Pacheco, H., Vargas, G., Oliveira Jr, G., Gonzalez-Peña, D., Vukasinovic, N. (2026) Genetic dissection of residual feed intake and methane production in dairy cattle using a large-scale single-step GWAS, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23765: https://doi.org/10.31274/wcgalp.23765 ©2026 Pacheco, et al. CC BY-NC-ND

Genetic dissection of residual feed intake and methane production in dairy cattle using a large-scale single-step GWAS

H. Pacheco[1*], G. Vargas[1], G. A. Oliveira Jr[1], D. Gonzalez-Peña[1], and N. Vukasinovic[1]

[1]*Zoetis Genetics, Kalamazoo, Michigan, 49007, USA; *Corresponding author: hendyel.pachecopowell@zoetis.com*

Introduction

Residual feed intake (RFI) and residual methane production (RMet) are key traits for improving feed efficiency and reducing greenhouse gas emissions in dairy cattle. RFI reflects differences between actual and expected feed intake, while RMet captures variation in methane emissions independent of production traits. It has been shown that these traits have a genetic component (Cavani *et al*., 2022, Ghavi Hossein-Zadeh, 2022), so improving them through genetics can boost profitability and sustainability, given that feed costs are a major expense, and methane is a potent contributor to climate change. Understanding their genetic basis is essential for effective selection, as both traits are complex and influenced by multiple genes and pathways.

Advances in genomic technologies, particularly single-step genome-wide association studies (ssGWAS), enable accurate identification of genomic regions associated with complex traits by integrating phenotypic, pedigree, and genomic data (Lourenco *et al*., 2015). Combined with the Algorithm for Proven and Young (APY), ssGWAS allows computationally efficient analyses of large populations while maintaining predictive power (Abdollahi-Arpanahi *et al*., 2022, Leite *et al*., 2024). These tools provide insights into biological mechanisms underlying RFI and RMet and support genomic selection strategies. Therefore, this study aimed to perform a large-scale ssGWAS using APY to identify genomic regions associated with RFI and RMet in Holstein cattle.

Material & Methods

Phenotypic and Genotypic data. Phenotypic data were collected in 5 commercial herds across the United States using the SmartFeed and GreenFeed (C-Lock Inc., Rapid City, SD, USA) equipment for RFI and RMet, respectively. RFI (kg/day) was defined as the residual of a linear regression of dry matter intake (DMI; kg/day) for each herd on energy-corrected milk, metabolic body weight, parity (primiparous or multiparous), age-at-calving category (4 levels), with stage of lactation modeled using week of lactation fitted as a spline interacting with parity and including the random effect of week of experiment. RMet was estimated by regressing methane production (g/day) on the previously mentioned energy sinks, including herd-year as a random effect. Weekly averages of daily RFI and RMet were available for 7,055 and 5,759 genotyped Holstein cows, respectively, resulting in 72,301 feed intake records and 42,273 methane emission measurements. Genomic data for 876,995 animals was used in the analysis, with 45,245 SNP markers. Genotyping was performed using SNP arrays of varying densities, as described by Vukasinovic *et al*. (2017). Genotypes underwent rigorous quality control following Wiggans *et al*. (2009). Animals genotyped with arrays containing fewer than 40K markers were imputed to a common set of 45,245 SNP using FImpute (Sargolzaei *et al*., 2011).

Statistical Models. A single-trait model was used to estimate SNP effects using the BLUPF90 suite (Misztal *et al*., 2014b) with the single-step genomic BLUP (ssGBLUP) method. In ssGBLUP, the inverse of the traditional pedigree relationship matrix, $\mathbf{A}^{-1}$, is replaced by the inverse of the **H** matrix, which integrates both pedigree (**A)** and the genomic relationship (**G**) information:

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}, \quad (1)$$

where $\mathbf{A}^{-1}$ is the inverse of the pedigree relationship matrix; $\mathbf{G}^{-1}$ is an inverse of the genomic relationship matrix (VanRaden, 2008); and $\mathbf{A}_{22}^{-1}$ is the inverse of the pedigree relationship matrix for genotyped animals only. To accommodate the large number of genotypes, the algorithm for proven and young animals (APY) was applied (Misztal *et al*., 2014a). The APY algorithm generates $\mathbf{G}^{-1}$ using the genomic recursion based on a subset of "core" animals, such that only the genomic relationship matrix for core animals is directly inverted, while relationships for all other animals are derived recursively, substantially reducing computational requirements. The core comprised 25,000 randomly selected animals, corresponding to the number of eigenvalues explaining approximately 99% of the genetic variance in **G**. Alternative core sizes were evaluated (results not shown), and results were consistent with previous studies showing that APY solutions are stable once the core captures the majority of the genetic variation, with limited additional benefit from further increasing core size (Pocrnic *et al*., 2022).

Gene mapping. Candidate genomic regions associated with RFI or RMet were identified based on the amount of genetic variance explained by a 30-SNP window. Given the genomic estimated breeding values (GEBVs), the SNP effects can be estimated as $\hat{\mathbf{m}} = \mathrm{DZ}'[\mathrm{ZDZ}']^{-1}\hat{\mathrm{a}}_{\mathrm{g}}$, where $\hat{\mathbf{m}}$ is the vector of SNP marker effects, D is a diagonal matrix of weights of SNPs, and $\hat{\mathrm{a}}_{\mathrm{g}}$ is the vector of GEBVs (Wang *et al*., 2012). Then, the proportion of genetic variance explained by a window of 30 adjacent SNPs was calculated using the program POSTGSF90 (Misztal *et al*., 2014b). Given the average marker density of the dataset, windows of 30 adjacent SNPs corresponded to an average physical size of approximately 1.6 Mb, although window size varied with local SNP spacing. Using the ARS-UCD 1.2 bovine genome assembly, a given SNP marker was assigned to a particular annotated gene if the SNP was located within the genomic sequence of the gene or at most 100kb upstream or downstream of the gene. A threshold of 0.5% of the total additive genetic variance captured by SNPs was used to define significant genomic regions, consistent with previous ssGWAS to identify biologically meaningful loci with moderate-to-large effects; candidate genes were defined as those containing at least one SNP within a window exceeding this threshold.

Results & Discussion

Significant genomic regions were identified for both traits, with several harboring candidate genes implicated in relevant biological processes. Figure 1 illustrates the GWAS results for RFI, where regions on chromosomes 10, 16, 26, and 29 each explained more than 0.5% of the genetic variance. Notable candidate genes include those involved in fatty acid and carbohydrate metabolism (*FADS1* and *TKFC*), and mitochondrial functions (*SDHAF2* and *CYB561A3*), membrane and transport of proteins (*NPC2, TMEM138, TMEM216, TMEM258*), transcription regulation (*CPSF7, U2, PPP1R32*), and developmental transcription factors (*EMX2* and *MYRF*). Among these, FADS1 is particularly relevant due to its role in fatty acid metabolism and its association with variation in feed efficiency across multiple cattle breeds (Yang *et al*., 2023). While most genes have not been

previously associated with feed efficiency, miRNAs, *bta-mir-2405* and *bta-mir-2885*, were identified in this study and in transcriptome analyses of Nelore cattle with different levels of RFI (De Oliveira *et al*., 2018). These miRNAs play regulatory roles in networks influencing metabolism and stress signaling, highlighting their potential importance for feed efficiency.
Figure 2 shows the GWAS results for RMet in which regions on BTA 1, 9, 19, and 23 explained more than 0.5% of the genetic variance. Candidate genes identified span diverse biological functions, including metabolic and enzymatic processes (*AFMID, NQO2, PGS1, TK1*), immune regulation and apoptosis (*SOCS3, RIPK1, SERPIN* family), RNA processing and gene regulation (*DHX36, TNRC6C, U1, U6*), and structural or transport roles (*DNAH17, TUBB2A, TMEM235*). While none of these genes have direct evidence linking them to methane emissions in cattle, some, such as *PGS1* (mitochondrial lipid metabolism) and *SOCS3* (immune-metabolic regulation), may indirectly influence rumen microbiology, which is fundamental to methane production (Fernando *et al*., 2025). Given the limited number of GWAS for RFI and methane traits, many identified regions represent novel associations and are discussed as hypothesis-generating based on known biological functions rather than on direct prior evidence.

Conclusion

This study presents one of the first large-scale ssGWAS of residual feed intake and residual methane production in Holstein cattle, identifying multiple genomic regions that explain a significant proportion of the genetic variance for both traits. Candidate genes were associated with key biological processes, including energy metabolism, mitochondrial function, immune regulation, and RNA processing. Additionally, miRNAs emerged as potential regulators of metabolic and stress pathways. Although most genes have not been previously associated with feed efficiency or methane emission traits, their functional roles suggest indirect effects on rumen fermentation and host-microbiome interactions, offering opportunities for further research. These findings provide valuable insights into the genetic architecture of sustainability-related traits and lay the groundwork for genomic selection strategies to enhance feed efficiency and reduce methane emissions in dairy cattle.

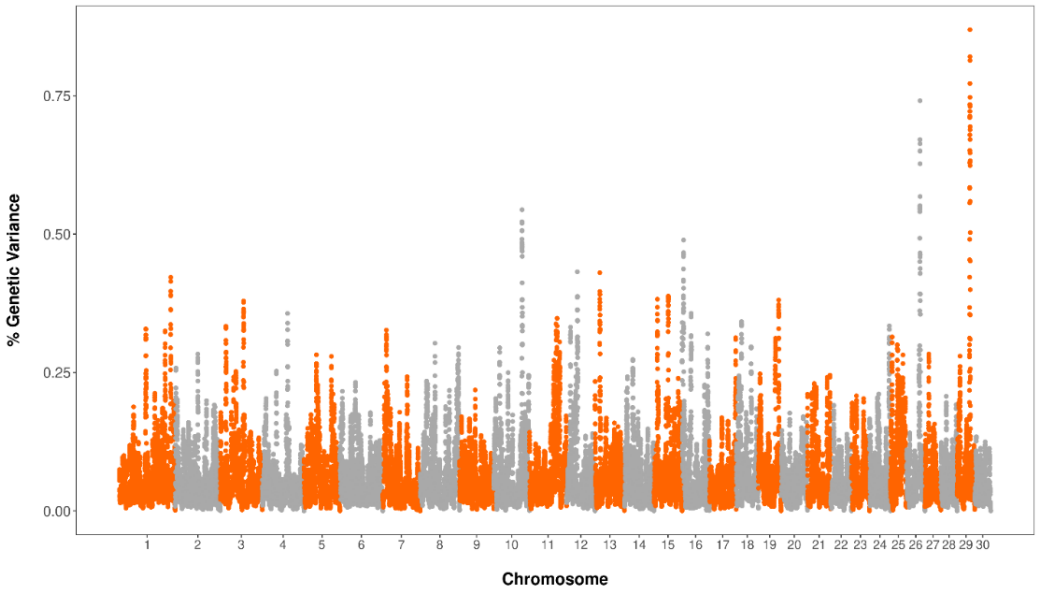

Figure 1. Manhattan plot of the percentage of variance explained for RFI.

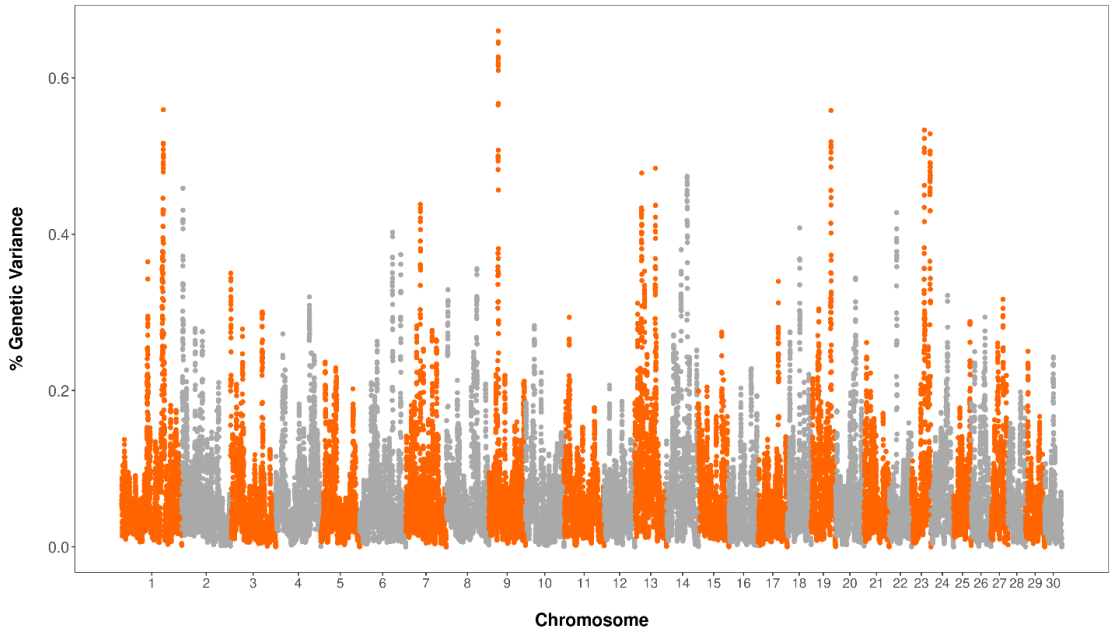

Figure 2. Manhattan plot of the percentage of variance explained for RMet.

Acknowledgments

This research was conducted as a part of the authors' employment[1] with Zoetis LLC.

References

Abdollahi-Arpanahi, R., D. Lourenco, and I. Misztal. (2022) Genet Sel Evol 54:34 doi.org/10.1186/s12711-022-00726-6.

Cavani, L., W. E. Brown, K. L. Parker Gaddis, *et al*. (2022) J Dairy Sci 105:7564-7574 doi.org/10.3168/jds.2022-22066.

De Oliveira, P. S. N., L. L. Coutinho, P. C. Tizioto, *et al*. (2018) Sci Rep 8:17072 doi.org/10.1038/s41598-018-35315-5.

Fernando, S. C., S. Adams, A. Lakamp, *et al*. (2025) J. Dairy Sci 108:7576-7590 doi.org/10.3168/jds.2024-25797.

Ghavi Hossein-Zadeh, N. (2022) Sci Rep 12:12352 doi.org/10.1038/s41598-022-16778-z.

Leite, N. G., M. Bermann, S. Tsuruta, *et al*. (2024) Genet Sel Evol 56:59 doi.org/10.1186/s12711-024-00925-3.

Lourenco, D. A., S. Tsuruta, B. O. Fragomeni, *et al*. (2015) J Anim Sci 93:2653-2662 doi.org/10.2527/jas.2014-8836.

Misztal, I., A. Legarra, and I. Aguilar. (2014a) J Dairy Sci 97:3943-3952 doi.org/10.3168/jds.2013-7752.

Misztal, I., S. Tsuruta, I. Aguilar, *et al*. (2014b) https://nce.ads.uga.edu/html/projects/programs/docs/blupf90_all8.pdf. Accessed Nov. 09, 2025.

Pocrnic I., D.A. Lourenco, C.-Y. Chen, *et al*. (2019) J Anim Sci 97:1513–1522. doi.org/10.1093/jas/skz042.

Sargolzaei M, J. Chesnais, and F. Schenkel. (2011) J Dairy Sci 94(Suppl. 1):421. (Abstr.)

VanRaden, P. M. (2008) J Dairy Sci 91:4414-4423 doi.org/10.3168/jds.2007-0980.

Vukasinovic, N., N. Bacciu, C.A. Przybyla, *et al*. (2017) J. Dairy Sci 704 100:428–438. doi.org/10.3168/jds.2016-11520.

Wang, H., I. Misztal, I. Aguilar, *et al*. (2012) Genetics Research 94:73-83 doi.org/10.1017/s0016672312000274.

Wiggans G.R., T.S. Sonstegard, P.M. VanRaden, *et al*. (2009) J. Dairy Sci 92, 3431-3436 doi.org/10.3168/jds.2008-1758.

Yang, C., Z. Huang, C. Pan, *et al*. (2023) PLoS One 18:e0289939 doi.org/10.1371/journal.pone.0289939.