J. Dairy Sci. 100:1223–1231
https://doi.org/10.3168/jds.2016-11587

Pathway-based genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield, and curd nutrient recovery in dairy cattle

C. Dadousis,* S. Pegolo,* G. J. M. Rosa,†‡ D. Gianola,†‡ G. Bittante,* and A. Cecchinato*[1]
*Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, Viale dell'Università 16, 35020 Legnaro, Italy
†Department of Animal Sciences, and
‡Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison 53706

ABSTRACT

It is becoming common to complement genome-wide association studies (GWAS) with gene-set enrichment analysis to deepen the understanding of the biological pathways affecting quantitative traits. Our objective was to conduct a gene ontology and pathway-based analysis to identify possible biological mechanisms involved in the regulation of bovine milk technological traits: coagulation properties, curd firmness modeling, individual cheese yield (CY), and milk nutrient recovery into the curd (REC) or whey loss traits. Results from 2 previous GWAS studies using 1,011 cows genotyped for 50k single nucleotide polymorphisms were used. Overall, the phenotypes analyzed consisted of 3 traditional milk coagulation property measures [RCT: rennet coagulation time defined as the time (min) from addition of enzyme to the beginning of coagulation; k_{20}: the interval (min) from RCT to the time at which a curd firmness of 20 mm is attained; a_{30}: a measure of the extent of curd firmness (mm) 30 min after coagulant addition], 6 curd firmness modeling traits [RCT_{eq}: RCT estimated through the CF equation (min); CF_P: potential asymptotic curd firmness (mm); k_{CF}: curd-firming rate constant (% × min^{-1}); k_{SR}: syneresis rate constant (% × min^{-1}); CF_{max}: maximum curd firmness (mm); and t_{max}: time to CF_{max} (min)], 3 individual CY-related traits expressing the weight of fresh curd ($\%CY_{CURD}$), curd solids ($\%CY_{SOLIDS}$), and curd moisture ($\%CY_{WATER}$) as a percentage of weight of milk processed and 4 milk nutrient and energy recoveries in the curd (REC_{FAT}, $REC_{PROTEIN}$, REC_{SOLIDS}, and REC_{ENERGY} calculated as the % ratio between the nutrient in curd and the corresponding nutrient in processed milk), milk pH, and protein percentage. Each trait was analyzed separately. In total, 13,269 annotated genes were used in the analysis. The Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway databases were queried for enrichment analyses. Overall, 21 Gene Ontology and 17 Kyoto Encyclopedia of Genes and Genomes categories were significantly associated (false discovery rate at 5%) with 7 traits (RCT, RCT_{eq}, k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS}, and REC_{ENERGY}), with some being in common between traits. The significantly enriched categories included calcium signaling pathway, salivary secretion, metabolic pathways, carbohydrate digestion and absorption, the tight junction and the phosphatidylinositol pathways, as well as pathways related to the bovine mammary gland health status, and contained a total of 150 genes spanning all chromosomes but 9, 20, and 27. This study provided new insights into the regulation of bovine milk coagulation and cheese ability that were not captured by the GWAS.

Key words: milk coagulation and curd firmness, cow cheese ability, genome-wide association, gene-set enrichment, pathway-based analysis

INTRODUCTION

Cheese manufacture is the main final target of dairy cattle milk production in many countries worldwide. Recently, exploitable additive genetic variation has been reported for different measures of individual bovine cheese yield (**CY**; Bittante et al., 2013). Moreover, milk coagulation properties (**MCP**) and curd firmness traits (**CF**) are used as indicators of cheese production. Although considerable additive genetic variation exists for a variety of direct or indirect cheese traits, high measurement costs and logistics place restrictions on the selection of cows for cheese productivity in breeding programs. A potential strategy is to identify and use genomic regions affecting the cow's ability to produce cheese that could enhance genomic breeding programs. Genome-wide association studies (**GWAS**) are widely used for this purpose and were proved to be effective in identifying genomic regions associated with the traits of interest. However, due to the stringent statistical

Received June 10, 2016.
Accepted October 20, 2016.
[1]Corresponding author: alessio.cecchinato@unipd.it

thresholds used to deal with multiple testing, a considerable number of important markers may remain undetected when dealing with polygenic traits (Peng et al., 2010). Moreover, with high SNP density panels, each gene might be represented by several proximal SNP, thus splitting its effect into parts that, in turn, might not be able to pass the defined GWAS threshold in a single marker regression (Ha et al., 2015). Additionally, linkage disequilibrium spans a wide region in the genome, especially in livestock species. As a result, a plethora of SNP might be in linkage disequilibrium with the causal genomic region, which creates extra difficulties in detecting the causal mutation (Hayes, 2013). Besides, although GWAS may be able to locate SNP significantly associated with the trait of interest, it does not make use of the fact that genes work together in biological pathways and are organized into networks. Further, the effect of a multi-allelic QTL may not be fully captured due to the bi-allelic nature of SNP. As a result, GWAS alone may provide a limited understanding of the complex nature of quantitative traits.

A solution to tackle the aforementioned problems, and deepen the understanding of the genetic background of complex traits, is to move up the analysis from the SNP to the gene and gene-set levels. In a gene-set analysis, a group of related genes (such as genes in a specific pathway or gene ontology) that harbor significant SNP previously identified in GWAS, is tested for over-representation in a specific pathway (Wang et al., 2011). Indeed, an increasing interest on pathway analysis has been recently observed in dairy cattle, to complement GWAS analyses of quantitative traits (Gambra et al., 2013; Peñagaricano et al., 2013; Iso-Touru et al., 2016; Abdalla et al., 2016).

Thus, the objective of this study was to conduct a gene ontology and pathway analysis to complement previously obtained GWAS results for phenotypes related to bovine MCP, curd firmness modeling ($\mathbf{CF_t}$), individual CY, and milk nutrient recovery into the curd (**REC**) or whey loss traits.

MATERIALS AND METHODS

Data

Phenotypes. Results of 2 recent GWAS analyses were used, consisting of 11 MCP and CF_t traits (Dadousis et al., 2016) as well as 7 individual CY traits (Dadousis et al., 2017). In brief, the milk MCP-CF_t data set contained the milk pH, milk protein percentage, 3 traditional MCP obtained from Formagraph [**RCT**: rennet coagulation time defined as the time (min) from addition of enzyme to the beginning of coagulation; k_{20}: the interval (min) from RCT to the time at which a curd firmness of 20 mm is attained; a_{30}: a measure of the extent of curd firmness (mm) 30 min after coagulant addition], 4 CF_t equation parameters [$\mathbf{RCT_{eq}}$: RCT estimated through the CF_t equation (min); $\mathbf{CF_P}$: potential asymptotical curd firmness (mm); $\mathbf{k_{CF}}$: curd-firming rate constant ($\% \times \text{min}^{-1}$); $\mathbf{k_{SR}}$: syneresis rate constant ($\% \times \text{min}^{-1}$)], and 2 derived traits [$\mathbf{CF_{max}}$: maximum curd firmness (mm) and $\mathbf{t_{max}}$: time to CF_{max} (**tmin**)]. The second GWAS data set included 3 individual CY traits expressing the weight of fresh curd ($\mathbf{\%CY_{CURD}}$), curd solids ($\mathbf{\%CY_{SOLIDS}}$), and curd moisture ($\mathbf{\%CY_{WATER}}$) as a percentage of weight of milk processed, and 4 milk nutrient and energy recoveries into the curd ($\mathbf{REC_{FAT}}$, $\mathbf{REC_{PROTEIN}}$, $\mathbf{REC_{SOLIDS}}$, and $\mathbf{REC_{ENERGY}}$), calculated as the % ratio between the nutrient in curd and the corresponding nutrient/energy in the processed milk. Details about the genotyping and the GWAS analyses are reported in (Dadousis et al., 2016, 2017).

Genotypic Data. Briefly, 1,152 cows were genotyped with the Illumina BovineSNP50 Bead Chip v.2 (Illumina Inc., San Diego, CA). After quality control [call rate >95%, minor allele frequency >0.05, and extreme deviation from Hardy-Weinberg proportions ($P > 0.001$, Bonferroni corrected)], 1,011 animals and 37,568 SNP, located on 29 autosomes and in the X-chromosome, were retained. Slight differences in the number of individuals and SNP between the 2 GWAS analyses are attributed to phenotypic editing.

Gene-Set Enrichment and Pathway-Based Analysis

The gene-set enrichment analysis workflow is represented in Figure 1. In brief, for each trait, nominal P-values < 0.05 from the GWAS analyses were used to identify significant SNP. Using the *biomaRt* R package (Durinck et al., 2005, 2009), the SNP were assigned to genes if they were within the genomic sequence of the gene or within a flanking region of 15 kb up- and downstream of the gene, to include SNP located in regulatory regions. The size of the flanking region was based on the finding that most SNP that affect the expression of a gene are located within 15 kb of the gene (Pickrell et al., 2010). The Ensembl *Bos taurus* UMD3.1 database was used as reference (Zimin et al., 2009). The background SNP represent all the SNP tested in the GWAS analyses, while the background genes were the genes associated with those SNP. For the assignment of the genes to functional categories, the Gene Ontology (**GO**; Ashburner et al., 2000) and Kyoto Encyclopedia of Genes and Genomes (**KEGG**) pathway (Ogata et al., 1999) databases were used. The GO database designates biological descriptors (GO terms) to genes based on attributes of their encoded products and it

is further partitioned into 3 components: biological process, molecular function, and cellular component. The KEGG pathway database contains metabolic and regulatory pathways, representing the actual knowledge on molecular interactions and reaction networks. To avoid testing narrow or broad functional categories, only GO and KEGG categories with more than 10 and less than 1,000 genes were tested. Finally, a Fisher's exact test was performed to test for overrepresentation of the significant genes for each gene-set (i.e., pathway/functional category). False discovery rate correction (controlled at 5%) was applied to account for multiple testing. The gene enrichment analysis was performed with the *goseq* R package (Young et al., 2010).

RESULTS AND DISCUSSION

In total, 17,006 SNP (out of the 37,568 tested) were located in annotated genes or in the 15 kb window (upstream or down-stream from a gene). The total number of background genes annotated in the *Bos taurus* UMD3.1 assembly was 13,269. Each trait was analyzed separately. On average, 1,700 SNP, ranging between 1,301 for $REC_{PROTEIN}$ to 1,899 for RCT_{eq}, were significantly associated with each trait in the GWAS analysis. For each trait, 585 significant SNP were assigned to 500 genes, on average (Figure 1, Table 1). The minimum number of mapped genes was found for $REC_{PROTEIN}$ (n = 399), whereas the maximum for RCT_{eq} (n = 574).

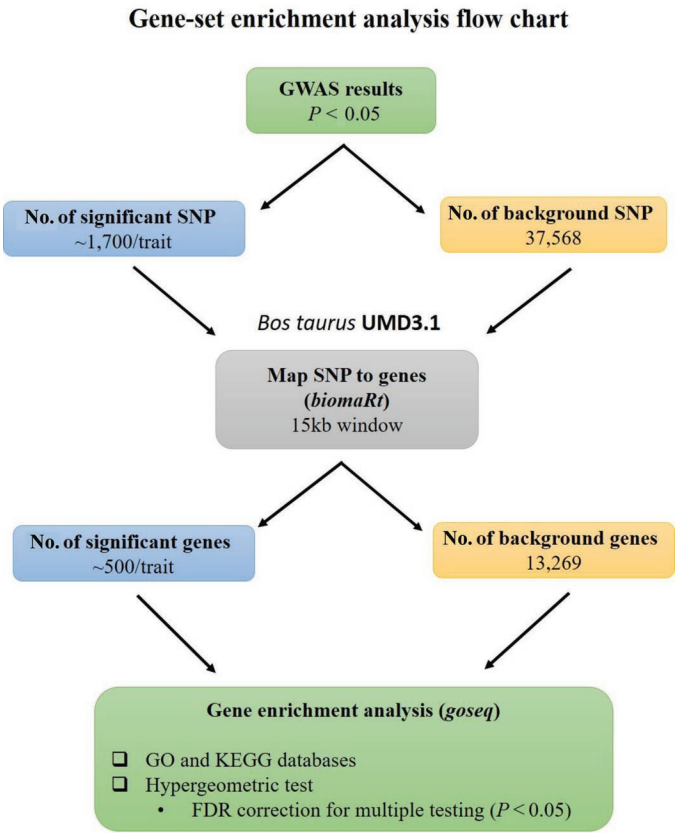

Figure 1. Flowchart for the gene-set enrichment analysis. GWAS = genome-wide association studies; GO = Gene Ontology; KEGG = Kyoto Encyclopedia of Genes and Genomes; FDR = false discovery rate. Color version available online.

Enrichment Pathway Analysis

After false discovery rate correction, 21 GO and 17 KEGG categories were found associated with 7 of the tested traits, namely RCT, RCT_{eq}, k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS}, and REC_{ENERGY}. Some of the categories were in common between traits. A total of 150 significant genes spanning all *Bos taurus* chromosomes (BTA) but 9, 20, and 27 were included into the significantly enriched GO and KEGG categories (Supplemental Table S1; https://doi.org/10.3168/jds.2016-11587). Table 2 summarizes all the significant pathways/ontologies, some of which were shared among the aforementioned traits. More precisely, the calcium signaling pathway (KEGG:bta04020) was associated with both RCT and REC_{FAT}; the arrhythmogenic right ventricular cardiomyopathy (**ARVC**) pathway (KEGG:bta05412) was enriched for both $\%CY_{SOLIDS}$ and REC_{SOLIDS}; the leucocyte transendothelial migration pathway (KEGG:bta04670) was in common between REC_{SOLIDS} and REC_{ENERGY}; and the synapse part cellular component (GO:0044456) was shared between RCT_{eq} and k_{CF}. Moreover, 6 GO biological process categories related to female sex characteristics and the ovulation cycle appeared significant for both RCT_{eq} and REC_{FAT} (Table 2). Not surprisingly, different pathways/functional categories were enriched for RCT and RCT_{eq}, reflecting the differences in their additive genetic variance found in the GWAS analysis.

Phosphatidylinositol Signaling Pathway. The phosphatidylinositol signaling pathway (KEGG:bta04070) was significantly enriched for the RCT trait. In milk, phospholipids are mainly present on the surface of the milk fat globules (**MFG**) and are responsible for the stabilization of the milk fat against coalescence (Rombaut et al., 2007; Walstra et al., 2014). Due to their technological and nutritional properties, previous studies focused on determining the phospholipid content of various dairy products (Rombaut et al., 2007). Recently, the evolution of the phospholipids during the quark cheese process from buttermilk was also examined (Ferreiro et al., 2016). Phosphatidylinositol represents a small fraction of the phospholipid components of milk. Among the significant genes included in the phosphatidylinositol pathway, 3 phospholipase C β (*PLCB*) isoforms were present: *PLCB1*, *PLCB3*,

Table 1. Number of significant[1] SNP identified from genome-wide association studies (GWAS) and genes mapped by trait

Trait[2]	No. of significant SNP	No. of significant SNP assigned to genes	No. of significant mapped genes[3]
Milk composition			
pH	1,848	624	552
Protein, %	1,808	641	563
Traditional MCP			
RCT	1,739	551	487
k_{20}	1,789	614	539
a_{30}	1,724	572	496
Curd firming			
RCT_{eq}	1,899	639	574
CF_P	1,423	486	422
CF_{max}	1,724	599	536
t_{max}	1,786	598	531
k_{CF}	1,822	603	545
k_{SR}	1,872	625	562
Cheese yields, %			
$\%CY_{CURD}$	1,817	621	538
$\%CY_{SOLIDS}$	1,796	590	533
$\%CY_{WATER}$	1,826	605	525
Recoveries, %			
REC_{SOLIDS}	1,797	581	527
REC_{FAT}	1,496	503	447
$REC_{PROTEIN}$	1,301	444	399
REC_{ENERGY}	1,800	627	548
Background[4]	37,568	17,006	13,269

[1]*P*-value < 0.05.

[2]pH = milk pH; Protein, % = milk protein (%); MCP = milk coagulation properties; RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; k_{20} = curd-firming time (min) of samples reaching 20 mm of firmness within 45 min from enzyme addition; a_{30} = curd firmness (mm) at 30 min after enzyme addition; RCT_{eq} = rennet coagulation time (min) estimated using the CF_t equation; CF_P = potential asymptotical curd firmness (mm); k_{CF} = curd-firming rate constant ($\% \times min^{-1}$); k_{SR} = syneresis rate constant ($\% \times min^{-1}$); CF_{max} = maximum curd firmness (mm); t_{max} = time to CF_{max} (min); %CY = weight of fresh curd, curd solids, and curd water as percentage of weight of milk processed; REC = protein, fat, solids, and energy of the curd as percentage of the protein, fat, solids, and energy of the milk processed.

[3]Ensembl *Bos taurus* UMD3.1 (http://www.ensembl.org/index.html); window: 15 kb.

[4]Background represents the total number of SNP used in the GWAS analyses, the number of SNP linked to genes and the genes mapped to those SNP.

and *PLCB4*. Phospholipases are responsible for the hydrolyses of phospholipids of the MFG membrane, thereby affecting the stability of the cream emulsion. In addition, the phospholipase treatment of milk was found to reduce fat losses in whey and cooking water and to increase CY by improving fat and moisture retention in the cheese curd in Mozzarella cheese (Lilbæk et al., 2006). Interestingly, an association between a SNP on *PLCB1* (rs41624761) and k_{CF} has been previously reported in a candidate gene analysis (Cecchinato et al., 2015). Moreover, studies related to the effect of MFG size on milk technological properties reported a significant relationship of MFG size with MCP (Bland et al., 2015), cheese ripening, and structure, as well as stability of dairy products (Lopez et al., 2011). Indeed, the biological explanation of the connection between phosphatidylinositol pathway and MCP can be found in the tight association between MFG size and phospholipids contents, with higher amounts of phospholipids in small versus large globules likely affecting, in turn, the technological properties of milk.

The salivary secretion pathway (KEGG:bta04970), which was also enriched for the RCT trait, shared 6 genes with the phosphatidylinositol signaling pathway including *PLCB1*, *PLCB3*, and *PLCB4*. Interestingly, histatherin (*HSTN*) was also present in the list of significant genes for this pathway (Supplemental Table S1; https://doi.org/10.3168/jds.2016-11587). Histatherin is a ruminant-specific gene that encodes for a host-defense related protein in the cow's oral cavity and milk, which may also be involved in the response to mastitis (Ju, 2014). The *HSTN* has been also proposed as a candidate gene related to MCP and CY traits in our previous GWAS analyses (Dadousis et al., 2016, 2017) and more precisely it was associated with the 3 MCP, CF_P, CF_{max}, and REC_{FAT}.

Calcium Signaling-Related Pathway. The calcium signaling pathway (KEGG:bta04020) was sig-

Table 2. Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways significantly enriched using genes associated with RCT, RCT_{eq}, k_{CF}, $\%CY_{SOLIDS}$, REC_{FAT}, REC_{SOLIDS}, and REC_{ENERGY}

Trait[1]	Category[2]	Term	No. of genes in the term	No. of significant genes[3]	FDR[4]
RCT	KEGG	bta04070:Phosphatidylinositol signaling	53	10	1.87×10^{-5}
		bta04730:Long-term depression	44	8	1.69×10^{-4}
		bta04540:Gap junction	57	9	2.07×10^{-4}
		bta04270:Vascular smooth muscle contraction	74	10	3.47×10^{-4}
		bta04020:Calcium signaling pathway	118	13	3.79×10^{-4}
		bta04970:Salivary secretion	55	8	8.18×10^{-4}
RCT_{eq}	GO_BP	GO:0048511~Rhythmic process	46	9	1.87×10^{-4}
		GO:0008585~Female gonad development	14	5	2.15×10^{-4}
		GO:0022602~Ovulation cycle process	14	5	2.15×10^{-4}
		GO:0042698~Ovulation cycle	14	5	2.15×10^{-4}
		GO:0046545~Development of primary female sexual characteristics	14	5	2.15×10^{-4}
		GO:0046660~Female sex differentiation	14	5	2.15×10^{-4}
	GO_CC	GO:0030425~Dendrite	39	11	5.01×10^{-7}
		GO:0044456~Synapse part	76	15	7.12×10^{-7}
		GO:0097458~Neuron part	162	22	1.84×10^{-6}
		GO:0045202~Synapse	102	16	7.30×10^{-6}
		GO:0036477~Somatodendritic compartment	62	11	6.15×10^{-5}
		GO:0043005~Neuron projection	110	15	7.58×10^{-5}
		GO:0008076~Voltage-gated potassium channel complex	12	5	9.15×10^{-5}
		GO:0034705~Potassium channel complex	13	5	1.44×10^{-4}
		GO:0008021~Synaptic vesicle	22	6	2.63×10^{-4}
		GO:0098793~Presynapse	22	6	2.63×10^{-4}
k_{CF}	GO_CC	GO:0044456~Synapse part	76	13	1.16×10^{-5}
		GO:0098794~Postsynapse	48	10	2.01×10^{-5}
$\%CY_{SOLIDS}$	KEGG	bta05412:Arrhythmogenic right ventricular cardiomyopathy (ARVC)	50	10	2.42×10^{-5}
		bta4260:Cardiac muscle contraction	41	8	1.90×10^{-4}
REC_{FAT}	GO_BP	GO:0008585~Female gonad development	14	6	3.37×10^{-6}
		GO:0022602~Ovulation cycle process	14	6	3.37×10^{-6}
		GO:0042698~Ovulation cycle	14	6	3.37×10^{-6}
		GO:0046545~Development of primary female sexual characteristics	14	6	3.37×10^{-6}
		GO:0046660~Female sex differentiation	14	6	3.37×10^{-6}
		GO:0001541~Ovarian follicle development	10	5	9.30×10^{-6}
		GO:0008406~Gonad development	21	6	4.98×10^{-5}
		GO:0045137~Development of primary sexual characteristics	21	6	4.98×10^{-5}
		GO:0007548~Sex differentiation	24	6	1.13×10^{-4}
		GO:0048511~Rhythmic process	46	8	1.32×10^{-4}
	KEGG	bta04020:Calcium signaling pathway	118	13	1.68×10^{-4}
REC_{SOLIDS}	KEGG	bta05412:Arrhythmogenic right ventricular cardiomyopathy (ARCV)	50	9	1.34×10^{-4}
		bta05218:Melanoma	44	8	2.94×10^{-4}
		bta05200:Pathways in cancer	195	18	7.73×10^{-4}
		bta01100:Metabolic pathways	635	42	8.10×10^{-4}
		bta04260:Cardiac muscle contraction	41	7	1.04×10^{-3}
		bta04670:Leucocyte transendothelial migration	66	9	1.13×10^{-3}
		bta05213:Endometrial cancer	31	6	1.20×10^{-3}
		bta04930:Type II diabetes mellitus	32	6	1.43×10^{-3}
		bta04973:Carbohydrate digestion and absorption	24	5	2.20×10^{-3}
REC_{ENERGY}	KEGG	bta04670:Leukocyte transendothelial migration	66	10	3.49×10^{-4}
		bta04514:Cell adhesion molecules	79	11	3.81×10^{-4}
		bta04530:Tight junction	83	11	5.87×10^{-4}

[1]RCT = rennet coagulation time (min) of samples coagulating within 45 min from enzyme addition; RCT_{eq} = rennet coagulation time (min) estimated using the CF_t equation; k_{CF} = curd-firming rate constant ($\% \times min^{-1}$); $\%CY_{SOLIDS}$ = weight of curd solids as percentage of weight of milk processed; REC = fat, solids, and energy of the curd as percentage of the fat, solids, and energy of the milk processed.

[2]KEGG: KEGG pathway; GO_BP: GO biological process; GO_CC: GO cellular component.

[3]Significant genes after mapping the significant SNP to genes using Ensembl *Bos taurus* UMD3.1 as reference (http://www.ensembl.org/index.html).

[4]False discovery rate (FDR) correction for multiple testing (P-value < 0.05).

nificantly enriched for both RCT and REC_{FAT}. It is widely known that Ca is one of the major components of the casein micelles. During the cheese process, after rennet addition, the casein reacts with Ca ions and precipitates. This phenomenon consists the basis of milk clotting (Walstra et al., 2014). Moreover, low content of the total and micellar Ca has been associated with noncoagulating milk (Gustavsson et al., 2014; Malacarne et al., 2014). Interestingly, transcriptomic analysis of mammary gland in mice showed that the calcium ion binding ontology was significantly over-represented among the differentially expressed genes associated with enhanced maternal performance phenotype (Ramanathan et al., 2008). Further analysis showed a positive correlation between the calcium signaling pathway and the lactation performance in mice (Wei et al., 2013).

The ARVC pathway (KEGG:bta05412) was enriched for both $\%CY_{SOLIDS}$ and REC_{SOLIDS} with 9 significantly enriched genes being in common. Moreover, this pathway shared 6 genes with the cardiac muscle contraction pathway (KEGG:bta04260; Supplemental Table S1; https://doi.org/10.3168/jds.2016-11587), which was also enriched for the $\%CY_{SOLIDS}$. Notably, genes encoding for several subunits of the voltage-dependent calcium channel complex were included in these pathways (e.g., calcium voltage-gated channel subunit α-1D and calcium voltage-gated channel auxiliary subunit α-2/delta-3].

Bovine Reproduction-Related Ontologies. The rennet coagulation time obtained from an extended CF testing time (RCT_{eq}) and the REC_{FAT} were associated with the GO terms related to female characteristics such as the ovulation cycle (GO:0042698) and female gonad development (GO:0008585). All the significant genes were shared among these biological processes (Supplemental Table S1; https://doi.org/10.3168/jds.2016-11587). In a similar gene-set enrichment and pathway analysis, the broader GO categories of reproduction (GO:0000003) and reproductive process (GO:002214) were associated with milk yield, milk fat and protein yields, and fertility index in the Nordic Red cattle (Iso-Touru et al., 2016). Indeed, a close link between the duration of estrus and multiple ovulation rate and milk production in dairy cattle was previously reported (Wiltbank et al., 2006).

When specifically looking at the significant genes involved in these pathways/processes (Supplemental Table S1; https://doi.org/10.3168/jds.2016-11587), the $\alpha\text{-casein}_{S1}$ (*CSN1S1*) and the luteinizing hormone receptor (*LHCGR*) were included. Interestingly, in a candidate gene approach, an association between an SNP on *CSN1S1* (rs109817504) and CF_P has been previously reported (Cecchinato et al., 2015). Moreover, *CSN1S1* genetic variants were shown to affect MCP of buffalo and goat milk (Caravaca et al., 2011; Devold et al., 2011; Bonfatti et al., 2012). Considering *LHCGR*, this gene has been significantly associated with milk composition and in particular with milk fat and total solid percentages (Molee et al., 2015). Notably, a positive effect of fat on milk coagulation properties was highlighted by Bland et al. (2015).

Additionally, ontologies related to the nervous tissue and more specifically to neuron parts (e.g., dendrite) and functions (e.g., synapse) were enriched for the RCT_{eq}. A possible interpretation can rely on the fact that during pregnancy and lactation many factors and signals (including the neuroendocrine signal of prolactin) act to adapt the pattern of neuronal responses to the lactating state (Grattan, 2002; Akers, 2016). Interestingly, in 3 of these GO terms (GO:0044456, GO:0097458, and GO:0045202), the vacuolar protein sorting-associated protein 35 gene (*VPS35*) was included. This gene has been recently proposed as a candidate gene strongly related to milk coagulation in Swedish Red cows (Duchemin et al., 2016).

Mammary Gland- and Mastitis-Related Pathways and Ontologies. Ontologies related to potassium channels (GO:0008076, GO:0034705) were significantly enriched for RCT_{eq}. The role of the voltage-gated potassium channels is to transfer ions across the cell membrane (Yellen, 2002). In milk, the concentrations of Na^+, K^+, and Cl^- are the most important ions for electrical conductivity (**EC**). It is well established that milk EC can be also used as an indicator of mastitis (Norberg, 2005; Viguier et al., 2009). Interestingly, the tight junction pathway (KEGG:bta04530) was significantly enriched for the REC_{ENERGY}. Tight junctions of the mammary epithelium control the movement of lactose and K^+ to the extracellular fluid, while Na^+ and Cl^- are moving into the milk. Tight junctions are known to be related to milk mammary gland development and milk secretion (Nguyen and Neville, 1998; Ramanathan et al., 2008; Stelwagen and Singh, 2014). More precisely, increased milk secretion is connected to a decrease in the tight junction permeability. After intramammary infection, destruction of tight junctions and of the ion-pumping system causes an increase in the concentration of Na^+ and Cl^- in the milk and consequently increases the milk EC (Norberg, 2005). It has been reported that the technological properties of milk (such as MCP and CF_t) are unfavorably influenced by mastitis indicators (Bittante et al., 2012; Bobbo et al., 2016). Indeed, other pathways related to the mastitis were significantly enriched. In particular, REC_{SOLIDS} and REC_{ENERGY} were associated with the leucocyte transendothelial migration (KEGG:bta04670). Leucocytes are typically present in milk and compose the majority of the SCC. Their concentration in milk increases

after bacterial infections and thus they are widely used as an indicator of mastitis (Dosogne et al., 2003). The leucocyte transendothelial migration pathway has been previously linked to milk and fat yield in dairy cattle (Edwards et al., 2015). Immune response-related categories [e.g., the immune system process (GO:0002376)], have been recently found to be related to fat yield, milk yield, and fertility (Iso-Touru et al., 2016). The tight junction and the leucocyte transendothelial migration pathways shared 3 significant genes, namely junctional adhesion molecule 2 (*JAM2*), actinin α 1 (*ACTN1*), and catenin α 3 (*CTNNA3*; Supplemental Table S1; https://doi.org/10.3168/jds.2016-11587). Interestingly, *JAM2* is located on BTA1 at ~10.1 Mb and a weak signal for RCT_{eq} at ~9.5 Mb has been detected in our GWAS analysis (Dadousis et al., 2016).

In addition to this, the cell adhesion molecule pathway (KEGG:bta04514) was enriched for REC_{ENERGY}. The cell adhesion molecule pathway is involved in a wide range of biologic processes, including immune response and neuronal cell adhesion. In the study of Ramanathan et al. (2008), differentially expressed genes belonging to this pathway were also enriched and related to the mammary development and milk secretion in mice. Moreover, the gap junction pathway (KEGG:bta04540), related to cell communication, was enriched for RCT. Functional analyses evidenced that the broader pathway of cell-cell signaling was significantly associated with MFG global gene expression during lactation in human (Maningat et al., 2009).

Interestingly, the ARVC pathway detected in our study has also been associated with the mammary gland functionality in pregnant sows in a study focusing in sow's mammary transcriptome in late gestation (Zhao et al., 2013).

Metabolism-Related Pathways. The broad category of "metabolic pathways" (KEGG:bta01100) was associated with REC_{SOLIDS}. Among the genes included in this specific pathway, 3 polypeptide N-acetylgalactosaminyltransferase (*GALNT*) isoforms were significant in our study (*GALNT1*, *GALNT13*, and *GALNT18*; Supplemental Table S1; https://doi.org/10.3168/jds.2016-11587). Although no association was detected in our previous GWAS analyses on BTA24, the *GALNT1* gene located on BTA24 has been reported as a candidate gene in another recent GWAS study related to MCP (Gregersen et al., 2015). This gene encodes for the GalNAc-T enzyme that is involved to κ-casein glycosylation (Holland et al., 2005). Higher content of glycosylated kappa casein has been linked to improved milk coagulation (Poulsen et al., 2016). In the recent gene-set enrichment study of Iso-Touru et al. (2016), the metabolic process ontology (GO:0008152) was significantly enriched for milk, fat, and protein yields and fertility in Nordic Red cattle. Moreover, in our study the carbohydrate digestion and absorption pathway (KEGG:bta04973) was enriched for REC_{SOLIDS}. The central carbohydrate of the milk is lactose. Although a strong influence of lactose on MCP has been recently reported (Bland et al., 2015), our knowledge on the exact mechanism is still limited. However, lactose is also related to SCC and mastitis. More precisely, a decrease of lactose is observed during mastitis (Kitchen, 1981). Not surprisingly, 2 genes (phosphoinositide-3-kinase, regulatory subunits 3 and 5; *PIK3R3* and *PIK3R5*) were in common between the carbohydrate digestion and absorption, and the leukocyte transendothelial migration pathways (Supplemental Table S1; https://doi.org/10.3168/jds.2016-11587).

Cancer-Related Pathways. Among the significantly enriched pathways detected for REC_{SOLIDS}, pathways in cancer (KEGG:bta05200), endometrial cancer (KEGG:bta05213), and melanoma (KEGG:bta05218) were present with some significant genes being in common, including also *PIK3R3*, *PIK3R5*, and AKT serine/threonine kinase 3 (*AKT3*; Supplemental Table S1; https://doi.org/10.3168/jds.2016-11587). These genes have a role in the PIK-Akt signaling pathway, which was associated with mammary development and breast cancer (Wickenden and Watson, 2010). Moreover, in a gene expression study in human mammary gland investigating the MFG transcriptome, one of the most significant networks associated with the top expressed genes was the cancer pathway (Maningat et al., 2009). Moreover, the aforementioned cancer-related KEGG pathways were associated with genetic variants in mammary development, prolactin signaling, and involution pathways, which were linked to bovine milk production traits (Raven et al., 2014). However, as highlighted by those authors, the significance of this information is challenged by the fact that KEGG database includes a large compendium of cancer-related gene sets.

Apparently, aside from pathways that could be strictly associated with milk technological properties (e.g., phosphatidylinositol signaling), pathways not directly related to the traits of our interest were also detected (e.g., pathways related to cancer). However, as publicly available ontologies and pathways in cattle are still limited (compared with human) and not all are well described, some of our results may be misleading, especially when the detected genes are involved in various biological processes (Fan et al., 2015). It is likely that, when more complete gene sets become available, more competitive pathways might be detected and the power to identify genomic regions influencing these traits might increase. In this respect, transcriptomic methods (e.g., RNA-seq) may represent a useful tool

to complement the present analysis and validate the achieved biological information.

Finally, it is worth noting that our gene-set enrichment analysis was conducted using a panel of SNP obtained from a single marker regression GWAS, which relies on a simplified theory of the genomic background of traits, without considering for instance the joint effect of SNP. Hence, other approaches (e.g., GWAS exploring SNP by SNP interactions) might provide a better basis for biological pathway analysis.

CONCLUSIONS

To our knowledge, this is the first pathway-based association analysis related to milk technological traits. In animal breeding, studies are generally focused on simple SNP-based associations with the traits of interest. The present pathway-based analysis provided new insights with respect to the previously conducted GWAS analyses, confirming that complex traits (i.e., milk technological properties) may be affected by the joint additive effect of several genes which cluster in specific biological pathways. In particular, calcium and phosphatidylinositol signaling, overall metabolism, carbohydrate digestion and absorption as well as pathways related to the bovine mammary gland health status were significantly enriched. The highlighted pathways and gene ontologies detected associated with technological traits may be useful in further studies on fine mapping of genes and development of marker-assisted breeding programs. However, further validation and replication of the most promising described pathways is needed to explore their role in relation to bovine milk coagulation and cheese-making characteristics.

ACKNOWLEDGMENTS

The authors thank the Trento Province (Italy), the Italian Brown Swiss Cattle Breeders Association (ANARB, Verona, Italy), and the Superbrown Consortium of Bolzano and Trento for financial and technical support. C. Dadousis benefitted from financial support of the CARIPARO (Cassa di Risparmio di Padova e Rovigo) Foundation (Padua, Italy). The authors also thank F. Peñagaricano (Department of Animal Sciences, University of Florida) for his help in setting up the analysis.

REFERENCES

Abdalla, E. A., F. Peñagaricano, T. Byrem, K. Weigel, and G. Rosa. 2016. Genome-wide association mapping and pathway analysis of leukosis incidence in a US Holstein cattle population. Anim. Genet.47:395–407.

Akers, R. M. 2016. Lactation and the Mammary Gland. John Wiley & Sons, Hoboken, NJ.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, and J. T. Eppig. 2000. Gene ontology: Tool for the unification of biology. Nat. Genet. 25:25–29.

Bittante, G., C. Cipolat-Gotet, and A. Cecchinato. 2013. Genetic parameters of different measures of cheese yield and milk nutrient recovery from an individual model cheese-manufacturing process. J. Dairy Sci. 96:7966–7979.

Bittante, G., M. Penasa, and A. Cecchinato. 2012. Invited review: Genetics and modeling of milk coagulation properties. J. Dairy Sci. 95:6843–6870.

Bland, J. H., A. S. Grandison, and C. C. Fagan. 2015. Evaluation of milk compositional variables on coagulation properties using partial least squares. J. Dairy Res. 82:8–14.

Bobbo, T., C. Cipolat-Gotet, G. Bittante, and A. Cecchinato. 2016. The nonlinear effect of somatic cell count on milk composition, coagulation properties, curd firmness modeling, cheese yield, and curd nutrient recovery. J. Dairy Sci.99:5104–5119.

Bonfatti, V., M. Giantin, M. Gervaso, A. Coletta, M. Dacasto, and P. Carnier. 2012. Effect of CSN1S1–CSN3 (α S1-κ-casein) composite genotype on milk production traits and milk coagulation properties in Mediterranean water buffalo. J. Dairy Sci. 95:3435–3443.

Caravaca, F., J. L. Ares, J. Carrizosa, B. Urrutia, F. Baena, J. Jordana, B. Badaoui, A. Sànchez, A. Angiolillo, and M. Amills. 2011. Effects of α s1-casein (CSN1S1) and κ-casein (CSN3) genotypes on milk coagulation properties in Murciano-Granadina goats. J. Dairy Res. 78:32–37.

Cecchinato, A., S. Chessa, C. Ribeca, C. Cipolat-Gotet, T. Bobbo, J. Casellas, and G. Bittante. 2015. Genetic variation and effects of candidate-gene polymorphisms on coagulation properties, curd firmness modeling and acidity in milk from Brown Swiss cows. Animal 9:1104–1112.

Dadousis, C., S. Biffani, C. Cipolat-Gotet, E. Nicolazzi, A. Rossoni, E. Santus, G. Bittante, and A. Cecchinato. 2016. Genome-wide association of coagulation properties, curd firmness modeling, protein percentage, and acidity in milk from Brown Swiss cows. J. Dairy Sci. 99:3654–3666.

Dadousis, C., S. Biffani, C. Cipolat-Gotet, E. Nicolazzi, G. J. M. Rosa, D. Gianola, A. Rossoni, E. Santus, G. Bittante, and A. Cecchinato. 2017. Genome-wide association study for cheese yield and curd nutrient recovery in dairy cows. J. Dairy Sci. https://doi.org/10.3168/jds.2016-11586.

Devold, T. G., R. Nordbø, T. Langsrud, C. Svenning, M. J. Brovold, E. S. Sørensen, B. Christensen, T. Ådnøy, and G. E. Vegarud. 2011. Extreme frequencies of the α_{s1}-casein "null" variant in milk from norwegian dairy goats—Implications for milk composition, micellar size and renneting properties. Dairy Sci. Technol. 91:39–51.

Dosogne, H., F. Vangroenweghe, J. Mehrzad, A. M. Massart-Leën, and C. Burvenich. 2003. Differential leukocyte count method for bovine low somatic cell count milk. J. Dairy Sci. 86:828–834.

Duchemin, S. I., M. Glantz, D. De Koning, M. Paulsson, and W. F. Fikse. 2016. Identification of QTL on chromosome 18 associated with non-coagulating milk in Swedish red cows. Front. Genet. 7:57.

Durinck, S., Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, and W. Huber. 2005. BioMart and bioconductor: A powerful link between biological databases and microarray data analysis. Bioinformatics 21:3439–3440.

Durinck, S., P. T. Spellman, E. Birney, and W. Huber. 2009. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. Nat. Protoc. 4:1184–1191.

Edwards, S. M., B. Thomsen, P. Madsen, and P. Sørensen. 2015. Partitioning of genomic variance reveals biological pathways associated with udder health and milk production traits in dairy cattle. Genet. Sel. Evol. 47:60.

Fan, H., Y. Wu, X. Zhou, J. Xia, W. Zhang, Y. Song, F. Liu, Y. Chen, L. Zhang, X. Gao, H. Gao, and J. Li. 2015. Pathway-based genome-wide association studies for two meat production traits in Simmental cattle. Sci. Rep. 5:18389.

Ferreiro, T., S. Martínez, L. Gayoso, and J. Rodríguez-Otero. 2016. Evolution of phospholipid contents during the production of quark cheese from buttermilk. J. Dairy Sci. 99:4154–4159.

Gambra, R., F. Peñagaricano, J. Kropp, K. Khateeb, K. Weigel, J. Lucey, and H. Khatib. 2013. Genomic architecture of bovine κ-casein and β-lactoglobulin. J. Dairy Sci. 96:5333–5343.

Grattan, D. R. 2002. Behavioural significance of prolactin signalling in the central nervous system during pregnancy and lactation. Reproduction 123:497–506.

Gregersen, V. R., F. Gustavsson, M. Glantz, O. F. Christensen, H. Stålhammar, A. Andrén, H. Lindmark-Månsson, N. A. Poulsen, L. B. Larsen, and M. Paulsson. 2015. Bovine chromosomal regions affecting rheological traits in rennet-induced skim milk gels. J. Dairy Sci. 98:1261–1272.

Gustavsson, F., M. Glantz, A. J. Buitenhuis, H. Lindmark-Månsson, H. Stålhammar, A. Andrén, and M. Paulsson. 2014. Factors influencing chymosin-induced gelation of milk from individual dairy cows: Major effects of casein micelle size and calcium. Int. Dairy J. 39:201–208.

Ha, N., J. J. Gross, A. van Dorland, J. Tetens, G. Thaller, M. Schlather, R. Bruckmaier, and H. Simianer. 2015. Gene-based mapping and pathway analysis of metabolic traits in dairy cows. PLoS One 10:e0122325.

Hayes, B. 2013. Overview of statistical methods for genome-wide association studies (GWAS). Pages 149–169 in Genome-Wide Association Studies and Genomic Prediction. Methods in Molecular Biology. G. Cedric, J. van der Werf, and B. Hayes, ed. Humana Press, New York, NY.

Holland, J. W., H. C. Deeth, and P. F. Alewood. 2005. Analysis of o-glycosylation site occupancy in bovine κ-casein glycoforms separated by two-dimensional gel electrophoresis. Proteomics 5:990–1002.

Iso-Touru, T., G. Sahana, B. Guldbrandtsen, M. Lund, and J. Vilkki. 2016. Genome-wide association analysis of milk yield traits in Nordic red cattle using imputed whole genome sequence variants. BMC Genet. 17:55.

Ju, Z. 2014. Single nucleotide polymorphisms haplotypes and combined genotypes of histatherin gene and their associations with mastitis in Chinese Holstein. Page 173 in The 34th International Society for Animal Genetics Conference, Xi′an, China.

Kitchen, B. J. 1981. Review of the progress of dairy science: Bovine mastitis: Milk compositional changes and related diagnostic test. J. Dairy Res. 48:167–188.

Lilbæk, H. M., M. Broe, E. Høier, T. Fatum, R. Ipsen, and N. Sørensen. 2006. Improving the yield of mozzarella cheese by phospholipase treatment of milk. J. Dairy Sci. 89:4114–4125.

Lopez, C., V. Briard-Bion, O. Ménard, E. Beaucher, F. Rousseau, J. Fauquant, N. Leconte, and B. Robert. 2011. Fat globules selected from whole milk according to their size: Different compositions and structure of the biomembrane, revealing sphingomyelin-rich domains. Food Chem. 125:355–368.

Malacarne, M., P. Franceschi, P. Formaggioni, S. Sandri, P. Mariani, and A. Summer. 2014. Influence of micellar calcium and phosphorus on rennet coagulation properties of cows milk. J. Dairy Res. 81:129–136.

Maningat, P. D., P. Sen, M. Rijnkels, A. L. Sunehag, D. L. Hadsell, M. Bray, and M. W. Haymond. 2009. Gene expression in the human mammary epithelium during lactation: The milk fat globule transcriptome. Physiol. Genomics 37:12–22.

Molee, A., C. Poompramun, and P. Mernkrathoke. 2015. Effect of casein genes—beta-LGB, DGAT1, GH, and LHR—on milk production and milk composition traits in crossbred Holsteins. Genet. Mol. Res. 14:2561–2571.

Nguyen, D. A., and M. C. Neville. 1998. Tight junction regulation in the mammary gland. J. Mammary Gland Biol. Neoplasia 3:233–246.

Norberg, E. 2005. Electrical conductivity of milk as a phenotypic and genetic indicator of bovine mastitis: A review. Livest. Prod. Sci. 96:129–139.

Ogata, H., S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27:29–34.

Peñagaricano, F., K. Weigel, G. J. Rosa, and H. Khatib. 2013. Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. Front. Genet. 3:307.

Peng, G., L. Luo, H. Siu, Y. Zhu, P. Hu, S. Hong, J. Zhao, X. Zhou, J. D. Reveille, and L. Jin. 2010. Gene and pathway-based second-wave analysis of genome-wide association studies. Eur. J. Hum. Genet. 18:111–117.

Pickrell, J. K., J. C. Marioni, A. A. Pai, J. F. Degner, B. E. Engelhardt, E. Nkadori, J. Veyrieras, M. Stephens, Y. Gilad, and J. K. Pritchard. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464:768–772.

Poulsen, N., H. Jensen, and L. Larsen. 2016. Factors influencing degree of glycosylation and phosphorylation of caseins in individual cow milk samples. J. Dairy Sci. 99:3325–3333.

Ramanathan, P., I. C. Martin, M. Gardiner-Garden, P. C. Thomson, R. M. Taylor, C. J. Ormandy, C. Moran, and P. Williamson. 2008. Transcriptome analysis identifies pathways associated with enhanced maternal performance in QSi5 mice. BMC Genomics 9:197.

Raven, L., B. G. Cocks, M. E. Goddard, J. E. Pryce, and B. J. Hayes. 2014. Genetic variants in mammary development, prolactin signalling and involution pathways explain considerable variation in bovine milk production and milk composition. Genet. Sel. Evol. 46:29.

Rombaut, R., K. Dewettinck, and J. Van Camp. 2007. Phospho- and sphingolipid content of selected dairy products as determined by HPLC coupled to an evaporative light scattering detector (HPLC–ELSD). J. Food Compos. Anal. 20:308–312.

Stelwagen, K., and K. Singh. 2014. The role of tight junctions in mammary gland function. J. Mammary Gland Biol. Neoplasia 19:131–138.

Viguier, C., S. Arora, N. Gilmartin, K. Welbeck, and R. O'Kennedy. 2009. Mastitis detection: Current trends and future perspectives. Trends Biotechnol. 27:486–493.

Walstra, P., P. Walstra, J. T. Wouters, and T. J. Geurts. 2014. Dairy Science and Technology. CRC Press, Boca Raton, FL.

Wang, L., P. Jia, R. D. Wolfinger, X. Chen, and Z. Zhao. 2011. Gene set analysis of genome-wide association studies: Methodological issues and perspectives. Genomics 98:1–8.

Wei, J., P. Ramanathan, I. C. Martin, C. Moran, R. M. Taylor, and P. Williamson. 2013. Identification of gene sets and pathways associated with lactation performance in mice. Physiol. Genomics 45:171–181.

Wickenden, J. A., and C. J. Watson. 2010. Key signalling nodes in mammary gland development and cancer. Signalling downstream of PI3 kinase in mammary epithelium: A play in 3 akts. Breast Cancer Res. 12:202.

Wiltbank, M., H. Lopez, R. Sartori, S. Sangsritavong, and A. Gümen. 2006. Changes in reproductive physiology of lactating dairy cows due to elevated steroid metabolism. Theriogenology 65:17–29.

Yellen, G. 2002. The voltage-gated potassium channels and their relatives. Nature 419:35–42.

Young, M. D., M. J. Wakefield, G. K. Smyth, and A. Oshlack. 2010. Method gene ontology analysis for RNA-seq: Accounting for selection bias. Genome Biol. 11:R14.

Zhao, W., K. Shahzad, M. Jiang, D. E. Graugnard, S. L. Rodriguez-Zas, J. Luo, J. J. Loor, and W. L. Hurley. 2013. Bioinformatics and gene network analyses of the swine mammary gland transcriptome during late gestation. Bioinform. Biol. Insights 7:193–216.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, and T. S. Sonstegard. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10:R42.