RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk

A. Cánovas,* G. Rincón,* A. Islas-Trejo,* R. Jimenez-Flores,† A. Laubscher,† and J. F. Medrano*¹

*Department of Animal Science, University of California-Davis, One Shields Ave., Davis 95616

†Dairy Products Technology Center, Department of Agriculture, California Polytechnic State University, San Luis Obispo 93407

ABSTRACT

The technological properties of milk have significant importance for the dairy industry. Citrate, a normal constituent of milk, forms one of the main buffer systems that regulate the equilibrium between Ca^{2+} and H^+ ions. Higher-than-normal citrate content is associated with poor coagulation properties of milk. To identify the genes responsible for the variation of citrate content in milk in dairy cattle, the metabolic steps involved in citrate and fatty acid synthesis pathways in ruminant mammary tissue using RNA sequencing were studied. Genetic markers that could influence milk citrate content in Holstein cows were used in a marker-trait association study to establish the relationship between 74 single nucleotide polymorphisms (SNP) in 20 candidate genes and citrate content in 250 Holstein cows. This analysis revealed 6 SNP in key metabolic pathway genes [isocitrate dehydrogenase 1 (NADP+), soluble (*IDH1*); pyruvate dehydrogenase (lipoamide) β (*PDHB*); pyruvate kinase (*PKM2*); and solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1 (*SLC25A1*)] significantly associated with increased milk citrate content. The amount of the phenotypic variation explained by the 6 SNP ranged from 10.1 to 13.7%. Also, genotype-combination analysis revealed the highest phenotypic variation was explained combining *IDH1*_23211, *PDHB*_5562, and *SLC25A1*_4446 genotypes. This specific genotype combination explained 21.3% of the phenotypic variation. The largest citrate associated effect was in the 3' untranslated region of the *SLC25A1* gene, which is responsible for the transport of citrate across the mitochondrial inner membrane. This study provides an approach using RNA sequencing, metabolic pathway analysis, and association studies to identify genetic variation in functional target genes determining complex trait phenotypes.

Key words: citrate, gene expression, RNA sequencing, single nucleotide polymorphism variation

INTRODUCTION

Technological properties of milk are of significant importance for the dairy industry. Citrate content, choline, and carnitine are associated with milk coagulation properties (Sundekilde et al., 2011). Citrate is a normal constituent of milk and forms one of the main buffer systems that regulates the equilibrium between Ca^{2+} and H^+ ions. It is an intermediate in the tricarboxylic acid cycle and, in ruminants, plays a role in de novo FA synthesis by providing reducing equivalents in the form of NADPH (Faulkner and Peaker, 1982). High levels of citrate content are associated with poor coagulation properties of milk under natural conditions (Sundekilde et al., 2011). Citrate affects milk-processing characteristics because it interacts with other milk constituents to influence the coagulation of milk proteins, and its fermentation products yield distinct aromatic flavors characteristic of fermented milk products (Rosenthal, 1991). In cheese, natural levels of citrate influence the coagulation properties of curds and the texture in ripened cheese (Brickley et al., 2009). In the highly competitive market of natural ripened cheeses, the importance of citrate in milk has long been recognized as a fundamental aspect of flavor development during aging by nonstarter bacteria (Faulkner and Peaker, 1982; Dudley and Steele, 2005). Central to the study of citrate in milk is the area of fermented milks and cheese ripening. Flavor development in many fermented milks and especially in ripening Cheddar cheese has been demonstrated as a complex microbial and biochemical process that is difficult to study but extremely important technologically (Faulkner and Peaker, 1982). It is agreed that a large part of the appealing flavor development depends on the use of citrate by the lactic acid bacteria involved in the food transformation, fermentation, and flavor development. Here, the natural content of citrate in milk for cheese is of great importance in the production of desired flavors (Medina de Figueroa et al., 2000; Dudley and Steele, 2005; Tan et al., 2012)

Received September 28, 2012.

Accepted December 11, 2012.

¹Corresponding author: jfmedrano@ucdavis.edu

Previous studies indicated that citrate content in milk varies according to season and stage of lactation, being higher during the grazing season for cows in seasonal production (Keogh and Pettingill, 1982) and during early lactation (Braunschweig and Puhan, 1999). However, Garnsworthy et al. (2006) described that stage of lactation and season have been confounded not only with each other, but also with changes in diet. In this context, those authors indicated that variation in milk citrate with stage of lactation is related to de novo synthesis of FA and that that relationship is independent of diet and milk yield (Garnsworthy et al., 2006).

The objective of the present study was to use RNA sequencing to study gene expression and SNP variation in metabolic pathways associated with citrate content in cow milk. The milk somatic cell transcriptome was analyzed using RNA sequencing technology to catalog the genes that are expressed during lactation in Holstein cows. Combining this information with the knowledge from citrate and FA synthesis pathways in ruminant mammary tissue, a list of candidate genes that are expressed in the mammary gland that could be responsible for the variation of citrate content in milk was developed. This was followed by a marker-trait association study to establish the relationship between 74 SNP in 20 candidate genes and citrate content in Holstein cows.

MATERIALS AND METHODS

Animals and Phenotypic Data

Milk and blood samples were collected from 250 Holstein cows from 4 freestall dairy farms in the California Central Valley. As described in Rincon et al. (2012), cows were selected in their first or second parity and 100 to 150 DIM to avoid the period of negative energy balance in early lactation and to maximize the period of de novo milk fat synthesis in the mammary gland. Therefore, only mature milk was analyzed in this study. Winter single morning milk samples were collected in triplicate during the routine US Department of Agriculture (USDA)/DHIA sampling procedure (National DHI Program, 2002) and kept on ice until the composition analysis was performed. Cows were carefully selected to maximize variation and to minimize inbreeding in samples with similar stage of lactation and DIM.

Milk composition analysis was performed at Silliker Laboratories Inc. (Modesto, CA). All milk samples were analyzed in duplicate using a MilkoScan FT-2 (Foss Analytical A/S, Hillerød, Denmark), under the raw milk calibration conditions. Readings were taken in the mid-infrared region, as described in the manual (Foss Analytical A/S, 2010). Further analysis of the

data was made using the spectra exported and stored in electronic files (Foss Analytical A/S, 2010). The equipment was calibrated monthly as recommended by the manufacturer. This, as well as a control sample, ensured the standardization of spectral acquisition throughout the experimental period (Hansen, 1998). Routine analyses by the MilkoScan FT-2 encompass TS, SNF, total protein, total fat, and lactose. We have a special calibration (internal program algorithm) for the MilkoScan FT-2 that allows the analysis of citrate in milk. Results of citrate are given in weight percent (% wt/wt). We have validated this method by comparison to published procedures using capillary electrophoresis (Izco et al., 2003). The capillary electrophoresis method was used with commercial milk samples (whole and skim milk) that spanned concentrations of citrate between 8.3 to 9.5 mM citrate. The results using the MilkoScan FT-2 are within 0.005% deviation of these values, in the range of commercial milk samples. This method is very accurate in the range between 0.15 to 0.25% citrate.

Milk yield information was collected from USDA/DHIA genetic evaluation data. The cows used in this trial were in the high-yielding or mid-lactation groups, or both. In all cases, feed was supplied as TMR as described in Rincon et al. (2012).

RNA Sequencing and Data Analysis

The RNA sequencing technology was used to obtain gene expression and SNP variation in coding regions of the bovine milk transcriptome. Seven milk samples were obtained from Holstein cows from the University of California-Davis herd. As described in Cánovas et al. (2010), after collection, milk was centrifuged at $2,000 \times g$ for 10 min to obtain a pellet of cells. The RNA was purified following TRIzol protocol (Invitrogen Corp., Carlsbad, CA) and mRNA was isolated and purified using a TruSeq RNA Sample Preparation Kit (Illumina Inc., San Diego, CA). The RNA integrity number (RIN) values ranged from 8.0 to 9.0 in milk somatic cell samples, indicating good quality of the RNA. Samples were sequenced on single lanes on an Illumina GA II sequence analyzer (Illumina Inc.).

Short sequence reads (36–40 bp) were assembled to the annotated bovine reference genome Btau_4.0 (<http://www.ncbi.nlm.nih.gov/genome/guide/cow/index.html>) using CLC bio workbench software (CLC bio, Aarhus, Denmark). Expression analysis was performed using the RNA sequencing application of CLC Genomics workbench and data was normalized by calculating the reads per kilobase per million mapped reads (RPKM) for each gene (Mortazavi et al., 2008).

To select expressed genes, a threshold of RPKM ≥ 0.2 was used (Wickramasinghe et al., 2012).

SNP Discovery and Genotyping

A SNP detection analysis was performed using sequencing reads from 7 Holstein cows to determine putative polymorphisms in genes involved in the citrate metabolism. The SNP detection was performed as described in Cánovas et al. (2010), considering the same criteria and quality filters to select only true SNP.

Using the RNA sequencing expression values and SNP discovery in genes involved in citrate and FA synthesis pathways in ruminant mammary tissue, 54 SNP located in 20 different genes were identified and selected for genotyping (Table 1). Additionally, to study the regulatory regions of 6 target genes in these pathways, conserved regulatory noncoding regions [2,000 bp upstream of the 5' untranslated region (UTR) and 2,000 bp downstream of the 3' UTR] were resequenced by Sanger methodology (Cánovas et al., 2010) in a panel of 8 unrelated Holstein animals. Target genes potentially associated with citrate content variation were selected based on gene expression, Gene Ontology (GO) and pathway analysis, prioritizing important metabolic steps of citrate metabolism. Twenty SNP were identified in ATP citrate lyase (*ACLY*); aconitase 1, soluble (*ACO1*); aconitase 2, mitochondrial (*ACO2*); citrate synthase (*CS*); isocitrate dehydrogenase 1 (NADP+), soluble (*IDH1*); and solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1 (*SLC25A1*) genes (Table 1) and added to a genotyping panel. A total of 74 SNP in 20 genes were selected to genotype 250 Holstein cows (54 SNP located in coding regions of 20 genes identified by RNA sequencing and 20 SNP located in conserved noncoding regulatory regions identified by Sanger resequencing). The DNA samples were genotyped using a MassARRAY matrix-assisted laser desorption-ionization time-of-flight mass spectrometry (MALDI-TOF MS) detection system at GeneSeek (Gene Seek Inc., Lincoln, NE).

Marker-Trait Association Analysis

Marker-trait association analysis was performed using a linear regression test under a dominant model assumption. Statistical analyses were performed using the genotype association and regression modules from SNP Variation Suite (SVS) version 7 (Golden Helix Inc., Bozeman MT) as described in Rincon et al. (2009a,b). The model used for regression analysis was as follows:

$$y_{ijkl} = \mu + f_i + t_{jk}g_k + e_{ijkl},$$

where y_{ijkl} represents phenotypic observation for each individual, μ is the overall mean, f_i is the systematic effect farm of origin (with 4 levels); t_{jk} is the regression coefficient (mean allele substitution effect) of phenotypes onto genotype for the j th SNP, g_k is the effect of individual genotype (3 levels: the 2 homozygous and the heterozygous), and e_{ijkl} is the residual effect.

Phenotypes were adjusted for systematic environmental effects (including covariates) using a full versus reduced model regression equation. The regression sum of squares was calculated both for the reduced and for the full model. An F-test was then performed to find the significance of the full versus the reduced model. A P -value threshold of 0.05 was used to establish significant associations. The linear regression was also performed including SNP interactions using the SVS version 7 regression module from Golden Helix Inc. as described in Rincon et al. (2012). Also, genotypes combinations were tested to obtain the best combination of markers significantly associated with citrate content using the haplotype module. False discovery rate was controlled according to the method of Storey (2002) and cut-off for significant association values was set at false discovery rate q -value < 0.1 .

Functional Analysis

To validate the functional relationship between the genes expressed in the mammary gland involved in citrate and FA synthesis pathways and citrate content, the 3 GO categories (biological process, molecular function, and cellular component) were analyzed using Blast2GO software (Götz et al., 2008). *Bos Taurus* coding sequences were downloaded from the ENSEMBL biomart martview application (<http://www.ensembl.org/biomart/martview>). These sequences were annotated with blastx and the Blast2GO mapping and GO annotation routines. The GO significance levels were computed following Fisher's exact test for multiple testing in Blast2GO. The Blast2GO software was also used to examine associated biological pathways involved in citrate metabolism in the list of expressed genes.

RESULTS AND DISCUSSION

The RNA sequencing approach was used to study gene expression and SNP variation associated with citrate content in cow milk. The workflow of the experimental approach used in this study (Figure 1) included (1) pathway development from the literature (Figure 2) and GO analysis of all genes involved in glycolysis and tricarboxylic acid pathways, (2) identification of expressed isoforms in each metabolic step, (3) identification and selection of SNP in the most highly

Table 1. The RNA sequencing (RNA-Seq) reads per kilobase per million mapped reads (RPKM) expression values and total number of SNP to genotype by gene

Gene	Gene name	Process in citrate and FA synthesis pathway ¹	RPKM gene expression value	SNP from RNA-Seq ²	SNP from resequencing ³
<i>ACLY</i>	ATP citrate lyase	Citrate-acetyl-CoA (cytosol)	24.53	6	4
<i>ACO1</i>	Aconitase 1, soluble	Citrate-isocitrate (cytosol)	94.21	4	3
<i>ACO2</i>	Aconitase 2, mitochondrial	Citrate-isocitrate	81.44	1	5
<i>ACSL1</i>	Acyl-CoA synthetase long-chain family	Acetate-acetyl-CoA	51.79	2	
<i>ACSL5</i>	Acyl-CoA synthetase long-chain family	Acetate-acetyl-CoA	43.74	4	
<i>ACSS1</i>	Acyl-CoA synthetase short-chain family	Acetate-acetyl-CoA	47.67	3	
<i>CS</i>	Citrate synthase	Acetyl-CoA-citrate	80.25	2	1
<i>DLAT</i>	Dihydrolipoamide S-acetyltransferase	Pyruvate-acetyl-CoA	11.78	1	
<i>DLD</i>	Dihydrolipoamide dehydrogenase	Pyruvate-acetyl-CoA	9.68	1	
<i>DLST</i>	Dihydrolipoamide S-succinyltransferase	Ketoglutarate-succinyl-CoA	61.51	2	
<i>IDH1</i>	Isocitrate dehydrogenase 1 (NADP+), soluble	Isocitrate-ketoglutarate (cytosol)	79.62	11	5
<i>IDH2</i>	Isocitrate dehydrogenase 2 (NADP+), mitochondrial	Isocitrate-ketoglutarate	47.30	1	
<i>IDH3A</i>	Isocitrate dehydrogenase 3 (NAD+) α	Isocitrate-ketoglutarate	11.58	1	
<i>IDH3B</i>	Isocitrate dehydrogenase 3 (NAD+) β	Isocitrate-ketoglutarate	51.04	1	
<i>IDH3G</i>	Isocitrate dehydrogenase 3 (NAD+) γ	Isocitrate-ketoglutarate	63.44	5	
<i>OGDH</i>	Oxoglutarate dehydrogenase (lipoamide)	Ketoglutarate-succinyl-CoA	78.07	4	
<i>PDHB</i>	Pyruvate dehydrogenase (lipoamide) β	Pyruvate-acetyl-CoA	34.42	1	
<i>PDHX</i>	Pyruvate dehydrogenase complex, component X	Pyruvate-acetyl-CoA	6.89	2	
<i>PKM2</i>	Pyruvate kinase	Pyruvate (cytosol)	715.44	1	
<i>SLC25A1</i>	Solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1	Transport mitochondria-cytosol	77.92	1	2

¹Information related to each metabolic step in citrate and FA synthesis pathway is detailed in Figure 2.

²A total of 54 SNP were genotyped in the coding regions of 20 genes identified by RNA-Seq.

³A total of 20 SNP were genotyped in promoters and conserved noncoding regions of selected target genes (*ACLY*, *ACO1*, *ACO2*, *CS*, *IDH1*, and *SLC25A1*) identified by Sanger resequencing (Cánovas et al., 2010).

Figure 1. Workflow to identify genes associated with a complex phenotype in milk using RNA sequencing, SNP discovery, and marker-trait association analysis. Single nucleotide polymorphism selection criteria followed the approach in Cánovas et al. (2010). TCA pathway = tricarboxylic acid cycle.

expressed isoforms to develop a SNP genotyping panel, (4) phenotyping and genotyping cows to perform a marker-trait association study, and (5) identification of target genes and potential regulators associated with the phenotype.

Functional Validation Between Candidate Genes Expressed in the Milk Somatic Cell Transcriptome and Citrate Content

To identify genes responsible for the variation of citrate content in milk in dairy cattle, all the metabolic steps involved in citrate and FA synthesis pathways in ruminant mammary tissue were examined. The milk somatic cell transcriptome has been analyzed to catalog the genes that are expressed during lactation in Holstein cows using RNA sequencing (Wickramasinghe et

al., 2012). Gene Ontology analysis was used to validate the functional relationship between the list of candidate genes expressed in milk somatic cells and citrate content (Table 1).

As expected, GO terms related to citrate and FA synthesis were particularly abundant in the chosen pathways. In this context, the biological process GO category included “NAD and NADP metabolic process” and “reductive tricarboxylic acid cycle” among GO terms significantly enriched by our list of genes. Additionally, a variety of biological processes related to FA synthesis, such as regulation of the triglyceride metabolic process, regulation of FA β -oxidation, and regulation of acetyl-CoA biosynthetic process from pyruvate was present in the list of most-significant GO terms. Concerning the molecular function GO category, “ATP binding,” “isocitrate dehydrogenase NAD⁺ and

Figure 2. Pathway of FA synthesis in ruminant mammary tissue (adapted from Bauman et al., 1973). Different gradients of color correspond to average expression values (reads per kilobase per million mapped reads, RPKM) measured by RNA sequencing in milk samples. Identified causal genes functionally related to citrate metabolism and which promoters and conserved noncoding regions were resequenced are shown in the large circle. Genes with significant associations to citrate content are indicated in bold-line circles. LDH = lactate dehydrogenase; PCK1 = phosphoenolpyruvate carboxykinase 1; MDH1 = malate dehydrogenase 1; CLYBL = citrate lyase beta like; PYC = pyruvate carboxylase; SDH = succinate dehydrogenase; SUCL = succinyl-CoA ligase; ACS = acyl-CoA synthetase. Gene names can be found in Table 1.

NADP⁺ activity,” and “AMP binding” were among the most significant GO terms. Other molecular functions, such as “S-acyltransferase activity,” “oxoglutarate dehydrogenase (NAD⁺) activity,” “succinyl transferase activity,” and “succinate-CoA ligase activity,” also appeared, although with less frequency. The cellular component GO classification indicated that most genes encoded cytosol, mitochondrial parts, mitochondria inner membrane, intracellular membrane-bounded organelle, and mitochondrial isocitrate dehydrogenase complex.

Pathway analyses of our list of genes (Table 1) was consistent with the GO analysis; pathways involved in the tricarboxylic acid cycle or glycolysis, or both, were among the most significant. These included the

carbon fixation and the pyruvate metabolism pathways on the top of the list. The tricarboxylic acid pathway contains 11 genes involved in the most important metabolic steps in the mitochondria [e.g., *ACO2*; *CS*; dihydrolipoamide S-acetyltransferase (*DLAT*); dihydrolipoamide dehydrogenase (*DLD*); dihydrolipoamide S-succinyltransferase (*DLST*); isocitrate dehydrogenase 2 (NADP⁺), mitochondrial (*IDH2*); oxoglutarate dehydrogenase (lipoamide) (*OGDH*); and pyruvate dehydrogenase (lipoamide) β (*PDHB*)] and in the cytosol (e.g., *ACLY*, *ACO1*, and *IDH1*). The glycolysis/gluconeogenesis pathway showed 5 genes. Among them, one gene [pyruvate kinase (*PKM2*)] was involved in one of the first steps of glycolysis, catalyzing the transfer of

phosphoenolpyruvate to pyruvate. Also, 3 genes related to conversion of pyruvate to acetyl-CoA, providing the primary link between glycolysis and the tricarboxylic acid cycle (e.g., *DLAT*, *DLD*, and *PDHB*), and 1 gene related to production of acetyl-CoA from acetate in a reaction that requires ATP [acyl-CoA synthetase short-chain family (*ACSS1*)] (Figure 2). Besides, carbon fixation and the pyruvate metabolism pathways appeared in the middle of the list, gathering several genes with roles in lipid synthesis, energy generation production, and regeneration of NADPH (e.g., *ACLY*, *ACO1*, *ACSS1*, and *IDH1*). Finally, the glyoxylate and dicarboxylate metabolism (e.g., *ACO1* and *CS*) and FA metabolism [e.g., acyl-CoA synthetase long-chain family (*ACSL1*)] pathways were also present in this list.

Identification of all Expressed Isoforms of Each Citrate Metabolic Step Using RNA Sequencing

Within the metabolic pathways affecting citrate content of milk, the most highly expressed isoforms in each metabolic step in glycolysis and the tricarboxylic acid cycle were chosen by manual curation to discover SNP and develop a SNP genotyping panel (Table 1 and Figure 2).

All expressed isoforms of each citrate metabolic step detected and their genomic location are shown in Table 2. Among the isoforms, isocitrate dehydrogenase presents an interesting pattern. Isocitrate dehydrogenases catalyze the oxidative decarboxylation of isocitrate to 2-oxoglutarate. These enzymes belong to 2 distinct subclasses, one of which utilizes NAD(+) as the electron acceptor and the other NADP(+). Five isocitrate dehydrogenases have been reported: 3 NAD(+)-dependent isocitrate dehydrogenases, which are localized in the mitochondrial matrix (*IDH3A*, *IDH3B*, and *IDH3G*), and 2 NADP(+)-dependent isocitrate dehydrogenases, one of which is mitochondrial (*IDH2*) and the other predominantly cytosolic (*IDH1*; Figure 2). The NAD(+)-dependent isocitrate dehydrogenases catalyze the allosterically regulated rate-limiting step of the tricarboxylic acid cycle. Among the isoforms located in mitochondrial matrix, *IDH3B* and *IDH3G* showed the highest RPKM values (51.04 and 63.44 RPKM, respectively), followed by *IDH2* with a 47.3 RPKM value. Similar expression values obtained in isoforms located in the mitochondrial matrix suggest that they are playing different important roles in this metabolic step due to the different use of NAD(+) or NADP(+) as the electron acceptor. For example, *IDH3G* is associated with periventricular heterotopias, whereas *IDH2* plays a role in intermediary metabolism and energy production. In addition, *IDH1* is the NADP(+)-dependent isocitrate dehydrogenase found in the cytoplasm. The

importance of cytoplasmic isocitrate dehydrogenase plays a significant role in cytoplasmic NADPH production to produce citrate. The key role of *IDH1* in citrate production is reflected in the gene expression results showing the highest RPKM value (79.62) of all mitochondria and cytosolic isocitrate dehydrogenases in Holstein cows. Therefore, for isocitrate dehydrogenase, SNP were chosen in *IDH1*, in the most highly expressed cytosolic isoform.

Also, different isoforms related to the FA synthesis pathways were found catalyzing the activation of acetate for lipid synthesis and energy generation. They act as a monomer and produce acetyl-CoA from acetate in a reaction that requires ATP. Among them, 3 different groups of isoforms can be identified: acyl-CoA synthetase long-chain family (*ACSL*), acyl-CoA synthetase medium-chain family (*ACSM*), and acyl-CoA synthetase short-chain family (*ACSS*). The RNA sequencing results suggested that *ACSL* and *ACSS* families may play an important role in this FA synthesis pathway showing high expression values (around a 50 RPKM value), whereas *ACSM* showed RPKM values lower than 1 (Figure 2). Although differing in substrate specificity, subcellular localization, and tissue distribution, all isoforms of *ACSL* convert free long-chain FA into fatty acyl-CoA esters, and thereby play a key role in lipid biosynthesis and FA degradation. In the case of the *ACSL4* isoform, the RPKM value was the lowest (17.03) compared with *ACSL1* (51.79) and *ACSL5* (43.74), maybe due to the fact that this isoform preferentially utilizes arachidonate as substrate. Related to *ACSS* isoforms, *ACSS2* showed the highest RPKM value (52.04). Expression of this gene is regulated by sterol regulatory element-binding proteins, transcription factors that activate genes required for the synthesis of cholesterol and unsaturated FA.

Citrate Target Genes Selection

Based on GO and pathway analysis, target genes potentially associated with citrate content variation were individually investigated in the GO term and pathway analysis. The most important metabolic steps were prioritized, such as the conversion of oxaloacetate and acetyl-CoA to citrate in the mitochondria. Citrate synthase acts in this step and is found in almost all cells capable of oxidative metabolism. According to the key role of this gene in the trait under study, *CS* was also highly expressed in the milk transcriptome (80.25 RPKM; Table 1). Aconitase 2, mitochondrial catalyzes the interconversion of citrate to isocitrate via *cis*-aconitate in the second step of the tricarboxylic acid cycle. Also, *ACO2* showed a high RPKM value (81.44) during lactation of analyzed Holstein cows.

Table 2. Isoforms of each citrate metabolic step¹

Isoform ²	Chromosome	Gene start (bp number)	Gene end (bp number)	RPKM ³	Total gene reads	Process/principal function in citrate and FA synthesis pathway
LDHA	29	26543861	26554318	410.5	62,689	Conversion of L-lactate and NAD to pyruvate and NADH
LDHB	5	88962679	88981219	260.86	27,799	
LDHC	29	26506380	26539353	0.19	52	
PDHA1	X	130874454	130897343	13.4	4,740	
PDHA2	6	30063480	30064806	0	0	Conversion of pyruvate to acetyl-CoA and provide the primary link between glycolysis and tricarboxylic acid cycle
PDHB	22	43461214	43466626	34.4	4,262	
PDHX	15	66231837	66307403	6.9	1,553	
SDHA	20	71940743	71964107	0	0	
SDHB	2	136125304	136154362	158.85	14,171	Oxidation of succinate and carries electrons from FADH to CoQ
SDHC	3	8195438	8226068	49.13	5,533	
SDHD	15	22722376	22733137	0	0	
SUCLA2	12	17884901	17930933	13.12	2,544	
SUCLG1	11	50400445	50453412	78.2	8,420	Catalyze the reversible reaction involving the formation of succinyl-CoA and succinate
SUCLG2	22	33977312	34251728	35.73	8,503	
IDH1	2	96941049	96962496	79.62	15,795	Catalyze the oxidative decarboxylation of isocitrate to 2-oxoglutarate
IDH2	21	21903847	21920153	47.3	7,169	
IDH3A	21	31063699	31079936	11.58	1,783	
IDH3B	13	52978729	52983685	51.04	6,625	
IDH3G	X	39916807	39925370	63.44	7,635	
ACSL1	27	14223449	14288333	51.79	17,205	
ACSL4	X	62476952	62508324	17.03	4,294	
ACSL5	26	33184653	33234956	43.74	9,733	Activation of acetate for lipid synthesis and energy generation
ACSM1	25	18349702	18413889	0.87	309	
ACSM2	25	18605635	18656582	0.04	0	
ACSM3	25	18605635	18656582	0.85	2,107	
ACSM5	25	18207129	18234451	0.38	74	
ACSS1	13	43029813	43073610	47.67	12,062	
ACSS2	13	64794728	64841518	52.04	13,318	
ACSS3	5	10731241	10905490	9.09	1,514	
ACSF3	18	14254106	14292739	3.85	1,059	

¹Expression levels were detected using RNA sequencing in bovine milk somatic cells.

²LDH = lactate dehydrogenase; PDH = pyruvate dehydrogenase; SDH = succinate dehydrogenase; SUCL = succinyl-CoA ligase; IDH = isocitrate dehydrogenase; ACS = acyl-CoA synthetase.

³Reads per kilobase per million mapped reads.

Another important citrate target gene was *SLC25A1*. This gene is responsible for the movement of citrate across the mitochondrial inner membrane (Kaplan et al., 1993). Figure 2 shows the key role of this target gene to citrate production in cytosol, having a 77.92 RPKM value in the milk transcriptome (Table 1). Following the movement of citrate across the mitochondria to cytosol by the *SLC25A1* gene-encoded transporter, we confirmed the metabolic step that involves the conversion of citrate to isocitrate by ACO1 (Figure 2). Compared with the selected target genes, *ACO1* showed the highest expression value (94.21 RPKM) in Holstein cows (Table 1). Aconitase 1, soluble also plays a central role in cellular iron homeostasis and is associated with repression of translation of ferritin mRNA, and inhibition of degradation of the otherwise rapidly degrading transferrin receptor mRNA.

As mentioned above, IDH1 is the most important isocitrate dehydrogenase in the citrate and FA synthesis pathways. It is located in the cytosol (in the isocitrate to ketoglutarate process), playing an important role in the production and regeneration of NADPH. The key role of IDH1 in citrate production is reflected in the RNA sequencing results, showing the highest RPKM value (79.62) of all mitochondria and cytosolic isocitrate dehydrogenases in Holstein cows (Table 1).

Adenosine triphosphate citrate lyase catalyzes the formation of acetyl-CoA and oxaloacetate from citrate and CoA, with a concomitant hydrolysis of ATP to ADP and phosphate for use in lipid synthesis and energy generation. The product, acetyl-CoA, serves several important biosynthetic pathways, including lipogenesis and cholesterol synthesis. Previous studies have indicated that in ruminants, mammary tissue acetate is one of the major substrates utilized by the ruminant mammary gland for FA synthesis. According to Bauman et al. (1973), the metabolic process from citrate to acetyl-CoA, catabolized by ACLY, is not present in the ruminant mammary gland. Our study indicates that *ACLY* is expressed in ruminant milk somatic cells showing a 24.53 RPKM value in the milk transcriptome. However, *ACSS*, *ACSM*, and *ACSL* family genes were more highly expressed in milk cells than *ACLY*, suggesting that acetate is the principal substrate utilized by the ruminant mammary gland for FA synthesis.

Association Between Target Gene Genotypes and Citrate Content

Previous studies indicated that citrate content in milk varies according to stage of lactation, being higher during early lactation compared with the rest of lactation (Braunschweig and Puhani, 1999). Taking this into account in our study, Holstein cows were care-

fully selected to maximize variation and to minimize inbreeding in samples with similar stages of lactation and DIM (Rincon et al., 2012). As mentioned above, a total of 74 SNP (Table 1) were genotyped in 20 genes using 250 Holstein cows.

The detailed description of SNP used in the marker-trait association study and marker statistics is shown in Supplemental Files 1 and 2 (available online at <http://www.journalofdairyscience.org/>). Marker-trait association results revealed 6 significant SNP associated with citrate content in Holstein cows (Table 3). Single nucleotide polymorphisms in the *IDH1*, *PDHB*, *PKM2*, and *SLC25A1* genes were associated with increments of citrate content of milk in Holstein cows. The percentage of phenotypic variation explained for the 6 significant markers ranged between 10.1 and 13.7% (Table 3). Individually, marker *SLC25A1_4446* showed the most significant association with citrate content in Holstein milk ($P = 0.001$), revealing 13.7% of the phenotypic variation. Markers *PDHB_5562* and *PKM2_24735* explained 11.1 and 11.4% of the phenotypic variation, respectively. The 3 SNP with significant associations located in the *IDH1* gene explained around 10.5% of the phenotypic variation ($P < 0.05$; Table 3). Also, genotype-combination analysis was performed to examine the combined effect of all possible genotypes of the 6 SNP in the 4 genes significantly associated with citrate content. As shown in Table 4, the highest phenotypic variation was explained combining *IDH1_23211*, *PDHB_5562*, and *SLC25A1_4446* genotypes. This specific genotype combination explained 21.3% of the phenotypic variation ($P = 0.00007$).

In milk processing, it has been shown that cows with higher levels of citrate content in milk have better flavor profiles in fermentation and cheese aging and heat stability; however, high levels of citrate are also associated with poor coagulation properties (Sundekilde et al., 2011). This explains the common practice in commercial cheese making of adding calcium chloride in cheese making before the enzymatic coagulation step so coagulation takes place normally, but aging cheeses with high citrate levels develop better flavors (Dudley and Steele, 2005; Brickley et al., 2009). Furthermore, an increasing concentration of lactose in milk has been shown to exert a beneficial effect on the formation of acid-induced milk gels (Hallén et al., 2010). In this context, lactose is the most important osmoregulator of milk synthesis and is, to some extent, correlated with the total protein and casein content of milk, whereby lower levels of citrate content in milk induced the caseins to become overloaded with calcium, and then precipitate (Hallén et al., 2007).

The average of citrate content in the cows under study was $0.156 \pm 0.0017\%$ (wt/wt; or $1,560 \pm 1.7$

Table 3. Allele substitution effects of markers associated with citrate content in Holstein cows

Marker ¹	Gene information		FvsR model ²		Allele substitution effect ³	
	Allele	MAF ⁴	P-value	R ²	Estimate	P-value
<i>IDH1</i> _23073	A	0.23	0.036	0.10	−0.007	0.031
<i>IDH1</i> _23211	G	0.33	0.040	0.10	−0.014	0.042
<i>IDH1</i> _23423	G	0.23	0.021	0.11	−0.007	0.012
<i>PDHB</i> _5562	G	0.19	0.008	0.11	0.011	0.039
<i>PKM2</i> _24735	G	0.14	0.040	0.11	0.008	0.046
<i>SLC25A1</i> _4446	T	0.11	0.001	0.14	0.021	0.001

¹*IDH1* = isocitrate dehydrogenase 1 (NADP+), soluble; *PDHB* = pyruvate dehydrogenase (lipoamide) β ; *PKM2* = pyruvate kinase; *SLC25A1* = solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1.

²FvsR model = full versus reduced model.

³Allele substitution effect values and P-values after correcting for false discovery rate.

⁴The minor allele frequency (MAF) referred to the allele of the previous column.

mg/L) suggesting that SNP located in *IDH1*, *PDHB*, and *SLC25A1* target genes could be causing an increment up to 0.19% (wt/wt) of the citrate content in Holstein cows. The major citrate associated marker was located in the 3' UTR of the *SLC25A1* gene (Figure 3). The significant association between *SLC25A1* and citrate content is of particular interest, as these genes encodes the transporter for the movement of citrate across the mitochondrial inner membrane, one of the key metabolic step in the final production of citrate (Kaplan et al., 1993). Thus, the *SLC25A1*_4446 marker suggests that the regulatory 3' UTR of the *SLC25A1* gene, most likely, plays an important role in the availability of this enzyme. Another interesting target gene in association with citrate content was *IDH1*, which plays an important role in the production and regeneration of NADPH. The key role of *IDH1* in citrate production is reflected in the fact that 3 SNP were significantly associated with citrate content. Also, the 3 significant SNP found in the *IDH1* gene were located in the 3' UTR of this target gene, identifying an interesting haplotype (Figure 3). Linkage disequilibrium between *IDH1*_23073, *IDH1*_23211, and *IDH1*_23423 SNP was examined. The correlation values among these SNP ranged from 0.98 to 0.99, suggesting strong linkage disequilibrium between markers. Related to the primary link between glycolysis and the tricarboxylic acid cycle, significant markers were found in *PDHB*

and *PKM2*. A significant SNP was found in exon 8 of *PDHB* mRNA. This gene plays an important role in the PDH complex, catalyzing the overall conversion of pyruvate to acetyl-CoA. Pyruvate kinase is involved in glycolysis, catalyzing the transfer of a phosphoryl group from phosphoenolpyruvate to ADP, generating ATP and pyruvate. A significant *PKM2*_24735 marker was located in exon 9 of *PKM2* mRNA (Figure 3). Also, the *PKM2* gene had the highest expression among all the genes in the analyzed pathways (715.44 RPKM; Table 1), supporting the importance of *PKM2* in the substrate availability between glycolysis and the tricarboxylic acid cycle and, in turn, in the citrate content of milk.

Marker-trait association results validate the power of RNA sequencing methodology to study complex traits in milk in dairy cattle. The combined analysis of gene expression in metabolic pathways associated with FA synthesis and citrate metabolism, SNP discovery, and a marker-trait association study identified 6 SNP associated with milk citrate content in *IDH1*, *PDHB*, *PKM2*, and *SLC25A1* genes. This study opens the door to the application of molecular markers to study and improve important milk coagulation properties that can be limiting factors in the milk manufacturing process. This study demonstrates that judicious selection of cows can result in an improved quality of milk for processing some dairy foods. Given the increase interest in natural

Table 4. Phenotypic variation of citrate content explained by *IDH1*_23211, *PDHB*_5562, and *SLC25A1*_4446 genotypes in Holstein cows¹

Marker	Haplotype	Frequency	Average	SE	R ²	P-value
<i>IDH1</i> _23211	A,A,C / A,G,T	0.671	0.158	0.002	0.21	7.30E-05
<i>PDHB</i> _5562	A,A,T / A,G,C	0.258	0.168	0.004		
<i>SLC25A1</i> _4446	G,A,C / G,A,T	0.100	0.154	0.005		

¹*IDH1* = isocitrate dehydrogenase 1 (NADP+), soluble; *PDHB* = pyruvate dehydrogenase (lipoamide) β ; *SLC25A1* = solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1.

Figure 3. Localization of genotyped SNP in the promoter, coding region and conserved noncoding regions of isocitrate dehydrogenase 1 (NADP+), soluble (*IDH1*); pyruvate dehydrogenase (lipoamide) β (*PDHB*); pyruvate kinase (*PKM2*); and solute carrier family 25 (mitochondrial carrier; citrate transporter), member 1 (*SLC25A1*) genes. Exon/intron distribution is taken from the *Bos taurus* gene. In circles: SNP significantly associated with citrate content in Holstein cows.

foods with no additives, milk with higher content of citrate will result in more flavorful cheeses, aromatic fermented milks, and perhaps faster processes.

CONCLUSIONS

We have demonstrated that RNA sequencing methodology is a powerful approach to study complex traits in milk in dairy cattle. Using RNA sequencing, the milk somatic cell transcriptome was analyzed to catalog the genes that are expressed during lactation in Holstein cows and their SNP variations. Combining this infor-

mation with the knowledge from citrate and FA synthesis pathways in ruminant mammary tissue, significant markers were identified that could influence milk citrate content in Holstein cows. Marker-trait association results revealed 6 SNP associated with increased citrate content in milk in *IDH1*, *PDHB*, *PKM2*, and *SLC25A1* genes. Individually, the percentage of phenotypic variation explained by the 6 significant markers ranged between 10.1 and 13.7%. However, genotype-combination analysis revealed that the highest phenotypic variation was explained by the combined effect of *IDH1*_23211, *PDHB*_5562, and *SLC25A1*_4446 genotypes. This

specific genotype combination explained 21.3% of the phenotypic variation ($P = 0.00007$). The major citrate associated effect was in the 3' UTR of the *SLC25A1* gene, which is responsible for the transport of citrate across the mitochondrial inner membrane. These findings provide an approach to identify genetic variation that can be used to improve important milk phenotypes affecting processing characteristics of milk and other metabolic complex trait phenotypes.

ACKNOWLEDGMENTS

This project was partially funded by the California State University Agricultural Research Initiative and by the University of California (Davis)-W. K. Kellogg Endowment. A. Cánovas was supported by a postdoctoral grant from Ministry of Education/Fulbright.

REFERENCES

- Bauman, D. E., R. W. Mellenberger, and R. G. Derrig. 1973. Fatty acid synthesis in sheep mammary tissue. *J. Dairy Sci.* 56:1312–1318.
- Braunschweig, M., and Z. Puhan. 1999. Association of casein haplotypes with casein composition and size of micelles in milks from Simmental $\times$ Red Holstein cows. *J. Dairy Res.* 66:619–625.
- Brickley, C. A., J. A. Lucey, and P. L. H. McSweeney. 2009. Effect of the addition of trisodium citrate and calcium chloride during salting on the rheological and textural properties of Cheddar-style cheese during ripening. *Int. J. Dairy Technol.* 62:527–534.
- 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.
- Dudley, E. G., and J. L. Steele. 2005. Succinate production and citrate catabolism by Cheddar cheese nonstarter lactobacilli. *J. Appl. Microbiol.* 98:14–23.
- Faulkner, A., and M. Peaker. 1982. Reviews of the progress of dairy science: Secretion of citrate into milk. *J. Dairy Res.* 49:159–169.
- Foss Analytical A/S. 2010. Foss Analytical Note 37: How to get started with abnormal spectrum screening module application 1–3. Foss Analytical A/S, Hillerød, Denmark.
- Garnsworthy, P. C., L. L. Masson, A. L. Lock, and T. T. Mottam. 2006. Variation of milk citrate with stage of lactation and de novo fatty acid synthesis in dairy cows. *J. Dairy Sci.* 89:1604–1612.
- Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talón, J. Dopazo, and A. Conesa. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36:3420–3435.
- Hallén, E., T. Allmere, J. Näslund, A. Andrén, and A. Lundén. 2007. Effect of genetic polymorphism of milk proteins on rheology of chymosin-induced milk gels. *Int. Dairy J.* 17:791–799.
- Hallén, E., A. Lundén, T. Allmere, and A. Andrén. 2010. Casein retention in curd and loss of casein into whey at chymosin-induced coagulation of milk. *J. Dairy Res.* 77:71–76.
- Hansen, P. W. 1998. Urea determination in milk using Fourier transform Infrared spectroscopy and multivariate calibration. *Milch-wissenschaft* 53:251–255.
- Izco, J. M., M. Tormo, A. Harris, P. S. Tong, and R. Jimenez-Flores. 2003. Optimization and validation of a rapid method to determine citrate and inorganic phosphate in milk by capillary electrophoresis. *J. Dairy Sci.* 86:86–95.
- Kaplan, R. S., J. A. Mayor, and D. O. Wood. 1993. The mitochondrial tricarboxylate transport protein: cDNA cloning, primary structure, and comparison with other mitochondrial transport proteins. *J. Biol. Chem.* 268:13682–13690.
- Keogh, B. P., and G. Pettingill. 1982. Possible role of coryneform bacteria in age gelation of ultrahigh-temperature-processed milk. *Appl. Environ. Microbiol.* 43:1495–1497.
- Medina de Figueroa, R., F. Alvarez, A. Pesce de Ruiz Holgado, G. Oliver, and F. Sesma. 2000. Citrate utilization by homo- and heterofermentative lactobacilli. *Microbiol. Res.* 154:313–320.
- 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.
- National, D. H. I. Program. 2002. Uniform operating procedures. <http://www.quality-certification.com/uop.pdf>.
- Rincon, G., E. A. Farber, C. R. Farber, J. D. Nkrumah, and J. F. Medrano. 2009a. Polymorphisms in the *STAT6* gene and their association with carcass traits in feedlot cattle. *Anim. Genet.* 40:878–882.
- Rincón, G., A. Islas-Trejo, J. Casellas, Y. Ronin, M. Soller, E. Lipkin, and J. F. Medrano. 2009b. Fine mapping and association analysis of a quantitative trait locus for milk production traits on *Bos taurus* autosome 4. *J. Dairy Sci.* 92:758–764.
- Rincon, G., A. Islas-Trejo, A. R. Castillo, D. E. Bauman, B. J. German, and J. F. Medrano. 2012. Polymorphisms in genes in the *SREBP1* signalling pathway and SCD are associated with milk fatty acid composition in Holstein cattle. *J. Dairy Res.* 79:66–75.
- Rosenthal, I. 1991. *Milk and Dairy Products: Properties and Processing*. VCH Publishers, New York, NY.
- Storey, J. D. 2002. A direct approach to false discovery rates. *J. R. Stat. Soc., B* 64:479–498.
- Sundekilde, U. K., P. D. Frederiksen, M. R. Clausen, L. B. Larsen, and H. C. Bertram. 2011. Relationship between the metabolite profile and technological properties of bovine milk from two dairy breeds elucidated by NMR-based metabolomics. *J. Agric. Food Chem.* 59:7360–7367.
- Tan, W. S., M. F. Budinich, R. Ward, J. R. Broadbent, and J. L. Steele. 2012. Optimal growth of *Lactobacillus casei* in a Cheddar cheese ripening model system requires exogenous fatty acids. *J. Dairy Sci.* 95:1680–1689.
- Wickramasinghe, S., G. Rincon, A. Islas-Trejo, and J. F. Medrano. 2012. Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* 13:45.