but can vary substantially depending on actual model settings. Factors such as herd size and distribution may interfere with its variance. For JD, grouping the animals by their herd-year-season combinations is conceptually different. The transmission of the disease is intrinsically linked to the environment and is known to be more active in larger herds (> 200 cows; Wolf et al., 2014; Corbett et al., 2018). While most studies do not report herd or herd-year-season sizes, and when present, the effect is included in the model as fixed, thus, direct comparisons between our study and previous ones may not be accurate. To the best of our knowledge, this is the largest data set to evaluate genetic parameters for JD. The total number of herds was larger than those reported in the literature, but the average herd size in this study was 349 (max = 19,351 animals). As larger herds tend to have more active transmission, we believe that the number of herds and herd size may display lower variance between groups. The fact that 97% of the animals originated from only three states with similar weather and management practices may contribute to less variable herd incidence. By adding year-season as a grouping factor, more variance is expected to be captured by the variable. For JD, however, changes in the prevalence of the

disease across time are not common, and once infected, the herd tends to continue to be positive (Corbett et al., 2018).

Reliabilities of (genomic) estimated breeding values

The reliabilities of (genomic) estimated breeding values for JD susceptibility based on the linear model (or liability to JD based on the threshold models) were generally low to moderate (Table 3). When considering all animals in the pedigree, the overall reliability means were 0.18 (THR), 0.21 (ssTHR), and 0.22 (ssLR) for IDX, and 0.14 (THR), 0.18 (ssTHR), and 0.18 (ssLR) for PCK. The average reliabilities of JD susceptibility in the present study were comparable to previously reported reliabilities for metabolic diseases (14.6 – 33.7%) but slightly lower than the genomic reliabilities for mastitis (65.8 – 76.6%) (Parker Gaddis et al., 2020). Sanchez et al. (2022) reported high (0.55) genomic reliabilities for JD resistance in Franch Holstein cattle. However, their results are not directly comparable to ours because, conceptually, JD resistance is not precisely the same as JD susceptibility. When coded as a binary trait, for example, the phenotypic coding for JD susceptibility has "1" standing for a positive and "0" standing for a negative phenotype, whereas the phenotypic coding for JD resistance takes the opposite values (1 = negative and 0 = positive). Hence, a low JD susceptibility indicates a high JD resistance. Besides that, Sanchez et al. (2022) used an approximation method to calculate reliabilities that differed from ours.

Including genomic information led to higher overall average reliabilities (0.21 for IDX and 0.18 for PCK) with the single-step threshold, compared to the pedigree-based threshold model (0.18 for IDX and 0.14 for PCK). These reliabilities were obtained based on animal models, which may differ from a sire model or a sire-maternal-grandsire model. Sire models are

widely used in genetic evaluations. Still, they may underestimate or overestimate estimated breeding values (EBVs) because assumptions such as no genetic relationships between sire and dam, and random mating are usually not met in dairy cattle breeding. In contrast, an animal model can provide more stability (Sun et al., 2009).

The distributions of reliabilities of JD with the IDX tests are shown in Figures 2 (IDX) and 3 (PCK). Based on the THR model, most animals had low to moderate (< 0.5) reliabilities, yet there were 0.5% of animals with 0.50 reliability or higher. With the ssTHR model, including genomic information led to a slightly larger percentage (4.1%) of animals with reliabilities higher than 0.50. The ssLR model had a similar pattern of reliability group distributions except fewer animals with reliabilities greater than 0.50. For the PCK tests, a considerable percentage (55%) of the animals had low (< 0.2) reliabilities. Again, including genomic information in the single-step models (ssTHR) led to more animals with reliabilities 0.20 or more compared to THR. Around 0.31% of the animals had reliabilities of 0.50 or higher with ssTHR, whereas the same category only accounted for 0.001% (THR) of the animals.

Furthermore, the reliabilities were calculated by subgroups of animals, that is, cows with their own records, sires (dams) with records measured in their daughters, and sires of cows with no phenotype (Table 3). Overall, the EBV of the liabilities to JD for sires were higher on average than those for cows or dams. As expected, sires of cows with phenotypes had higher reliabilities than sires of cows with no phenotypes. Across models, sires of cows with no phenotypes for IDX seemed to be under the more significant impact (≅ 4% of the change in reliability), suggesting that the shift in overall reliability observed in Figure 2 may be linked with this category. For PCK, on the other hand, only dams of cows with phenotypes appeared not to be influenced by the change in models and a more stable category. In contrast, other categories, especially cows

with phenotypes, were sensitive to both different phenotype formats and the inclusion of genotypes in the evaluation. In other words, the shift of animals from 0.05 – 0.1 to 0 – 0.05 category between THR and ssTHR models, for instance, could result from an impact on all subgroups collectively.

The addition of genomic information in a genetic analysis aims to better estimate breeding values for animals that lack genotypes through an improved relationship matrix. Although for some categories (i.e. sires of cows with phenotype) there was a minor reduction in reliability with the addition of genomic information, sires of cows with no phenotype had, had considerably higher reliabilities. Conceptually, this category is the one to have more impact of the single-step approach. As sires of cows with no phenotypes have a more distant connection to the phenotypes per se and are connected to it only by the relationship matrix, the blending of genomic and pedigree information is a more substantial source change than for the other categories.

Between the two JD tests, the reliabilities of JD breeding values were higher with the IDX tests than those with the PCK tests because the sample sizes (and the family sizes) with the IDX tests were larger than those with the PCK tests. By adding genomic information in the single-step model (ssTHR), an increase in the reliability for all categories was expected. Yet, the single-step approach did not significantly change reliability values when the trait has many phenotypes. These results coincided with the observations on feed intake by Harder et al. (2020). For the IDX and PCK tests, the evaluations on JD susceptibility did not vary substantially, whether using pedigree information or including genomic information. Regarding genomic prediction, Kizilkaya et al. (2014) showed that the accuracies for ordinal categorical scores

analyzed by the Bayes Cπ threshold model were 20% to 50% lower as much as the accuracies obtained using a linear model.

Reliability corresponds to the correlation between the estimated and actual breeding values of the animals. Consequently, it would be expected that models with higher heritability values will also result in higher reliabilities. However, the results presented in this work partially contradict this assumption. Despite using the same dataset, the threshold models presented higher heritabilities while higher reliabilities were derived from the linear model. It is important to note that regarding these two types of models, direct comparisons should be considered carefully as they are in different scales (liability and linear). The effect of different scales on the comparing of reliabilities of the same animals has been reported with higher values for linear models as a consequence (Tsuruta et al., 2017).

Genetic trends

Annual averages of breeding values for JD susceptibility (ssLR) or liabilities to JD susceptibility (ssTHR) showed temporal decreasing trends. For example, with the IDX tests, the annual means of JD breeding values increased during 2002 and 2004, remained relatively stable from 2008 to 2009, and dropped after 2009 (Figure 4). Genetic trends of JD were estimated jointly for the two JD test results obtained from linear (ssLR) and threshold (ssTHR) single-step models, respectively (Table 4), assuming that the differences in observed JD incidences between the two JD tests were non-genetic. The genetic trends were relatively flat from 1994 to 2009 for PCK tests (Figure 6) and from 2002 to 2009 for IDX tests (Figure 5) and began to drop afterward (Figure 5). The extent of the decrease in genetic trend for IDX tests after 2009 was more significant with the single-step threshold model (ssTHR) than with the single-step linear model

(ssLR). In contrast, PCK test results did not show apparent changes in the genetic trend over the years. The visual fluctuations in the genetic trends between 2009 and 2017 were attributable to fewer data analyzed. There were between 5 and 124 PCK records per year for this period, which were insufficient to accurately estimate yearly means of JD incidence rate, thus leading to inaccurately estimated genetic trends.

Although there is no official genetic evaluation and selection program implemented for Johne's disease in the United States, this study observed significantly (P-value < 0.01) negative, favorable trends based on the EBV/GEBV obtained from the linear (ssLR) and the threshold (ssTHR) models, respectively. The estimated genetic trends from the EBV/GEBV obtained from both models are not comparable because they were not assessed on the same scales. The threshold model estimated (genomic) EBV on the unobserved liability scale, whereas the linear model estimated (genomic) EBV was on the observed binary scale. Still, the results from both models consistently suggested a negative, favorable trend on the IDX test results for JD, likely starting around 2004. Coincidentally, the genomic selection program in the U.S. Holstein cattle began in 2009 and accelerated after 2010 (Wiggans et al., 2017). The introduction of genomic selection in the U.S. dairy cattle improvement programs in 2008 has led to a significant increase in the rates of genetic gain, particularly for traits with low heritabilities, such as fertility, longevity, and udder health (García-Ruiz et al., 2016). Possibly, the negative, favorable trend can be attributed to indirect genetic gain from genomic selection of highly genetically correlated traits, such as milk yield, somatic cell score, and lifespan. For example, Neupane et al. (2021) showed that moderate correlation with productive life helped decrease replacement heifer cost resulting from the correlated response in heifer livability and assisted in genetic improvement programs through added selection intensity resulting from beneficial genetic correlations. It is

possible that the favorable, negative genetic trend was associated with net merit (NM) used in the U.S. cattle genetic evaluations. NM is a selection index used to identify animals that are expected to have the greatest genetic merit for a combination of economically important traits. It takes into account multiple traits, including milk yield, fat and protein yield, somatic cell score, productive life, and daughter pregnancy rate. The weights assigned to these traits in the calculation of NM are determined based on their relative economic importance to the dairy industry. Recently, Sanchez et al. (2022) reported a slightly favorable genetic trend in resistance to JD in Holstein cattle raised in northwestern France over the last two decades, also possibly resulting from the indirect selection because of a low positive genetic correlation (0.06) between resistance to JD and total merit index.

Conclusion

Johne's disease records have been collected in the U.S. since 2002, which consisted of test results with two kits, IDX and PCK. The records have accumulated substantially over the past two decades. The addition of genotypes has enabled the use of single-step genetic evaluation to assess JD susceptibility. The overall positive rate of JD in the U.S. Holstein cattle was 2.80% with the IDX tests and 6.64% with PCK tests. Estimated heritabilities were 0.05 to 0.09 for JD susceptibilities on the observed binary scale based on the single-step linear model, and 0.12 to 0.17 for the liabilities to JD on the unobserved liability scale based on the two threshold models. Our results assert that Johne's disease has a genetic component upon which selection could operate to achieve the expected genetic improvement. The reliabilities of (genomic) breeding values for JD susceptibility (or liability to JD) were low to moderate. There were approximately 0.001 to 4.1% of the animals with > 0.5 reliabilities. Thus, implementing routine genetic evaluations of Johne's disease in the U.S. Holstein dairy population is technically feasible in terms of data sufficiency and statistical model capability, and direct selection on JD will further reduce JD infection in the long run. The protocol used for diagnosing JD should be considered because varied protocols led to systematic biases in the estimated JD incidence rate. The health statuses of animals can be modeled linearly as a binary variable or described by the underlying continuous liability variable. Adding genomic information resulted in elevated reliabilities of estimated genomic breeding values. The present study utilized animal models in three forms, allowing all relatives to contribute to the assessment of an animal, and permitting a cow to be evaluated based on its own phenotypes. Compared to a sire model or S-MGS model, an animal model is more computationally intensive but can offer greater accuracy and more flexibility.

Finally, this preliminary study aimed to leverage the current data repositories toward the official implementation of genetic evaluation of Johne's disease in the Holstein dairy breed. Follow-up studies are expected as more JD phenotypes and genotypes become available.

Table 1. Overall Johne's disease incidence rate (positive % versus negative %) and positive rate (%) by lactation observed in the U.S. Holstein cattle[1]

Diagnostic test	N	Overall		Positive rate (%) by lactation				
		Negative %	Positive %	1	2	3	4	5+
PCK	94,639	93.35	6.64	4.83	7.56	9.11	7.91	6.30
IDX	271,341	97.20	2.80	1.75	2.80	3.84	4.32	3.59
Weighted average	-	96.18	3.82	1.76	3.17	4.48	5.04	5.17

[1] PCK: Parachek tests; IDX: Idexx tests; Weighted average: average of results weighted by the number of records per diagnostic test.

Table 2. (Co)variance components, heritability estimates, and standard deviations for Johne's disease (JD) susceptibility (or liabilities to JD susceptibility) in U.S. Holstein dairy cattle diagnosed by two different test kits: Idexx and Parachek. [1,2]

	THR		ssTHR		ssLR	
	IDX	PCK	IDX	PCK	IDX	PCK
h^2	0.13 (0.01)	0.16 (0.12)	0.12 (0.003)	0.17 (0.01)	0.05 (0.003)	0.09 (0.008)
σ^2_g	0.17 (0.02)	0.23 (0.03)	0.16 (0.02)	0.25 (0.05)	0.001 (0.00-09)	0.005 (0.0002)
σ^2_{hys}	0.18 (0.01)	0.24 (0.01)	0.18 (0.009)	0.25 (0.02)	0.001 (0.0004)	0.003 (0.0005)
σ^2_e	1 (0.003)	1 (0.003)	1 (0.003)	1 (0.003)	0.02 (0.0009)	0.05 (0.0005)

[1] PCK: Parachek tests; IDX: Idexx tests;

[2] Variance and covariance components were estimated using three statistical models: a pedigree-based threshold model (THR), a single-step threshold model (ssTHR), and a single-step linear model (ssLR).

[3] Numbers in parentheses are standard deviation.

[4] h^2 = heritability; σ^2_g = genetic variance; σ^2_{hys} = herd-year-season variance; σ^2_e = residual (error) variance.

Table 3. Reliability estimates (standard errors) of genomic estimated breeding values for Johne's disease susceptibility in U.S. Holstein dairy cattle [1,2]

	N		THR		ssTHR		ssLR	
	IDX	PCK	IDX	PCK	IDX	PCK	IDX	PCK
Overall	638,316	247,031	0.18 (0.14)	0.14 (0.11)	0.21 (0.12)	0.18 (0.13)	0.22 (0.14)	0.18 (0.13)
Cows with phenotype	271,341	94,639	0.31 (0.07)	0.30 (0.08)	0.31 (0.11)	0.27 (0.08)	0.30 (0.11)	0.25 (0.10)
Sires of cows with phenotypes	13,000	8,658	0.41 (0.17)	0.35 (0.15)	0.38 (0.17)	0.32 (0.14)	0.41 (0.15)	0.36 (0.13)
Dams of cows with phenotypes	196,846	60,212	0.24 (0.10)	0.21 (0.11)	0.23 (0.11)	0.19 (0.11)	0.25 (0.12)	0.20 (0.12)
Sires of cows with no phenotype	25,693	14,927	0.18 (0.13)	0.18 (0.15)	0.20 (0.14)	0.17 (0.11)	0.22 (0.14)	0.21 (0.18)

[1] Johne's disease was diagnosed with two test kits: Parachek tests (PCK) and Idexx tests (IDX);
[2] Reliabilities were calculated for (genomic) estimated breeding values based on three statistical models: a pedigree-based threshold model (THR), a single-step threshold model (ssTHR), and a single-step linear model (ssLR)

Table 4. Estimated intercepts and slopes from linear regression of genetic trends for genomic breeding values of Johne's disease obtained from a threshold and a linear single-step models [1,2]

	Heterogeneous intercept		Genetic trend (slope)			
Model	IDX	PCK	Estimate	SE	t-value	Probability
ssTHR	-0.0918 (0.0190)	-0.1440 (0.0163)	-0.0091	0.0018	-5.226	<0.001
ssLR	-0.005 (0.0016)	-0.1149 (0.0014)	-0.0006	0.0001	-3.810	<0.001

[1] Johne's disease was diagnosed with two test kits: PCK: Parachek and IDX: Idexx;
[2] Genetic trends were estimated for annual means obtained from two statistical models: a single-step threshold model (ssTHR) and a single-step linear model (ssLR).

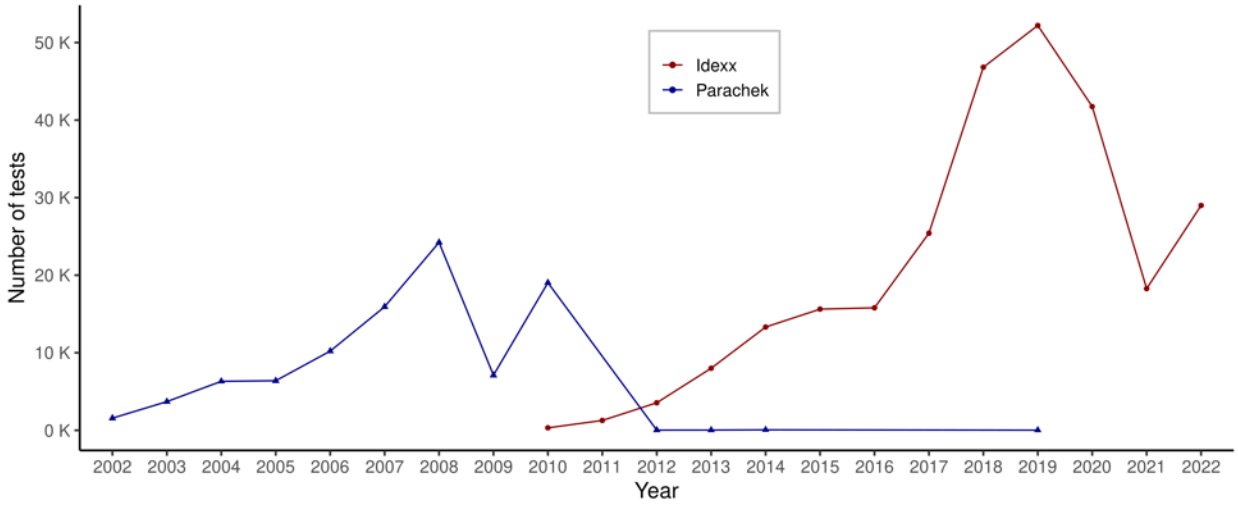

Figure 4. Yearly distribution of two Enzyme-Linked Immunosorbent Assay (ELISA) kits, IDX (Idexx) and PCK (Parachek), testing for Johne's Disease.

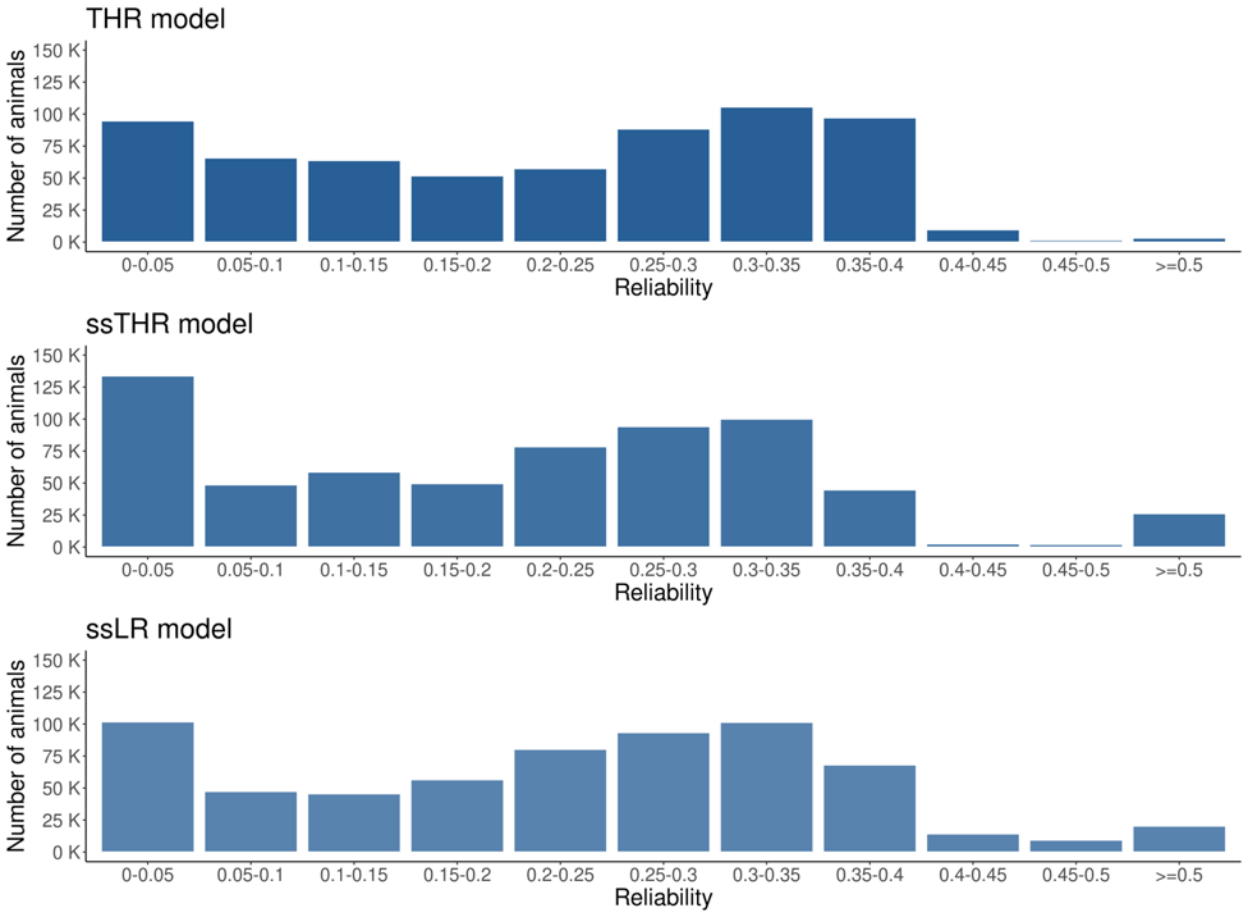

Figure 2. Overall reliabilities for IDX (Idexx) for Johne's Disease in U.S. Holstein dairy cattle under three models: pedigree-based threshold (THR), single-step threshold (ssTHR), and single-step linear (ssLR) models.

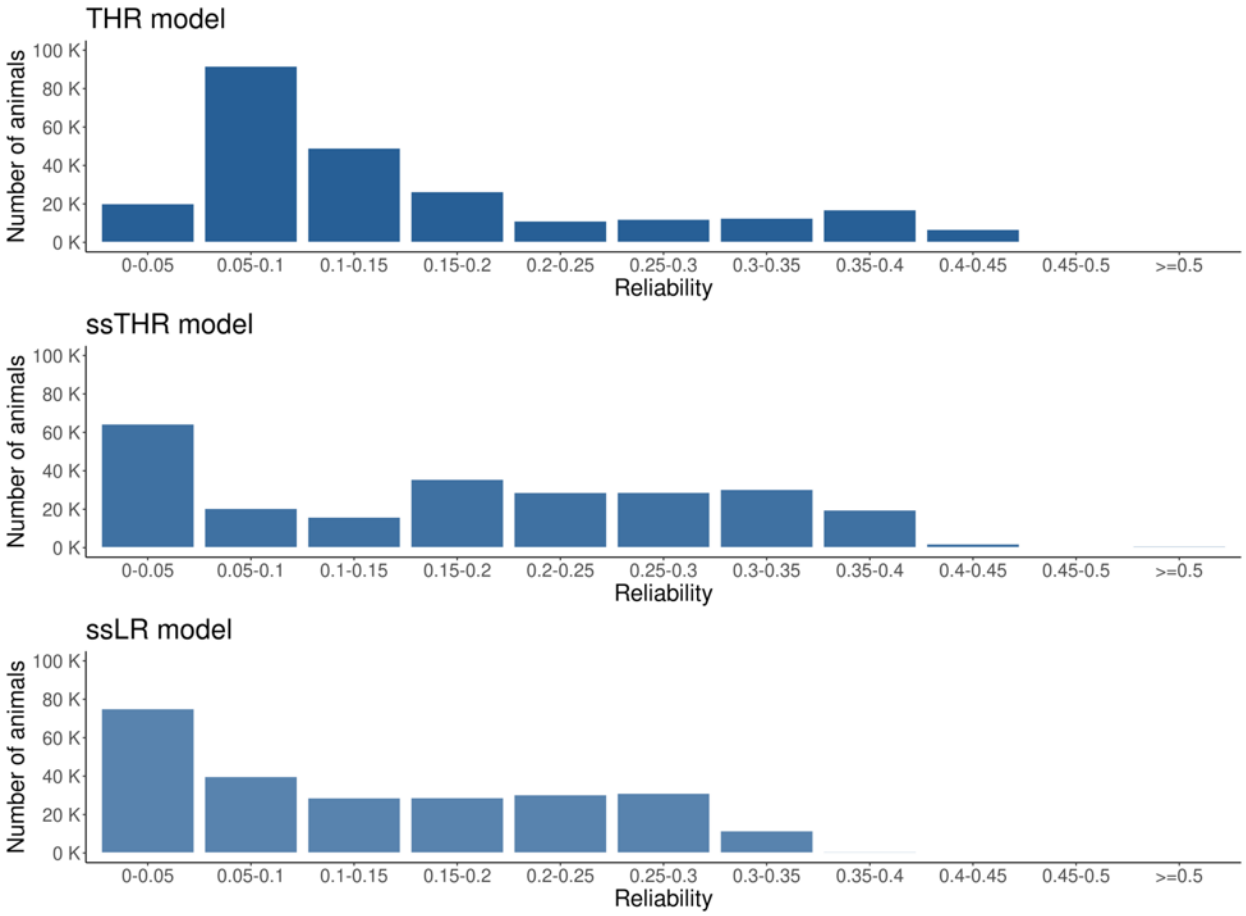

Figure 3. Overall reliabilities for PCK (Parachek), test for Johne's Disease in U.S. Holstein dairy cattle under three models: pedigree-based threshold (THR), single-step threshold (ssTHR), and single-step linear (ssLR) models.

Figure 4. Phenotypic trend as percentage of positive results relative to total tests with two test kits for Johne's Disease: IDX (Idexx) and PCK (Parachek).

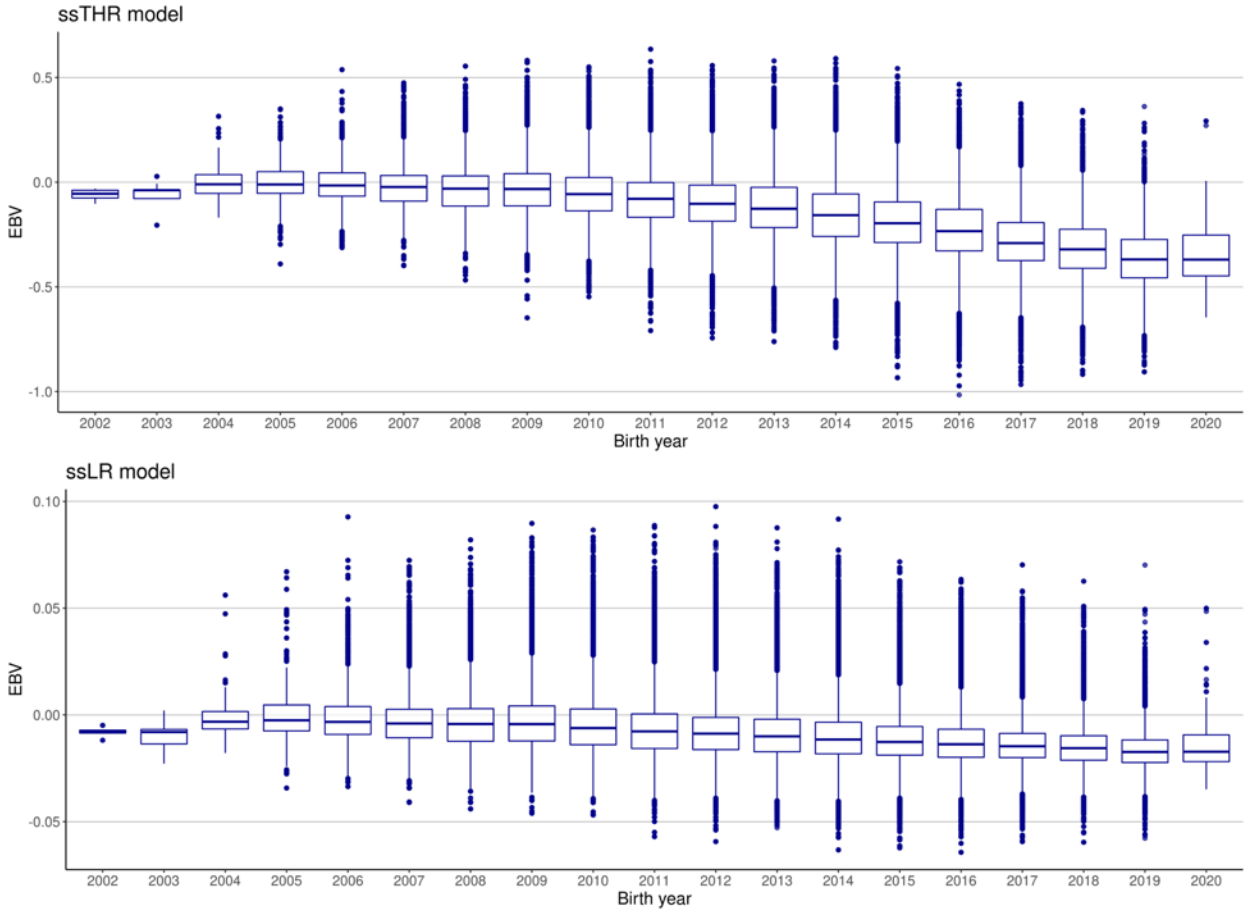

Figure 5. Genetic trend for IDX (Idexx) test for susceptibility for Johne's Disease in U.S. Holstein cows under single-step threshold (ssTHR) and single-step linear (ssLR) models.

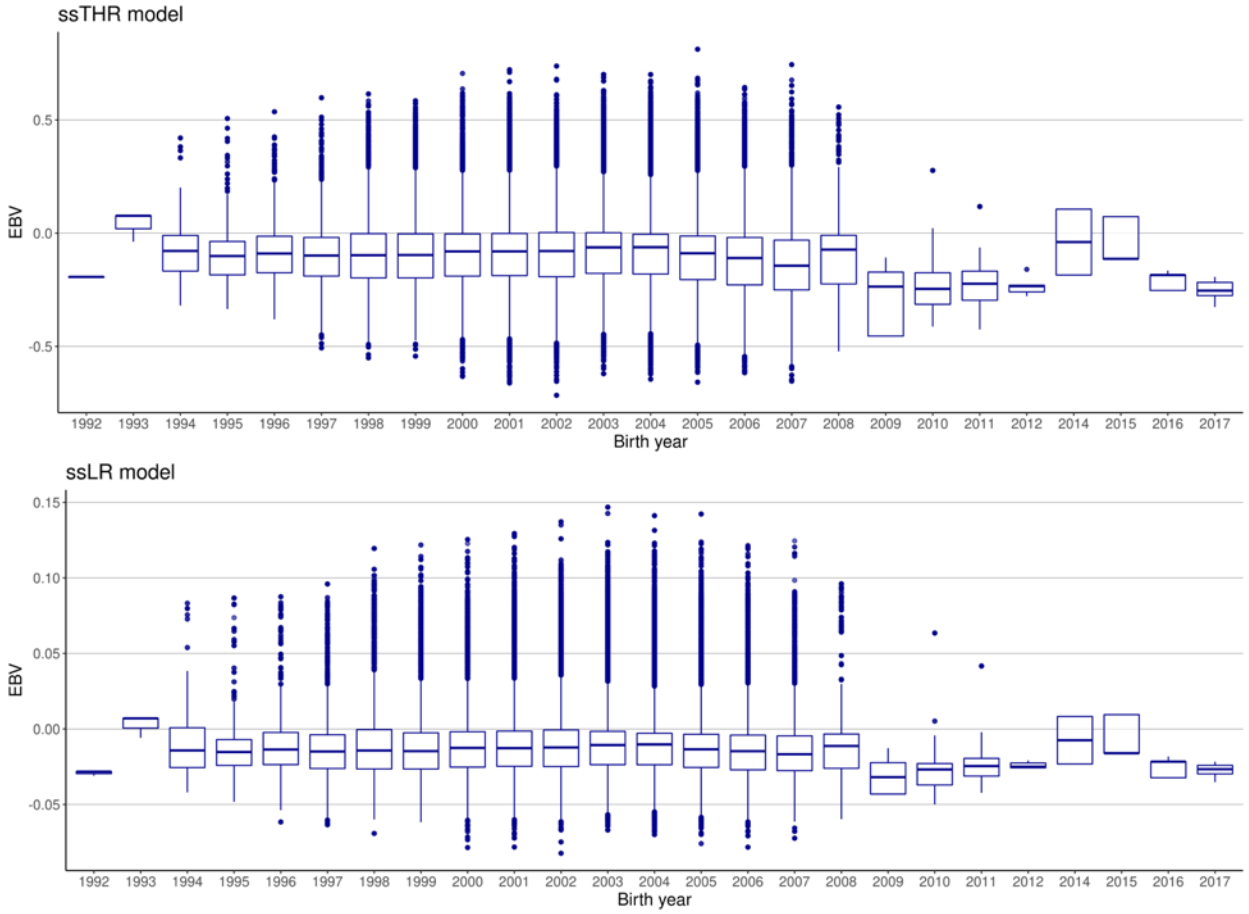

Figure 6. Genetic trend PCK (Parachek) test for susceptibility for Johne's Disease in U.S. Holstein cows under single-step threshold (ssTHR) and single-step linear (ssLR) models.

References

Attalla, S. A., Seykora, A. J., Cole, J. B., and Heins, B. J. (2010). Genetic parameters of milk ELISA scores for Johne's disease. *J. Dairy Sci.* 93, 1729–1735. doi:10.3168/jds.2009-2625.

Bannantine, J. P., Hines, M. E., Bermudez, L. E., Talaat, A. M., Sreevatsan, S., Stabel, J. R., et al. (2014). Rational framework for evaluating the next generation of vaccines against mycobacterium avium subspecies paratuberculosis. *Front. Cell. Infect. Microbiol.* 4, 1–11. doi:10.3389/fcimb.2014.00126.

Berry, D. P., Good, M., Mullowney, P., Cromie, A. R., and More, S. J. (2010). Genetic variation in serological response to Mycobacterium avium subspecies paratuberculosis and its association with performance in Irish Holstein-Friesian dairy cows. *Livest. Sci.* 131, 102–107. doi:10.1016/j.livsci.2010.03.007.

Brito, L. F., Mallikarjunappa, S., Sargolzaei, M., Koeck, A., Chesnais, J., Schenkel, F. S., et al. (2018). The genetic architecture of milk ELISA scores as an indicator of Johne's disease (paratuberculosis) in dairy cattle. *J. Dairy Sci.* 101, 10062–10075. doi:10.3168/jds.2017-14250.

Cole, J. B., Eaglen, S. A. E., Maltecca, C., Mulder, H. A., and Pryce, J. E. (2020). The future of phenomics in dairy cattle breeding. *Anim. Front.* 10, 37–44. doi:10.1093/af/vfaa007.

Corbett, C. S., Naqvi, S. A., De Buck, J., Kanevets, U., Kastelic, J. P., and Barkema, H. W. (2018). Environmental sample characteristics and herd size associated with decreased herd-level prevalence of Mycobacterium avium ssp. paratuberculosis. *J. Dairy Sci.* 101, 8092–8099. doi:10.3168/jds.2018-14661.

Davis, W. C., and Park, K. T. (2018). Progress Towards Control of a Mycobacterial Pathogen, Mycobacterium avium subsp. paratuberculosis, the Causative Agent of Johne's Disease in

Cattle and Humans. *J. Food Hyg. Saf.* 33, 221–228. doi:10.13103/jfhs.2018.33.4.221.

García-Ruiz, A., Cole, J. B., VanRaden, P. M., Wiggans, G. R., Ruiz-López, F. J., and Van Tassell, C. P. (2016). Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U. S. A.* 113, E3995–E4004. doi:10.1073/pnas.1519061113.

Garcia, A. B., and Shalloo, L. (2015). Invited review: The economic impact and control of paratuberculosis in cattle. *J. Dairy Sci.* 98, 5019–5039. doi:10.3168/jds.2014-9241.

Geweke, J. (1992). "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics*, eds. J. O. Berger, J. M. Bernardo, A. P. Dawid, and A. F. M. Smith (Oxford, UK: Oxford University Press), 164–193.

Gianola, D. (1982). Theory and Analysis of Threshold Characters. *J. Anim. Sci.* 54, 1079–1096. doi:10.2527/jas1982.5451079x.

Gonda, M. G., Chang, Y. M., Shook, G. E., Collins, M. T., and Kirkpatrick, B. W. (2006). Genetic variation of Mycobacterium avium ssp. paratuberculosis infection in US Holsteins. *J. Dairy Sci.* 89, 1804–1812. doi:10.3168/jds.S0022-0302(06)72249-4.

Gonda, M. G., Chang, Y. M., Shook, G. E., Collins, M. T., and Kirkpatrick, B. W. (2007). Effect of Mycobacterium paratuberculosis infection on production, reproduction, and health traits in US Holsteins. *Prev. Vet. Med.* 80, 103–119. doi:10.1016/j.prevetmed.2007.01.011.

Harder, I., Stamer, E., Junge, W., and Thaller, G. (2020). Estimation of genetic parameters and breeding values for feed intake and energy balance using pedigree relationships or single-step genomic evaluation in Holstein Friesian cows. *J. Dairy Sci.* 103, 2498–2513. doi:10.3168/jds.2019-16855.

Hinger, M., Brandt, H., and Erhardt, G. (2008). Heritability estimates for antibody response to

Mycobacterium avium subspecies paratuberculosis in German Holstein cattle. *J. Dairy Sci.* 91, 3237–3244. doi:10.3168/jds.2008-1021.

Kirkpatrick, B. W., Cooke, M. E., Frie, M., Sporer, K. R. B., Lett, B., Wells, S. J., et al. (2022). Genome-wide association analysis for susceptibility to infection by Mycobacterium avium ssp. paratuberculosis in US Holsteins. *J. Dairy Sci.* 105, 4301–4313. doi:10.3168/jds.2021-21276.

Kirkpatrick, B. W., and Lett, B. M. (2018). Short communication: Heritability of susceptibility to infection by Mycobacterium avium ssp. paratuberculosis in Holstein cattle. *J. Dairy Sci.* 101, 11165–11169. doi:10.3168/jds.2018-15021.

Kirkpatrick, B. W., Shi, X., Shook, G. E., and Collins, M. T. (2011). Whole-Genome association analysis of susceptibility to paratuberculosis in Holstein cattle. *Anim. Genet.* 42, 149–160. doi:10.1111/j.1365-2052.2010.02097.x.

Kizilkaya, K., Fernando, R. L., and Garrick, D. J. (2014). Reduction in accuracy of genomic prediction for ordered categorical data compared to continuous observations. *Genet. Sel. Evol.* 46, 1–13. doi:10.1186/1297-9686-46-37.

Korsgaard, I. R., Lund, M. S., Sorensen, D., Gianola, D., Madsen, P., and Jensen, J. (2003). Multivariate Bayesian analysis of Gaussian, right censored Gaussian, ordered categorical and binary traits using Gibbs sampling. *Genet. Sel. Evol.* 35, 159–183. doi:10.1186/1297-9686-35-2-159.

Lourenco, D., Legarra, A., Tsuruta, S., Masuda, Y., Aguilar, I., and Misztal, I. (2020). Single-step genomic evaluations from theory to practice: using snp chips and sequence data in blupf90. *Genes (Basel).* 11, 1–32. doi:10.3390/genes11070790.

Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Genetics*. Sunderland,

MA: Sinauer Associates Inc.

Minozzi, G., Buggiotti, L., Stella, A., Strozzi, F., Luini, M., and Williams, J. L. (2010). Genetic loci involved in antibody response to Mycobacterium avium ssp. paratuberculosis in cattle. *PLoS One* 5. doi:10.1371/journal.pone.0011117.

Misztal, I., Tsuruta, S., Lourenco, D., Aguilar, I., Legarra, A., and Vitezica, Z. (2014). Manual for BLUPF90 family of programs. *Univ. Georg. Athens, USA*, 1–125. Available at: http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all2.pdf.

Neupane, M., Hutchison, J. L., Van Tassell, C. P., and VanRaden, P. M. (2021). Genomic evaluation of dairy heifer livability. *J. Dairy Sci.* 104, 8959–8965. doi:10.3168/jds.2020-19687.

Pant, S. D., Verschoor, C. P., Schenkel, F. S., You, Q., Kelton, D. F., and Karrow, N. A. (2011). Bovine PGLYRP1 polymorphisms and their association with resistance to Mycobacterium avium ssp. Paratuberculosis. *Anim. Genet.* 42, 354–360. doi:10.1111/j.1365-2052.2010.02153.x.

Parker Gaddis, K. L., VanRaden, P. M., Cole, J. B., Norman, H. D., Nicolazzi, E., and Dürr, J. W. (2020). Symposium review: Development, implementation, and perspectives of health evaluations in the United States. *J. Dairy Sci.* 103, 5354–5365. doi:10.3168/jds.2019-17687.

Pinedo, P. J., Buergelt, C. D., Donovan, G. A., Melendez, P., Morel, L., Wu, R., et al. (2009a). Association between CARD15/NOD2 gene polymorphisms and paratuberculosis infection in cattle. *Vet. Microbiol.* 134, 346–352. doi:10.1016/j.vetmic.2008.09.052.

Pinedo, P. J., Buergelt, C. D., Donovan, G. A., Melendez, P., Morel, L., Wu, R., et al. (2009b). Candidate gene polymorphisms (BoIFNG, TLR4, SLC11A1) as risk factors for paratuberculosis infection in cattle. *Prev. Vet. Med.* 91, 189–196.

doi:10.1016/j.prevetmed.2009.05.020.

Pryce, J. E., Esslemont, R. J., Thompson, R., Veerkamp, R. F., Kossaibati, M. A., and Simm, G. (1998). Estimation of genetic parameters using health, fertility and production data from a management recording system for dairy cattle. *Anim. Sci.* 66, 577–584. doi:10.1017/S1357729800009152.

Rasmussen, P., Barkema, H. W., Mason, S., Beaulieu, E., and Hall, D. C. (2021). Economic losses due to Johne's disease (paratuberculosis) in dairy cattle. *J. Dairy Sci.* 104, 3123–3143. doi:10.3168/jds.2020-19381.

Rauw, W. M., Kanis, E., Noordhuizen-Stassen, E. N., and Grommers, F. J. (1998). Undesirable side effects of selection for high production efficiency in farm animals: A review. *Livest. Prod. Sci.* 56, 15–33. doi:10.1016/S0301-6226(98)00147-X.

Sanchez, M. P., Tribout, T., Fritz, S., Guatteo, R., Fourichon, C., Schibler, L., et al. (2022). New insights into the genetic resistance to paratuberculosis in Holstein cattle via single-step genomic evaluation. *Genet. Sel. Evol.* 54, 1–10. doi:10.1186/s12711-022-00757-z.

Settles, M., Zanella, R., McKay, S. D., Schnabel, R. D., Taylor, J. F., Whitlock, R., et al. (2009). A whole genome association analysis identifies loci associated with Mycobacterium avium subsp. paratuberculosis infection status in US holstein cattle. *Anim. Genet.* 40, 655–662. doi:10.1111/j.1365-2052.2009.01896.x.

Shook, G. E., Chaffer, M., Wu, X. L., and Ezra, E. (2012). Genetic parameters for paratuberculosis infection and effect of infection on production traits in Israeli Holsteins. *Anim. Genet.* 43, 56–64. doi:10.1111/j.1365-2052.2012.02349.x.

Sigdel, A., Wu, X. L., Parker Gaddis, K. L., Norman, H. D., Carrillo, J. A., Burchard, J., et al. (2022). Genetic Evaluations of Stillbirth for Five United States Dairy Breeds: A Data-

Resource Feasibility Study. *Front. Genet.* 13, 1–12. doi:10.3389/fgene.2022.819678.

Sun, C., Madsen, P., Nielsen, U. S., Zhang, Y., Lund, M. S., and Su, G. (2009). Comparison between a sire model and an animal model for genetic evaluation of fertility traits in Danish Holstein population. *J. Dairy Sci.* 92, 4063–4071. doi:10.3168/jds.2008-1918.

Tsuruta, S., Lourenco, D. A. L., Misztal, I., and Lawlor, T. J. (2017). Genomic analysis of cow mortality and milk production using a threshold-linear model. *J. Dairy Sci.* 100, 7295–7305. doi:10.3168/jds.2017-12665.

Tsuruta, S., Misztal, I., and Strandén, I. (2001). Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. *J. Anim. Sci.* 79, 1166–1172. doi:10.2527/2001.7951166x.

van Hulzen, K. J. E., Schopen, G. C. B., van Arendonk, J. A. M., Nielen, M., Koets, A. P., Schrooten, C., et al. (2012). Genome-wide association study to identify chromosomal regions associated with antibody response to Mycobacterium avium subspecies paratuberculosis in milk of Dutch Holstein-Friesians. *J. Dairy Sci.* 95, 2740–2748. doi:10.3168/jds.2011-5005.

VanRaden, P. M. (2017). Net merit as a measure of lifetime profit: 2017 revision. *AIP Res. Rep. NM$6*. Available at: https://aipl.arsusda.gov/reference/nmcalc-2014.htm.

VanRaden, P. M., and Wiggans, G. R. (1995). Productive Life Evaluations: Calculation, Accuracy, and Economic Value. *J. Dairy Sci.* 78, 631–638. doi:10.3168/jds.S0022-0302(95)76674-7.

Whittington, R. J., Begg, D. J., de Silva, K., Purdie, A. C., Dhand, N. K., and Plain, K. M. (2017). Case definition terminology for paratuberculosis (Johne's disease). *BMC Vet. Res.* 13, 1–13. doi:10.1186/s12917-017-1254-6.

Wiggans, G. R., Cole, J. B., Hubbard, S. M., and Sonstegard, T. S. (2017). Genomic Selection in Dairy Cattle: The USDA Experience. *Annu. Rev. Anim. Biosci.* 5, 309–327. doi:10.1146/annurev-animal-021815-111422.

Wolf, R., Barkema, H. W., De Buck, J., Slomp, M., Flaig, J., Haupstein, D., et al. (2014). High herd-level prevalence of Mycobacterium avium subspecies paratuberculosis in Western Canadian dairy farms, based on environmental sampling. *J. Dairy Sci.* 97, 6250–6259. doi:10.3168/jds.2014-8101.

Wu, X. L., Heringstad, B., and Gianola, D. (2008). Exploration of lagged relationships between mastitis and milk yield in dairy cows using a Bayesian structural equation Gaussian-threshold model. *Genet. Sel. Evol.* 40, 333–357. doi:10.1051/gse:2008009.

Yao, C., Weigel, K. A., and Cole, J. B. (2014). Short communication: Genetic evaluation of stillbirth in US Brown Swiss and Jersey cattle. *J. Dairy Sci.* 97, 2474–2480. doi:10.3168/jds.2013-7320.

Zare, Y., Shook, G. E., Collins, M. T., and Kirkpatrick, B. W. (2014). Genome-wide association analysis and genomic prediction of Mycobacterium avium subspecies paratuberculosis infection in US Jersey cattle. *PLoS One* 9. doi:10.1371/journal.pone.0088380.

Zwald, N. R., Weigel, K. A., Chang, Y. M., Welper, R. D., and Clay, J. S. (2004). Genetic selection for health traits using producer-recorded data. I. Incidence rates, heritability estimates, and sire breeding values. *J. Dairy Sci.* 87, 4287–4294. doi:10.3168/jds.S0022-0302(04)73573-0.

CHAPTER SEVEN: CONCLUSIONS

This work addressed different aspects involved in the development and implementation of novel traits in dairy cattle breeding. The development of a new phenotype, such as body temperature, can be favored by automating the process but requires the handling and analysis of high-frequency data. The first research study of this thesis investigated novel body temperature phenotypes by using temperature loggers and found that body temperature can be used as an indicator for different feed efficiency traits, including milk energy and residual feed intake. The results indicated, for example, that less consistent body temperature is associated with more energy directed to milk production and higher body temperature is associated with less feed efficiency. The second research study provides a better understanding of the biological mechanisms underlying visceral fat accumulation, showing that inflammation, insulin resistance, and immune response are all related to abdominal obesity. This same study also revealed that visceral fat and displaced abomasum are genetically linked, and this genetic link deserves further investigation. The third research study identifies candidate variants and genes for postpartum blood calcium concentration using whole-genome sequence data. All the putative causal variants were located in non-coding regions of genes involved in vitamin D metabolism, calcium homeostasis, calcium and potassium channels, mitochondrial energy metabolism, and the immune response. The fourth and last study evaluated the implementation of a national genetic evaluation for Johne's Disease, a potential novel health trait in U.S. Holsteins. Overall, this thesis has performed a comprehensive investigation of the different steps required for the development and implementation of novel traits in dairy cattle.