J. Dairy Sci. 104:1928–1950
https://doi.org/10.3168/jds.2020-18241

Identification of functional candidate variants and genes for feed efficiency in Holstein and Jersey cattle breeds using RNA-sequencing

S. Lam,[1] F. Miglior,[1] P. A. S. Fonseca,[1] I. Gómez-Redondo,[1] J. Zeidan,[1] A. Suárez-Vega,[1] F. Schenkel,[1] L. L. Guan,[2] S. Waters,[3] P. Stothard,[2] and A. Cánovas[1]*

[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, ON, Canada N1G 2W1
[2]Department of Agriculture, Food & Nutritional Science, University of Alberta, Edmonton, Canada T6H 2P5
[3]Teagasc, Animal & Grassland Research and Innovation Centre, Grange, Dunsany, Ireland C15 PW93

ABSTRACT

The identification of functional genetic variants and associated candidate genes linked to feed efficiency may help improve selection for feed efficiency in dairy cattle, providing economic and environmental benefits for the dairy industry. This study used RNA-sequencing data obtained from liver tissue from 9 Holstein cows [n = 5 low residual feed intake (RFI), n = 4 high RFI] and 10 Jersey cows (n = 5 low RFI, n = 5 high RFI), which were selected from a single population of 200 animals. Using RNA-sequencing, 3 analyses were performed to identify: (1) variants within low or high RFI Holstein cattle; (2) variants within low or high RFI Jersey cattle; and (3) variants within low or high RFI groups, which are common across both Holstein and Jersey cattle breeds. From each analysis, all variants were filtered for moderate, modifier, or high functional effect, and co-localized quantitative trait loci (QTL) classes, enriched biological processes, and co-localized genes related to these variants, were identified. The overlapping of the resulting genes co-localized with functional SNP from each analysis in both breeds for low or high RFI groups were compared. For the first two analyses, the total number of candidate genes associated with moderate, modifier, or high functional effect variants fixed within low or high RFI groups were 2,810 and 3,390 for Holstein and Jersey breeds, respectively. The major QTL classes co-localized with these variants included milk and reproduction QTL for the Holstein breed, and milk, production, and reproduction QTL for the Jersey breed. For the third analysis, the common variants across both Holstein and Jersey breeds, uniquely fixed within low or high RFI groups were identified, revealing a total of 86,209 and 111,126 functional variants in low and high RFI groups, respectively. Across all 3 analyses for low and high RFI cattle, 12 and 31 co-localized genes were overlapping, respectively. Among the overlapping genes across breeds, 9 were commonly detected in both the low and high RFI groups (*INSRR*, *CSK*, *DYNC1H1*, *GAB1*, *KAT2B*, *RXRA*, *SHC1*, *TRRAP, PIK3CB*), which are known to play a key role in the regulation of biological processes that have high metabolic demand and are related to cell growth and regeneration, metabolism, and immune function. The genes identified and their associated functional variants may serve as candidate genetic markers and can be implemented into breeding programs to help improve the selection for feed efficiency in dairy cattle.

Key words: feed efficiency, Holstein, Jersey, RNA-sequencing

INTRODUCTION

Feed costs are a highly variable expense in cattle production and represent up to 75% of production costs (Food and Agriculture Organization, 2017), emphasizing the importance in improving genetic selection for feed efficiency in cattle production. In addition, the correlation between feed efficiency and methane emission traits suggests that feed efficiency contributes to variation in the environmental footprint of the dairy industry (Connor et al., 2012). Therefore, there is a demand for improving the accuracy of selection for superior feed efficiency, and thus milk production efficiency, in dairy cattle, which can lead to economic benefits, reduced production costs, and improved environmental sustainability for the dairy industry (Connor et al., 2012; Seymour et al., 2019).

A strategy to better understand the genetic architecture of feed efficiency traits is using high throughput RNA-sequencing (**RNA-seq**). The RNA-seq allows for the identification of differentially expressed genes in specific tissues across phenotypically divergent groups and has been performed to identify differentially ex-

Received January 21, 2020.
Accepted August 29, 2020.
*Corresponding author: acanovas@uoguelph.ca

pressed genes between divergent feed efficiency groups in cattle (Kern et al., 2016; Salleh et al., 2017; Keel, et al., 2018; Higgins et al., 2019). In addition to differentially expressed gene analysis, RNA-seq has been applied to identify mRNA isoforms and functional genetic variants such as SNP and insertions and deletions (**INDEL**) associated with desirable traits (Mortazavi et al., 2008; Cánovas et al., 2013; Cardoso et al., 2017; Wickramasinghe et al., 2014). Accordingly, it has successfully been applied to large scale SNP discovery analyses in livestock, leading to the identification of functional genetic markers, a deeper understanding of the genetic architecture, and a reduction in genome complexity of important production traits such as feed efficiency, health, fertility, and meat quality traits in ruminants (Cánovas et al., 2010, 2014a; Dias et al., 2017; Weber et al., 2016). To explain and understand the genetic variation related to feed efficiency or metabolic status, genetic markers influencing the regulation of these traits must be identified. In addition to identifying putative genetic markers, the functional study of positional candidate genes, along with the identification of QTL co-localized with functional variants, can be performed to generate a greater understanding of feed efficiency and to identify associated genomic regions (Pareek et al., 2016). Currently, the identification of genetic variants associated with feed efficiency using RNA-seq and their subsequent functional annotation in dairy cattle has not been fully completed. This could provide insight into the biology underlying genetic markers that may influence the regulation of feed efficiency in dairy cattle and ultimately improve selection strategies for feed efficiency, resulting in economic and environmental benefits for dairy production.

This study aimed to improve the understanding of the genetic architecture underlying genetic variants and candidate genes that may influence the regulation of this trait in dairy cattle. This was done by using Holstein and Jersey cattle, which were previously selected for extreme feed efficiency groups based on residual feed intake (**RFI**; kg/d) from a larger population. The objectives of this study were to: (1) identify SNP and INDEL uniquely fixed within low or high RFI Holstein or Jersey cattle (independent breeds), and SNP and INDEL uniquely fixed within low or high RFI groups common across both Holstein and Jersey cattle (both breeds); (2) determine functional information of the genetic variants by evaluating variant functional effect, co-localized QTL classes, significantly enriched biological processes, and functional candidate genes associated with genetic variants related to feed efficiency; and (3) determine the overlapping positional candidate genes associated with low or high RFI groups across all 3 analyses (Holstein analysis, Jersey analysis, across-breeds analysis).

MATERIALS AND METHODS

RNA-Seq Data Set Information

The data used for this study were obtained from National Center for Biotechnology Information Gene Expression Omnibus public database (liver RNA-seq data) accession number: GSE92398 (Salleh et al., 2017). The RNA-seq raw sequence reads from liver tissue of 9 Holstein (n = 5 low RFI, n = 4 high RFI) and 10 Jersey cattle (n = 5 low RFI, n = 5 high RFI) divergent for feed efficiency were analyzed. Feed efficiency groups were classified based on RFI (kg/d), which is a calculation for feed efficiency that accounts for body weight and growth (Koch et al., 1963). The RFI classification groups in this study included low RFI, which represents the high feed efficient group, and high RFI, which represents the low feed efficient group. The 9 Holstein (5 low RFI and 4 high RFI) and 10 Jersey cows (5 low RFI and 5 high RFI) were selected from a larger research herd of 200 dairy cattle. Detailed animal management and sampling information has been previously described (Salleh et al., 2017). Briefly, animals in this study with the lowest (more feed efficient) and highest RFI (less feed efficient) values were used, classifying them as extreme RFI animals with the intention to capture the greatest phenotypic and genetic differences to facilitate characterization of the RFI trait. The RFI values were calculated using one-step approach and random animal solutions were extracted from random regression model. The RFI was adjusted for stage of lactation, age, management group, breed, and parity. Two liver tissue samples were collected per animal via biopsy (Salleh et al., 2017). The RNA was extracted from liver samples using QIAzol, RNeasy Mini Kit (Qiagen, Hilden, Germany), and the cDNA was paired-end sequenced using Illumina HiSeq 2500 machine (Illumina, San Diego, CA), generating paired-end reads (100 bp length).

Variant Site Detection and Analysis Workflow

The workflow used for SNP and INDEL detection for each analysis is summarized in Figure 1 and was adapted from a study that determined an optimized RNA-seq pipeline to detect genetic variants for RNA-seq data from multiple samples per animal (Lam et al., 2020). Additionally, the overall study workflow is shown in Figure 2, which was performed for the independent breeds analysis, as well as the across-breeds analysis,

in which the detected SNP and INDEL unique to the RFI groups (low or high RFI) were pooled for both Holstein and Jersey breeds. Quality of sequence reads were verified using FastQC (Andrews, 2014) to identify sequencing read artifacts including sites with low quality Phred scores, duplicated reads, uncalled bases (N sequences) and potential contamination (Cardoso et al., 2018; Cánovas et al., 2014b). Reads were trimmed to remove Illumina adapters and low quality bases at the start and end of reads (sites were removed if lower than Phred score = 30) using Trimmomatic (Bolger et al., 2014). Additionally, using Trimmomatic, reads with an average quality score below 20 within a sliding window of 5 nucleotides were removed, and reads with

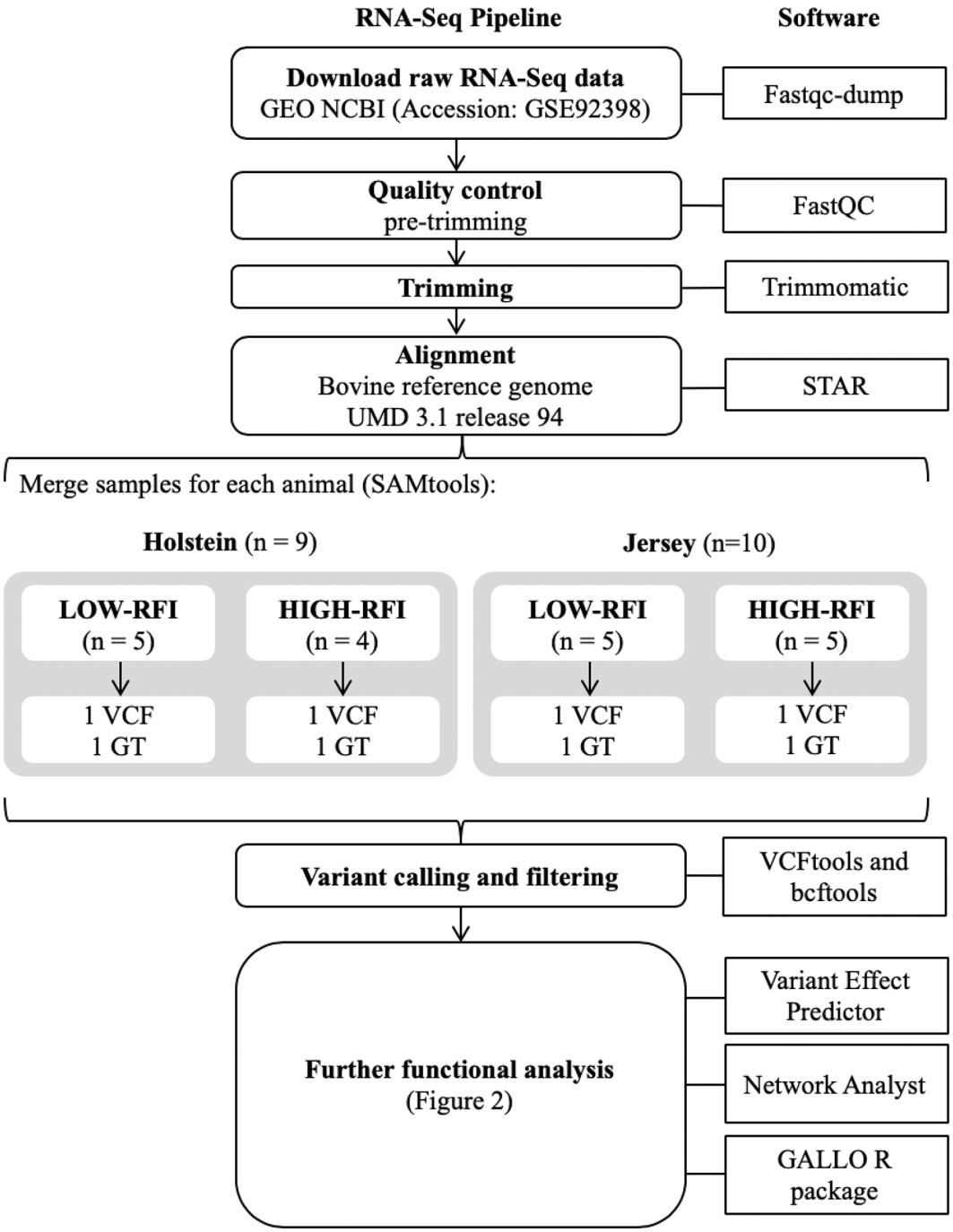

Figure 1. Workflow diagram to identify unique functional SNP and INDEL within low or high RFI groups and predict corresponding QTL regions and types in Holstein and Jersey cattle.

a minimum length of 75 bp following trimming were retained. Quality of sequence reads were re-evaluated after trimming of reads using FastQC (Andrews, 2014). Trimmed reads were individually aligned to the *Bos Taurus* reference genome (Assembly UMD3.1. release 94), using STAR (Dobin et al., 2013). The alignment statistics of mapped reads are reported in Table 1. ReadGroups were then added to each sample using SAMtools (Version 1.4; Li, 2011) consequently allowing the differentiation of samples by assigning the origin of the read (low or high RFI group) and assignment of SNP to a specific genotype. Additionally, PCR duplicates were marked and removed using Picard tools (Version 2.18.25.; Picard tools; http://broadinstitute.github.io/picard/). Then, multisample calling was performed to enhance the sensitivity of variant detection and accuracy of genotype calling over calling each sample independently (Brouard et al., 2019). This was due to the availability of 2 replicate samples for each animal, which included replicate 1 representing the tissue sample collected during low concentrate diet treatment and replicate 2 representing the tissue sample collected during the high concentrate diet (Salleh et al., 2017). Therefore, BAM files of animals in the same RFI group were merged for each tissue sample and were called for variants using multisample calling, resulting in 1 variant calling format (**VCF**) file for each tissue, each containing 2 genotypes (1 genotype per RFI group; Lam et al., 2020). Variant calling was performed to identify SNP and INDEL using the mpileup and call commands from BCFtools (Version 1.9–77-gd0cf724+; Danecek and McCarthy, 2017). After BAM files with the corresponding merging methodology was created for each approach, variant filtering was performed using VCFtools to remove variants with a minimum read depth below 10 and a minimum of 2 supporting reads for the alternative allele, as well as to filter SNP within 3 bp surrounding a gap (Cánovas et al., 2010). The BCFtools filter was then used to remove variants with quality values below 30 (based on Phred scaled scores for the assertion made in the alternative allele), filter SNP within 5 bp of an INDEL, and filter any alternative allele with a lower frequency of 20% in the population.

To categorize the functional genetic variants, the unique SNP fixed within low or high RFI groups were identified using the VCF files containing filtered SNP. SnpSift (Version 4.0; Cingolani et al., 2012) was used to filter variants present in one RFI condition, and not present in the other. The VCF file was then split using VCFtools vcf-subset to create one VCF file with all low RFI variants and one VCF file with all high RFI variants, which were then compared using the BCFtools isec command to determine the intersection of the files and create 3 files: SNP or INDEL exclusive to low RFI VCF file, SNP or INDEL exclusive to high RFI VCF file, and SNP or INDEL shared between both low and high RFI groups VCF file.

Comparison of Variant Calling Performance Using Merged Replicate Samples Compared with Nonmerged Individual Replicate Samples Approach

Additional analyses were performed to ensure merging the technical replicate samples by RFI group improved the performance of variant calling for SNP and INDEL for the purpose to identify potential candidate variants associated with the trait of interest. Optimization of RNA-Seq pipelines using different sample merging approaches to improve power and accuracy of variant calling is described in detail by Lam et al. (2020). BCFtools isec was used to determine the percentage of unique and shared SNP or INDEL between the nonmerged (individual replicate samples) and merged (merged replicate samples by RFI group) approach. The percent proportion of SNP and INDEL detected uniquely by each variant calling approach and detected commonly between both variant calling approaches was performed to determine the intersection of the files. This resulted in 3 files: SNP or INDEL exclusive to replicate sample 1 (representing the tissue sample collected during low concentrate diet treatment), SNP or INDEL exclusive to replicate sample 2 (representing the tissue sample collected during the high concentrate diet), and SNP or INDEL shared between both replicate samples.

To compare the performance of the nonmerged (individual replicate samples) and merged (merged replicate samples by RFI group) approaches, the quality of detected variants was analyzed. The variant quality distribution was evaluated for SNP and INDEL for each Holstein and Jersey population, separately. This resulted in violin plots illustrating the variant quality for detected SNP and INDEL in the Holstein population and detected SNP and INDEL in the Jersey population.

Identification of Functional Variants and Percentage of Variants Overlapping/Co-localized with QTL Classes and QTL Related to Specific Traits

Functional variants based on the variant effect prediction tool and the percentage of variants overlapping/co-localized with QTL classes and QTL related to specific traits were identified. In this study, QTL classes are defined as QTL categorized into a major trait category, including milk, meat and carcass, production, health, exterior, and reproduction. Furthermore, this study also discusses trait-specific QTL, which are defined as

QTL related to specific traits, for example, milk yield, milk protein percent, calving ease, and pregnancy rate, among others (Supplemental Figures S1–S12, https://figshare.com/s/dedb14cd92bd854fa7d5). The unique SNP and INDEL fixed within low or high RFI groups for each Holstein or Jersey breed (Figure 3), and common SNP and INDEL within low or high RFI groups common across both Holstein and Jersey cattle (Figure 4), were analyzed for their variant functional consequences using variant effect predictor (McLaren et al., 2010). Variants were selected based on moderate, modifier, or high functional effect, which included missense, stop gain, stop loss, intergenic, splice, 5′or 3′ untranslated region, upstream or downstream gene variant for SNPs, and frameshift deletion, inframe deletion, inframe insertion, 5′ or 3′ untranslated region, upstream or downstream gene variant for INDEL. The chromosome number and start and end position information of these variants of the unique variants fixed within low or high RFI groups and identified as a moderate, modifier, or high functional effect variant, were used to determine the percentage of variants co-localized with QTL regions and classes (Asselstine et al., 2019). This analysis allowed for the characterization of potential cattle QTL classes and QTL related to specific traits that may be influenced by the functional variants with moderate, modifier, or high functional effect. This was performed using R (R Version 3.6.0.; R Foundation for Statistical Computing, Vienna, Austria) and the R package: Genomic functional annotation in livestock for positional candidate loci, also known as GALLO (https://github.com/pablobio/GALLO) with previously known QTL class and trait information from the Cattle QTL Database that were filtered for dairy cattle breeds (Hu et al., 2007), sourced from the Animal QTL Database (Hu et al., 2019). To evaluate if the QTL classes and traits identified around the selected variants were significantly overrepresented, the qtl_enrich() function from GALLO was used. Briefly, this function performs a bootstrap analysis though the random sampling of each QTL class or trait (depending on the user's choice) present in the list of annotated QTL co-localized with the candidate variants. This random sampling is performed respecting the number of record proportions of each QTL class or trait by chromosome in the input list. In this study, after 1,000 iterations of random sampling, the number of observed and expected QTL classes and traits, per chromosome, were compared and the enrichment status was defined for all QTL classes and traits.

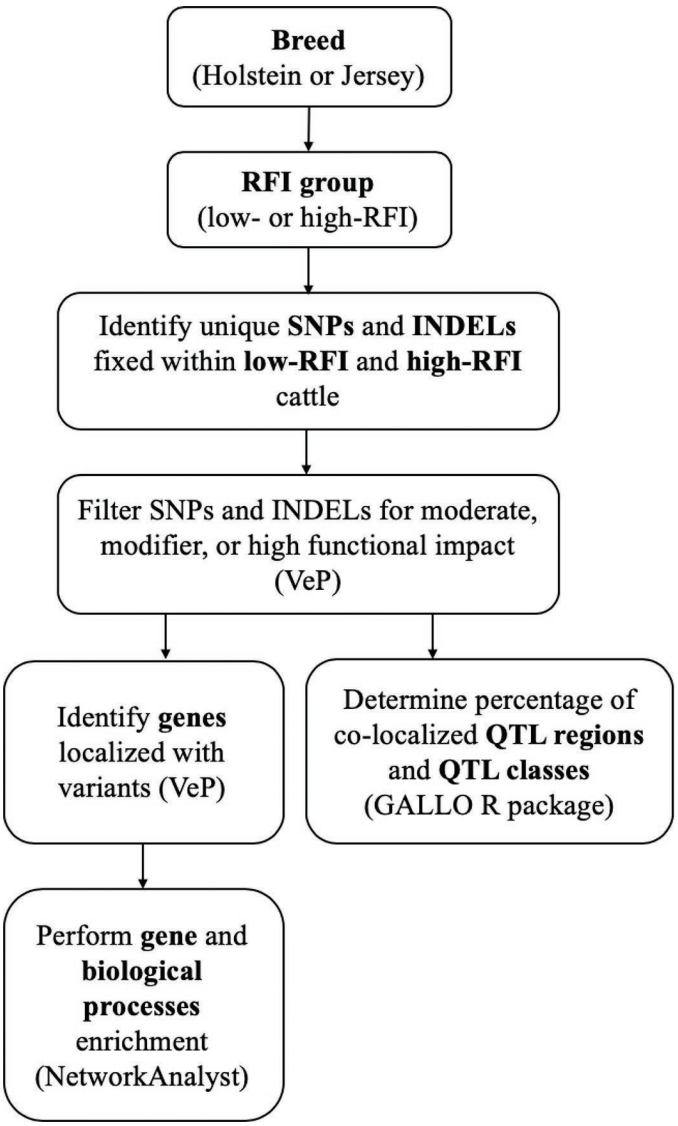

Figure 2. Workflow diagram of genetic variant identification and functional analyses used to analyze Holstein and Jersey dairy cattle breeds.

Functional Analysis of Positional Candidate Genes Associated with Variants Fixed Within Low or High RFI Groups

To compare the overlapping positional candidate genes unique to low or high RFI groups from each analysis, the positional variants (i.e., SNP and INDEL) uniquely detected in low or high RFI groups from each analysis (independent analysis for Holstein and Jersey breeds and across-breeds analysis) were used. Using variant effect predictor tool (McLaren et al., 2010), genes co-localized with these variants were identified. Duplicate genes were removed from the gene lists and Venn diagrams were generated using R (R Version 3.6.0.) VennDiagram-package to compare the overlapping of the resulting genes from each analysis. The resulting genes were used to perform the network analysis

using NetworkAnalyst software (Xia et al., 2014) for each low or high RFI group for each analysis.

The gene network analysis was performed using the gene names detected for each RFI group, using protein-protein interaction network analysis on NetworkAnalyst software (Xia et al., 2014, 2015; http://www.networkanalyst.ca). This approach used the pre-defined protein-protein interaction network available in the STRING database to build a gene network composed of the list of candidate genes used as input and other proteins, which creates a broader overview of the biological processes involved with the evaluated phenotype. To build this network, the option Minimum Interaction Network was used to construct a minimally connected network that contains all of the genes (seeds) uploaded for the analysis, which is performed by keeping only those additional nodes (proteins not presented in the input files) that are necessary to connect the seed nodes. A biological processes analysis (GO:BP explorer) was performed using Overrepresentation Analysis to identify significantly enriched biological processes associated with the submitted gene lists (and other related genes added by the software) based on Gene Ontology (**GO**) terms through the NetworkAnalyst software. An example of the use of GO terms to better understand the functional characteristics of positional variants identified by RNA-seq in livestock was done by Cánovas et al. (2012). The following biological processes that were most related to metabolic efficiency from the GO:BP functional explorer were selected to construct a subnetwork: lipid biosynthetic process, carbohydrate transport, regulation of protein metabolic process, en-

Table 1. Sample information regarding feed efficiency group, total reads, number of uniquely mapped reads, and percent uniquely mapped reads for Holstein and Jersey cattle

Sample accession number by breed	Feed efficiency group	Total reads	No. uniquely mapped reads	% uniquely mapped reads
Holstein				
SRR5110641	high	25,321,432	23,666,950	93.47
SRR5110642	high	25,057,354	23,3612,00	93.23
SRR5110605	high	23,502,530	21,729,637	92.46
SRR5110606	high	24,403,058	22,759,681	93.27
SRR5110615	high	24,427,969	22,783,258	93.27
SRR5110616	high	25,282,382	23,533,724	93.08
SRR5110625	high	24,847,652	23,400,261	94.17
SRR5110626	high	25,618,485	23,928,688	93.40
SRR5110635	high	24,671,570	22,878,009	92.73
SRR5110636	high	24,043,757	22,098,566	91.91
SRR5110607	low	24,840,679	23,216,873	93.46
SRR5110608	low	23,491,361	21,874,159	93.12
SRR5110617	low	21,658,806	20,216,184	93.34
SRR5110618	low	23,109,961	21,646,222	93.67
SRR5110621	low	24,537,168	22,615,327	92.17
SRR5110622	low	24,578,437	23,023,871	93.68
SRR5110633	low	24,050,940	22,508,736	93.59
SRR5110634	low	22,914,036	21,297,319	92.94
Average ± SD		24,242,088 ± 996,513.08	22,539,851 ± 964,518.76	93.16 ± 0.56
Jersey				
SRR5110613	high	22,869,282	21,080,678	92.18
SRR5110614	high	23,556,898	21,848,994	92.75
SRR5110619	high	22,850,990	21,461,451	93.92
SRR5110620	high	24,563,768	22,976,995	93.54
SRR5110623	high	22,817,590	20,803,374	91.17
SRR5110624	high	24,200,729	22,434,967	92.70
SRR5110631	high	25,221,268	23,604,244	93.59
SRR5110632	high	30,716,358	27,750,531	90.34
SRR5110637	high	22,828,294	21,510,319	94.23
SRR5110638	high	24,049,084	22,886,706	95.17
SRR5110609	low	23,388,117	21,712,208	92.83
SRR5110610	low	22,751,232	21,271,977	93.50
SRR5110611	low	27,078,436	25,325,368	93.53
SRR5110612	low	24,580,784	23,098,526	93.97
SRR5110627	low	25,438,093	23,890,611	93.92
SRR5110628	low	26,824,275	24,885,483	92.77
SRR5110629	low	26,255,195	24,536,291	93.45
SRR5110630	low	24,637,330	22,757,648	92.37
SRR5110639	low	27,771,572	25,726,479	92.64
SRR5110640	low	24,431,226	22,834,599	93.46
Average ± SD		24,841,526 ± 2,052,720.86	23,119,872 ± 1,801,865.61	93.00 ± 1.08

ergy reserve metabolic process, cellular carbohydrate metabolic process, fatty acid oxidation, lipid catabolic process, regulation of growth, skeletal muscle tissue development. The reported enriched GO:BP terms were obtained using the subnetwork composed by those nodes associated with the most biologically relevant GO:BP terms. Consequently, resulting *P*-values may by slightly inflated but still are representative of the functional profile of our candidate genes. The genes associated with the latter biological processes were used to construct the subnetwork using only the nodes associated with the previously reported biological processes. An enrichment analysis was then performed on this subnetwork for the biological processes. The top 50% most significantly enriched biological processes of the first quartile, corrected for False Discovery Rate (FDR <0.05), were reported in Table 2. This resulted in 16 and 18 significantly enriched biological pathways in the low and high RFI group for the Holstein analysis, 16 and 15 significantly enriched biological pathways in the low and high RFI group in the Jersey analysis, and 16 and 17 significantly enriched biological pathways in the low and high RFI group in the across-breeds analysis (Table 2).

RESULTS AND DISCUSSION

Alignment Statistics

The number of total reads, total uniquely mapped reads, and percentage of uniquely mapped reads are shown in Table 1. For Holstein low RFI samples, the

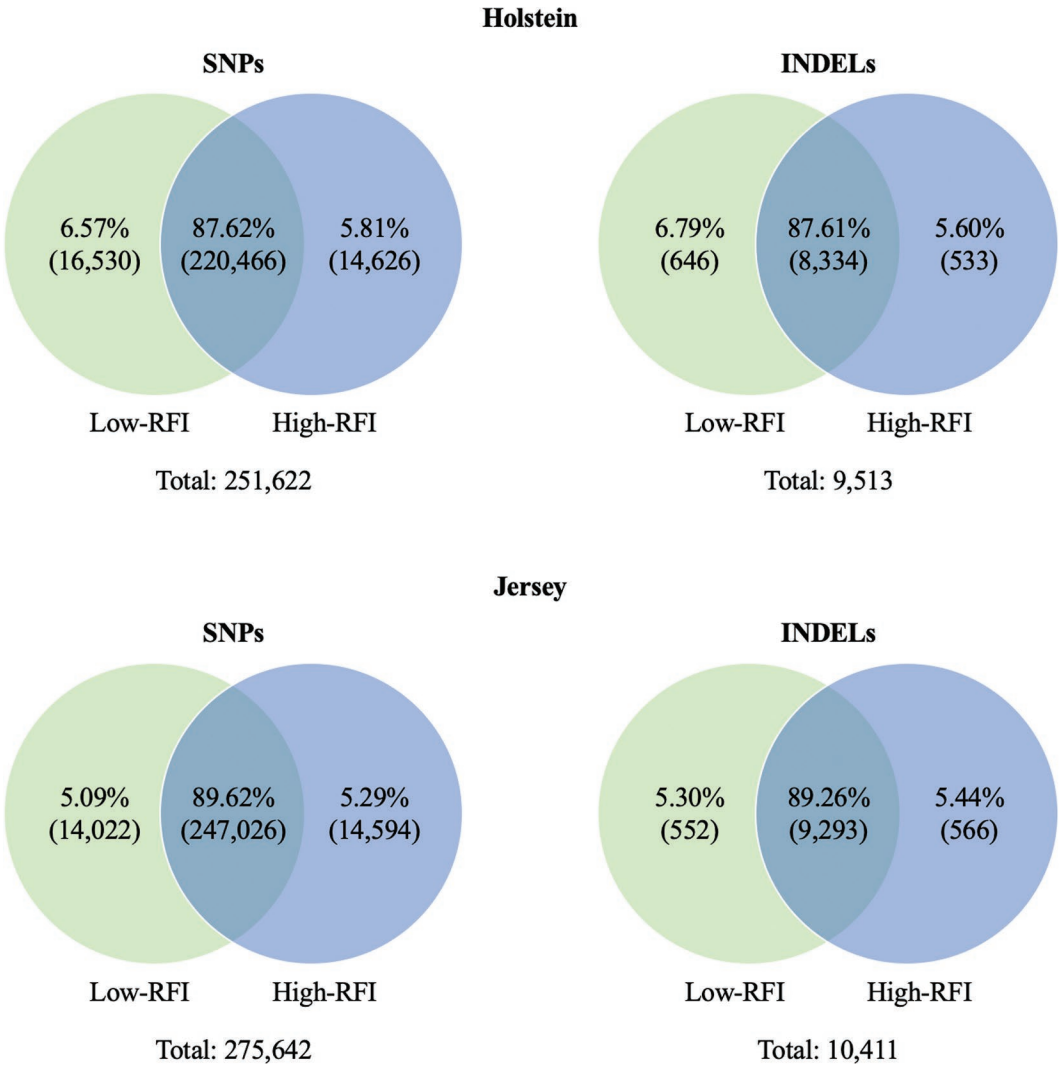

Figure 3. Total and proportion (%) of variants identified as unique or shared across low or high RFI groups in Holstein or Jersey cattle breeds.

average number of total reads, total uniquely mapped, and percent uniquely mapped was 24,717,619 ± 644,888.7, 22,905,631 ± 750,453.8, and 93.0 ± 0.6%, respectively. For Holstein high RFI samples, the average number of total reads, total uniquely mapped, and percent uniquely mapped was 23,647,674 ± 1,072,433.4, 22,049,836 ± 997,383.5, and 93.0 ± 0.5%, respectively. For Jersey low RFI samples, the average number of total reads, total uniquely mapped, and percent uniquely mapped was 24,367,426 ± 2,384,616.2, 22,635,826 ± 2,009,950.7, and 93.0 ± 1.5%, respectively. For Jersey high RFI samples, the average number of total reads, total uniquely mapped, and percent uniquely mapped was 25,315,626 ± 1,646,094.0, 2,635,826 ± 2,009,950.7, and 93.0 ± 0.6%, respectively. Overall, on average, 93.13% reads from each sample were uniquely mapped to the UMD3.1 bovine reference genome, release 94 (Table 1). This alignment statistic was expected, as in vitro and in vivo bovine embryo samples have shown 91 and 92% alignment of uniquely mapped reads to the bovine reference (btau 4.0), respectively (Driver et al., 2012), whereas Salleh et al. (2017) showed 91% of uniquely mapped reads to the bovine reference genome release 82. Two technical replicate samples were available for each animal, as the previous study collected 2 samples at different time points. The availability of replicate samples allowed for the merging of sample RNA-seq data by each RFI group, resulting in increased read quality and improving accuracy of variant calling (Supplemental Tables S7–S9, https://figshare.com/s/dedb14cd92bd854fa7d5).

The percentage of shared and unique variants between the replicate samples were evaluated for the detected SNP (Supplemental Table S5, https://figshare.com/

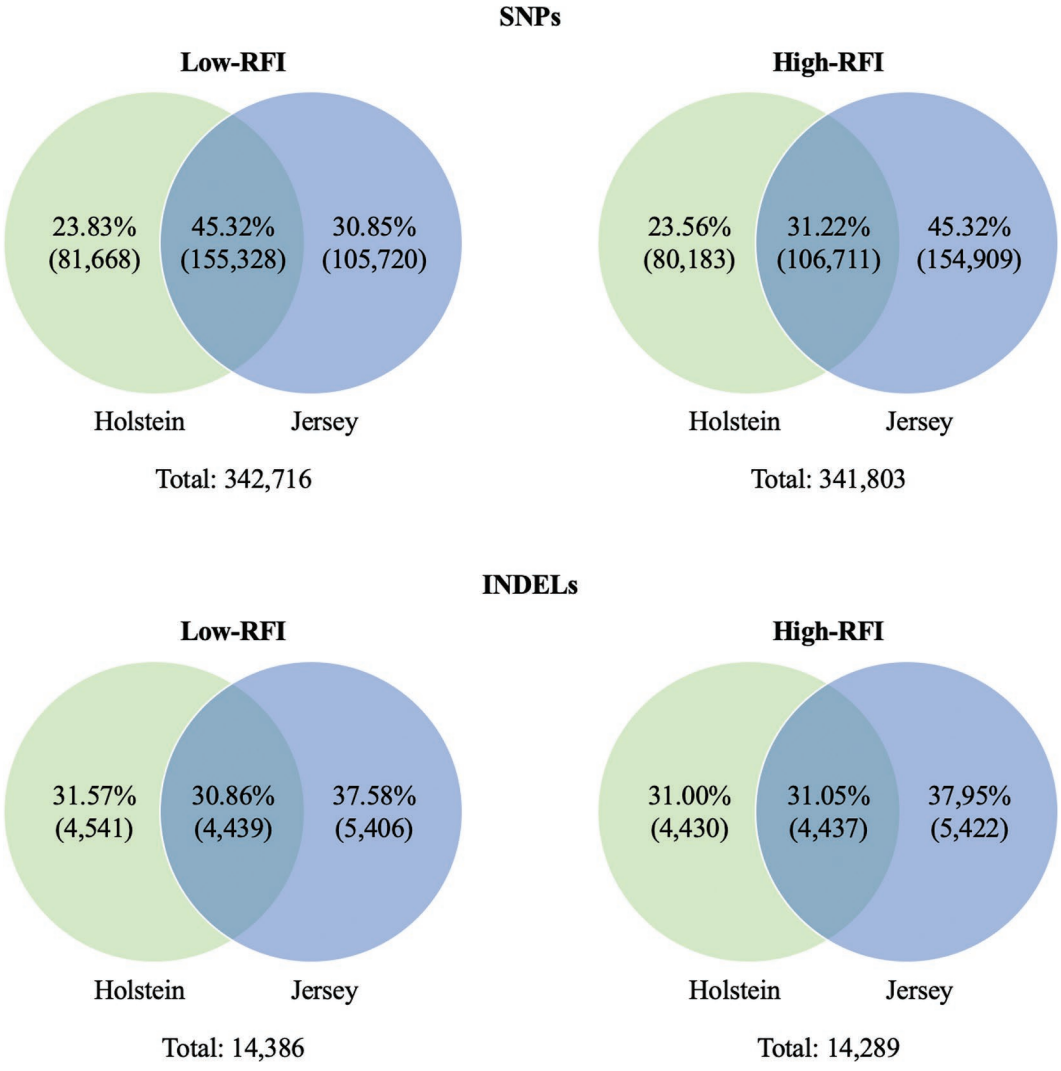

Figure 4. Proportion (%) of SNP and INDEL detected across Holstein or Jersey cattle breeds for low or high RFI groups.

Table 2. Significantly enriched biological processes identified for the independent breed analyses (Holstein and Jersey breeds) and the across-breeds analysis (Holstein and Jersey breeds)

Pathway[1]	Total genes[2]	Hits[3]	*P*-value	FDR[4]
Holstein low RFI				
Regulation of transcription from RNA polymerase II promoter	498	63	$2.71 \cdot 10^{-27}$	$1.80 \cdot 10^{-24}$
Lipid biosynthetic process	374	51	$1.53 \cdot 10^{-23}$	$5.07 \cdot 10^{-21}$
Regulation of cytokine biosynthetic process	708	55	$4.95 \cdot 10^{-14}$	$1.09 \cdot 10^{-11}$
Actin filament-based process	161	25	$2.99 \cdot 10^{-13}$	$4.95 \cdot 10^{-11}$
Immune response	28	12	$1.26 \cdot 10^{-12}$	$1.67 \cdot 10^{-10}$
Energy reserve metabolic process	206	27	$2.12 \cdot 10^{-12}$	$2.34 \cdot 10^{-10}$
Carbohydrate transport	146	22	$1.56 \cdot 10^{-11}$	$1.48 \cdot 10^{-09}$
Positive regulation of T cell proliferation	126	19	$3.92 \cdot 10^{-10}$	$3.24 \cdot 10^{-08}$
Cell maturation	426	35	$7.72 \cdot 10^{-10}$	$5.68 \cdot 10^{-08}$
Response to hypoxia	105	17	$1.09 \cdot 10^{-09}$	$7.21 \cdot 10^{-08}$
Nitrogen compound metabolic process	227	24	$3.30 \cdot 10^{-09}$	$1.99 \cdot 10^{-07}$
Fatty acid oxidation	292	27	$6.34 \cdot 10^{-09}$	$3.50 \cdot 10^{-07}$
DNA damage checkpoint	487	36	$7.48 \cdot 10^{-09}$	$3.81 \cdot 10^{-07}$
DNA damage response, signal transduction by p53 class mediator	317	26	$1.41 \cdot 10^{-07}$	$6.45 \cdot 10^{-06}$
Holstein high RFI				
Regulation of transcription from RNA polymerase II promoter	498	64	$4.91 \cdot 10^{-30}$	$3.25 \cdot 10^{-27}$
Lipid biosynthetic process	374	56	$1.18 \cdot 10^{-29}$	$3.89 \cdot 10^{-27}$
Regulation of cytokine biosynthetic process	708	59	$5.70 \cdot 10^{-18}$	$1.26 \cdot 10^{-15}$
Immune response	28	12	$5.44 \cdot 10^{-13}$	$9.01 \cdot 10^{-11}$
Actin filament-based process	161	23	$3.81 \cdot 10^{-12}$	$4.22 \cdot 10^{-10}$
Carbohydrate transport	146	22	$3.83 \cdot 10^{-12}$	$4.22 \cdot 10^{-10}$
Cell maturation	426	37	$5.82 \cdot 10^{-12}$	$5.51 \cdot 10^{-10}$
Response to hypoxia	105	18	$4.07 \cdot 10^{-11}$	$3.37 \cdot 10^{-09}$
Energy reserve metabolic process	206	23	$6.20 \cdot 10^{-10}$	$4.56 \cdot 10^{-08}$
Positive regulation of T cell proliferation	126	18	$9.10 \cdot 10^{-10}$	$6.03 \cdot 10^{-08}$
Fatty acid oxidation	292	27	$1.35 \cdot 10^{-09}$	$8.12 \cdot 10^{-08}$
Regulation of protein metabolic process	147	19	$1.74 \cdot 10^{-09}$	$9.59 \cdot 10^{-08}$
DNA damage checkpoint	487	35	$3.98 \cdot 10^{-09}$	$1.98 \cdot 10^{-07}$
Nitrogen compound metabolic process	227	23	$4.19 \cdot 10^{-09}$	$1.98 \cdot 10^{-07}$
Regulation of binding	83	14	$7.87 \cdot 10^{-09}$	$3.47 \cdot 10^{-07}$
Regulation of transcription from RNA polymerase II promoter	498	64	$4.91 \cdot 10^{-30}$	$3.25 \cdot 10^{-27}$
Jersey low RFI				
Lipid biosynthetic process	374	78	$9.76 \cdot 10^{-47}$	$6.46 \cdot 10^{-44}$
Regulation of transcription from RNA polymerase II promoter	498	81	$5.16 \cdot 10^{-40}$	$1.71 \cdot 10^{-37}$
Positive regulation of T cell proliferation	126	32	$5.09 \cdot 10^{-22}$	$1.12 \cdot 10^{-19}$
Regulation of binding	83	27	$6.97 \cdot 10^{-22}$	$1.15 \cdot 10^{-19}$
Inflammatory response	80	24	$1.23 \cdot 10^{-18}$	$1.62 \cdot 10^{-16}$
Carbohydrate transport	146	28	$6.05 \cdot 10^{-16}$	$6.68 \cdot 10^{-14}$
Energy reserve metabolic process	206	32	$2.68 \cdot 10^{-15}$	$2.54 \cdot 10^{-13}$
Fatty acid oxidation	292	37	$1.31 \cdot 10^{-14}$	$1.08 \cdot 10^{-12}$
Regulation of cytokine biosynthetic process	708	59	$3.40 \cdot 10^{-14}$	$2.50 \cdot 10^{-12}$
Regulation of protein metabolic process	147	25	$4.14 \cdot 10^{-13}$	$2.74 \cdot 10^{-11}$
Response to hypoxia	105	21	$1.26 \cdot 10^{-12}$	$7.59 \cdot 10^{-11}$
Nitrogen compound metabolic process	227	29	$9.20 \cdot 10^{-12}$	$5.07 \cdot 10^{-10}$
Cell maturation	426	39	$7.94 \cdot 10^{-11}$	$4.04 \cdot 10^{-09}$
Immune response	28	11	$1.14 \cdot 10^{-10}$	$5.40 \cdot 10^{-09}$
Reciprocal meiotic recombination	145	21	$7.37 \cdot 10^{-10}$	$3.25 \cdot 10^{-08}$
Jersey high RFI				
Lipid biosynthetic process	374	68	$8.93 \cdot 10^{-39}$	$5.91 \cdot 10^{-36}$
Regulation of transcription from RNA polymerase II promoter	498	71	$2.37 \cdot 10^{-33}$	$7.84 \cdot 10^{-31}$
Positive regulation of T cell proliferation	126	29	$1.08 \cdot 10^{-19}$	$2.38 \cdot 10^{-17}$
Energy reserve metabolic process	206	34	$4.54 \cdot 10^{-18}$	$7.52 \cdot 10^{-16}$
Regulation of binding	83	23	$8.55 \cdot 10^{-18}$	$1.13 \cdot 10^{-15}$
Fatty acid oxidation	292	36	$6.53 \cdot 10^{-15}$	$6.68 \cdot 10^{-13}$
Carbohydrate transport	146	26	$7.06 \cdot 10^{-15}$	$6.68 \cdot 10^{-13}$
Inflammatory response	80	18	$1.76 \cdot 10^{-12}$	$1.46 \cdot 10^{-10}$
Regulation of protein metabolic process	147	23	$4.56 \cdot 10^{-12}$	$3.35 \cdot 10^{-10}$
Nitrogen compound metabolic process	227	27	$4.49 \cdot 10^{-11}$	$2.97 \cdot 10^{-09}$
Regulation of cytokine biosynthetic process	708	50	$9.73 \cdot 10^{-11}$	$5.86 \cdot 10^{-09}$
Immune response	28	10	$1.18 \cdot 10^{-09}$	$6.49 \cdot 10^{-08}$
Response to hypoxia	105	17	$1.77 \cdot 10^{-09}$	$9.01 \cdot 10^{-08}$
Positive regulation of lymphocyte activation	33	10	$7.38 \cdot 10^{-09}$	$3.49 \cdot 10^{-07}$

Continued

Table 2 (Continued). Significantly enriched biological processes identified for the independent breed analyses (Holstein and Jersey breeds) and the across-breeds analysis (Holstein and Jersey breeds)

Pathway[1]	Total genes[2]	Hits[3]	*P*-value	FDR[4]
Across-breeds low RFI				
Lipid biosynthetic process	374	94	$3.23 \cdot 10^{-48}$	$2.13 \cdot 10^{-45}$
Regulation of transcription from RNA polymerase II promoter	498	103	$1.77 \cdot 10^{-44}$	$5.87 \cdot 10^{-42}$
Energy reserve metabolic process	206	47	$3.74 \cdot 10^{-22}$	$8.25 \cdot 10^{-20}$
Positive regulation of T cell proliferation	126	36	$1.04 \cdot 10^{-20}$	$1.71 \cdot 10^{-18}$
Regulation of binding	83	29	$1.28 \cdot 10^{-19}$	$1.69 \cdot 10^{-17}$
Fatty acid oxidation	292	51	$1.61 \cdot 10^{-18}$	$1.78 \cdot 10^{-16}$
Nitrogen compound metabolic process	227	43	$4.18 \cdot 10^{-17}$	$3.95 \cdot 10^{-15}$
Regulation of cytokine biosynthetic process	708	80	$1.49 \cdot 10^{-16}$	$1.23 \cdot 10^{-14}$
Regulation of protein metabolic process	147	33	$1.22 \cdot 10^{-15}$	$8.95 \cdot 10^{-14}$
Inflammatory response	80	24	$9.80 \cdot 10^{-15}$	$6.48 \cdot 10^{-13}$
Carbohydrate transport	146	30	$3.03 \cdot 10^{-13}$	$1.82 \cdot 10^{-11}$
Response to hypoxia	105	25	$9.19 \cdot 10^{-13}$	$5.07 \cdot 10^{-11}$
Cell maturation	426	51	$9.90 \cdot 10^{-12}$	$5.04 \cdot 10^{-10}$
Immune response	28	12	$4.45 \cdot 10^{-10}$	$2.11 \cdot 10^{-08}$
Actin filament-based process	161	27	$6.34 \cdot 10^{-10}$	$2.80 \cdot 10^{-08}$
Across-breeds high RFI				
Lipid biosynthetic process	374	93	$4.02 \cdot 10^{-48}$	$2.66 \cdot 10^{-45}$
Regulation of transcription from RNA polymerase II promoter	498	101	$1.13 \cdot 10^{-43}$	$3.72 \cdot 10^{-41}$
Regulation of cytokine biosynthetic process	708	90	$7.97 \cdot 10^{-23}$	$1.76 \cdot 10^{-20}$
Fatty acid oxidation	292	54	$3.33 \cdot 10^{-21}$	$5.51 \cdot 10^{-19}$
Positive regulation of T cell proliferation	126	35	$4.92 \cdot 10^{-20}$	$6.30 \cdot 10^{-18}$
Energy reserve metabolic process	206	44	$5.71 \cdot 10^{-20}$	$6.30 \cdot 10^{-18}$
Regulation of binding	83	28	$9.10 \cdot 10^{-19}$	$8.60 \cdot 10^{-17}$
Cell maturation	426	59	$9.09 \cdot 10^{-17}$	$7.52 \cdot 10^{-15}$
Carbohydrate transport	146	31	$2.57 \cdot 10^{-14}$	$1.89 \cdot 10^{-12}$
Regulation of protein metabolic process	147	31	$3.14 \cdot 10^{-14}$	$2.08 \cdot 10^{-12}$
Inflammatory response	80	23	$6.21 \cdot 10^{-14}$	$3.74 \cdot 10^{-12}$
Nitrogen compound metabolic process	227	38	$9.57 \cdot 10^{-14}$	$5.28 \cdot 10^{-12}$
Immune response	28	14	$7.89 \cdot 10^{-13}$	$4.02 \cdot 10^{-11}$
Actin filament-based process	161	30	$2.49 \cdot 10^{-12}$	$1.18 \cdot 10^{-10}$
Response to hypoxia	105	24	$4.27 \cdot 10^{-12}$	$1.89 \cdot 10^{-10}$
Reciprocal meiotic recombination	145	25	$9.68 \cdot 10^{-10}$	$4.01 \cdot 10^{-08}$

[1]Pathways that were in the top 50% biological processes of the first quartile of the most significantly enriched biological pathways ($P < 0.05$, FDR < 0.05) were reported in this table.

[2]Total genes = total number of genes from gene list associated with that specific biological pathway.

[3]Hits = gene hits within the network.

[4]FDR = false discovery rate < 0.05.

s/dedb14cd92bd854fa7d5) and INDEL (Supplemental Table S6). This revealed a proportion of detected SNP (27.46%) and INDEL (34.01%) that were unique to each replicate sample. The use of RNA-seq as a tool for variant calling is novel in livestock research and is continuously improving. These uniquely detected variants to each replicate sample may represent low frequency variants or errors associated sequencing errors during library preparation in RNA-seq. This is possible due to highly differential coverage among different genes, resulting in different variants detected in the replicate samples, leading to allele-specific expression (Han et al., 2015). However, this study aimed to identify positional candidate markers associated with the trait of interest that involves capturing allele-specific expression. Further analysis was performed to validate the variant calling approach used in this study.

To validate merging the replicate samples by RFI group for the purpose of identifying potential candidate variants associated with the trait of interest, in comparison to traditionally calling from individual samples, the variant quality distribution and variant density at those variants was evaluated (Supplemental Figures S13–S16, https://figshare.com/s/dedb14cd92bd854fa7d5), along with the descriptive statistics of the variant quality (Supplemental Table S7). It was observed that the unique SNP for the nonmerged (individual replicate samples) approach in both the Holstein and Jersey population, showed a smaller median and larger density of variants with low quality values (Supplemental Table S7). Additionally, the quality values in the nonmerged VCF file, for the variants shared between the merged and nonmerged approaches showed a larger median and highest values for the third quartile of quality distribu-

tion (Supplemental Table S7). These results may suggest that the variants which are shared between the merged and nonmerged approach are those with the highest probability to be true variants as opposed to false positives. Additionally, it is important to highlight that the quality distribution of the variants uniquely identified in the merged approach was higher than the quality distribution of the unique SNP for the nonmerged approach (Supplemental Figures S13 and S15).

Interestingly, the described quality distribution and read density patterns are observed in the opposite direction for INDEL, with the nonmerged groups displaying greater medians, means, and third quartile values in both Holstein and Jersey populations (Supplemental Table S7, https://figshare.com/s/dedb14cd92bd854fa7d5). This may be explained by the higher heterogeneity of reads and consequently increased difficulty to accurately align reads to perform the INDEL calling. However, it is important to highlight that the majority of INDEL detected in this study are shared between merged and nonmerged approaches. Additionally, the minority of INDEL uniquely identified in the approach used in this study (merging samples by RFI group), were subjected to a quality filtering where only those INDEL that qualified for minimum read depth of 10, minimum of 2 supporting reads for the alternative allele, quality values greater than 30 (based on Phred scaled scores for the assertion made in the alternative allele), and a minimum frequency of 20% in the population, were retained for the functional analysis.

In addition to variant quality and variant distribution, this study also compared variant calling from nonmerged samples and merged samples approach by evaluating the average total reads overlapping the alternative allele of each variant (SNP or INDEL) in the nonmerged and merged approach for each Holstein and Jersey population. The results are displayed in violin plots in Supplemental Figures S17–S20 (https://figshare.com/s/dedb14cd92bd854fa7d5) and corresponding descriptive statistics are shown in Supplemental Table S8. This illustrated that the average number of reads overlapping the alternative alleles for the uniquely detected SNP is higher in the merged approach in both Holstein and Jersey populations, compared with the nonmerged approach. Additionally, the third quartile of the alternative allele count for the unique SNP was higher in the merged approach compared with the nonmerged approach in both Holstein and Jersey populations. Notably, the average reads overlapping the unique SNP are higher in the Holstein population, which may suggest read coverage variability between the population data, which would influence coverage of the alternative allele. This may be explained by different overall coverage, library preparation, among other factors.

As similarly observed from the quality distribution analysis (Supplemental Table S7, https://figshare.com/s/dedb14cd92bd854fa7d5), the INDEL showed an opposing pattern indicating higher coverage of the alternative alleles in the nonmerged approach, which, as stated, can be a consequence of the limitations and challenges associated with INDEL calling caused by the increase in read sequence variability. On the other hand, in the Jersey population, this effect appears negligible with all groups showing similar distributions for the number of reads overlapping the alternative allele of the uniquely called INDEL. Interestingly, the third quartile of the distribution for the unique INDEL called in the merged approach is higher than all the other groups. These results reinforce that INDEL calling is complex and is affected by different variables.

Furthermore, correlation values were calculated between the average number of reads overlapping the alternative allele and the quality values for uniquely detected and commonly detected variants in each nonmerged and merged sample approaches (Supplemental Table S9, https://figshare.com/s/dedb14cd92bd854fa7d5). The results revealed a higher correlation between the average number of reads overlapping the alternative allele and quality in the uniquely detected groups regarding SNP regardless of the approaches in both Holstein and Jersey populations. These results suggest that a higher number or reads covering the alternative allele is associated with higher quality SNP calling. Additionally, highly significant correlations were observed in the shared groups, suggesting the shared variants have a stronger positive correlation between the number of reads overlapping the alternative allele and variant detection quality (Supplemental Table S9). Furthermore, when considering both the distribution of reads overlapping the alternative allele (Supplemental Figures S17–20), and the correlation results (Supplemental Table S9), the results suggest that calling variants from merged samples produce more reliable SNP callings when the unique SNP are considered.

Regarding the INDEL, the highest correlations were observed for the shared groups. However, it is observed that the unique INDEL from the merged groups also show higher correlation values between the average reads overlapping the alternative allele and quality when compared with the uniquely detected INDEL from the nonmerged groups.

Overall, these results reinforce the hypothesis that merging replicate samples for variant calling, increase the variant quality, and therefore reduce the false-positive detection rate. This has been previously described in exome data (Zhang et al., 2014). In addition, the use of multisample calling (or joint calling) method has shown to enhance the sensitivity of SNP detection and

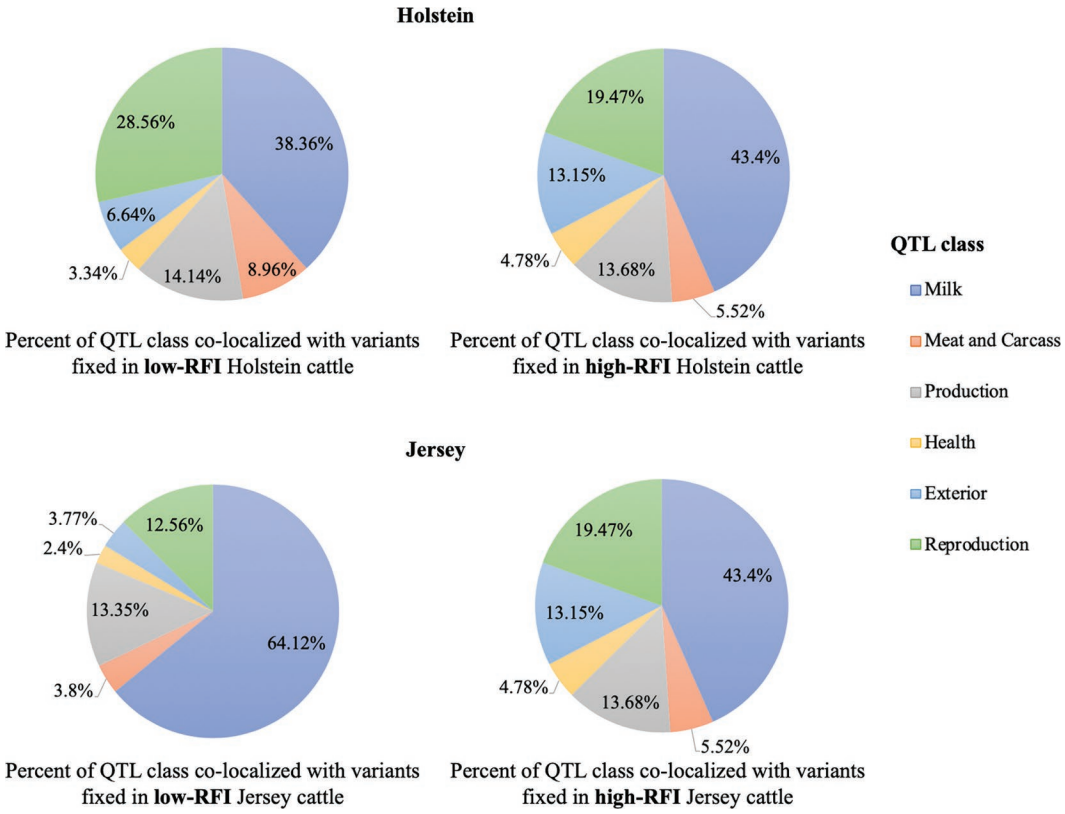

Figure 5. Percent (%) proportion of QTL classes co-localized with variants including both SNP and INDEL for low or high RFI groups in Holstein and Jersey cattle breeds.

accuracy of genotype calling over calling each sample independently in data sets with low sequencing depth (Nielsen et al., 2011). Furthermore, Brouard et al. (2019) has previously demonstrated the improved sensitivity of joint genotype calling using GATK compared with individual calling, supporting these results.

SNP and INDEL Uniquely Identified Within Low or High RFI Groups for Each Analysis

For the independent breeds analyses, functional SNP and INDEL within low or high RFI groups for each Holstein and Jersey cattle breeds were identified (Figure 2). In total, 251,622 SNP and 9,513 INDEL were identified for the Holstein cattle breed (Figure 3). For the Jersey cattle breed, 275,642 SNP and 10,411 INDEL were identified (Figure 3). A greater number of SNP were identified compared with INDEL across feed efficiency groups for both breeds (Figure 3). This was expected as INDEL occur less frequently in the genome; however, they are still capable of causing substantial genetic and phenotypic variation (Mullaney et al., 2010). The majority of variants were observed to be shared among low and high RFI groups for both Holstein and Jersey cattle (Figure 3). On average, 5.94%, and 5.54% of variants were uniquely fixed within low RFI and high RFI groups respectively, which could be contributing to the biological regulation of feed efficiency. The total number of uniquely detected and commonly detected SNP and INDEL for low or high RFI groups for each breed, are shown in Figure 3.

For across-breeds analysis, functional SNP and INDEL within low or high RFI groups that are common across both Holstein and Jersey cattle breeds were identified, revealing a total of 159,767 in low RFI and 111,148 in high RFI groups (Figure 4). Identification of common putative genetic variants across breeds is important to build supportive evidence for selection of variants that may serve as potential genetic markers to select for production traits in livestock (Pareek et al., 2016).

Variants Associated with Feed Efficiency that are Co-localized with Dairy Cattle QTL Classes and QTL Related to Specific Traits

It was observed that variants unique to more feed efficient Holstein cattle are mainly co-localized with

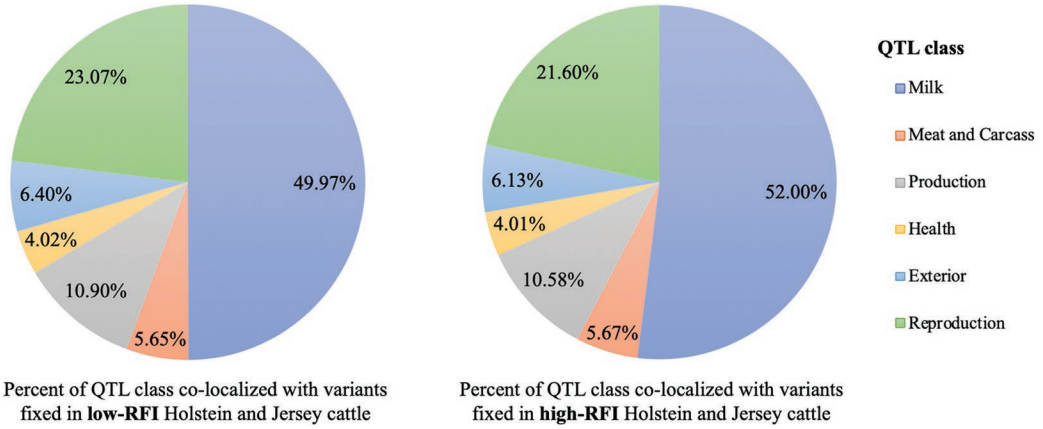

Figure 6. Percent (%) proportion of QTL classes co-localized with variants including both SNP and INDEL for low or high RFI groups that were common across both Holstein and Jersey cattle breeds.

milk (38.36%) and reproduction (28.56%) QTL classes (Figure 5). Variants unique to more feed efficient Jersey cattle are mainly co-localized with the milk QTL class (64.12%; Figure 5). When examining less feed efficient Holstein and Jersey cattle, milk and reproduction are also the major QTL classes co-localized with SNP and INDEL unique to these groups (Figure 5). This may suggest that variants unique to more or less feed efficient animals may be regulating the same QTL but in different ways to influence the regulation of feed efficiency in dairy cattle. In addition, it was expected that these variants are co-localized mainly within milk QTL, as feed efficiency in dairy cattle refers to milk production efficiency, which is related to milk production and milk composition traits. In addition, feed efficiency in cattle is known to be correlated with reproduction traits (Hurley et al., 2018), suggesting the large percentage of variants unique to low or high RFI would be co-localized with both milk and reproduction QTL. It is known that there is a negative correlation between fertility traits, in both females (Ferreira Júnior et al., 2018) and bulls (Awda et al., 2013) with feed efficiency. However, it is important to note that there is a higher probability that co-localized QTL are mainly in QTL associated with milk traits, due to the Cattle QTL database having information predominantly in milk QTL, leading to a bias toward overlapping with milk related QTL. This is due to the larger number of existing studies that have evaluated milk traits in dairy cattle, in comparison with other traits in other breeds. However, it is shown that the QTL classes and QTL related to specific traits are significantly enriched for the variants unique to low or high RFI groups detected commonly across breeds (Figure 7, Figure 8). Results from the latter QTL enrichment analysis are shown in Supplemental Tables S3 and S4 (https://figshare.com/s/dedb14cd92bd854fa7d5), which report the trait related to the QTL, chromosome, number of QTL, *P*-value, and adjusted *P*-value, for all QTL co-localized with variants uniquely fixed to low or high RFI Holstein and Jersey cattle. The QTL related to specific traits that the variants were co-localized were identified to perform a more in-depth analysis, which are shown in Supplemental Figures S1–12 (https://figshare.com/s/dedb14cd92bd854fa7d5). The Holstein low and high RFI groups were both mainly associated with milk fat yield and milk fat percentage QTL. Additionally, the reproductive traits related to the QTL identified were also of the same order for both RFI groups for Holstein as well, including calving ease, and interval to first estrous after calving QTL classes. This may further suggest that these variants are influencing similar QTL differently, which could either improve or reduce the metabolic status or level of feed efficiency in Holstein cattle. In addition, as feed efficiency is a complex trait, regulated by multiple genes, pleiotropic effects should be considered, as multiple loci could be simultaneously regulating feed efficiency (Fonseca et al., 2018).

When evaluating the QTL related to specific traits associated with the Jersey cattle variants (Figure 5), it is observed that the milk QTLs related to the more feed efficient Jersey cattle included milk fat yield and milk yield. Additionally, the major milk QTL type associated with less feed efficient Jersey cattle is milk yield and milk fatty acid index. Reproduction QTLs related to both Jersey feed efficiency groups included calving ease and interval to first estrous after calving. Comparably, the milk and reproduction QTL types were fairly similar across efficiency groups and dairy breeds.

The substantial overlapping of milk and reproduction QTL classes with divergent feed efficiency groups was expected, as metabolic energy utilization influences

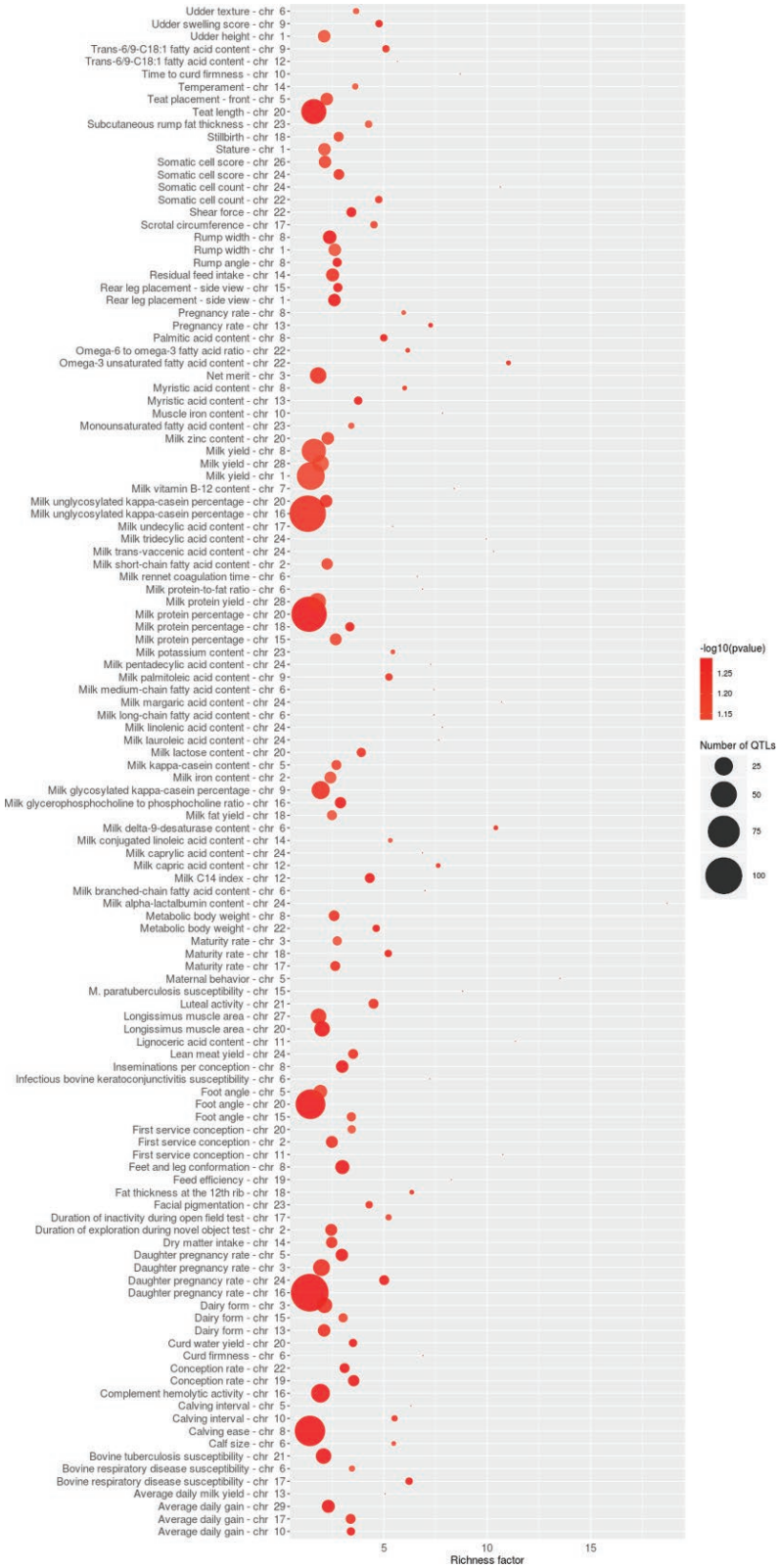

Figure 7. Significantly enriched ($P < 0.005$) QTLs co-localized with variants (i.e., SNP and INDEL) unique to low RFI animals across both Holstein and Jersey cattle breeds. The area of the bubbles represents the number of observed QTL for that QTL class, and the color represents the P-value scale (darker color = smaller P-value). The richness factor for each QTL represents the ratio of the number of QTL and the expected number of QTL.

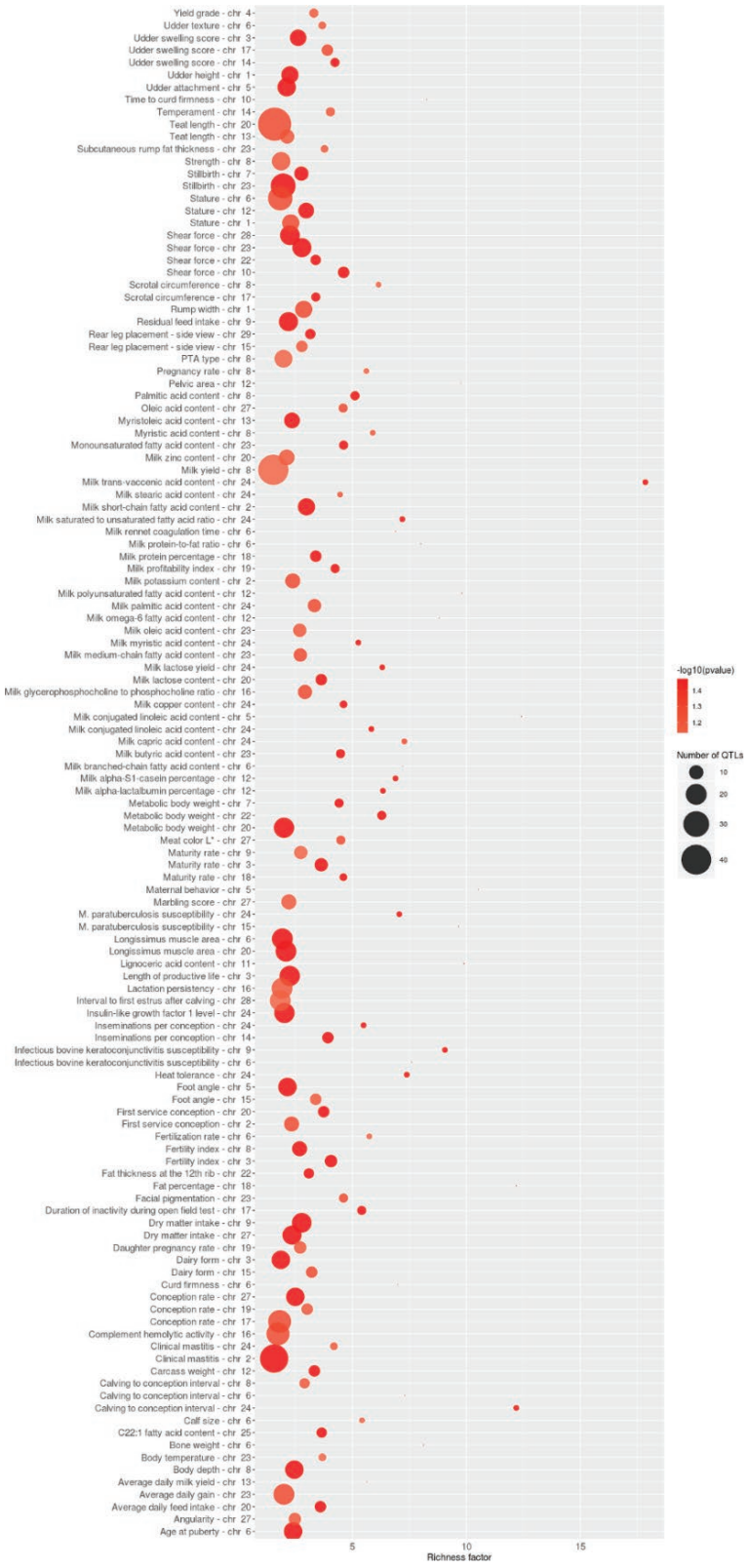

Figure 8. Significantly enriched ($P < 0.005$) QTL co-localized with variants (i.e., SNP and INDEL) unique to high RFI animals across both Holstein and Jersey cattle breeds. The area of the bubbles represents the number of observed QTL for that QTL class, and the color represents the P-value scale (darker color = smaller P-value). The richness factor for each QTL represents the ratio of the number of QTL and the expected number of QTL.

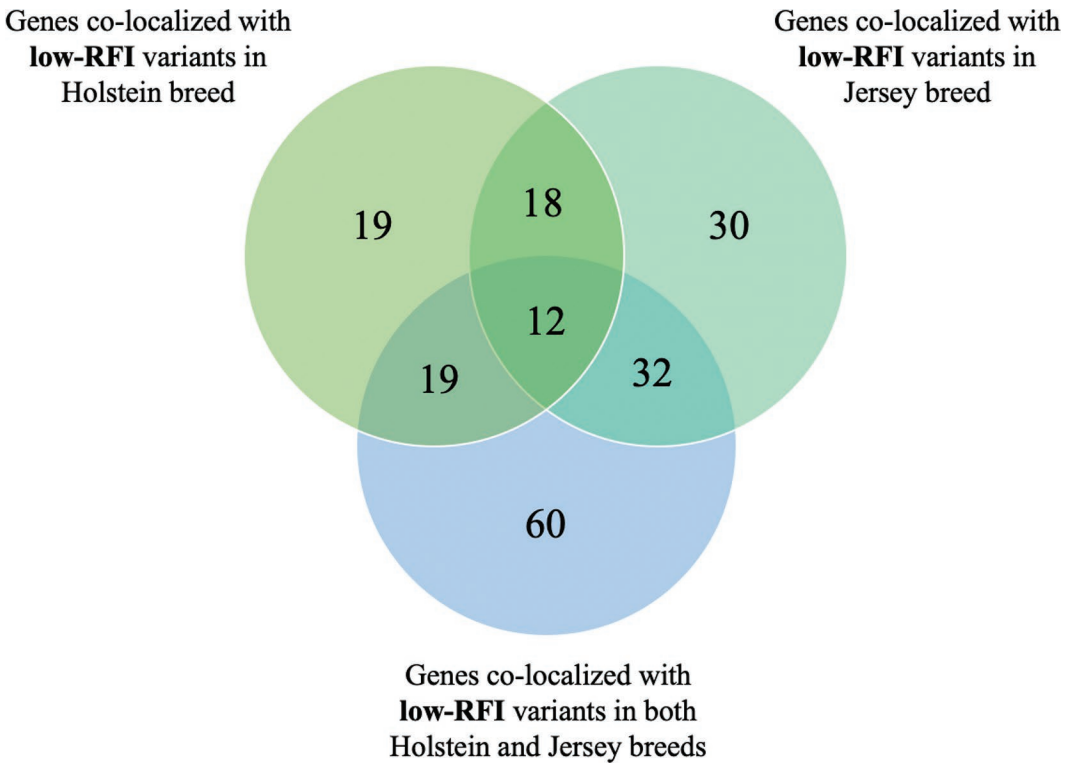

Figure 9. Venn diagram of genes co-localized with low RFI variants (SNP and INDEL) from each analysis (Holstein breed, Jersey breed, and both Holstein and Jersey breeds).

milk traits (Berry and Crowley, 2013) and reproductive conditions (Awda et al., 2013; Ferreira Júnior et al., 2018). It is known that increased milk yield is associated with improved energetic efficiency due to the "dilution of maintenance effect" in which high producing dairy cows may use less energy for body maintenance relative to a low producing cow of the same body size (Bauman et al., 1985; de Vries et al., 2000; Collard et al., 2000). When considering feed efficiency and reproduction, it is known that greater milk yield is associated with poor fertility conditions and metabolic imbalance (de Vries and Veerkamp, 2000; Oltenacu and Broom, 2010; Wathes et al., 2014). The strong correlation between reproduction, milk production, and metabolic efficiency supports our results of the substantial overlap of these QTLs within each feed efficiency group (Figure 6).

In the across-breed analysis, positional candidate genes co-localized with SNP and INDEL that were common across both Holstein and Jersey breeds were identified. Then, co-localized QTL classes and QTL related to specific traits associated with these variants were determined. This revealed that the variants unique to low RFI or more efficient dairy cattle were mainly co-localized with QTL classes related to milk (49.97%) and reproduction (23.07%; Figure 4). From this, the most highly significant QTLs related to milk and reproduction enriched for variants uniquely fixed to low RFI (Figure 7) and high RFI (Figure 8) groups across breeds were identified, revealing that specific QTLs are more prevalent or significant than others. This includes a larger number of QTLs enriched for milk yield, milk protein yield, and milk fat yield associated with low RFI cattle, with higher significant enrichment for specific QTLs such as milk yield, teat length, milk protein percentage, foot angle, daughter pregnancy rate, calving ease QTLs (Figure 7). Regarding the high RFI group, QTL classes associated with milk (52.00%) and reproduction (21.60%) QTLs were also overlapping in high RFI, or less feed efficient animals (Figure 4), with variants being more significantly enriched in specific QTL classes (Figure 8). This includes a larger number of QTLs enriched for teat length, milk yield, and clinical mastitis (Figure 8). Similar QTLs were significantly enriched, which was expected, as the majority of the overlapping QTL classes were associated with the similar trait-specific QTLs when evaluating the dairy breeds individually. However, when evaluating the common variants across both breeds, more similar proportion sizes of overlapping for each QTL class was observed when comparing high and low feed efficiency groups, which can be observed in Figure 6. This may support the hypothesis that although the genetic vari-

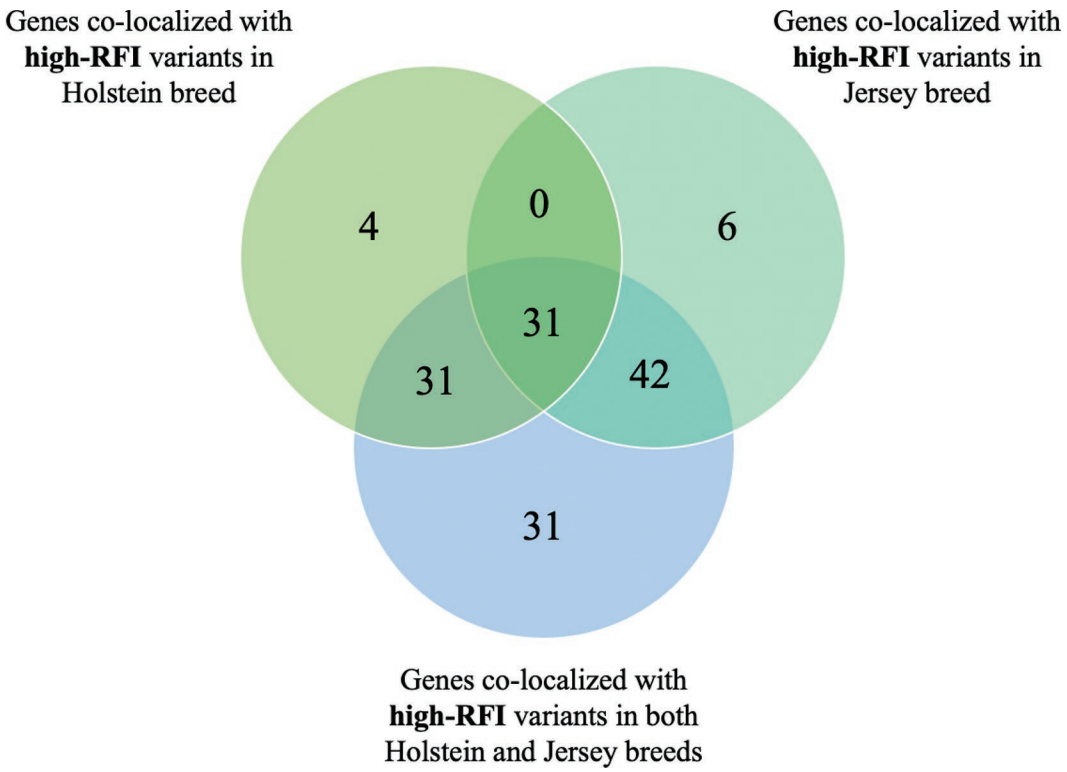

Figure 10. Venn diagram of genes co-localized with high RFI variants (SNP and INDEL) from each analysis (Holstein breed, Jersey breed, and both Holstein and Jersey breeds).

ants identified within low or high RFI cattle are different; they are still influencing similar genes but in a different manner, leading to an animal being more or less feed efficient. This is further supported as the 2 most significant QTL classes for both the low and high RFI cattle groups were the same for both milk and reproduction QTL classes. These milk trait QTLs included milk fat yield and milk kappa-casein percentage, whereas the trait-specific QTLs for reproduction involved calving ease and interval to first calving.

Biological Processes and Positional Candidate Genes Associated with Variants Uniquely Found in High or Low Feed Efficient Cattle

After filtering positional candidate genes co-localized with functional variants fixed within low or high RFI groups, we identified 1,444 and 1,366 genes for low RFI and high RFI, respectively within the Holstein breed. For the Jersey breed analysis, 1,683 and 1,707 positional genes were identified for low RFI and high RFI, respectively. Analysis across breeds revealed 2,268, and 2,248 genes for low RFI and high RFI, respectively. Using RNA-seq technology, detection of positional genetic variants associated with an extreme phenotype can be identified from mRNA reads (Cánovas et al., 2010). However, the mRNA reads are not necessarily differentially expressed or highly expressed in the tissue. The discussion of the following positional variants and their associated functional information (co-localized genes, enriched biological processes, and overlapping QTL classes), are approached by considering the positional variants that are unique or fixed within RFI groups and how their functional characteristics could be related to the feed efficiency trait.

The resulting significantly enriched biological processes associated with the genes co-localized with the variants fixed within low or high RFI groups from each analysis are reported in Table 2. Due to the substantial amount of significantly enriched biological processes detected, only the top 50% biological processes of the first quartile of the most significantly enriched biological processes ($P < 0.05$, FDR <0.05) were reported (Table 2). Many reported biological processes associated with the genes co-localized with variants fixed within RFI groups were related to metabolic function (Table 2). The study of metabolic energy, partitioned into multiple physiological functions and systems has shown that major metabolic activities contributing to the variation of RFI includes protein turnover, tissue metabolism, and stress (37%), physical activity (8%), and body composition (5%; Herd and Arthur, 2009).

Table 3. Genes co-localized with variants fixed within low RFI or high RFI groups overlapping between the 3 analyses (Holstein breed, Jersey breed, and across-breeds analysis)

			Gene region		
RFI group	ENSEMBL ID	Gene name	BTA	Start (bp)	End (bp)
Low RFI (high feed efficiency)					
	ENSBTAG00000003054	*INSRR*	3	13,991,619	14,008,734
	ENSBTAG00000021424	*CSK*	21	33,873,489	33,892,160
	ENSBTAG00000016598	*DYNC1H1*	21	66,850,740	66,911,778
	ENSBTAG00000002813	*GAB1*	17	14,421,800	14,549,193
	ENSBTAG00000000746	*KAT2B*	1	157,277,141	157,374,828
	ENSBTAG00000006909	*PIK3CB*	1	130,293,099	130,471,254
	ENSBTAG00000010989	*PIK3R1*	20	11,397,628	11,480,087
	ENSBTAG00000008063	*PPARA*	5	116,438,987	116,507,065
	ENSBTAG00000017851	*RXRA*	11	105,021,683	105,114,871
	ENSBTAG00000010627	*SF3B3*	18	1,578,539	1,618,405
	ENSBTAG00000019838	*SHC1*	3	15,616,454	15,626,916
	ENSBTAG00000007113	*TRRAP*	25	37,265,876	37,353,871
High RFI (low feed efficiency)					
	ENSBTAG00000021181	*BUB1*	11	1,585,412	1,618,575
	ENSBTAG00000001700	*CDC42*	2	130,732,620	130,787,969
	ENSBTAG00000010109	*CDK1*	28	16,488,142	16,582,565
	ENSBTAG00000026403	*CREBBP*	25	3,054,344	3,173,309
	ENSBTAG00000021424	*CSK*	21	33,873,489	33,892,160
	ENSBTAG00000016598	*DYNC1H1*	21	66,850,740	66,911,778
	ENSBTAG00000005676	*EHMT2*	23	27,467,178	27,480,388
	ENSBTAG00000017355	*ERCC1*	18	53,024,536	53,039,604
	ENSBTAG00000002813	*GAB1*	17	14,421,800	14,549,193
	ENSBTAG00000004736	*GRB2*	19	56,117,146	56,181,857
	ENSBTAG00000012698	*HDAC1*	2	121,225,584	121,257,548
	ENSBTAG00000006270	*HSP90AA1*	21	66,936,747	66,945,064
	ENSBTAG00000001354	*INPP5E*	11	103,876,330	103,886,130
	ENSBTAG00000003054	*INSRR*	3	13,991,619	14,008,734
	ENSBTAG00000000746	*KAT2B*	1	157,277,141	157,374,828
	ENSBTAG00000013790	*MAP3K1*	20	22,340,163	22,417,428
	ENSBTAG00000020783	*MAPK14*	23	9,969,009	10,044,336
	ENSBTAG00000030965	*MCM7*	25	36,349,409	36,356,836
	ENSBTAG00000006909	*PIK3CB*	1	130,293,099	130,471,254
	ENSBTAG00000014453	*PLK1*	25	21,334,598	21,345,432
	ENSBTAG00000053390	*RAC1*	25	38,278,343	38,297,376
	ENSBTAG00000011043	*RAC2*	5	75,656,456	75,673,313
	ENSBTAG00000004279	*RHOA*	22	50,701,226	50,751,132
	ENSBTAG00000015473	*RPS27A*	11	37,970,354	37,972,652
	ENSBTAG00000017851	*RXRA*	11	105,021,683	105,114,871
	ENSBTAG00000019838	*SHC1*	3	15,616,454	15,626,916
	ENSBTAG00000009985	*SIN3A*	21	33,383,914	33,450,894
	ENSBTAG00000019220	*SMARCA4*	7	15,424,807	15,515,086
	ENSBTAG00000011643	*SOS1*	11	21,308,259	21,434,854
	ENSBTAG00000008938	*SRC*	13	66,295,099	66,349,583
	ENSBTAG00000007113	*TRRAP*	25	37,265,876	37,353,871

Table 4. Common genes identified overlapping between low and high RFI groups that were common across 3 analyses (Table 3; Holstein breed, Jersey breed, and both Holstein and Jersey breed)

		Gene region		
ENSEMBL ID	Gene name	BTA	Start (bp)	End (bp)
ENSBTAG00000003054	*INSRR*	3	13,991,619	14,008,734
ENSBTAG00000021424	*CSK*	21	33,877,662	33,877,662
ENSBTAG00000016598	*DYNC1H1*	21	66,870,569	66,870,569
ENSBTAG00000002813	*GAB1*	17	14,417,132	14,417,132
ENSBTAG00000000746	*KAT2B*	1	157,273,917	157,273,917
ENSBTAG00000006909	*PIK3CB*	1	144,408,494	144,408,494
ENSBTAG00000017851	*RXRA*	11	105,114,221	105,114,221
ENSBTAG00000019838	*SHC1*	3	15,452,192	15,452,195
ENSBTAG00000007113	*TRRAP*	25	3,051,535	3,051,535

This supports the gene network analysis results in this study, as several major metabolic processes related to feed efficiency were present in the analyses including lipid biosynthetic process, immune response, carbohydrate transport, fatty acid oxidation, and energy reserve metabolic process (Table 2). Additionally, the observation of similar biological processes detected across the analyses and RFI groups may suggest the greater influence of these functions on RFI, and further support the previous results suggesting multiple variants may be influencing similar QTLs in low and high RFI groups (Figure 6). These results highly support the study of positional variants which may be influencing the same gene regions and biological processes to regulate feed efficiency in cattle.

The associated candidate genes were identified and compared for overlapping among the 3 analyses (Holstein breed, Jersey breed, and both Holstein and Jersey breeds) for low RFI (Figure 9) and for high RFI (Figure 10). Table 3 shows all 12 (Figure 9) and 31 (Figure 10) genes identified as common across all 3 analyses for low and high RFI analyses, respectively. The associated variants identified as co-localized with these genes are found in Supplemental Table S1 (https://figshare.com/s/dedb14cd92bd854fa7d5). Additionally, many of the genes identified were similar across low and high RFI analyses including *INSRR*, *CSK*, *DYNC1H1*, *GAB1*, *KAT2B*, *RXRA*, *SHC1*, *TRRAP*, and *PIK3CB* which are shown in Table 4, with the corresponding gene region information and corresponding variants shown in Supplemental Table S2.

The *INSRR* gene, also known as *IRR* (insulin receptor-related receptor), is a member of the insulin receptor family, which includes its homologs insulin-like growth factor receptor (*IGF-IR*) and the insulin receptor (IR) (Tatulian, 2015). The *IRR* gene is found in cells of highly functioning metabolic organs including the kidney, pancreas, and stomach (Petrenko et al., 2013), and works in synchrony with liver function to regulate glucose (Raile et al., 2005). Specifically, *IRR* has shown to be expressed in liver tissue (Mathi et al., 1995), suggesting possible IRR signaling involved with the regulation of glucose homeostasis. Similarly, Lysine Acetyltransferase 2B (*KAT2B*), identified in this study, plays a key role in stimulating hepatic gluconeogenesis, by which its concentration affects the regulation of blood glucose concentrations (Ravnskjaer et al., 2013).

The C-Terminal Src Kinase (*CSK)* gene along with Phosphoinositide-3-Kinase Regulatory Subunit 3 *(PIK3R3)* gene, which leads to the quaternary structure *PIK3CB* found in the present study, was identified in Cobb chickens as differentially expressed across feed efficiency groups in breast muscle and liver samples (Liu et al., 2018). In this study, both genes revealed upregulated expression in low RFI Cobb chickens compared with high RFI, suggesting their importance in the regulation of feed efficiency. The *CSK* gene is additionally known to influence the Hippo signaling pathway, also known as Salvador-Warts-Hippo pathway, which regulates organ size and tissue growth (Kwon, et al., 2015). The *PIK3R3* was also identified in a study on Duroc boars ear tissue where it was significantly associated with average daily feed intake (Ding et al., 2017). Specifically, *PIK3CB* has been shown to be upregulated in breast muscle of high feed efficiency chickens, suggesting increased activity in breast muscle of higher feed efficient animals (Zhou et al., 2015), and also has been found to be associated with leukocyte trans-endothelial migration and melanoma KEGG pathways (Taye et al., 2017).

The *DYNC1H1* gene known as Cytoplasmic dynein 1 heavy chain 1, results in the production of a protein that is part of a group (complex) of proteins called dynein. This gene has not yet been identified in association with feed efficiency traits in livestock. However, it has been shown that mRNA levels of dynein increase as chickens are induced with cardiotoxicity, suggesting a relationship between dynein production and cardiac function (Li et al., 2018). In addition, dynein plays a major functional role in energy production for bovine sperm motility, suggesting its importance in reproductive function in cattle (McConnell et al., 1987; Lorch et al., 2008).

The *GAB1* gene also known as growth factor receptor bound protein 2-associated protein 1, is known to play a key role in endothelial cell migration, blood capillary formation, and pathways associated with vascular endothelial growth factor (Laramée et al., 2007). More recently, evaluation of metabolic pathways related to hepatic growth in high feed efficient pigs has shown overexpression of *GAB1* suggesting its role in liver function in increased feed efficiency (Horodyska et al., 2019). In the same study, *GAB1* was significantly enriched in immune response pathways, which was also a commonly enriched biological process across the 3 analyses in this study. Based on the role of *GAB1* gene in hepatic function, it may serve as a good candidate gene for feed efficiency, as the gut and liver contribute 38% to total energy expenditure in cattle, which is related to metabolic efficiency (Ortigues-Marty et al., 2017). In support, a study examining hepatic tissue transcriptome in beef cattle divergent for feed efficiency identified genes differentially expressed across RFI groups, which were related to immune function (Higgins et al., 2019).

The *RXRA* is a transcription factor gene, which is part of the retinoid X receptors (RXR). It is known that RXR initiates tocopherol metabolism pathways in

mammals (Azzi et al., 2004). Many beneficial effects result from tocopherol (for example vitamin E) supplementation in livestock feed, including tocopherols acting as an antioxidant and improving immune function and health in dairy cattle, which has been shown to reduce incidence of mastitis and retained placenta in transition dairy cattle (Spears and Weiss, 2008).

The *SHC1* (SHC Adaptor Protein 1) is a scaffold protein coding gene, which helps regulate the epidermal growth factor (EGF) signaling pathway by directing the signal information after the EGF is stimulated, regulating cell proliferation (Zheng, et al., 2013). In addition, a prior study has shown that *SHC1* was upregulated in the liver, which functioned to promote cell division and growth as a hepatic response to realimentation in feed-restricted steers (Connor et al., 2010).

The *TRRAP* is a cofactor transformation/transcription domain-associated protein which, in conjunction with Histone Acetyltransferase 1 (*HAT*), has an important role in liver regeneration after toxic shock by helping with cell cycle progression and cell proliferation (Shukla et al., 2011).

An animal's efficiency to use feed for production is dependent on metabolic efficiency, which is depicted by metabolic tissue function and capacity at the whole body and cellular level (Cantalapiedra-Hijar et al., 2018). The liver, specifically, elicits high metabolic activity, being the primary site of gluconeogenesis and producing 90% of glucose to the host (Nafikov and Beitz, 2007), which is a major energetic product needed to support meat and milk production (Connor et al., 2010). As several candidate genes were associated with hepatic function, this reinforces the importance of the positional candidate genes identified in this study. The genes described may serve as potential candidate genes for selecting for feed efficiency, as they are involved in many metabolically demanding biological processes related to cell growth and regeneration, metabolism, and immune function. These genes may provide insight on which animals exhibit lower energy requirements to respond to high metabolic needs. In addition, the relationship of these genes with metabolic processes that have been studied in previous experiments, may suggest that the associated functional variants that affect these genes (Table 4) may be valuable functional variants (Supplemental Table S2, https://figshare.com/s/dedb14cd92bd854fa7d5) for selecting animals for feed efficiency. Currently, the Canadian dairy industry implements genomic breeding strategies to select for desirable milk production and composition traits. However, feed efficiency is a novel trait that is recently being implemented into genomic selection. With further validation studies, these functional variants can be implemented into genotype panels in aim to help select for more feed efficient dairy cattle.

In addition to the across-breeds analysis, the genes co-localized with the SNP and INDEL unique to low and high RFI groups were analyzed for overlapping with the differentially expressed genes between RFI groups identified by the prior study, which used the same animal population (Salleh et al., 2017). It was observed that 2 genes including *ABR* and *SOCS2* were found to be overlapping between the studies. Active breakpoint cluster region-related protein (*ABR*) plays a key role in GTPase activation, which modulates pathway and signaling activation (Tan et al., 1993). Interestingly, suppressor of cytokine signaling 2 (*SOCS2*) is a key regulator of growth hormone receptor sensitivity and has been shown to greatly influence body growth (Vesterlund et al., 2011). Salleh et al. (2017) observed that both *ABR* (FC = 0.329, *P*-value = 0.032) and *SOCS2* (FC = 0.422, *P*-value = 0.044) were upregulated in high RFI Holstein, which may suggest the higher energy expenditure for metabolic function and growth in less feed efficient cattle.

The similar detection of these genes in both studies may be explained by their critical roles in growth and metabolic function. However, this lack of overlapping results may be due to the fewer amount of significantly differentially expressed genes detected by Salleh et al. (2017) which were in total, 69 for Holstein and 19 for Jersey. Contrastingly, our study identified 2,268 and 2,248 genes co-localized with the variants fixed within low or high RFI groups, respectively. This highlights the importance of using RNA-seq to identify positional candidate variants in addition to the study of differentially expressed genes between 2 extreme phenotypic groups to better understand the functional genetics underlying a desirable trait.

CONCLUSIONS

The findings obtained in this study were important for identifying functional genetic variants which may be candidate genetic markers for feed efficiency, due to their supporting positional and functional evidence and their co-localization with relevant cattle QTL. Additionally, this study improved the knowledge of genetic variants and associated biological processes and candidate genes, which are largely related to highly metabolically demanding processes, underlying feed efficiency in Holstein and Jersey dairy cattle. Further validation analyses in additional independent cattle populations of these functional genetic variants and positional candidate genes may lead to their use in improving selection strategies for feed efficiency, leading

to the improvement of the economic and environmental sustainability of the dairy cattle industry.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the Ontario Ministry of Agriculture, Food, and Rural Affairs (OMAFRA; Ontario, Canada), Ontario Ministry of Research and Innovation, Agriculture and Agri-Food Canada (AAFC, Ontario, Canada), and Efficient Dairy Genome Project, funded by Genome Canada (Ottawa), Genome Alberta (Calgary, Canada), Ontario Genomics (Toronto, Canada), Alberta Ministry of Agriculture (Edmonton, Canada), Ontario Ministry of Research and Innovation (Toronto, Canada), Ontario Ministry of Agriculture, Food and Rural Affairs (Guelph, Canada), Canadian Dairy Network (Guelph), GrowSafe Systems (Airdrie, Canada), Alberta Milk (Edmonton, Canada). This study (FDE.13.17) was also supported by the Sustainable Beef and Forage Science Cluster funded by the Canadian Beef Cattle Check-Off, Beef Cattle Research Council (BCRC; Ontario, Canada), Alberta Beef Producers (Alberta, Canada), Alberta Cattle Feeders' Association (Alberta, Canada), Beef Farmers of Ontario (Ontario, Canada), La Fédération des Productuers de bovins du Québec (Québec, Canada), and Agriculture and Agri-Food Canada's Canadian Agricultural Partnership (Canada). The authors have not stated any conflicts of interest.

REFERENCES

Andrews, S. 2014. FastQC A quality control tool for high throughput sequence data. Accessed Jan. 27, 2019. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Asselstine, V., F. Miglior, A. Suárez-Vega, P. A. S. Fonseca, B. Mallard, N. Karrow, A. Islas-Trejo, J. F. Medrano, and A. Cánovas. 2019. Genetic mechanisms regulating the host response during mastitis. J. Dairy Sci. 102:9043–9059. https://doi.org/10.3168/jds.2019-16504.

Awda, B. J., S. P. Miller, Y. R. Montanholi, G. Vander Voort, T. Caldwell, M. M. Buhr, and K. C. Swanson. 2013. The relationship between feed efficiency traits and fertility in young beef bulls. Can. J. Anim. Sci. 93:185–192. https://doi.org/10.4141/cjas2012-092.

Azzi, A., R. Gysin, P. Kempná, A. Munteanu, Y. Negis, L. Villacorta, T. Visarius, and J.-M. Zingg. 2004. Vitamin E mediates cell signaling and regulation of gene expression. Ann. N. Y. Acad. Sci. 1031:86–95. https://doi.org/10.1196/annals.1331.009.

Bauman, D. E., S. N. McCutcheon, W. D. Steinhour, P. J. Eppard, and S. J. Sechen. 1985. Sources of variation and prospects for improvement of productive efficiency in the dairy cow: A review. J. Anim. Sci. 60:583–592. https://doi.org/10.2527/jas1985.602583x.

Berry, D. P., and J. J. Crowley. 2013. Cell Biology Symposium: Genetics of feed efficiency in dairy and beef cattle. J. Anim. Sci. 91:1594–1613. https://doi.org/10.2527/jas.2012-5862.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brouard, J.-S., F. Schenkel, A. Marete, and N. Bissonnette. 2019. The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. J. Anim. Sci. Biotechnol. 10:44. https://doi.org/10.1186/s40104-019-0359-0.

Cánovas, A., R. N. Pena, D. Gallardo, O. Ramírez, M. Amills, and R. Quintanilla. 2012. Segregation of regulatory polymorphisms with effects on the gluteus medius transcriptome in a purebred pig population. PLoS One 7:e35583. https://doi.org/10.1371/journal.pone.0035583.

Cánovas, A., A. Reverter, K. L. DeAtley, R. L. Ashley, M. L. Colgrave, M. R. S. Fortes, A. Islas-Trejo, S. Lehnert, L. Porto-Neto, G. Rincón, G. A. Silver, W. M. Snelling, J. F. Medrano, and M. G. Thomas. 2014a. Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. PLoS One 9:e102551. https://doi.org/10.1371/journal.pone.0102551.

Cánovas, A., G. Rincón, C. Bevilacqua, A. Islas-Trejo, P. Brenaut, R. C. Hovey, M. Boutinaud, C. Morgenthaler, M. K. VanKlompenberg, P. Martin, and J. F. Medrano. 2014b. Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. Sci. Rep. 4:5297. https://doi.org/10.1038/srep05297.

Cánovas, A., G. Rincón, A. Islas-Trejo, R. J. Jimenez-Flores, A. Laubscher, and J. F. Medrano. 2013. RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk. J. Dairy Sci. 96:2637–2648. https://doi.org/10.3168/jds.2012-6213.

Cánovas, A., G. Rincon, A. Islas-Trejo, S. Wickramasinghe, and J. F. Medrano. 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mamm. Genome 21:592–598. https://doi.org/10.1007/s00335-010-9297-z.

Cantalapiedra-Hijar, G., M. Abo-Ismail, G. E. Carstens, L. L. Guan, R. Hegarty, D. A. Kenny, M. McGee, G. Plastow, A. Relling, and I. Ortigues-Marty. 2018. Review: Biological determinants of between-animal variation in feed efficiency of growing beef cattle. Animal 12(s2):s321–s335. https://doi.org/10.1017/S1751731118001489.

Cardoso, T. F., A. Cánovas, O. Canela-Xandri, R. González-Prendes, M. Amills, and R. Quintanilla. 2017. RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles. Sci. Rep. 7:40005. https://doi.org/10.1038/srep40005.

Cardoso, T. F., R. Quintanilla, A. Castelló, R. González-Prendes, M. Amills, and A. Cánovas. 2018. Differential expression of mRNA isoforms in the skeletal muscle of pigs with distinct growth and fatness profiles. BMC Genomics 19:145. https://doi.org/10.1186/s12864-018-4515-2.

Cingolani, P., V. M. Patel, M. Coon, T. Nguyen, S. J. Land, D. M. Ruden, and X. Lu. 2012. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. Front. Genet. 3. https://doi.org/10.3389/fgene.2012.00035.

Collard, B. L., P. J. Boettcher, J. M. Dekkers, D. Petitclerc, and L. R. Schaeffer. 2000. Relationships between energy balance and health traits of dairy cattle in early lactation. J. Dairy Sci. 83:2683–2690. https://doi.org/10.3168/jds.S0022-0302(00)75162-9.

Connor, E. E., J. L. Hutchison, K. M. Olson, and H. D. Normal. 2012. Opportunities for improving milk production efficiency in dairy cattle. J. Anim. Sci. 90:1687–1694. https://doi.org/10.2527/jas.2011-4528.

Connor, E. E., S. Kahl, T. H. Elsasser, J. S. Parker, R. W. Li, C. P. Van Tassell, R. L. Baldwin, and S. M. Barao. 2010. Enhanced mitochondrial complex gene function and reduced liver size may mediate improved feed efficiency of beef cattle during compensatory growth. Funct. Integr. Genomics 10:39–51. https://doi.org/10.1007/s10142-009-0138-7.

Danecek, P., and S. A. McCarthy. 2017. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics 33:2037–2039. https://doi.org/10.1093/bioinformatics/btx100.

de Vries, M. J., and R. F. Veerkamp. 2000. Energy balance of dairy cattle in relation to milk production variables and fertility. J. Dairy Sci. 83:62–69. https://doi.org/10.3168/jds.S0022-0302(00)74856-9.

Dias, M. M., A. Cánovas, C. Mantilla-Rojas, D. G. Riley, P. Luna-Nevarez, S. J. Coleman, S. E. Speidel, R. M. Enns, A. Islas-Trejo, J. F. Medrano, S. S. Moore, M. R. Fortes, L. T. Nguyen, B. Venus,

I. S. D. P. Diaz, F. R. P. Souza, L. F. S. Fonseca, F. Baldi, L. G. Albuquerque, M. G. Thomas, and H. N. Oliveira. 2017. SNP detection using RNA-sequences of candidate genes associated with puberty in cattle. Genet. Mol. Res. 16:gmr16019522. https://doi.org/10.4238/gmr16019522.

Ding, R., J. Quan, M. Yang, X. Wang, E. Zheng, H. Yang, D. Fu, Y. Yang, L. Yang, Z. Li, D. Liu, G. Cai, Z. Wu, and J. Yang. 2017. Genome-wide association analysis reveals genetic loci and candidate genes for feeding behavior and eating efficiency in Duroc boars. PLoS One. 12:e0183244. https://doi.org/10.1371/journal.pone.0183244.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

Driver, A. M., F. Penagaricano, W. Huand, K. R. Ahmed, K. S. Hackbart, and M. C. Wiltbank. 2012. RNA-Seq analysis uncovers transcriptomic variations between morphologically similar in vivo- and in vitro-derived bovine blastocysts. BMC Genomics. 13:118. https://doi.org/10.1186/1471-2164-13-118.

Ferreira Júnior, R. J., S. F. M. Bonilha, F. M. Monteiro, J. N. S. G. Cyrillo, R. H. Branco, J. A. Silva, and M. E. Z. Mercadante. 2018. Evidence of negative relationship between female fertility and feed efficiency in Nellore cattle. J. Anim. Sci. 96:4035–4044. https://doi.org/10.1093/jas/sky276.

Fonseca, P. A. S., S. Id-Lahoucine, A. Reverter, J. F. Medrano, M. S. Fortes, J. Casellas, F. Miglior, L. Brito, M. R. S. Carvalho, F. S. Schenkel, L. T. Nguyen, L. R. Porto-Neto, M. G. Thomas, and A. Cánovas. 2018. Combining multi-OMICs information to identify key-regulator genes for pleiotropic effect on fertility and production traits in beef cattle. PLoS One 13:e0205295. https://doi.org/10.1371/journal.pone.0205295.

Food and Agriculture Organization. 2017. The State of Food and Agriculture 2016: Climate Change, Agriculture and Food Security. Food and Agriculture Organization of the United Nations, Rome, Italy.

Han, Y., S. Gao, K. Muegge, W. Zhang, and B. Zhou. 2015. Advanced applications of RNA sequencing and challenges. Bioinform. Biol. Insights 9(S1):29–46. https://doi.org/10.4137/BBI.S28991.

Herd, R. M., and P. F. Arthur. 2009. Physiological basis for residual feed intake. J. Anim. Sci. 87(suppl_14):E64–E71. https://doi.org/10.2527/jas.2008-1345.

Higgins, M. G., D. A. Kenny, C. Fitzsimons, G. Blackshields, S. Coyle, C. McKenna, M. McGee, D. W. Morris, and S. M. Waters. 2019. The effect of breed and diet type on the global transcriptome of hepatic tissue in beef cattle divergent for feed efficiency. BMC Genomics 20:525. https://doi.org/10.1186/s12864-019-5906-8.

Horodyska, J., R. M. Hamill, H. Reyer, N. Trakooljul, P. G. Lawlor, U. M. McCormack, and K. Wimmers. 2019. RNA-Seq of liver from pigs divergent in feed efficiency highlights shifts in macronutrient metabolism, hepatic growth and immune response. Front. Genet. 10:117. https://doi.org/10.3389/fgene.2019.00117.

Hu, Z.-L., E. R. Fritz, and J. M. Reecy. 2007. AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. Nucleic Acids Res. 35(Database):D604–D609. https://doi.org/10.1093/nar/gkl946.

Hu, Z.-L., C. A. Park, and J. M. Reecy. 2019. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. Nucleic Acids Res. 47(D1):D701–D710. https://doi.org/10.1093/nar/gky1084.

Hurley, A. M., N. Lopez-Villalobos, S. McParland, E. Lewis, E. Kennedy, M. O'Donovan, J. L. Burke, and D. P. Berry. 2018. Characteristics of feed efficiency within and across lactation in dairy cows and the effect of genetic selection. J. Dairy Sci. 101:1267–1280. https://doi.org/10.3168/jds.2017-12841.

Keel, B. N., C. M. Zarek, J. W. Keele, L. A. Kuehn, W. M. Snelling, W. T. Oliver, H. C. Freetly, and A. K. Lindholm-Perry. 2018. RNA-Seq Meta-analysis identifies genes in skeletal muscle associated with gain and intake across a multi-season study of crossbred beef steers. BMC Genomics 19:430. https://doi.org/10.1186/s12864-018-4769-8.

Kern, R. J., A. K. Lindholm-Perry, H. C. Freetly, W. M. Snelling, J. W. Kern, J. W. Keele, J. R. Miles, A. P. Foote, W. T. Oliver, L. A. Kuehn, and P. A. Ludden. 2016. Transcriptome differences in the rumen of beef steers with variation in feed intake and gain. Gene 586:12–26. https://doi.org/10.1016/j.gene.2016.03.034.

Koch, R. M., L. A. Swiger, D. Chambers, and K. E. Gregory. 1963. Efficiency of feed use in beef cattle. J. Anim. Sci. 22:486–494. https://doi.org/10.2527/jas1963.222486x.

Kwon, H. J., I. Waghmare, S. Werghese, A. Singh, A. Singh, and M. Kango-Signh. 2015. Drosophila C-terminal Src kinase regulates growth via the Hippo signaling pathway. Dev. Biol. 397:67–76. https://doi.org/http://dx.doi.org/10.1016/j.ydbio.2014.10.010.

Lam, S., J. Zeidan, F. Miglior, A. Suárez-Vega, I. Gómez-Redondo, P. A. S. Fonseca, L. L. Guan, S. Waters, and A. Cánovas. 2020. Development and comparison of RNA-sequencing pipelines for more accurate SNP identification: Practical example of functional SNP detection associated with feed efficiency in Nellore beef cattle. BMC Genomics 21:703. https://doi.org/10.1186/s12864-020-07107-7.

Laramée, M., C. Chabot, M. Cloutier, R. Stenne, M. Holgado-Madruga, A. J. Wong, and I. Royal. 2007. The scaffolding adapter Gab1 mediates vascular endothelial growth factor signaling and is required for endothelial cell migration and capillary formation. J. Biol. Chem. 282:7758–7769. https://doi.org/10.1074/jbc.M611327200.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

Li, S., H. Zhao, Y. Wang, Y. Shao, B. Wang, Y. Wang, and M. Xing. 2018. Regulation of autophagy factors by oxidative stress and cardiac enzymes imbalance during arsenic or/and copper induced cardiotoxicity in *Gallus gallus*. Ecotoxicol. Environ. Saf. 148:125–134. https://doi.org/10.1016/j.ecoenv.2017.10.018.

Liu, J., R. Liu, J. Wang, Y. Zhang, S. Xing, M. Zheng, H. Cui, Q. Li, P. Li, X. Cui, W. Li, G. Zhao, and J. Wen. 2018. Exploring genomic variants related to residual feed intake in local and commercial chickens by whole genomic resequencing. Genes (Basel) 9:57. https://doi.org/10.3390/genes9020057.

Lorch, D. P., C. B. Lindemann, and A. J. Hunt. 2008. The motor activity of mammalian axonemal dynein studied in situ on doublet microtubules. Cell Motil. Cytoskeleton 65:487–494. https://doi.org/10.1002/cm.20277.

Mathi, S. K., J. Chan, and V. M. Watt. 1995. Insulin receptor-related receptor messenger ribonucleic acid: Quantitative distribution and localization to subpopulations of epithelial cells in stomach and kidney. Endocrinology 136:4125–4132. https://doi.org/10.1210/endo.136.9.7649121.

McConnell, J. D., D. K. Stone, L. Johnson, and J. D. Wilson. 1987. Partial purification and characterization of dynein adenosine triphosphatase from bovine sperm. Biol. Reprod. 37:385–393. https://doi.org/10.1095/biolreprod37.2.385.

McLaren, W., B. Pritchard, D. Rios, Y. Chen, P. Flicek, and F. Cunningham. 2010. Deriving the consequences of genomic variants with the Ensembl API and SNP effect predictor. Bioinformatics 26:2069–2070. https://doi.org/10.1093/bioinformatics/btq330.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5:621–628. https://doi.org/10.1038/nmeth.1226.

Mullaney, J. M., R. E. Mills, W. S. Pittard, and S. E. Devine. 2010. Small insertions and deletions (INDELs) in human genomes. Hum. Mol. Genet. 19(R2):R131–R136. https://doi.org/10.1093/hmg/ddq400.

Nafikov, R. A., and D. C. Beitz. 2007. Carbohydrate and lipid metabolism in farm animals. J. Nutr. 137:702–705. https://doi.org/10.1093/jn/137.3.702.

Nielsen, R., J. S. Paul, A. Albrechtsen, and Y. S. Song. 2011. Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12:443–451. https://doi.org/10.1038/nrg2986.

Oltenacu, P. A., and D. M. Broom. 2010. The impact of genetic selection for increased milk yield on the welfare of dairy cows. Anim. Welf. 19:39–49.

Ortigues-Marty, I., G. Cantalapiedra-Hijar, J. Vernet, and P. Nozière. 2017. From dietary energy to nutrient utilisation in ruminants: Which role for splanchnic tissues? Cah. Nutr. Diét. 52:45–53. https://doi.org/10.1016/j.cnd.2016.09.005.

Pareek, C. S., R. Smoczyński, H. N. Kadarmideen, P. Dziuba, P. Błaszczyk, M. Sikora, P. Walendzik, T. Grzybowski, M. Pierzchała, J. Horbańczuk, A. Szostak, M. Ogluszka, L. Zwierzchowski, U. Czarnik, L. Fraser, P. Sobiech, K. Wąsowicz, B. Gelfand, Y. Feng, and D. Kumar. 2016. Single nucleotide polymorphism discovery in bovine pituitary gland using RNA-seq technology. PLoS One 11:e0161370. https://doi.org/10.1371/journal.pone.0161370.

Petrenko, A. G., S. A. Zozulya, I. E. Deyev, and D. Eladari. 2013. Insulin receptor-related receptor as an extracellular pH sensor involved in the regulation of acid-base balance. Biochim. Biophys. Acta 1834:2170–2175. https://doi.org/10.1016/j.bbapap.2012.11.011.

Raile, K., J. Klammt, S. Laue, A. Garten, M. Blüher, S. Kralisch, N. Klöting, and W. Kiess. 2005. Glucose concentration and AMP-dependent kinase activation regulate expression of insulin receptor family members in rat islets and INS-1E beta cells. Diabetologia 48:1798–1809. https://doi.org/10.1007/s00125-005-1860-x.

Ravnskjaer, K., M. F. Hogan, D. Lackey, L. Tora, S. Y. R. Dent, J. Olefsky, and M. Montminy. 2013. Glucagon regulates gluconeogenesis through KAT2B- and WDR5-mediated epigenetic effects. J. Clin. Invest. 123:4318–4328. https://doi.org/10.1172/JCI69035.

Salleh, M. S., G. Mazzoni, J. K. Höglund, D. W. Olijhoek, P. Lund, P. Løvendahl, and H. N. Kadarmideen. 2017. RNA-seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high and low residual feed intake in Nordic dairy cattle. BMC Genomics 18:258. https://doi.org/10.1186/s12864-017-3622-9.

Seymour, D. J., A. Cánovas, T. C. S. Chud, J. P. Cant, V. R. Osborne, C. F. Baes, F. S. Schenkel, and F. Miglior. 2020. The dynamic behavior of feed efficiency in primiparous dairy cattle. J. Dairy Sci. 103:1528–1540. https://doi.org/10.3168/jds.2019-17414.

Shukla, V., C. Cuenin, N. Dubey, and Z. Herceg. Z. 2011. Loss of histone acetyltransferase cofactor transformation/transcription domain-associated protein impairs liver regeneration after toxic injury. Hepatology 53:954–963. https://doi.org/https://doi-org.subzero.lib.uoguelph.ca/10.1002/hep.24120.

Spears, J. W., and W. P. Weiss. 2008. Role of antioxidants and trace elements in health and immunity of transition dairy cows. Vet. J. 176:70–76. https://doi.org/10.1016/j.tvjl.2007.12.015.

Tan, E. C., T. Leung, E. Manser, and L. Lim. 1993. The human active breakpoint cluster region-related gene encodes a brain protein with homology to guanine nucleotide exchange proteins and GTPase-activating proteins. J. Biol. Chem. 268:27291–27298.

Tatulian, S. A. 2015. Structural dynamics of insulin receptor and transmembrane signaling. Biochemistry 54:5523–5532. https://doi.org/10.1021/acs.biochem.5b00805.

Taye, M., J. Kim, S. H. Yoon, W. Lee, O. Hanotte, T. Dessie, S. Kemp, O. A. Mwai, K. Caetano-Anolles, S. Cho, S. J. Oh, H. K. Lee, and H. Kim. 2017. Whole genome scan reveals the genetic signature of African Ankole cattle breed and potential for higher quality beef. BMC Genet. 18:11. https://doi.org/10.1186/s12863-016-0467-1.

Vesterlund, M., F. Zadjali, T. Persson, M. L. Nielsen, B. M. Kessler, G. Norstedt, and A. Flores-Morales. 2011. The SOCS2 ubiquitin ligase complex regulates growth hormone receptor levels. PLoS One 6:e25358. https://doi.org/10.1371/journal.pone.0025358.

Wathes, D. C., G. E. Pollott, K. F. Johnson, H. Richardson, and J. S. Cooke. 2014. Heifer fertility and carry over consequences for life time production in dairy and beef cattle. Animal 8(s1):91–104. https://doi.org/10.1017/S1751731114000755.

Weber, K. L., B. T. Welly, A. L. Van Eenennaam, A. E. Young, L. R. Porto-Neto, A. Reverter, and G. Rincon. 2016. Identification of gene networks for residual feed intake in Angus cattle using genomic prediction and RNA-seq. PLoS One 11:e0152274–e0152274. https://doi.org/10.1371/journal.pone.0152274.

Wickramasinghe, S., A. Cánovas, G. Rincón, and J. F. Medrano. 2014. RNA-sequencing: A tool to explore new frontiers in animal genetics. Livest. Sci. 166:206–216. https://doi.org/10.1016/j.livsci.2014.06.015.

Xia, J., M. J. Benner, and R. E. W. Hancock. 2014. NetworkAnalyst – integrative approaches for protein-protein interaction network analysis and visual exploration. Nucleic Acids Res. 42(W1):W167–W174. https://doi.org/10.1093/nar/gku443.

Xia, J., E. E. Gill, and R. E. W. Hancock. 2015. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nat. Protoc. 10:823–844. https://doi.org/10.1038/nprot.2015.052.

Zhang, Y., B. Li, C. Li, Q. Cai, W. Zheng, and J. Long. 2014. Improved variant calling accuracy by merging replicates in whole-exome sequencing studies. BioMed Res. Int. 2014:319534. https://doi.org/10.1155/2014/319534.

Zheng, Y., C. Zhang, D. R. Croucher, M. A. Soliman, N. St-Denis, A. Pasculescu, L. Taylor, S. A. Tate, W. R. Hardy, K. Colwill, A. Y. Dai, R. Bagshaw, J. W. Dennis, A. Gingras, R. J. Daly, and T. Pawson. 2013. Temporal regulation of EGF signalling networks by the scaffold protein Shc1. Nature 499:166–171. https://doi.org/10.1038/nature12308.

Zhou, N., W. R. Lee, and B. Abasht. 2015. Messenger RNA sequencing and pathway analysis provide novel insights into the biological basis of chickens' feed efficiency. BMC Genomics 16:195. https://doi.org/10.1186/s12864-015-1364-0.

ORCIDS

S. Lam https://orcid.org/0000-0003-2949-1651
F. Miglior https://orcid.org/0000-0003-2345-8842
I. Gómez-Redondo https://orcid.org/0000-0002-9319-7580
A. Suárez-Vega https://orcid.org/0000-0002-7726-4288
F. Schenkel https://orcid.org/0000-0001-8700-0633
L. L. Guan https://orcid.org/0000-0002-8420-4371
P. Stothard https://orcid.org/0000-0003-4263-969X
A. Cánovas https://orcid.org/0000-0002-0036-0757