



J. Dairy Sci. TBC

<https://doi.org/10.3168/jds.2025-27058>

© TBC, The Authors. Published by Elsevier Inc. on behalf of the American Dairy Science Association®.
This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

Genome-wide association study and fine-mapping using imputed sequences to prioritize candidate genes for 30 complex traits in 50,309 Holstein bulls

Junjian Wang,¹ Yahui Gao,^{2*} Sajjad Toghiani,³ John B. Cole,^{2,4,5,6} Christian Maltecca,¹ Li Ma,² and Jicai Jiang^{1†}

¹Department of Animal Science, North Carolina State University, Raleigh, NC 27695

²Department of Animal and Avian Sciences, University of Maryland, College Park, MD 20742

³Agricultural Research Service, Animal Genomics and Improvement Laboratory, USDA, Beltsville, MD 20705

⁴Council on Dairy Cattle Breeding, Bowie, MD 20716

⁵Department of Animal Sciences, Donald Henry Barron Reproductive and Perinatal Biology Research Program, and the Genetics Institute, University of Florida, Gainesville, FL 32611-0910

⁶Department of Animal Biosciences, University of Guelph, Ontario, Canada N1G 2W1

ABSTRACT

Identifying causal genetic variants underlying economically important traits in dairy cattle is essential for understanding their genetic basis and optimizing breeding programs. The growing availability of sequenced reference genomes and individuals with both phenotypic and genotypic data notably enhances our ability to detect genetic associations and further pinpoint causal effects. This comprehensive GWAS of dairy cattle used deregressed breeding values as phenotypes and analyzed 11,292,243 quality-controlled, imputed sequence variants from 50,309 Holstein bulls. The number of bulls with available phenotypes ranged from 23,121 to 50,309 across 30 complex traits categorized into production and yield, type, and longevity and health. We performed GWAS using our SLEMM-GWA approach, which accounts for the varying reliability of deregressed breeding values across individuals and demonstrates computational efficiency for large sample sizes and sequence data. This analysis identified 381 significant association peaks, of which 126 are novel findings. Subsequent Bayesian fine-mapping provided statistical prioritization by assigning posterior conditional inclusion probabilities to individual variants and genes, yielding a list of credible candidate genes—an advancement over conventional GWAS reporting of all proximal genes. This prioritization offered direct statistical support for previously reported genes

and, more importantly, identified credible candidate genes within the 126 newly discovered peaks for specific traits, including *AOPEP*, *GC*, *E2F6*, *MGST1*, *VPS13B*, *ZNF652*, *ASPH*, *SFMBT1*, and *MAPRE2*. These findings enhance the understanding of the genetic architecture of these complex dairy traits and provide valuable insights for the refinement of genomic selection strategies and breeding programs in Holstein cattle.

Key words: GWAS, fine-mapping, dairy cattle, candidate genes

INTRODUCTION

The genetic improvement of economically important traits in livestock represents a cornerstone of modern animal agriculture, yielding significant implications for production efficiency, animal welfare, and environmental sustainability. Over the past century, the spectrum of traits considered for genetic selection within dairy cattle populations has been expanded to address the evolving demands of both industry and society. Selective breeding efforts in recent decades have resulted in significant advancements in traits such as milk production, body conformation, reproductive performance, and disease resistance (Miglior et al., 2017). The integration of genomic data into breeding programs has greatly improved the ability to make informed selection decisions, thereby accelerating genetic gains and improving the efficiency of breeding strategies (Meuwissen et al., 2001; VanRaden, 2008; García-Ruiz et al., 2016).

A comprehensive understanding of the genetic architecture underlying these traits provides significant advantages for selective breeding programs (Weller et al., 2017). The identification of causal genes, such as *DGAT1* (Grisart et al., 2002; Winter et al., 2002) and *ABCG2* (Cohen-Zinder et al., 2005), has advanced our knowl-

Received June 8, 2025.

Accepted August 12, 2025.

*Current address: State Key Laboratory of Swine and Poultry Breeding Industry, National Engineering Research Center for Breeding Swine Industry, Guangdong Provincial Key Laboratory of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou 510642, China

†Corresponding author: jjjiang26@ncsu.edu

edge of complex traits and informed targeted breeding strategies. Within this framework, animal genotyping using SNP panels or whole-genome sequencing followed by GWAS has been presented as the gold standard approach to link phenotypes of interest to their underlying genetics in livestock species (Korte and Farlow, 2013). The GWAS have been widely used to identify genetic variants associated with complex traits, providing deeper insights into their genetic basis (McCarthy et al., 2008). Studies conducted in various dairy cattle breeds have identified numerous loci associated with dairy traits. As of release 55 (December 23, 2024), the Cattle QTLdb has compiled 192,336 QTL or associations covering 553 different traits (Hu et al., 2022).

However, GWAS typically identify associations rather than causal variants, facing limitations in precisely pinpointing causal variants due to factors such as linkage disequilibrium and population structure (Stram, 2004; Goddard and Hayes, 2009). A conventional GWAS often yields numerous significant associations, complicating the prioritization of variants most likely to be causal. Furthermore, typical GWAS interpretation often relies on nominating candidate genes based simply on proximity to association peaks, which limits statistical confidence in identifying truly causal genes. These challenges impede a more granular understanding of the genetic mechanisms underlying the complex traits and hinder the effective application of these findings in breeding programs. Fortunately, the increasing availability of high-density SNP chip genotyping and whole-genome sequencing data from larger cattle populations has substantially enhanced both statistical power and genomic coverage. This combination provides a critical opportunity to move beyond identifying associations toward pinpointing causal variants through fine-mapping.

Fine-mapping offers a way to prioritize the causal variants underlying the associations identified by GWAS. Many fine-mapping methods have been developed, including CAVIARBF (Chen et al., 2015), FINEMAP (Benner et al., 2016), and SuSiE (Zou et al., 2022). However, these methods are primarily designed for samples of unrelated individuals, and direct application to populations with extensive relatedness, common in livestock, can significantly compromise fine-mapping power and precision (Wang et al., unpublished data, 2025). In dairy cattle, intensive use of artificial insemination and strong selection practices have resulted in populations characterized by strong, long-range linkage disequilibrium and large half-sib families (de Roos et al., 2008; Kim and Kirkpatrick, 2009), requiring different approaches than those typically applied in human studies. We specifically developed BFMAP for samples of related individuals (Jiang et al., 2019a). The BFMAP applies a linear mixed model framework to account for the relatedness among individuals and polygenic

effects, thereby improving the accuracy of causal variant identification in livestock populations.

The US dairy industry has been collecting and evaluating economically important traits in dairy cattle for over a century, and the traits considered for genetic selection in dairy cattle populations have evolved to meet the demands of industry and market (Weigel et al., 2017; Guinan et al., 2023). This large-scale collection of phenotypic records and genotype data for a series of key dairy traits, including production, conformation, and health (VanRaden, 2016), provides a unique opportunity to investigate the genetic basis of complex traits in dairy cattle.

In this study, we use GWAS and Bayesian fine-mapping to identify genetic associations and prioritize candidate variants and genes for 30 complex traits in Holstein cattle. By analyzing a large cohort of Holstein bulls ($n = 50,309$) with over 11 million quality-controlled, imputed sequence variants, the aim was to detect robust associations and identify causal genetic elements influencing production, conformation, and health traits. These findings should provide valuable biological insights into these complex traits, thereby informing genetic improvement strategies and enhancing the efficiency of breeding programs in the dairy industry.

MATERIALS AND METHODS

Phenotype Data

This study used data for 30 traits with sample sizes ranging from 23,121 to 50,309 Holstein bulls, accessed through the Council on Dairy Breeding (CDCB). The PTA derived from traditional evaluations were deregressed following VanRaden et al. (2009) to remove the contribution of parent information and reduce dependence among animals. This deregression process generated a deregressed PTA and its corresponding reliability for each bull, which reflects the amount of information from its own records and progeny. For subsequent analyses, these deregressed PTA were used as pseudophenotypes, and their reliabilities were used to derive error variance weights (VanRaden, 2008; VanRaden et al., 2011). The 30 traits were classified into 3 categories: production and yield (PY), longevity and health (LH), and type (TY), with net merit treated as a distinct trait. Details on the number of Holstein bulls and the mean and SD of deregressed PTA reliabilities for each trait are provided in Table 1.

Genotype Data

Genotypes for the 50,309 Holstein bulls were originally obtained from various SNP arrays, encompassing

Table 1. Sample size (N) and reliability statistics (mean and SD) for deregressed PTA across 30 dairy traits in Holstein bulls

Trait name	Abbreviation	Group ¹	N	Reliability	
				Mean	SD
Milk Yield	Milk	PY	50,309	0.793	0.186
Fat Yield	Fat	PY	50,309	0.785	0.191
Protein Yield	Protein	PY	50,308	0.774	0.189
Fat Percentage	Fat_Percent	PY	50,309	0.785	0.191
Protein Percentage	Pro_Percent	PY	50,308	0.774	0.189
Final Score	Final_score	TY	40,793	0.652	0.174
Stature	Stature	TY	40,839	0.841	0.085
Strength	Strength	TY	40,833	0.718	0.153
Dairy Form	Dairy_form	TY	40,562	0.714	0.152
Foot Angle	Foot_angle	TY	40,633	0.613	0.229
Rear Legs (Side View)	Rear_legs(side)	TY	40,837	0.742	0.140
Body Depth	Body_depth	TY	40,837	0.724	0.167
Rump Angle	Rump_angle	TY	40,837	0.819	0.096
Rump Width	Rump_width	TY	40,015	0.755	0.117
Fore Udder Attachment	Fore_udder_att	TY	40,838	0.758	0.126
Rear Udder Height	Rear_ud_height	TY	40,839	0.708	0.151
Rear Udder Width	Rear_ud_width	TY	40,831	0.616	0.224
Udder Depth	Udder_depth	TY	40,834	0.832	0.088
Udder Cleft	Udder_cleft	TY	40,799	0.663	0.198
Front Teat Placement	Front_teat_pla	TY	40,838	0.753	0.120
Teat Length	Teat_length	TY	40,824	0.813	0.092
Rear Legs (Rear View)	Rear_legs(rear)	TY	39,855	0.574	0.193
Feet and Legs	Feet_and_legs	TY	35,247	0.564	0.242
Rear Teat Placement	Rear_teat_pla	TY	39,670	0.770	0.107
Net Merit	Net_Merit	—	50,309	0.605	0.183
Productive Life	Prod_Life	LH	49,389	0.624	0.229
Somatic Cell Score	SCS	LH	50,193	0.714	0.223
Livability	Livability	LH	49,574	0.359	0.226
Calf Livability	Calf_Livability	LH	23,121	0.126	0.195
Gestation Length	Gestleng	LH	49,599	0.454	0.416

¹Abbreviations for trait groups: PY = production and yield; TY = type; LH = longevity and health.

over 50 different SNP panels validated by the CDCB (Bowie, MD) for genomic evaluations (ranging from low-density arrays with a few thousand SNPs to high-density arrays with over 60,000 SNPs). These genotypes were then imputed to a common set of 78,965 SNPs to standardize the genomic data across diverse genotyping platforms. After position and allele matching, exclusion of sex chromosomes, and quality control procedures, ~70,000 autosomal SNPs remained and were subsequently imputed to the whole-genome sequence level using IMPUTE5 (Rubinacci et al., 2020). This imputation leveraged a reference panel consisting of ~2,800 bulls (including Holsteins and other breeds) from both Run8 and Run9 (European Nucleotide Archive accessions PRJEB42783 and PRJEB56689, respectively) of the 1000 Bull Genomes Project (Hayes and Daetwyler, 2019) and 491 dairy bulls (including 318 Holsteins) from the Cooperative Dairy DNA Repository (CDDR; Madison, WI). The sequence data from CDDR were provided to USDA-AGILand made available for this project with CDDR approval. Variants were retained if they had a minor allele frequency ≥ 0.01 , satisfied Hardy-Weinberg equilibrium criteria at a P -value threshold

$\geq 1 \times 10^{-9}$, and achieved an IMPUTE5 INFO score ≥ 0.8 . Only biallelic SNPs on autosomes were retained, excluding sex chromosomal and mitochondrial variants. After quality control, 11,292,243 sequence variants remained for subsequent analyses, and variant positions were annotated based on the ARS-UCD1.2 genome assembly (Rosen et al., 2020).

GWAS

A GWAS was performed using the SLEMM-GWA function in SLEMM v0.89.5 software (Cheng et al., 2023; available at <https://github.com/jiang18/slemm>), which can model the varying reliability of pseudophenotypes (e.g., deregressed PTA) across individuals. The SLEMM-GWA uses the following linear mixed model:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{z}_j a_j + g + \mathbf{e} \text{ with } g \sim \mathcal{N}(0, \mathbf{G}\sigma_g^2) \\ \text{and } \mathbf{e} \sim \mathcal{N}(0, \mathbf{R}\sigma_e^2),$$

where $\mathbf{y}$ represents a vector of (pseudo-)phenotypes, $\mathbf{b}$ denotes a vector of nongenetic fixed effects (including

the intercept) with corresponding design matrix $\mathbf{X}$, $\mathbf{z}_j$ is a vector of genotypes for the j -th variant (coded as 0, 1, or 2) with additive effect a_j , g is a random-effect term accounting for relatedness and polygenic effects, and $\mathbf{e}$ is a vector of residuals. Additionally, $\mathbf{G}$ is a genomic relationship matrix (GRM) constructed using the second method of VanRaden (2008). Also, $\mathbf{R}$ is a diagonal matrix modeling individual reliability, with diagonal elements set to 1 for directly measured phenotypes and for pseudophenotypes computed as $R_{ii} = 1 / r_i^2 - 1$, where r_i^2 denotes the reliability of the i -th individual's pseudophenotype. In this study, we used deregressed PTA as pseudophenotypes and included only the intercept in fixed effects. For each variant j , a score test chi-squared statistic was computed to test the null hypothesis that $a_j = 0$.

To optimize SNP selection for constructing the GRM, we compared GWAS results using GRM derived from different SNP sources and densities. Specifically, we evaluated the GRM built using the full ~70,000 autosomal chip SNPs, alongside GRM constructed from randomly selected subsets without replacement of 30,000, 50,000, and 70,000 SNPs drawn from the 11,292,243 quality-controlled imputed sequence variants. Based on a comparison of genomic inflation factors and the number of significant associations, the GRM built using 70,000 randomly selected sequence SNPs was chosen for subsequent analyses.

Genome-wide significant associations were initially declared using a genome-wide significance threshold of $P < 5 \times 10^{-8}$. Associated regions were defined around each significant peak, encompassing the contiguous genomic segment containing the cluster of significant SNPs. Association peaks were identified by visual inspection of the Manhattan plot for each trait. The boundaries of each associated region spanned from the position of the first to the last significant SNP within each visually identified cluster. For comparison with previously reported associations, only regions containing at least 3 variants with $P < 5 \times 10^{-7}$ were retained to reduce potential false positives. To determine whether these associated regions had been previously reported, we cross-referenced our results against the Cattle QTLdb release 55 (Hu et al., 2022) and extracted traits with definitions closely matching the 30 traits analyzed in our study (Supplemental Table S1, see Notes; Wang et al., 2025a). Additionally, we compared our findings with our previous GWAS results from Jiang et al. (2019a), a study that analyzed 35 traits in 27,214 individuals (27 traits overlapping with this study) and whose results were not included in the Cattle QTLdb. Genomic coordinates from our previous study were converted from UMD 3.1 to ARS-UCD 1.2 using liftOver (Kent et al., 2002), and peaks located on unplaced genomic scaffolds were excluded from comparison. A peak was considered consistent with existing findings if its

associated region intersected with the genomic interval extending ± 500 kb from a previously reported association for the same trait in either the Cattle QTLdb or our previous study, whereas those falling outside this interval were classified as novel discoveries.

To broadly capture candidate regions for fine-mapping, we further identified associations using an inclusive significance threshold ($P < 5 \times 10^{-5}$). Significant SNPs were grouped into contiguous clusters to define initial candidate regions. Adjacent clusters spaced by less than 5 Mb were merged, ensuring subsequent candidate regions were at least 5 Mb apart to avoid redundancy. To reduce potential false positives, only regions containing at least 3 variants with $P < 5 \times 10^{-4}$ were retained for further analysis. For each retained region, we established initial minimal boundaries encompassing all constituent variants meeting the significance threshold ($P < 5 \times 10^{-5}$). Following our previous procedure (Jiang et al., 2019a), the region boundaries were then adjusted based on the position of the variant with the minimum P -value within that region. Specifically, the boundaries were extended outwards where necessary to ensure a minimum distance of 1 Mb both upstream and downstream from this minimum P -value variant's position. This ensured each final candidate region spanned at least 2 Mb and was conceptually centered around the location of the minimum P -value variant, thereby increasing the likelihood of including potential causal variant(s).

Fine-Mapping

To identify potential causal variants and candidate genes, each candidate region identified by GWAS was fine-mapped using the forward selection method in BFMAP v.0.65 (Jiang et al., 2019a). The BFMAP is a Bayesian fine-mapping software tool designed for samples of related individuals, such as livestock, and it employs a linear mixed model framework incorporating a GRM to account for polygenic effects or relatedness, or both. The forward selection procedure in BFMAP sequentially adds independent signals to the model, repositions them for refinement, and identifies variants related to each signal. It computes posterior conditional inclusion probability (PCIP) for each variant in a signal, which quantifies the probability of being included conditional on the other signals in the model. For each signal, BFMAP generates a credible set of variants, defined as the smallest set whose cumulative PCIP reaches a specified confidence level. Detailed documentation and guidelines for BFMAP are available at <https://github.com/jiang18/bfmap>. In our study, we set the confidence level to 95%, thereby deriving a 95% credible set of variants for each signal. To reduce computational

Table 2. Genomic inflation factor (λ) and number of peaks identified using GRM constructed with different SNP sets across 3 representative traits¹

Trait	Parameter	Seq 30,000	Seq 50,000	Seq 70,000	Chip 70,000
Milk	λ	0.905	0.821	0.833	0.825
	No. of peaks	14	13	12	12
Foot Angle	λ	1.093	1.006	1.002	0.923
	No. of peaks	8	7	8	3
Livability	λ	0.999	0.952	0.955	0.922
	No. of peaks	8	8	9	6

¹GRM were constructed using SNPs randomly selected from imputed sequence data (Seq) at different densities (30,000, 50,000, 70,000) or using ~70,000 autosomal chip SNPs. Reliability was incorporated in the model. Number of peaks defined using genome-wide significance threshold of $P < 5 \times 10^{-8}$ and identified by visualization.

burden for fine-mapping, only variants with a GWAS $P < 0.05$ were retained for each candidate region. The BFMAP was run separately for each candidate region. We also computed gene-level PCIP by aggregating the variant-level PCIP for all variants located within a gene, including its 3 kb upstream and downstream flanks to capture potential regulatory regions, based on gene locations extracted from Ensembl release112.

RESULTS

GWAS Optimization

To optimize the GWAS model, we compared results obtained using GRM constructed from the ~70,000 autosomal chip SNPs and from random subsets of 30,000, 50,000, and 70,000 SNPs from the imputed sequence data. For this comparison, we selected 3 representative traits: Milk (from the PY trait group), Livability (from the LH trait group), and Foot Angle (from the TY trait group). As illustrated in Supplemental Figure S1, see Notes (Wang et al., 2025a), GWAS using the GRM from ~70,000 autosomal chip SNPs detected slightly fewer significant peaks at the P -value threshold of 5×10^{-5} than GWAS using GRM from randomly selected imputed sequence SNPs. The number of significant peaks was similar across GWAS using the 3 GRM constructed from randomly selected sequence SNPs. We further evaluated the control of genomic inflation (λ) across these different GRM constructions. As shown in Table 2, the GRM from ~70,000 autosomal chip SNPs resulted in deflated λ values for all 3 traits ($\lambda < 0.95$), whereas the GRM from 30,000 randomly selected sequence SNPs produced an inflated λ for Foot Angle ($\lambda > 1.05$). While the difference in λ values between GRM from 50,000 and 70,000 randomly selected sequence SNPs was minimal, the latter produced λ values marginally closer to the ideal value of 1.0 for all 3 representative traits. Based on these comparisons,

we chose to use the 70,000 randomly selected SNPs from the imputed sequence data to construct the GRM that was applied consistently across all traits in all subsequent analyses to ensure methodological consistency.

Additionally, we assessed the effect of incorporating reliability into our GWAS by comparing 2 models: one with a reliability-weighted residual term ($R_{ii} = 1 / r_i^2 - 1$), and one with an identity residual term ($\mathbf{R} = \mathbf{I}$). Accounting for reliability allowed us to down-weight pseudophenotypes with lower accuracy, potentially reducing statistical noise. Log-likelihood values in Supplemental Table S2 (see Notes; Wang et al., 2025a) demonstrate substantially improved model fitting when incorporating reliability information. For Milk Yield (Figure 1A vs. 1B), a trait characterized by a high mean (0.793) and low variation (SD = 0.186) in reliability, the overall association pattern remained largely consistent between the 2 models, with slight differences in peak prominence. In contrast, for Foot Angle (Figure 1C vs. 1D), which exhibits an intermediate mean and higher variation in reliability (mean = 0.613, SD = 0.229), modeling reliability enhanced the significance of association peaks. The effect was most notable for Livability (Figure 1E vs. 1F), which has a substantially lower mean (0.359) and high variation (SD = 0.226) in reliability. Without modeling reliability, no signals exceeded the genome-wide significance threshold (5×10^{-8}); however, the GWAS incorporating reliability identified 9 peaks. Therefore, we incorporated reliability for all 30 complex traits in subsequent GWAS analyses to optimize model performance.

GWAS

We conducted GWAS for 30 complex traits in 50,309 Holstein bulls using the optimized model, with sample sizes for individual traits ranging from 23,121 to 50,309. These bulls possess extensive daughter records, with deregressed PTA reliability ranging from 0.126 for Calf Livability (a typical low- h^2 trait) to 0.841 for Stature (a typical high- h^2 trait; Table 1). Our mixed model GWAS approach effectively controlled population structure and familial relatedness, as indicated by genomic control factors (λ) smaller than 1.03 for all the 30 traits (Supplemental Table S3, see Notes; Wang et al., 2025a). The GWAS Manhattan plots for the 30 traits are provided in Supplemental Figure S2 (see Notes; Wang et al., 2025a), showing readily identifiable association peaks. Using a genome-wide significance threshold of $P < 5 \times 10^{-8}$, 381 significant peaks were identified for the 30 complex traits. Significant peak counts varied by trait, ranging from 0 for Calf Livability to 28 for Protein Percentage (Supplemental Table S4, see Notes; Wang et al., 2025a).

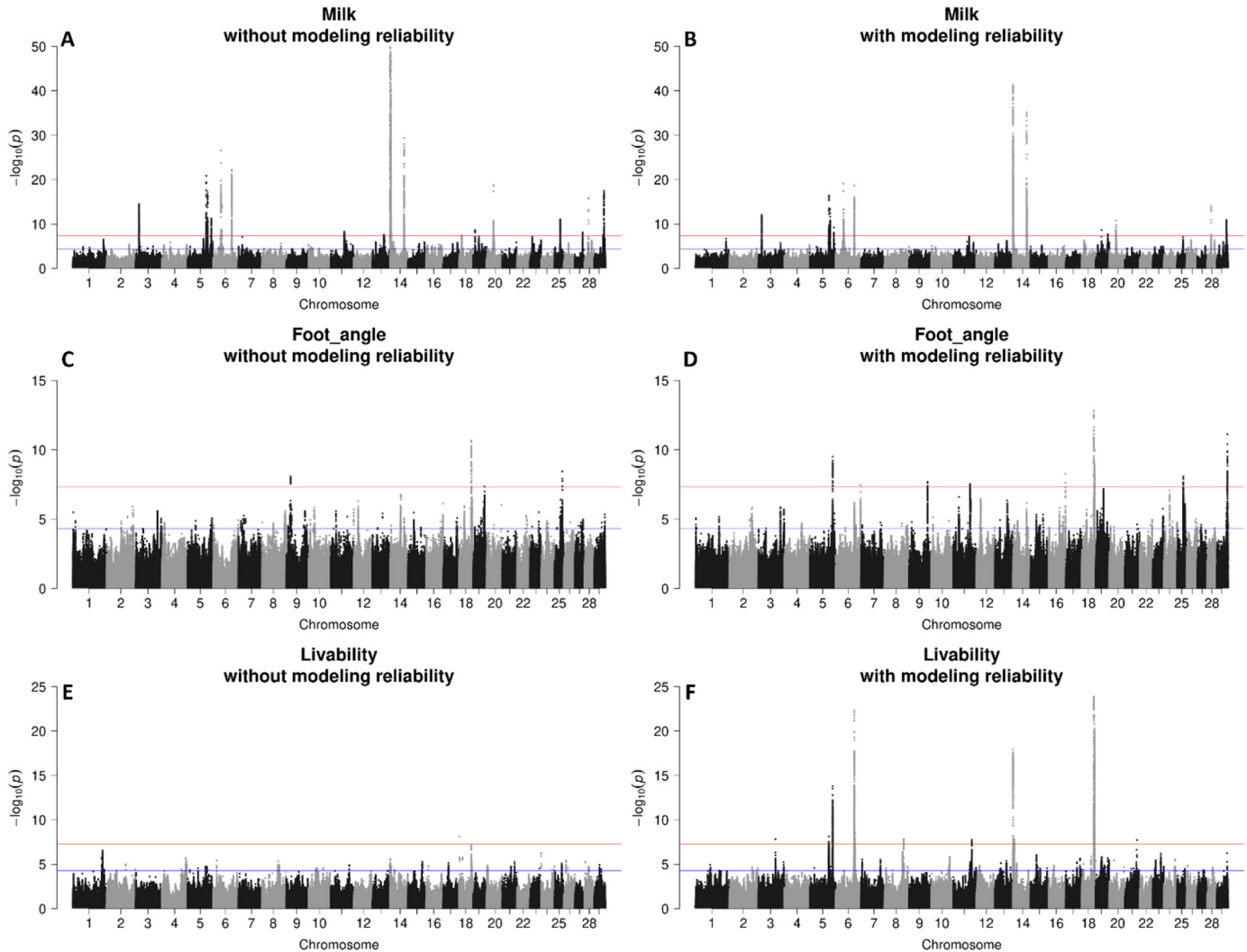


Figure 1. Comparison of GWAS results with (right panels B, D, and F) and without (left panels A, C, and E) modeling reliability for Milk Yield, Foot Angle, and Livability. The red and blue horizontal lines represent genome-wide significance thresholds of 5×10^{-8} and 5×10^{-5} , respectively. For Milk Yield (A and B), the y-axis is truncated at $-\log_{10}(P) = 50$ for better visualization of secondary peaks; the strongest signal on chromosome 14, reaches a $-\log_{10}(P)$ value exceeding 250.

Comparison with existing studies revealed that of the 381 significant peaks identified in this study, our analysis newly identified 126 peaks (33%) that have not been previously reported in the Cattle QTLdb or our earlier work (Figure 2; Jiang et al., 2019a). The remaining 255 peaks (67%) had been previously documented, with 250 of these specifically reported in Holstein cattle (Supplemental Table S5, see Notes; Wang et al., 2025a). No novel peaks were identified for PY traits in this study. In contrast, we discovered 97 and 21 novel significant peaks for TY and LH traits, respectively. The remaining 8 novel peaks were associated with Net Merit. Furthermore, compared with our previous study using data from ~27k Holstein bulls (Jiang et al., 2019a), the increased scale in the present analysis led to

the discovery of 206 additional peaks, 80 of which had been previously documented in the Cattle QTLdb. Only 47 peaks from our previous study were not replicated at the genome-wide significance threshold ($P < 5 \times 10^{-8}$) in the current study, with 11 of these not replicated at the more lenient threshold ($P < 5 \times 10^{-5}$), likely due to sample changes and methodological differences in imputation and association procedures.

Fine-Mapping

The fine-mapping process using BFMAP involved identifying independent signals within each candidate region, computing PCIP for variants within these signals, and generating a 95% credible variant set for each

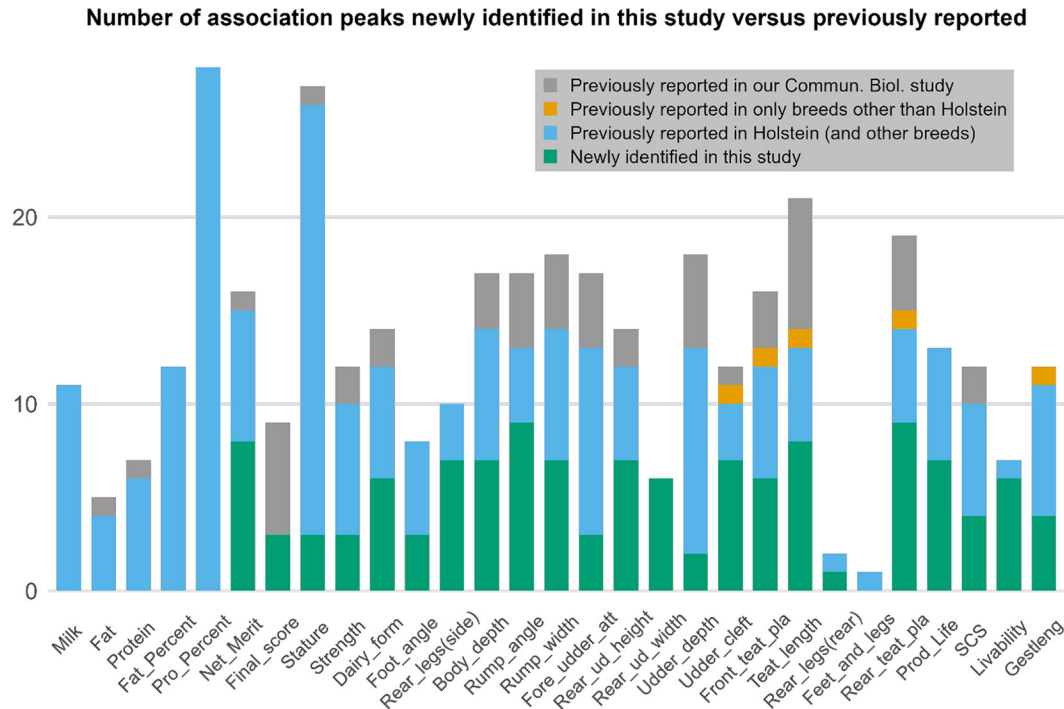


Figure 2. Comparison of peaks identified using a genome-wide significance threshold ($P < 5 \times 10^{-8}$) with previously reported associations across 29 traits. Calf Livability was excluded because no associations passed this threshold. Bars represent the total number of significant peaks, subdivided into 4 categories based on reporting status: (blue) reported in Cattle QTLdb for Holstein (alone or with other breeds); (orange) reported in Cattle QTLdb for non-Holstein breeds only; (gray) identified in our previous study (Jianget al., 2019) but absent from Cattle QTLdb; and (green) newly discovered in this study.

independent signal. Applying this process to the 2,113 candidate regions (identified by the inclusive GWAS significance threshold of $P < 5 \times 10^{-5}$) yielded 4,023 independent signals (Supplemental Tables S6–S8, see Notes; Wang et al., 2025a). We further filtered these signals using a fine-mapping P -value threshold of $< 5 \times 10^{-8}$, resulting in 792 high-confidence signals (Supplemental Table S4; Wang et al., 2025a). Among these signals, 376 SNP-trait pairs exhibited a PCIP > 0.5 , and 174 pairs had a PCIP > 0.9 , representing high-confidence candidate variants.

Candidate Genes

For each of the independent fine-mapped signals where the signal's lead variant fine-mapping P -value was $< 5 \times 10^{-5}$, we computed gene-level PCIP by aggregating variant PCIP for all variants within each associated gene (including 3 kb upstream/downstream flanks based on Ensembl locations) signals (Supplemental Table S9, see Notes; Wang et al., 2025a). This analysis yielded 2,038 gene-trait pairs with a gene-level PCIP > 0.5 .

To identify the most promising candidate genes, stringent filtering criteria were applied: a gene-level

PCIP of > 0.8 and, for the associated signal's lead variant, a fine-mapping P -value of $< 5 \times 10^{-8}$. This selection yielded a focused list of 229 unique candidate genes (Table 3), including several well-studied or repeatedly reported genes in cattle such as *DGAT1*, *ABCG2*, *GHR*, *GPIHBP1*, *ZNF623*, *ZC3H3*, *PLEC*, and *HSF1* (Arranz et al., 1998; Grisart et al., 2002; Cohen-Zinder et al., 2005; Viitala et al., 2006) for PY traits; *ABCC9*, *CCND2*, *ARRDC3*, *TMTC2*, and *IGF2* for TY traits (Saatchi et al., 2014; Weng et al., 2016; Seabury et al., 2017; Ghoreishifar et al., 2020; Cai et al., 2023a; Gualdrón Duarte et al., 2023; Schmidtman et al., 2023); and *BTBD9* for LH traits (Cai et al., 2023b). Notably, 9 candidate genes were identified within the newly discovered peaks for specific traits: *AOPEP* (Fore Udder Attachment), *GC* (Livability), *E2F6* (Livability), *MGST1* (Net Merit), *VPS13B* (Rear Udder Height), *ZNF652* (Rump Angle), *ASPH* (Rump Width), *SFMBT1* (Rump Width), and *MA-PRE2* (Teat Length). While these stringent criteria were applied for the prioritized candidate genes presented herein, less restrictive standards could also be applied to identify additional candidate genes with potentially weaker statistical evidence. The complete list is available in Supplemental Table S9 (Wang et al., 2025a).

Table 3. Highly credible candidate genes meeting stringent criteria (gene-level PCIP >0.8 and lead variant *P*-value <5 × 10⁻⁸)

Trait	Candidate genes (posterior conditional inclusion probability [PCIP])
Milk Yield	<i>ABCC9</i> (0.96), <i>ABCG2</i> (1.00), <i>ACO2</i> (0.92), <i>ARHGAP39</i> (0.97), <i>COX6C</i> (0.81), <i>DGAT1</i> (0.80), <i>GHR</i> (0.96), <i>GNAT3</i> (0.98), <i>KCNK3</i> (0.95), <i>KIAA0930</i> (0.93), <i>LPO</i> (0.94), <i>LRP5</i> (0.96), <i>RECQL</i> (1.00), <i>SLC4A4</i> (0.99), <i>SMPD5</i> (0.93), <i>SPATC1</i> (0.93), <i>VPS13B</i> (0.97)
Fat Yield	<i>DGAT1</i> (1.00), ENSBTAG0000004596 (1.00), ENSBTAG00000049400 (1.00), <i>FANCC</i> (0.95), <i>FASN</i> (0.97), <i>GPIHBP1</i> (0.86), <i>HSF1</i> (1.00), <i>IGF2</i> (0.99), <i>MGST1</i> (0.97), <i>PLEC</i> (1.00), <i>ST3GAL4</i> (0.93), <i>ZC3H3</i> (0.82)
Protein Yield	<i>ASTN1</i> (0.95), <i>C19H17orf49</i> (0.96), <i>CUX2</i> (1.00), ENSBTAG00000048091 (0.91), <i>FAM13A</i> (0.96), <i>FANCC</i> (0.96), <i>GCK</i> (0.95), <i>KCNN3</i> (0.96), <i>RECQL</i> (0.97), <i>TBC1D32</i> (0.98), <i>bta-mir-195</i> (0.96), <i>bta-mir-497</i> (0.96)
Fat Percentage	<i>ABCG2</i> (1.00), <i>ADGRB1</i> (1.00), <i>ARHGAP39</i> (1.00), <i>COX6C</i> (0.98), ENSBTAG00000049400 (0.95), <i>EPPK1</i> (0.98), <i>EPS8</i> (0.95), <i>GHR</i> (1.00), <i>GPIHBP1</i> (0.97), <i>KCNK3</i> (0.92), <i>MGST1</i> (0.99), <i>PLEC</i> (1.00), <i>PUF60</i> (1.00), <i>RNF217</i> (0.95), <i>SCRIB</i> (1.00), <i>TRAPPC9</i> (0.97), <i>VPS13B</i> (0.98), <i>ZC3H3</i> (0.98), <i>ZNF623</i> (0.98), <i>bta-mir-2285be</i> (0.91)
Protein Percentage	<i>ABO</i> (0.95), <i>ADCY6</i> (1.00), <i>C6</i> (0.98), <i>CAPN1</i> (1.00), <i>COX6C</i> (1.00), <i>CYHR1</i> (1.00), <i>DGAT1</i> (1.00), <i>EFNA1</i> (0.99), <i>GHR</i> (1.00), <i>GMD5</i> (0.96), <i>GPIHBP1</i> (1.00), <i>HERC3</i> (1.00), <i>HSD11B1</i> (0.99), <i>HSF1</i> (1.00), <i>IRF6</i> (0.95), <i>JAK2</i> (0.84), <i>MEPE</i> (0.99), <i>NNT</i> (1.00), <i>PAIP1</i> (1.00), <i>PKD2</i> (1.00), <i>PLEC</i> (0.95), <i>POP1</i> (0.84), <i>RAVER2</i> (1.00), <i>RNF217</i> (0.96), <i>RNF43</i> (1.00), <i>SLC35B4</i> (0.95), <i>SMIM13</i> (1.00), <i>SPX</i> (1.00), <i>TAF6L</i> (0.94), <i>TBC1D22A</i> (1.00), <i>TRIM46</i> (0.98), <i>VPS13B</i> (1.00), <i>WNT10B</i> (1.00), <i>ZNF250</i> (0.95), <i>ZNF623</i> (0.83)
Final Score	<i>ATAD2</i> (0.87), <i>CDYL2</i> (0.95), ENSBTAG00000054384 (0.96), <i>RDH8</i> (0.94), <i>TAFAI</i> (0.95)
Stature	<i>CCND2</i> (0.95), <i>CPEB3</i> (0.96), <i>DIS3L2</i> (0.95), ENSBTAG00000054384 (0.96), <i>ERICH4</i> (0.96), <i>ESR1</i> (0.96), <i>FSTL1</i> (0.98), <i>KMT5B</i> (0.99), <i>LHPP</i> (0.96), <i>NCOR2</i> (0.82), <i>NPM1</i> (1.00), <i>NRTN</i> (1.00), <i>RABEP1</i> (0.83), <i>RBFOX3</i> (1.00), <i>RNASEH2B</i> (0.96), <i>TMCO4</i> (0.94), <i>TTC32</i> (1.00)
Strength	<i>ABCC9</i> (0.97), <i>ARRDC3</i> (1.00), <i>CCND2</i> (0.88), <i>DHX34</i> (1.00), ENSBTAG0000004608 (0.84), ENSBTAG00000039491 (0.97), <i>KCNK9</i> (0.85), <i>PHF19</i> (0.80), <i>ST3GAL1</i> (0.96)
Dairy Form	<i>ABCC9</i> (0.96), <i>GC</i> (0.85), <i>NREP</i> (1.00)
Foot Angle	<i>DSC3</i> (0.87), <i>STAG3</i> (0.93), <i>TNNI2</i> (0.90)
Rear Legs (Side View)	<i>DNMT3A</i> (0.93), <i>MYO1D</i> (1.00), <i>SSH2</i> (0.82), <i>TSPAN9</i> (0.96)
Body Depth	<i>ARRDC3</i> (1.00), <i>CCND2</i> (0.96), <i>CHSY3</i> (1.00), ENSBTAG0000004608 (0.94), <i>GHRH</i> (0.93), <i>IGF1</i> (0.90)
Rump Angle	<i>ABCA5</i> (0.94), <i>CDH4</i> (0.98), <i>COLEC12</i> (0.97), ENSBTAG00000052955 (0.98), <i>FBN1</i> (0.96), <i>LIPE</i> (0.98), <i>MPP7</i> (0.95), <i>NAV3</i> (0.95), <i>ZNF652¹</i> (0.95)
Rump Width	<i>CCDC77</i> (0.80), <i>CCND2</i> (0.96), <i>DNAI2</i> (0.95), <i>GPRC5C</i> (0.95), <i>ZNF677</i> (0.96), <i>ASPH*</i> (0.96), <i>SFMBT1*</i> (0.95)
Fore Udder Attachment	<i>ABCC9</i> (0.99), <i>ARRDC3</i> (1.00), ENSBTAG00000053615 (0.95), <i>NEATC2</i> (0.95), <i>RUNX1</i> (0.99), <i>SOX5</i> (0.95), <i>WDR88</i> (0.97), <i>AOPEP¹</i> (0.92)
Rear Udder Height	<i>ARRDC3</i> (0.99), <i>CLIP2</i> (1.00), <i>KCNMA1</i> (0.95), <i>KLHL29</i> (0.99), <i>ODAD2</i> (0.95), <i>PRRX1</i> (1.00), <i>SLC24A3</i> (0.95), <i>ZNF423</i> (1.00), <i>VPS13B¹</i> (0.84)
Rear Udder Width	<i>ARID5B</i> (0.99), ENSBTAG00000054109 (0.80), <i>RIMS1</i> (0.95), <i>TMTC2</i> (0.94)
Udder Depth	<i>ABCC9</i> (0.97), <i>ARID4B</i> (0.96), <i>ARRDC3</i> (1.00), <i>CHD3</i> (0.88), <i>DOCK1</i> (1.00), ENSBTAG00000044837 (0.88), ENSBTAG00000050669 (0.95), ENSBTAG00000053793 (1.00), ENSBTAG00000054384 (0.98), <i>ESR1</i> (0.95), <i>FANCC</i> (0.96), <i>FDFI1</i> (0.96), <i>FOXP1</i> (0.95), <i>GC</i> (0.96), <i>IGF2</i> (1.00), <i>KIAA0930</i> (0.99), <i>LRP5</i> (1.00), <i>LSM14A</i> (0.92), <i>SEC23IP</i> (0.95), <i>SHANK2</i> (0.97)
Udder Cleft	<i>EIF3A</i> (1.00), ENSBTAG00000043641 (1.00), ENSBTAG00000049502 (0.95), <i>IQCAIL</i> (0.89), <i>SNORA19</i> (1.00), <i>TAF44</i> (0.95), <i>TSHZ3</i> (0.96), <i>ZMIZ1</i> (0.95)
Front Teat Placement	<i>AKAP8L</i> (0.85), <i>TMTC2</i> (0.95)
Teat Length	<i>AKAP10</i> (0.96), <i>AOPEP</i> (0.95), <i>ARRDC3</i> (0.81), <i>CARS1</i> (0.87), <i>CPED1</i> (0.89), <i>DEPDC5</i> (1.00), <i>JCAD</i> (0.96), <i>PHF12</i> (0.90), <i>PRKCZ</i> (0.91), <i>SFXN5</i> (0.87), <i>SLC35D2</i> (0.95), <i>TMTC2</i> (1.00), <i>WNT7A</i> (0.96), <i>MAPRE2*</i> (0.95)
Rear Legs (Rear View)	<i>CALCOCO2</i> (1.00)
Rear Teat Placement	<i>ACTN1</i> (0.95), <i>ADAM12</i> (0.96), <i>CHN1</i> (0.97), <i>IRX2</i> (0.99), <i>LMF1</i> (0.89), <i>SUSD6</i> (0.95), <i>VPS13B</i> (0.96)
Net Merit	<i>AGTR1</i> (0.95), <i>ARRDC3</i> (0.99), <i>CAPN7</i> (0.95), ENSBTAG00000048639 (0.95), <i>EXOC2</i> (0.92), <i>OLFM1</i> (0.99), <i>OTOGL</i> (0.85), <i>PLEC</i> (1.00), <i>TBX2</i> (0.83), <i>MGST1¹</i> (0.95)
Productive Life	<i>ABCC9</i> (0.95), <i>BTBD9</i> (0.96), <i>CHST8</i> (1.00), <i>DYRK4</i> (0.95), <i>ERCC2</i> (0.96), <i>FRMD5</i> (0.96), <i>GC</i> (0.96), <i>HELB</i> (0.95), <i>KLC3</i> (0.96), <i>MFS14B</i> (1.00), <i>SEMA4D</i> (0.89), <i>bta-mir-2285cr-2</i> (0.95)
SCS	<i>APCDD1L</i> (0.97), <i>CMIP</i> (0.97), <i>CTIF</i> (0.95), ENSBTAG00000055157 (0.92), <i>INHCA</i> (0.97), <i>LARS2</i> (1.00), <i>MTSSI</i> (1.00), <i>NPNT</i> (0.96), <i>ROCK2</i> (0.97), <i>SEPTIN9</i> (0.95), <i>SLC7A9</i> (0.94)
Livability	<i>BTBD9</i> (0.91), ENSBTAG0000004608 (0.96), ENSBTAG00000049494 (0.96), <i>LDHB</i> (0.99), <i>MGC137454</i> (0.97), <i>PDE4B</i> (0.97), <i>TRIP11</i> (0.91), <i>E2F6¹</i> (0.85), <i>GC¹</i> (0.97)
Gestation Length	<i>ADAMTS2</i> (1.00), <i>AKT1</i> (1.00), <i>BCO1</i> (0.91), <i>CFAP61</i> (0.87), <i>CHD3</i> (1.00), <i>CMIP</i> (0.95), ENSBTAG00000044837 (1.00), ENSBTAG00000051766 (0.98), <i>GPS1</i> (0.97), <i>HNRNP1</i> (0.88), <i>KDM7A</i> (0.99), <i>MGC137454</i> (0.90), <i>NDFIPI</i> (1.00), <i>SIVA1</i> (1.00), <i>TRAM2</i> (1.00), <i>ZNF532</i> (0.90)

¹Denotes candidate genes within newly reported peaks in the current study.

DISCUSSION

In this study, we performed GWAS on a large cohort of 50,309 Holstein bulls using imputed sequence data comprising 11,292,243 quality-controlled variants and subsequently applied Bayesian fine-mapping to identify causal variants and genes underlying 30 complex

traits. This large-scale analysis significantly enhanced statistical power for detecting genetic associations and improved the precision of genetic-mapping in identifying causal variants. The absence of novel peaks for PY traits is likely due to extensive prior research on these economically critical traits, while numerous novel peaks were discovered for TY and LH traits, which have his-

torically received less research attention. Our GWAS optimization revealed important considerations for conducting association studies in dairy cattle populations. The GRM construction analysis demonstrated that using SNPs from genomic evaluation-centric chip panels can lead to deflated association statistics, likely due to the ascertainment bias of these markers toward large-effect loci for economic traits (VanRaden et al., 2017). As demonstrated by our findings, a randomly selected SNP subset from imputed sequence data of an appropriate size offered a better balance between maintaining statistical power and controlling genomic inflation in test statistics.

A key aspect of our GWAS model was the incorporation of deregressed PTA reliability into the residual term. By weighting residuals inversely by reliability, pseudophenotypes with lower accuracy exert less influence on the association statistics. The effect of modeling reliability varied with trait characteristics, specifically the variation in reliability among individuals. While traits exhibiting low variance in reliability across animals (often those with universally high reliability) showed minimal differences in association patterns between the 2 models, traits with higher variance in reliability (where individuals differ substantially in PTA accuracy) demonstrated substantial improvements in peak detection when reliability was incorporated. Incorporating reliability into GWAS improved model fitting and increased statistical power. These findings underscore the necessity of modeling reliability in GWAS when deregressed PTA serve as pseudophenotypes, particularly for traits where individual PTA reliabilities vary considerably, such as in analyses that include both bulls and cows.

Most existing fine-mapping methods (Chen et al., 2015; Benner et al., 2016; Zou et al., 2022) are based on linear regression, which assumes samples are largely unrelated. This assumption is suitable for human datasets, and the linear regression framework simplifies the utilization of summary statistics for fine-mapping. These methods have often been used in farm animal fine-mapping studies (Li et al., 2020; Gualdrón Duarte et al., 2023; Qiu et al., 2024); unfortunately, such use violates the assumption in livestock populations where individuals are generally closely related, which can lead to poor fine-mapping performance, as demonstrated in our other study (Wang et al., 2025b). To address this limitation, the current study employed BFMAP, which uses a linear mixed model framework that accounts for the relatedness among individuals. Our analysis produced a comprehensive list of candidate variants and genes that serve as promising targets for future functional validation studies. In contrast to most conventional GWAS, which typically reported associated SNPs or genes near a peak without formal statistical prioritization, our study offers a systematic prioritization derived from Bayesian

fine-mapping, yielding statistically supported candidate variants and genes.

Traditional GWAS in farm animals typically nominate candidate genes by proximity to the lead SNP, without statistical prioritization of causality (Lung et al., 2019; Jiang et al., 2019b; Gao et al., 2023). This strategy typically yields broad candidate lists, which can hinder efficient functional follow-up investigations. In contrast, our fine-mapping strategy substantially narrows down this range by providing direct statistical evidence for candidate variants and genes. For example, *MGST1* (Chr5:93,497,064-93,521,047) has been repeatedly reported to be associated with milk composition traits, but conventional GWAS would report it alongside numerous neighboring genes (Cai et al., 2020; Teng et al., 2023). Our study provides direct statistical evidence through gene prioritization showing that *MGST1* emerges as the most credible candidate gene (PCIP = 0.99 for Fat Percent) in this genomic region. By calculating gene-level PCIP, a direct probabilistic link to genes was established, effectively transitioning from variant-level evidence to gene-level inferences. Our previous simulation analysis indicated that accurately fine-mapping individual causal variants in farm animals remains challenging in maintaining a low false discovery rate while preserving statistical power, largely due to the strong linkage disequilibrium. Gene-level analysis, however, achieves a more favorable balance between power and precision, making it a more robust approach for causal inference in farm animal populations (Wang et al., 2025b). The stringent criteria employed herein (gene-level PCIP > 0.8 and lead variant P -value < 5×10^{-8}) facilitated the prioritization of the candidate genes with strong statistical support, resulting in a narrow set of high-confidence candidate genes rather than the typically extensive lists generated by proximity-based approaches. Despite our GWAS detecting 126 novel peaks, only 9 novel genes met our stringent criteria, indicating that many newly discovered loci exhibit weaker signals that do not firmly establish causal relationships. It is worth noting that the criteria used to select candidate genes were not intended as an absolute gold standard but rather as a conservative approach to minimize false positives. Employing more lenient criteria could potentially identify additional candidate genes with weaker evidence, offering a broader scope for further investigation.

The 9 novel candidate genes identified through our stringent fine-mapping approach represent newly identified candidate genes for their respective traits and warrant further investigation for their roles in dairy traits. Amino-peptidase O (*AOPEP*), a candidate gene for Fore Udder Attachment, encodes a zinc-dependent metallopeptidase, and specific pathogenic biallelic loss-of-function variants in this gene cause autosomal recessive dystonia in

humans (Garavaglia et al., 2022; Zech et al., 2022). Vacuolar protein sorting 13 homolog B (*VPS13B*), identified as a candidate gene for Rear Udder Height, encodes a large transmembrane protein involved in vesicular trafficking. In humans, pathogenic mutations of this gene cause Cohen syndrome (Duplomb et al., 2019; Momtazmanesh et al., 2020), and in mice, *Vps13b* knockouts cause male infertility through Golgi disruption (Nagata et al., 2018). Our study also identified this gene as the candidate gene for milk production traits, suggesting pleiotropic effects or involvement in related functional pathways influencing these traits in dairy cattle. Microsomal glutathione S-transferase 1 (*MGST1*) was identified in our study as a candidate gene for Fat Yield, Fat Percentage, and Net Merit. It was detected in a newly identified peak for Net Merit, suggesting the economic significance of this gene in dairy cattle breeding programs. Vitamin D binding protein (*GC*) and E2F transcription factor 6 (*E2F6*) were prioritized for Livability. The *GC* gene encodes the vitamin D binding protein involved in vitamin D transport and immune function (White and Cooke, 2000). While previous fine-mapping studies have implicated *GC* as a candidate gene for clinical mastitis resistance in cattle (Olsen et al., 2016; Freebern et al., 2020; Lee et al., 2021), our study provides statistical evidence supporting *GC* as a novel candidate gene for Livability. The *E2F6* gene encodes a transcriptional repressor. Mice lacking *E2f6* exhibit homeotic axial skeletal transformations and testicular histological abnormalities but retain fertility (Storre et al., 2002). Zinc finger protein 652 (*ZNF652*), candidate gene for Rump Angle, encodes a C2H2-type zinc finger transcriptional repressor that regulates gene networks in cell differentiation (Kumar et al., 2010) and functions as a tumor suppressor in human lung adenocarcinoma by arresting cells in G₁ and thus reducing cell proliferation (Xie et al., 2024). For Rump Width, 2 genes, aspartate β -hydroxylase (*ASPH*) and scm like with 4 mbt domains 1 (*SFMBT1*), were prioritized. The *SFMBT1* gene encodes a chromatin-binding repressor that partners with histone-modifying enzymes to silence key histone genes, regulate myogenic differentiation, and function during spermatogenesis (Lin et al., 2013; Zhang et al., 2013). The *ASPH* gene encodes an enzyme in the endoplasmic reticulum that modifies calcium-binding proteins to help cells maintain proper calcium balance. Pathogenic *ASPH* mutations in humans cause Traboulsi syndrome, which features lens dislocation and facial dysmorphism (Pfeffer et al., 2019; Jones et al., 2022). Microtubule-associated protein RP/EB family member 2 (*MAPRE2*), a candidate gene for Teat Length, encodes a conserved microtubule plus-end tracking protein that stabilizes spindle microtubules during oocyte meiosis. In mouse oocytes, knockdown of *Mapre2* using siRNA-disrupts kinetochore-microtubule attachments, activates

the spindle-assembly checkpoint, and prevents first polar body extrusion (Li et al., 2022).

While this study benefited from a large sample size and sequence-level variant data, certain limitations should be acknowledged. First, our analysis focused exclusively on quality-controlled, imputed biallelic SNPs. Consequently, we did not investigate the potential contributions of other important classes of genetic variation, such as multiallelic SNPs, insertions-deletions (indels), and larger structural variants, which may also influence the complex traits studied and warrant investigation in future research. Second, while imputation enables genome-wide coverage, the accuracy of imputed genotypes may vary, particularly for lower-frequency variants or variants located in regions with complex linkage disequilibrium patterns or limited coverage in the reference panel. This inherent uncertainty associated with imputation could potentially influence the precision of fine-mapping results, potentially affecting the definitive identification of causal variants or the exact composition of credible sets. Acknowledging these limitations is crucial for interpreting and contextualizing the study's findings accurately.

CONCLUSIONS

Our study underscores the necessity of optimizing GRM construction and incorporating reliability information in GWAS using deregressed PTA, particularly for traits with high variation in reliability across individuals. By leveraging a large cohort of Holstein bulls and imputed sequence data, our comprehensive GWAS and Bayesian fine-mapping analysis identified many novel associations and prioritized candidate variants and genes for 30 complex traits in dairy cattle. These findings provide new insights into the genetic architecture of production, conformation, and health traits in dairy cattle and serve as a valuable resource for further functional validation and enhancement of breeding strategies to improve dairy productivity and health.

NOTES

We thank the CDCB (Bowie, MD), CDDR (Madison, WI), and dairy industry contributors for providing data access. We also thank Paul M. VanRaden for his valuable comments. This work is supported by the Agriculture and Food Research Initiative (AFRI) Foundational and Applied Science Program, project award no. 2021-67015-33409, 2023-67015-39260, 2024-67015-42295, and the Research Capacity Fund (HATCH), project award no. 7008128, from the USDA's National Institute of Food and Agriculture (Washington, DC). This research used resources provided by the SCINet project and the AI Center of Excellence of the USDA Agricultural Research

Service, ARS project 0201-88888-003-000D and 0201-88888-002-000D. The sole ownership and rights of the data remain with the producer, and we express our gratitude to US dairy producers for sharing their data for research purposes. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. The USDA is an equal opportunity provider and employer. Fine-mapping summary statistics of all 30 dairy traits, Supplemental Tables, and Supplemental Figures are available in the Dryad Digital Repository (<https://doi.org/10.5061/dryad.vmcvndnd3q>). This repository is temporarily private during peer review. Reviewers can access these materials using the following link: http://datadryad.org/share/tTt2BC_5m1NRqjxv1AulFbIDzSIs85hDUZnlA525TRw. No human or animal subjects were used, so this analysis did not require approval by an Institutional Animal Care and Use Committee or Institutional Review Board. The authors have not stated any conflicts of interest.

Nonstandard abbreviations used: CDCB = Council on Dairy Breeding; CDDR = Cooperative Dairy DNA Repository; GRM = genomic relationship matrix; LH = longevity and health; PCIP = posterior conditional inclusion probability; PY = production and yield; TY = type.

REFERENCES

- Arranz, J.-J., W. Coppieters, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marq, L. Moreau, C. Mezer, J. Riquet, P. Simon, P. Vanmanshoven, D. Wagenaar, and M. Georges. 1998. A QTL affecting milk yield and composition maps to bovine chromosome 20: A confirmation. *Anim. Genet.* 29:107–115. <https://doi.org/10.1046/j.1365-2052.1998.00307.x>.
- Benner, C., C. C. A. Spencer, A. S. Havulinna, V. Salomaa, S. Ripatti, and M. Pirinen. 2016. FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* 32:1493–1501. <https://doi.org/10.1093/bioinformatics/btw018>.
- Cai, Z., M. Duszka, B. Gulbrandsen, M. S. Lund, and G. Sahana. 2020. Distinguishing pleiotropy from linked QTL between milk production traits and mastitis resistance in Nordic Holstein cattle. *Genet. Sel. Evol.* 52:19. <https://doi.org/10.1186/s12711-020-00538-6>.
- Cai, W., Y. Zhang, T. Chang, Z. Wang, B. Zhu, Y. Chen, X. Gao, L. Xu, L. Zhang, H. Gao, J. Song, and J. Li. 2023a. The eQTL colocalization and transcriptome-wide association study identify potentially causal genes responsible for economic traits in Simmental beef cattle. *J. Anim. Sci. Biotechnol.* 14:78. <https://doi.org/10.1186/s40104-023-00876-7>.
- Cai, Z., X. Wu, B. Thomsen, M. S. Lund, and G. Sahana. 2023b. Genome-wide association study identifies functional genomic variants associated with young stock survival in Nordic Red Dairy Cattle. *J. Dairy Sci.* 106:7832–7845. <https://doi.org/10.3168/jds.2023-23252>.
- Chen, W., B. R. Larrabee, I. G. Ovsyannikova, R. B. Kennedy, I. H. Haralambieva, G. A. Poland, and D. J. Schaid. 2015. Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. *Genetics* 200:719–736. <https://doi.org/10.1534/genetics.115.176107>.
- Cheng, J., C. Maltecca, P. M. VanRaden, J. R. O'Connell, L. Ma, and J. Jiang. 2023. SLEMM: Million-scale genomic predictions with window-based SNP weighting. *Bioinformatics* 39:btad127. <https://doi.org/10.1093/bioinformatics/btad127>.
- Cohen-Zinder, M., E. Seroussi, D. M. Larkin, J. J. Loo, A. E. Wind, J.-H. Lee, J. K. Drackley, M. R. Band, A. G. Hernandez, M. Shani, H. A. Lewin, J. I. Weller, and M. Ron. 2005. Identification of a missense mutation in the bovine *ABCG2* gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* 15:936–944. <https://doi.org/10.1101/gr.3806705>.
- de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein–Friesian, Jersey and Angus cattle. *Genetics* 179:1503–1512. <https://doi.org/10.1534/genetics.107.084301>.
- Duplomb, L., J. Rivière, G. Jego, R. Da Costa, A. Hammann, J. Racine, A. Schmitt, N. Droin, C. Capron, M.-A. Gougerot-Pocidalo, L. Dubrez, B. Aral, A. Lafon, P. Edery, J. Ghoumid, E. Blair, S. El Chehadeh-Djebbar, V. Carmignac, J. Thevenon, J. Guy, F. Girodon, J.-N. Bastie, L. Delva, L. Faivre, C. Chauvin-Robinet, and E. Solary. 2019. Serpin B1 defect and increased apoptosis of neutrophils in Cohen syndrome neutropenia. *J. Mol. Med. (Berl.)* 97:633–645. <https://doi.org/10.1007/s00109-019-01754-4>.
- Freebern, E., D. J. A. Santos, L. Fang, J. Jiang, K. L. Parker Gaddis, G. E. Liu, P. M. VanRaden, C. Maltecca, J. B. Cole, and L. Ma. 2020. GWAS and fine-mapping of livability and six disease traits in Holstein cattle. *BMC Genomics* 21:41. <https://doi.org/10.1186/s12864-020-6461-z>.
- Gao, Y., A. Marceau, V. Iqbal, J. A. Torres-Vázquez, M. Neupane, J. Jiang, G. E. Liu, and L. Ma. 2023. Genome-wide association analysis of heifer livability and early first calving in Holstein cattle. *BMC Genomics* 24:628. <https://doi.org/10.1186/s12864-023-09736-0>.
- Garavaglia, B., S. Vallian, L. M. Romito, G. Straccia, M. Capecci, F. Invernizzi, E. Andrenelli, A. Kazemi, S. Boesch, R. Kopajtich, N. Olfati, M. Shariati, A. Shoeibi, A. Sadr-Nabavi, H. Prokisch, J. Winkelmann, and M. Zech. 2022. *AOPEP* variants as a novel cause of recessive dystonia: Generalized dystonia and dystonia-parkinsonism. *Parkinsonism Relat. Disord.* 97:52–56. <https://doi.org/10.1016/j.parkreldis.2022.03.007>.
- García-Ruiz, A., J. B. Cole, P. M. VanRaden, G. R. Wiggins, F. J. Ruiz-López, and C. P. Van Tassell. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. USA* 113:E3995–E4004. <https://doi.org/10.1073/pnas.1519061113>.
- Ghoreishifar, S. M., S. Eriksson, A. M. Johansson, M. Khansefid, S. Moghaddaszadeh-Ahrabi, N. Parna, P. Davoudi, and A. Javanmard. 2020. Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five Swedish cattle breeds. *Genet. Sel. Evol.* 52:52. <https://doi.org/10.1186/s12711-020-00571-5>.
- Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.* 10:381–391. <https://doi.org/10.1038/nrg2575>.
- Grisart, B., W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Res.* 12:222–231. <https://doi.org/10.1101/gr.224202>.
- Gualdrón Duarte, J. L., C. Yuan, A.-S. Gori, G. C. M. Moreira, H. Takeda, W. Coppieters, C. Charlier, M. Georges, and T. Druet. 2023. Sequenced-based GWAS for linear classification traits in Belgian Blue beef cattle reveals new coding variants in genes regulating body size in mammals. *Genet. Sel. Evol.* 55:83. <https://doi.org/10.1186/s12711-023-00857-4>.
- Guinan, F. L., G. R. Wiggins, H. D. Norman, J. W. Dürr, J. B. Cole, C. P. Van Tassell, I. Misztal, and D. Lourenco. 2023. Changes in genetic trends in US dairy cattle since the implementation of genomic selection. *J. Dairy Sci.* 106:1110–1129. <https://doi.org/10.3168/jds.2022-22205>.
- Hayes, B. J., and H. D. Daetwyler. 2019. 1000 Bull Genomes Project to map simple and complex genetic traits in cattle: Applications and outcomes. *Annu. Rev. Anim. Biosci.* 7:89–102. <https://doi.org/10.1146/annurev-animal-020518-115024>.

- Hu, Z.-L., C. A. Park, and J. M. Reecy. 2022. Bringing the Animal QTLdb and CorrDB into the future: Meeting new challenges and providing updated services. *Nucleic Acids Res.* 50(D1):D956–D961. <https://doi.org/10.1093/nar/gkab1116>.
- Jung, L. H. S., J. Petrini, J. Ramirez-Diaz, M. Salvian, G. A. Rovadoscki, F. Pilonetto, B. D. Dauria, P. F. Machado, L. L. Coutinho, G. R. Wiggans, and G. B. Mourão. 2019. Genome-wide association study for milk production traits in a Brazilian Holstein population. *J. Dairy Sci.* 102:5305–5314. <https://doi.org/10.3168/jds.2018-14811>.
- Jiang, J., J. B. Cole, E. Freebern, Y. Da, P. M. VanRaden, and L. Ma. 2019a. Functional annotation and Bayesian fine-mapping reveals candidate genes for important agronomic traits in Holstein bulls. *Commun. Biol.* 2:212. <https://doi.org/10.1038/s42003-019-0454-y>.
- Jiang, J., L. Ma, D. Prakapenka, P. M. VanRaden, J. B. Cole, and Y. Da. 2019b. A large-scale genome-wide association study in U.S. Holstein cattle. *Front. Genet.* 10:412. <https://doi.org/10.3389/fgene.2019.00412>.
- Jones, G., K. Johnson, J. Eason, M. Hamilton, D. Osio, F. Kanani, J. Baptista, and M. Suri. 2022. Traboulsi syndrome caused by mutations in ASPH: An autosomal recessive disorder with overlapping features of Marfan syndrome. *Eur. J. Med. Genet.* 65:104572. <https://doi.org/10.1016/j.ejmg.2022.104572>.
- Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler. 2002. The human genome browser at UCSC. *Genome Res.* 12:996–1006. <https://doi.org/10.1101/gr.229102>.
- Kim, E.-S., and B. W. Kirkpatrick. 2009. Linkage disequilibrium in the North American Holstein population. *Anim. Genet.* 40:279–288. <https://doi.org/10.1111/j.1365-2052.2008.01831.x>.
- Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* 9:29. <https://doi.org/10.1186/1746-4811-9-29>.
- Kumar, R., K. M. Cheney, P. M. Neilsen, R. B. Schulz, and D. F. Callen. 2010. CBFA2T3–ZNF651, like CBFA2T3–ZNF652, functions as a transcriptional corepressor complex. *FEBS Lett.* 584:859–864. <https://doi.org/10.1016/j.febslet.2010.01.047>.
- Lee, Y.-L., H. Takeda, G. Costa Monteiro Moreira, L. Karim, E. Mullaart, W. Coppieters, R. Appeltant, R. F. Veerkamp, M. A. M. Groenen, M. Georges, M. Bosse, T. Druet, A. C. Bouwman, and C. Charlier. 2021. A 12 kb multi-allelic copy number variation encompassing a GC gene enhancer is associated with mastitis resistance in dairy cattle. *PLoS Genet.* 17:e1009331. <https://doi.org/10.1371/journal.pgen.1009331>.
- Li, Y., B. Li, M. Yang, H. Han, T. Chen, Q. Wei, Z. Miao, L. Yin, R. Wang, J. Shen, X. Li, X. Xu, M. Fang, and S. Zhao. 2020. Genome-wide association study and fine mapping reveals candidate genes for birth weight of Yorkshire and Landrace pigs. *Front. Genet.* 11:183. <https://doi.org/10.3389/fgene.2020.00183>.
- Li, Y.-Y., W.-L. Lei, C.-F. Zhang, S.-M. Sun, B.-W. Zhao, K. Xu, Y. Hou, Y.-C. Ouyang, Z.-B. Wang, L. Guo, Q.-Y. Sun, and Z. Han. 2022. MAPRE2 regulates the first meiotic progression in mouse oocytes. *Exp. Cell Res.* 416:113135. <https://doi.org/10.1016/j.yexcr.2022.113135>.
- Lin, S., H. Shen, J.-L. Li, S. Tang, Y. Gu, Z. Chen, C. Hu, J. C. Rice, J. Lu, and L. Wu. 2013. Proteomic and functional analyses reveal the role of chromatin reader SFMBT1 in regulating epigenetic silencing and the myogenic gene program. *J. Biol. Chem.* 288:6238–6247. <https://doi.org/10.1074/jbc.M112.429605>.
- McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. A. Ioannidis, and J. N. Hirschhorn. 2008. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9:356–369. <https://doi.org/10.1038/nrg2344>.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. <https://doi.org/10.1093/genetics/157.4.1819>.
- Miglior, F., A. Fleming, F. Malchiodi, L. F. Brito, P. Martin, and C. F. Baes. 2017. A 100-Year Review: Identification and genetic selection of economically important traits in dairy cattle. *J. Dairy Sci.* 100:10251–10271. <https://doi.org/10.3168/jds.2017-12968>.
- Momtazmanesh, S., E. Rayzan, S. Shahkarami, M. Rohlf, C. Klein, and N. Rezaei. 2020. A novel VPS13B mutation in Cohen syndrome: A case report and review of literature. *BMC Med. Genet.* 21:140. <https://doi.org/10.1186/s12881-020-01075-1>.
- Nagata, O., M. Nakamura, H. Sakimoto, Y. Urata, N. Sasaki, N. Shio-kawa, and A. Sano. 2018. Mouse model of chorea-acanthocytosis exhibits male infertility caused by impaired sperm motility as a result of ultrastructural morphological abnormalities in the mitochondrial sheath in the sperm midpiece. *Biochem. Biophys. Res. Commun.* 503:915–920. <https://doi.org/10.1016/j.bbrc.2018.06.096>.
- Olsen, H. G., T. M. Knutsen, A. M. Lewandowska-Sabat, H. Grove, T. Nome, M. Svendsen, M. Arnyasi, M. Sodeland, K. K. Sundsaasen, S. R. Dahl, B. Heringstad, H. H. Hansen, I. Olsaker, M. P. Kent, and S. Lien. 2016. Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the *group-specific component (GC)* gene in clinical mastitis and milk production. *Genet. Sel. Evol.* 48:79. <https://doi.org/10.1186/s12711-016-0257-2>.
- Pfeffer, I., L. Brewitz, T. Krojer, S. A. Jensen, G. T. Kochan, N. J. Kershaw, K. S. Hewitson, L. A. McNeill, H. Kramer, M. Münzel, R. J. Hopkinson, U. Oppermann, P. A. Handford, M. A. McDonough, and C. J. Schofield. 2019. Aspartate/asparagine- β -hydroxylase crystal structures reveal an unexpected epidermal growth factor-like domain substrate disulfide pattern. *Nat. Commun.* 10:4910. <https://doi.org/10.1038/s41467-019-12711-7>.
- Qiu, Z., W. Cai, Q. Liu, K. Liu, C. Liu, H. Yang, R. Huang, P. Li, and Q. Zhao. 2024. Unravelling novel and pleiotropic genes for cannon bone circumference and bone mineral density in Yorkshire pigs. *J. Anim. Sci.* 102:skae036. <https://doi.org/10.1093/jas/skae036>.
- Rosen, B. D., D. M. Bickhart, R. D. Schnabel, S. Koren, C. G. Elsik, E. Tseng, T. N. Rowan, W. Y. Low, A. Zimin, C. Couldrey, R. Hall, W. Li, A. Rhie, J. Ghurye, S. D. McKay, F. Thibaud-Nissen, J. Hoffman, B. M. Murdoch, W. M. Snelling, T. G. McDanel, J. A. Hammond, J. C. Schwartz, W. Nandolo, D. E. Hagen, C. Dreischer, S. J. Schulteis, S. G. Schroeder, A. M. Phillippy, J. B. Cole, C. P. Van Tassel, G. Liu, T. P. L. Smith, and J. F. Medrano. 2020. De novo assembly of the cattle reference genome with single-molecule sequencing. *Gigascience* 9:giaa021. <https://doi.org/10.1093/gigascience/giaa021>.
- Rubinacci, S., O. Delaneau, and J. Marchini. 2020. Genotype imputation using the Positional Burrows Wheeler Transform. *PLoS Genet.* 16:e1009049. <https://doi.org/10.1371/journal.pgen.1009049>.
- Saatchi, M., R. D. Schnabel, J. F. Taylor, and D. J. Garrick. 2014. Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics* 15:442. <https://doi.org/10.1186/1471-2164-15-442>.
- Schmidtman, C., D. Segelke, J. Bennewitz, J. Tetens, and G. Thaller. 2023. Genetic analysis of production traits and body size measurements and their relationships with metabolic diseases in German Holstein cattle. *J. Dairy Sci.* 106:421–438. <https://doi.org/10.3168/jds.2022-22363>.
- Seabury, C. M., D. L. Oldeschulte, M. Saatchi, J. E. Beever, J. E. Decker, Y. A. Halley, E. K. Bhattarai, M. Molaei, H. C. Freely, S. L. Hansen, H. Yampara-Iquise, K. A. Johnson, M. S. Kerley, J. Kim, D. D. Loy, E. Marques, H. L. Neibergs, R. D. Schnabel, D. W. Shike, M. L. Spangler, R. L. Weaver, D. J. Garrick, and J. F. Taylor. 2017. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* 18:386. <https://doi.org/10.1186/s12864-017-3754-y>.
- Storre, J., H.-P. Elsässer, M. Fuchs, D. Ullmann, D. M. Livingston, and S. Gaubatz. 2002. Homeotic transformations of the axial skeleton that accompany a targeted deletion of *E2f6*. *EMBO Rep.* 3:695–700. <https://doi.org/10.1093/embo-reports/kvf141>.
- Stram, D. O. 2004. Tag SNP selection for association studies. *Genet. Epidemiol.* 27:365–374. <https://doi.org/10.1002/gepi.20028>.
- Teng, J., D. Wang, C. Zhao, X. Zhang, Z. Chen, J. Liu, D. Sun, H. Tang, W. Wang, J. Li, C. Mei, Z. Yang, C. Ning, and Q. Zhang. 2023. Longitudinal genome-wide association studies of milk production traits in Holstein cattle using whole-genome sequence data imputed from medium-density chip data. *J. Dairy Sci.* 106:2535–2550. <https://doi.org/10.3168/jds.2022-22277>.
- VanRaden, P. M., C. Van Tassel, G. Wiggans, T. Sonstegard, R. Schnabel, J. Taylor, and F. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92:16–24. <https://doi.org/10.3168/jds.2008-1514>.

- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>.
- VanRaden, P. M. 2016. Practical implications for genetic modeling in the genomics era. *J. Dairy Sci.* 99:2405–2412. <https://doi.org/10.3168/jds.2015-10038>.
- VanRaden, P. M., J. R. O’Connell, G. R. Wiggans, and K. A. Weigel. 2011. Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43:10. <https://doi.org/10.1186/1297-9686-43-10>.
- VanRaden, P. M., M. E. Tooker, J. R. O’Connell, J. B. Cole, and D. M. Bickhart. 2017. Selecting sequence variants to improve genomic predictions for dairy cattle. *Genet. Sel. Evol.* 49:32. <https://doi.org/10.1186/s12711-017-0307-4>.
- Viitala, S., J. Szyda, S. Blott, N. Schulman, M. Lidauer, A. Mäki-Tanila, M. Georges, and J. Vilkkii. 2006. The role of the bovine growth hormone receptor and prolactin receptor genes in milk, fat and protein production in Finnish Ayrshire dairy cattle. *Genetics* 173:2151–2164. <https://doi.org/10.1534/genetics.105.046730>.
- Wang, J., Y. Gao, L. Ma, and J. Jiang. 2025a. Supplemental material for: Genome-Wide Association Study and Bayesian Fine-Mapping for 30 Production, Conformation and Health Traits in 50,309 Holstein Cattle.
- Weigel, K. A., P. M. VanRaden, H. D. Norman, and H. Grosu. 2017. *A 100-Year Review: Methods and impact of genetic selection in dairy cattle—From daughter–dam comparisons to deep learning algorithms*. *J. Dairy Sci.* 100:10234–10250. <https://doi.org/10.3168/jds.2017-12954>.
- Weller, J. I., E. Ezra, and M. Ron. 2017. *Invited review: A perspective on the future of genomic selection in dairy cattle*. *J. Dairy Sci.* 100:8633–8644. <https://doi.org/10.3168/jds.2017-12879>.
- Weng, Z., H. Su, M. Saatchi, J. Lee, M. G. Thomas, J. R. Dunkelberger, and D. J. Garrick. 2016. Genome-wide association study of growth and body composition traits in Brangus beef cattle. *Livest. Sci.* 183:4–11. <https://doi.org/10.1016/j.livsci.2015.11.011>.
- White, P., and N. Cooke. 2000. The multifunctional properties and characteristics of vitamin D-binding protein. *Trends Endocrinol. Metab.* 11:320–327. [https://doi.org/10.1016/S1043-2760\(00\)00317-9](https://doi.org/10.1016/S1043-2760(00)00317-9).
- Winter, A., W. Krämer, F. A. O. Werner, S. Kollers, S. Kata, G. Durstewitz, J. Buitkamp, J. E. Womack, G. Thaller, and R. Fries. 2002. Association of a lysine-232/alanine polymorphism in a bovine gene encoding acyl-CoA:diacylglycerol acyltransferase (*DGAT1*) with variation at a quantitative trait locus for milk fat content. *Proc. Natl. Acad. Sci. USA* 99:9300–9305. <https://doi.org/10.1073/pnas.142293799>.
- Xie, C., X. Zhou, J. Wu, W. Chen, D. Ren, C. Zhong, Z. Meng, Y. Shi, and J. Zhu. 2024. ZNF652 exerts a tumor suppressor role in lung cancer by transcriptionally downregulating cyclin D3. *Cell Death Dis.* 15:792. <https://doi.org/10.1038/s41419-024-07197-1>.
- Zech, M., K. R. Kumar, S. Reining, J. Reunert, M. Tchan, L. G. Riley, A. P. Drew, R. J. Adam, R. Berutti, S. Biskup, N. Derive, S. Bakhtiari, S. C. Jin, M. C. Kruer, T. Bardakjian, P. Gonzalez-Alegre, I. J. Keller Sarmiento, N. E. Mencacci, S. J. Lubbe, M. A. Kurian, F. Clot, A. Méneret, J.-M. de Sainte Agathe, V. S. C. Fung, M. Vidailhet, M. Baumann, T. Marquardt, J. Winkelmann, and S. Boesch. 2022. Biallelic *AOPEP* loss-of-function variants cause progressive dystonia with prominent limb involvement. *Mov. Disord.* 37:137–147. <https://doi.org/10.1002/mds.28804>.
- Zhang, J., R. Bonasio, F. Strino, Y. Kluger, J. K. Holloway, A. J. Modzelewski, P. E. Cohen, and D. Reinberg. 2013. SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors. *Genes Dev.* 27:749–766. <https://doi.org/10.1101/gad.210963.112>.
- Zou, Y., P. Carbonetto, G. Wang, and M. Stephens. 2022. Fine-mapping from summary data with the “Sum of Single Effects” model. *PLoS Genet.* 18:e1010299. <https://doi.org/10.1371/journal.pgen.1010299>.

ORCID

- Sajjad Toghiani,  <https://orcid.org/0000-0002-5090-728X>
 John B. Cole,  <https://orcid.org/0000-0003-1242-4401>
 Christian Maltecca,  <https://orcid.org/0000-0002-9996-4680>
 Li Ma,  <https://orcid.org/0000-0003-1038-1081>
 Jicai Jiang  <https://orcid.org/0000-0001-6890-7539>