Front. Agr. Sci. Eng. 2014, 1(4): 321–330
DOI: 10.15302/J-FASE-2014037

RESEARCH ARTICLE

A post-GWAS replication study confirming the association of *C14H8orf33* gene with milk production traits in dairy cattle

Shaohua YANG*, Chao QI*, Yan XIE, Xiaogang CUI, Yahui GAO, Jianping JIANG, Li JIANG, Shengli ZHANG, Qin ZHANG, Dongxiao SUN (✉)

Key Laboratory of Animal Genetics and Breeding of the Ministry of Agriculture, National Engineering Laboratory of Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

Abstract Genome-wide association studies with an Illumina Bovine50K chip have detected 105 SNPs associated with one or multiple milk production traits in the Chinese Holstein population. Of these, 38 significant SNPs detected with high confidence by both L1-TDT and MMRA methods were selected to further mine potential key genes affecting milk yield and milk composition. By blasting the flanking sequences of these 38 SNPs with the bovine genome sequence combined with comparative genomics analysis, 26 genes were found to contain or be near to such SNPs. Among them, the *C14H8orf33* gene is merely 87 bp away from the significant SNP, Hapmap30383-BTC-005848. Hence, we report herein genotype-phenotype associations to further validate the genetic effects of the *C14H8orf33* gene. By pooled DNA sequencing of 14 unrelated Holstein sires, a total of 18 with seven novel SNPs were identified. Among them, nine SNPs were in the 5′ regulatory region, one in exon 6 and the other in the 3′ UTR and 3′ regulatory region. A total of nine of these identified SNPs were successfully genotyped and analyzed by mass spectrometry for association with five milk production traits in an independent resource population. The results showed that these SNPs were statistically significant for more than two traits [$P < (0.0001 - 0.0267)$]. In addition, mRNA expression analyses revealed that *C14H8orf33* was ubiquitous in eight different tissues, with a relatively higher expression level in the mammary gland than in other tissues. These findings, therefore, provide strong evidence for association of *C14H8orf33* variants with milk yield and milk composition traits and may be applied in Chinese Holstein breeding programs.

Keywords GWAS, functional annotation, Chinese Holstein, milk production traits, *C14H8orf33* gene, single nucleotide polymorphisms, association study

1 Introduction

QTL linkage analyses and fine mapping studies have achieved remarkable results in recent decades [1–3]. However, the low density markers of the genetic variation in the complex economic traits cannot be captured using this method [4–7]. Genome wide association study (GWAS), which utilizes a large number of high-density genetic markers throughout the entire genome, provides a new approach to detect causal variations underlying complex traits [8,9]. So far, GWAS has been successfully applied to identify genes involved in human diseases [10,11], economical traits and various complex traits in animals [12,13]. Our previous GWAS with an illumina 50K chip detected 105 SNPs which were significantly associated with one or multiple milk production traits in dairy cattle [14]. As the first step in gene discovery [15,16], the results from GWAS still need further functional annotation and validation by use of genetic association studies. Thus, we selected 38 highly significant SNPs detected with high confidence by two statistical methods from these 105 significant SNPs. Through bioinformatics and comparative genomics analysis, a total of 26 genes were found to contain or be near to at least one of 38 significant SNPs, including the well-known *DGAT1* and *GHR* genes [17,18]. Of these, the chromosome 14 open reading frame 33 ortholog (*C14H8orf33*) gene had the nearest location to the significant SNP, Hapmap30383-BTC-005848 [14], and was considered as a promising candidate gene for milk production traits.

The *C14H8orf33* gene is located on BTA14, which includes a large number of QTLs for milk production traits, i.e., *DGAT1* [19–22]. The bovine *C14H8orf33* gene spans 2054 bp and contains 6 exons and 5 introns. The cDNA

Received November 1, 2014; accepted December 25, 2014

Correspondence: sundx@cau.edu.cn

*These authors contributed to this work equally

consists of 1220 bp with an open reading frame encoding a 188-amino acid protein. It is 313 kb away from the causal mutation K232A of the *DGAT1* gene. However, until recently, almost no relevant reports have been available for the *C14H8orf33* gene. In this research, an association study was conducted to confirm our previous GWAS result and to search for potential variants of the *C14H8orf33* gene affecting milk production traits in dairy cattle.

2 Materials and methods

2.1 Bioinformatics and comparative genomics analysis

To further validate the exact physical location of the 38 SNPs selected from the 105 significant SNPs, we separately compared each of the 60 bp upstream and downstream nucleotide sequences with NCBI (http://www.ncbi.nlm.nih.gov) and UCSC (http://genome.ucsc.edu/) website Btau 3.1 databases. From the exact physical location, we inferred the gene that the SNP was located within or near to.

The potential biochemistry and physiology of the gene based on its genome sequence was predicted by searching for the homologous and similar sequences from cattle, human and mouse. For the purpose of precise and accurate prediction, we used the websites; NCBI (http://www.ncbi.nlm.nih.gov), Ensemble (http://asia.ensembl.org/index.html), Uniprot (http://www.uniprot.org), KEGG (http://www.genome.jp/KEGG), GeneCards (http://www.genecards.org) and wikipathways (http://www.wikipathways.org) to achieve functional annotations for each gene.

2.2 Animal resource and DNA extraction

A daughter design was employed in this study. A total of 742 daughters from 14 corresponding sires were selected to construct the study population. The numbers of daughters for each of the 14 sires ranged from 22 to 125. These daughters were from 15 dairy farms in Beijing Sanyuanlvhe Dairy Farming Center. The official estimated breeding values (EBVs) for the five milk production traits, including milk yield (MY), fat yield (FY), protein yield (PY), fat percentage (FP) and protein percentage (PP) were provided by the Dairy Data Center of the Dairy Association of China (DAC) (http://www.holstein.org.cn). Genomic DNA was isolated from whole blood samples of cows and frozen semen of sires. A DNA pool was constructed from the DNA of 14 sires at the same concentration of 50 $ng \cdot \mu L^{-1}$.

2.3 SNP identification and genotyping

A total of 18 pairs of PCR primers (Appendix A, Table S1) were designed with Primer Premier 3 (Premier, Canada), according to the genomic sequence of the bovine *C14H8orf33* gene, to amplify all exons plus 5′ and 3′ franking regions. The SNPs identified using the pooled DNA from daughters of 14 sires, further SNPs were genotyped for all experimental cows using the iPLEX MassArray system (Sequenom Inc.). In addition, the SNP Hapmap30383-BTC-005848 from a previous GWAS [14] was genotyped for the purpose of replication in this study.

2.4 Statistical analyses

Allele and genotype frequencies were compared between the mutant and wild type through a chi-square test. The chi-square tests were also used to determine whether individual variants were in equilibrium at each locus by comparing the expected and observed genotype frequencies (Hardy–Weinberg equilibrium). Pedigrees of the population were traced back for three generations to create the relationship matrix. We calculated linkage disequilibrium between all pairs of biallelic loci using HAPLOVIEW 4.2. For single locus and haplotype analyses, the mixed procedure in SAS 9.1.3 with the animal model was fitted as follows:

$$y = 1\mu + bx + Za + e$$

where y is the vector of EBVs for each trait, μ is the overall mean, b is the regression coefficient of EBVs on SNP genotypes, x is the fixed effect vector, Z is the 742×742 indicator matrix, a is the vector of polygenetic effects with $a \sim N(0, A\delta_a^2)$ (where A is the additive kinship matrix and δ_a^2 is the additive variance), and e is the vector of residual errors distributed as $e \sim N(0, W\delta_e^2)$ [23].

2.5 Total RNA isolation and cDNA synthesis

Total RNA from 8 different tissues, i.e., heart, liver, small intestine, kidney, mammary gland, ovary, uterus and gluteus, was extracted using Trizol Reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocols and DNA contamination removed from RNA extracts with RNase-free DNase I for 30 min at 37°C. RNA integrity was checked by 1% agarose gels and the quantity was detected with NANODROP 2000 (Thermo Scientific, DE, USA). One microgram of total RNA for each tissue was reverse transcribed by PrimeScript® RT reagent Kit (TaKaRa, Ostn, Japan) to obtain the cDNA. Each cDNA sample was amplified to ascertain its quality with a pair of specific primers for *GAPDH*, which covers two partial adjacent exons and the whole intron between those two exons.

2.6 Real-time quantitative RT-PCR

With the primers as shown in Appendix A (Table S1), quantitative real-time RT-PCR was carried out with a LightCycler 480 Real-Time PCR System (Roche, Hercules, CA, USA). The reaction condition were as follows:

pre-denaturation at 95°C for 10 s; amplification 45 cycles of 95°C for 10 s, 60°C for 10 s, and 72°C for 10 s. The relative expression level was normalized by the *GAPDH* with $2^{-\Delta\Delta C_T}$ method as described previously (Livak and Schmittgen, 2001). All the measurements of *C14H8orf33* gene expression in different tissues were performed in triplicate, and the average values obtained. These data were analyzed by a *t*-test using the SAS9.0 program (SAS Institute, Inc., Cary, NC, USA), with a P value of < 0.05 considered significant.

3 Results

3.1 Function annotation

Based on the 38 SNPs selected with high confidence from the 105 significant SNPs identified by our initial GWAS, a total of 26 genes were obtained through bioinformatics and comparative genomics analysis, and their functions placed into seven major categories: including body metabolism and nutrient balance; cytoskeleton or extracellular matrix components; regulation of cell proliferation and apoptosis; cell signal transduction and salt ion channel composition; kinase activity; mRNA transcription and translation regulation (Table 1).

3.2 SNP identification and selection

By sequencing of pooled DNA from daughters of 14 unrelated sires in a Chinese Holstein population, a total of 18, including seven novel SNPs were identified (Table 2). Among them, nine SNPs were in the 5′ regulatory region, one in exon 6 and the remainder in the 3′ UTR and 3′ regulatory region. The SNP in exon 6 was a non-synonymous SNP with the amino acid alteration from proline (CCC) to histidine (CAC). Nine out of these identified SNPs were successfully genotyped by mass spectrometry and analyzed for association with five milk production traits in an independent resource population. Chi-square test showed all nine SNPs were in Hardy–Weinberg equilibrium ($P > 0.05$). The genotypic and allele frequencies are shown in Table 3.

3.3 Associations analyses

The association results are shown in Table 4. These SNPs were significantly associated with protein yield [$P < (0.0001 - 0.0267)$] but not associated with other milk production traits ($P > 0.05$) in present study. In addition, the SNP Hapmap30383-BTC-005848 identified in our initial GWAS [14] was successfully confirmed to have significant associations with MY, PY, FP and PP in this independent dairy cattle population. This provided convincing statistical evidence for our previous study.

3.4 Linkage disequilibrium analysis

The LD block generated by all nine SNPs within 5 kb (Fig. 1), consisted of three haplotypes, TCACCGTTT, AACTGACAC and TCACCGCTT with frequencies of 0.44, 0.39 and 0.15, respectively. The statistical analysis of the haplotypes with EBVs of five milk production traits showed that the haplotypes were associated with PY ($P = 2.31 \times 10^4$) (Table 5). The results were consistent with the associations of single SNPs.

3.5 Expression analysis of the bovine *C14H8orf33* gene

The relative mRNA expression of *C14H8orf33* in eight different tissues was determined by quantitative real-time PCR. The results revealed that *C14H8orf33* was ubiquitous in these eight tissues, and at a relatively higher expression level in the mammary gland than in other tissues. In addition, the expression of *C14H8orf33* in small intestine, kidney, ovary and uterus was also relatively higher than in three other tissues (Fig. 2).

4 Discussion

In this study, we annotated the function of 26 genes that correspond to 31 of 38 SNPs identified as highly significant via bioinformatics and comparative genomics analysis, and identified several novel *C14H8orf33* variants associated with milk production traits.

In previous studies, the known functional genes *DGAT1* [24], *ABCG2* [25] and *SCD1* for milk production traits [26] have been observed to have high expression in mammary tissue of mammals. In this study, we found that the *C14H8orf33* gene was also expressed at a relatively higher level in the mammary gland compared with seven other tissues, indicating its importance in mammary biologic processing in dairy cattle. Furthermore, our association data showed that the nine identified SNPs, including the SNP Hapmap30383-BTC-005848 identified by our initial GWAS [14], in the *C14H8orf33* gene were significantly associated with at least one milk trait. Therefore, it was inferred that the *C14H8orf33* gene showed relatively independent effects on the milk traits. At the same time, our findings provided convincing evidence for our previous GWAS study by a replication study. In conjunction with association analyses, the SNP Hapmap30383-BTC-005848 in the 3′ UTR in the *C14H8orf33* gene could be selected to examine whether this mutation is involved in interaction with some miRNA in a follow-up investigation.

In addition, the *C14H8orf33* gene is 313 kb away from the causal mutation K232A of the true QTL for milk composition, i.e., the *DGAT*1 gene. Although the significant associations of the *C14H8orf33* gene with milk production traits were associated with higher expression in

Table 1 Detailed information and functional annotation of 26 positional candidate genes corresponding to the 31 most significant SNPs in previous GWAS

Gene	Chr.	Name of SNP near gene	Position of SNP in gene	Related trait for SNP	Functional annotation
PDE9A	1	BTA-55340-no-rs	Intron 14	PY	The encoded protein plays a role in signal transduction by regulating the intracellular concentration of cyclic nucleotides
DIP2A	1	ARS-BFGL-NGS-113002	Intron 15	PY	Provides positional cues for axon path finding and patterning in the central nervous system, catalytic activity, transcription factor binding
KBTBD10	2	Hapmap39717-BTA-112973	Intron 5	FP	Mediates the ubiquitination and subsequent proteasomal degradation of target proteins
KCND3	3	INRA-701	Intron 1	PY	Regulates neurotransmitter release, heart rate, insulin secretion, neuronal excitability, epithelial electrolyte transport, smooth muscle contraction, and cell volume
FGGY	3	Hapmap38643-BTA-95454	Intron 8	MY	Carbohydrate metabolic process definition, neuron homeostasis definition; kinase activity definition, phosphotransferase activity, alcohol group as acceptor definition, transferase activity definition
ITPR2	5	Hapmap51303-BTA-74377	3′ 8578 bp	FP	Calcium ion transport, ion transport, transmembrane transport
HERC3	6	Hapmap24324-BTC-062449	3′ 5127 bp	PP	E3 ubiquitin-protein ligase
PKD2	6	BTA-121739-no-rs	Intron 2	PP	ATPase binding, calcium channel activity, calcium ion binding, channel activity, potassium channel activity, receptor binding
LPHN3	6	Hapmap41083-BTA-76098	Intron 21	PP	G-protein coupled receptor activity, sugar binding
C14H8orf33	14	Hapmap30383-BTC-005848	3′ 87 bp	FP MY PY	-
CYHR1	14	ARS-BFGL-NGS-34135	Exon 1	FP FY MY PP	Metal ion binding, protein binding
VPS28	14	ARS-BFGL-NGS-94706	Intron 6	FP FY MY PP	Component of the ESCRT-I complex, a regulator of vesicular trafficking process
DGAT1	14	ARS-BFGL-NGS-4939	5′ 160 bp	MY PY FP PP	Encodes a multipass transmembrane protein that functions as a key metabolic enzyme The encoded protein catalyzes the conversion of diacylglycerol and fatty acyl CoA to triacylglycerol. This enzyme can also transfer acyl CoA to retinol. Activity of this protein may be associated with obesity and other metabolic diseases
MAPK15	14	Hapmap24715-BTC-001973	Intron 8	FP	ATP binding, MAP kinase activity,SH3 domain binding, nucleotide binding, protein serine/threonine kinase activity, transferase activity
EEF1D	14	ARS-BFGL-NGS-101653	Intron 1	FP	Protein binding, signal transducer activity, translation elongation factor activity
ZC3H3	14	ARS-BFGL-NGS-26520	Intron 4	FP	Nucleic acid binding, zinc ion binding

(*Continued*)

Gene	Chr.	Name of SNP near gene	Position of SNP in gene	Related trait for SNP	Functional annotation
GML	14	Hapmap25486-BTC-072553	Intron 3	FP MY	May play a role in the apoptotic pathway or cell-cycle regulation induced by TP53/p53 after DNA damage
GPIHBP1	14	Hapmap30646-BTC-002054	5′ 1295 bp	FP	Plays a key role in the lipolytic processing of chylomicrons
EIF2C2	14	UA-IFASA-7269	5′ 2768 bp	FP	RNA 7-methylguanosine cap binding, endonuclease activity, cleaving siRNA-paired mRNA, hydrolase activity, metal ion binding, protein binding, siRNA binding
TRAPPC9	14	ARS-BFGL-NGS-100480 UA-IFASA-5306 Hapmap27703-BTC-053907	Intron 16 Intron 17 Intron 20	FP PP MY	Activates of NF-kappa-B through increased phosphorylation of the IKK complex, neuronal cells differentiation, vesicular transport
COL22A1	14	ARS-BFGL-NGS-3571	Intron 15	FP	Acts as a cell adhesion ligand for skin epithelial cells and fibroblasts
KHDRBS3	14	UA-IFASA-6647 Hapmap32948-BTC-047992	Intron 1 Intron 3	FP	May play a role as a negative regulator of cell growth, inhibits cell proliferation, involved in splice site selection of vascular endothelial growth factor
PARP8	20	ARS-BFGL-BAC-27914	Intron 2	PP	ADP-ribosyltransferase activity, transferase activity, transferase activity, transferring glycosyl groups
GHR	20	BFGL-NGS-118998	Intron1	PP	A transmembrane receptor for growth hormone. Binding of growth hormone to the receptor leads to receptor dimerization and the activation of an intra-and intercellular signal transduction pathway leading to growth
RICTOR	20	ARS-BFGL-NGS-38482	Intron 3	PP	Subunit of mTORC2, which regulates cell growth and survival in response to hormonal signals
NIPBL	20	BTB-00782435 BTA-13793-rs29018751 BTB-01842107	Intron 1 Intron 6 Intron 34	PP	Regulates the activity of certain genes for normal development, involved in the repair of damaged DNA

Table 2 SNPs detected by sequencing in the *C14H8orf33* gene

No.	Position on UMD_3.1	Gene region	Mutation	Polymorphism type	Amino acid substitution
1	72576	5′ flanking region	C/A	SNP	-
2	72793	5′ flanking region	T/A	SNP	-
3	73189	5′ flanking region	-/T	ins/del	-
4	73306	5′ flanking region	T/G	SNP	-
5	73466	5′ flanking region	T/A	SNP	-
6	73551	5′ flanking region	A/C	SNP	-
7	73552	5′ flanking region	T/C	SNP	-
8	74412	5′ flanking region	-/ACC	ins/del	-
9	74484	5′ flanking region	T/C	SNP	-
10	74654	Intron 1	T/C	SNP	-
11	75863	Exon 6	C/A	SNP	A/D
12	76100	3′ UTR	C/T	SNP	-
13	76321	3′ UTR	A/G	SNP	-
14	76378	3′ UTR	C/G	SNP	-
15	76589	3′ UTR	A/G	SNP	-
16	76703	3′ flanking region	C/T	SNP	-
17	77377	3′ flanking region	A/T	SNP	-
18	78386	3′ flanking region	C/T	SNP	-

Table 3 Genotypes and allele frequencies of the nine identified SNPs in the *C14H8orf33* gene

SNP ID	Location	SNP position	Genotype	Genotype frequency	Allele	Allelic frequency
SNP1	5′ flanking region	72793	TT (244)	0.368	T	0.609
			TA (319)	0.481		
			AA (100)	0.151	A	0.391
SNP2	5′ flanking region	73306	AA (131)	0.174	A	0.429
			CA (383)	0.509		
			CC (238)	0.316	C	0.571
SNP3	5′ flanking region	73551	AA (239)	0.318	A	0.573
			CA (384)	0.511		
			CC (129)	0.172	C	0.427
SNP4	Intron 1	74654	TT (131)	0.174	T	0.428
			CT (381)	0.507		
			CC (240)	0.319	C	0.572
SNP5	3′ UTR	76100	GG (114)	0.174	G	0.406
			GC (304)	0.464		
			CC (237)	0.362	C	0.594
SNP6	3′ UTR	76321	GG (242)	0.322	G	0.575
			GA (380)	0.506		
			AA (129)	0.172	A	0.425
SNP7	3′ UTR	76703	TT (122)	0.162	T	0.402
			CT (361)	0.480		
			CC (269)	0.358	C	0.598
SNP8	3′ flanking region	77377	TT (237)	0.315	T	0.570
			AT (384)	0.511		
			AA (131)	0.174	A	0.430

(*Continued*)

SNP ID	Location	SNP position	Genotype	Genotype frequency	Allele	Allelic frequency
SNP9	3′ flanking region	78386	TT (239)	0.318	T	0.573
			TC (382)	0.509		
			CC (130)	0.173	C	0.427

Table 4 The association analysis between *C14H8orf33* and EBVs of 5 milk production traits

SNPs	Genotype	MY	FY	FP	PY	PP
SNP1	AA	230.36±111.63	4.51±2.68	−0.06±0.023	13.76^{A}±3.26	0.030±0.010
	TA	203.11±66.05	1.35±2.92	−0.05±0.014	5.38^{B}±1.91	0.007±0.006
	TT	212.72±77.87	3.15±6.26	−0.03±0.019	4.18^{B}±2.26	−0.012±0.007
	P value	0.9781	0.7021	0.4739	0.0003**	0.2864
SNP2	AA	276.57±97.61	2.09±6.43	−0.05±0.023	16.73^{A}±2.72	−0.022±0.009
	CA	220.88±59.18	0.71±2.94	−0.05±0.014	5.46^{B}±1.62	−0.007±0.006
	CC	206.90±78.91	3.23±2.68	−0.03±0.087	4.43^{B}±2.24	−0.012±0.008
	P value	0.8465	0.7073	0.4901	< 0.0001**	0.3655
SNP3	AA	204.38±64.67	3.05±2.89	−0.03±0.018	1.48^{A}±2.01	−0.011±0.008
	CA	226.90±60.74	1.01±2.72	−0.05±0.014	5.66^{A}±1.91	−0.007±0.006
	CC	265.01±83.64	1.56±3.55	−0.06±0.023	15.91^{B}±2.48	−0.025±0.010
	P value	0.8909	0.7994	0.5159	0.0003**	0.2618
SNP4	CC	211.62±78.40	3.50±2.51	−0.03±0.019	8.44^{a}±2.99	−0.009±0.008
	CT	218.27±59.41	−0.05±2.14	−0.06±0.014	8.84^{a}±2.93	−0.008±0.006
	TT	276.57±97.61	1.21±2.98	−0.05±0.033	12.34^{b}±2.79	0.027±0.010
	P value	0.8531	0.6227	0.4540	0.0193*	0.3587
SNP5	CC	236.77±80.33	3.32±2.68	−0.04±0.019	2.14^{A}±2.28	−0.014±0.008
	GC	209.75±68.08	6.09±2.92	−0.06±0.016	6.04^{A}±1.86	−0.002±0.006
	GG	282.54±107.78	0.99±3.37	−0.07±0.025	12.48^{B}±3.05	−0.027±0.010
	P value	0.8479	0.5916	0.5983	0.0003**	0.0948
SNP6	AA	265.01±98.62	1.56±3.07	−0.06±0.023	13.91^{A}±2.78	0.005±0.008
	GA	216.68±59.41	0.63±1.85	−0.05±0.014	5.27^{B}±1.63	0.004±0.009
	GG	222.39±77.91	3.66±2.43	−0.03±0.017	2.26^{B}±2.21	−0.035±0.025
	P value	0.9130	0.6115	0.5092	0.0004**	0.2711
SNP7	CC	299.27^{A}±70.23	0.19±2.68	-0.08^{B}±0.016	11.13^{A}±1.98	-0.025^{A}±0.007
	CT	144.81^{B}±61.55	1.80±2.92	-0.02^{A}±0.014	3.48^{B}±1.73	0.001^{B}±0.006
	TT	309.96^{A}±106.82	4.94±6.41	-0.04^{AB}±0.025	3.89^{B}±3.08	-0.014^{AB}±0.010
	P value	0.0074**	0.4932	0.0267*	0.0106*	0.0078**
SNP8	AA	276.57±97.60	4.74±2.64	−0.06±0.023	11.28^{A}±2.74	−0.024±0.009
	AT	226.75±59.07	5.89±3.92	−0.05±0.014	5.49^{B}±1.63	−0.007±0.006
	TT	196.26±79.17	1.99±6.11	−0.03±0.018	1.37^{B}±2.24	−0.011±0.008
	P value	0.8153	0.7299	0.6034	0.0001**	0.3281
SNP9	CC	262.45±98.11	1.79±1.05	−0.06±0.027	11.82^{A}±2.76	−0.025±0.008
	TC	227.33±59.30	0.93±1.84	−0.05±0.010	5.59^{B}±1.63	−0.008±0.009
	TT	205.17±78.41	3.21±2.44	−0.03±0.018	4.62^{B}±2.23	−0.010±0.007
	P value	0.9012	0.7570	0.4739	0.0003**	0.3108

Note: ^{a,b}: within the same column with different superscripts means $P < 0.05$, ^{A,B}: within the same column with different superscripts means $P < 0.01$; *: $P < 0.05$; **: $P < 0.01$; MY: milk yield; FY: fat yield; FP: fat percentage; PY: protein yield; PP: protein percentage.

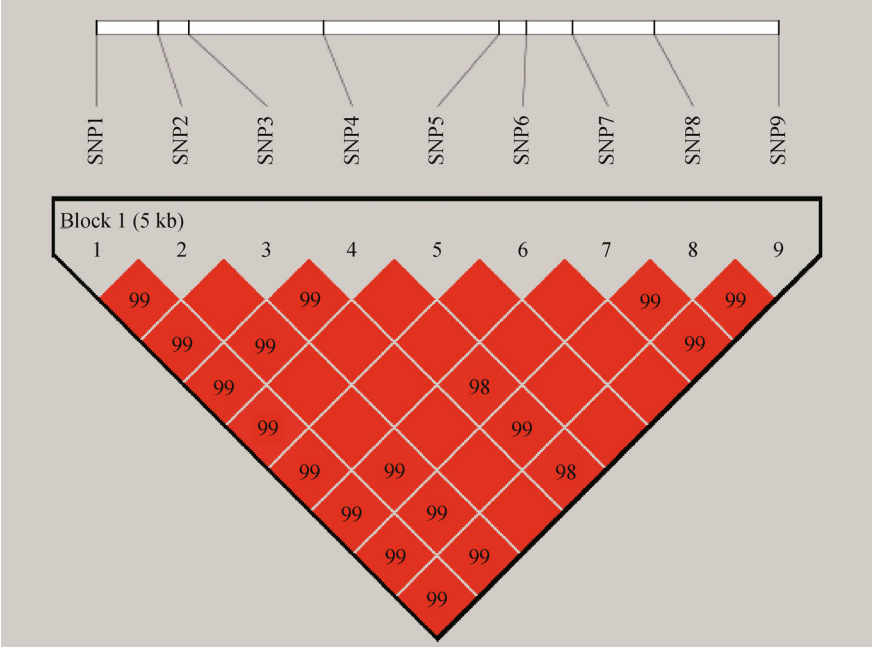

Fig. 1 The haplotype block and LD pattern for nine SNPs in the *C14H8orf33* gene. The length of the block is provided in kilobases (kb), and pairwise linkage disequilibrium (LD) is given for each SNP combination. The darker shading indicates higher linkage disequilibrium.

Table 5 Main haplotypes of the *C14H8orf33* gene, their frequencies and significant associations with the EBVs of protein yield in dairy cattle

Haplotypes	SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	Frequency/%	PY(*P*-value)
TCACCGTTT	T	C	A	C	C	G	T	T	T	43.9	2.31×10^{4}
AACTGACAC	A	A	C	T	G	A	C	A	C	38.9	
TCACCGCTT	T	C	A	C	C	G	C	T	T	14.9	

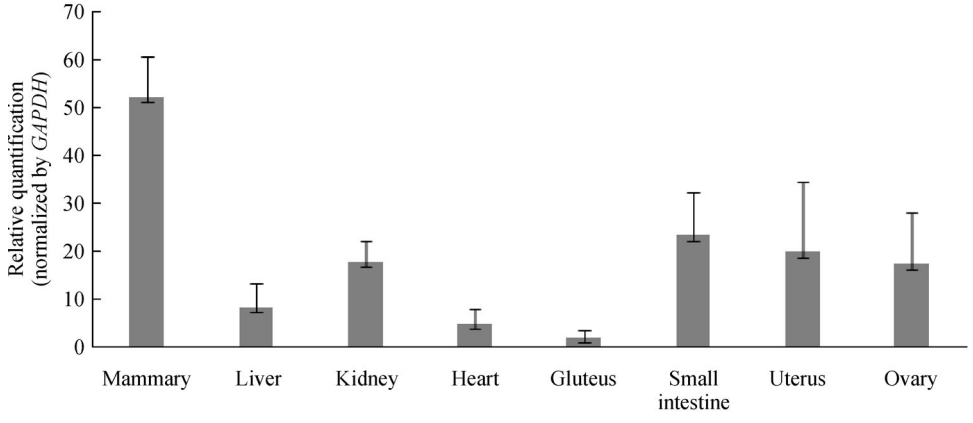

Fig. 2 Relative quantification of the *C14H8orf33* mRNA in eight tissues

mammary gland of lactating cows, it is suspected that these significant associations could be due to the linkage disequilibrium (LD) between *C14H8orf33* and *DGAT1*. We found a total of 25 genes between *C14H8orf33* and *DGAT1* and further investigations are needed in order to verify whether the strong LD exist between these two genes.

5 Conclusions

This study provided strong evidence for association of *C14H8orf33* variants with milk yield and milk composition traits and may be applied in Chinese Holstein breeding programs.

Acknowledgements This work was supported by the National High-tech R&D Program of China (2013AA102504), the National Key Technology R&D Program (2011BAD28B02), the National Transgenic Major Project (2014ZX08009-053B), the Beijing Innovation Team of Technology System in the National Dairy Industry, the Beijing Research and Technology Program (D121100003312001), the earmarked fund for Modern Agro-industry Technology Research System (CARS-37) and the Program for Changjiang Scholar and Innovation Research Team in University (IRT1191). The authors also express appreciation for the kind help from the official Dairy Data Center of China in providing the official EBVs data.

Supplementary material The online version of this article at http://dx.doi.org/10.15302/J-FASE-2014037 contains supplementary material (Appendix A).

Compliance with ethics guidelines Shaohua Yang, Chao Qi, Yan Xie, Xiaogang Cui, Yahui Gao, Jianping Jiang, Li Jiang, Shengli Zhang, Qin Zhang and Dongxiao Sun declare that they have no conflict of interest or financial conflicts to disclose.

All applicable institutional and national guidelines for the care and use of animals were followed.

References

1. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino A T, Sargeant L S, Sorensen A, Steele M R, Zhao X, Womack J E, Hoeschele I. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, 1995, 139(2): 907–920
2. Khatkar M S, Thomson P C, Tammen I, Raadsma H W. Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genetics Selection Evolution*, 2004, 36(2): 163–190
3. Smaragdov M. Genetic mapping of loci responsible for milk production traits in dairy cattle. *Russian Journal of Genetics*, 2006, 42(1): 1–15
4. Zhang H, Wang Z, Wang S, Li H. Progress of genome wide association study in domestic animals. *Journal of Animal Science and Biotechnology*, 2012, 3(1): 26
5. Olsen H G, Lien S, Gautier M, Nilsen H, Roseth A, Berg P R, Sundsaasen K K, Svendsen M, Meuwissen T H. Mapping of a milk production quantitative trait locus to a 420 kb region on bovine chromosome 6. *Genetics*, 2005, 169(1): 275–283
6. Gutiérrez-Gil B, Williams J L, Homer D, Burton D, Haley C S, Wiener P. Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle. *Journal of Animal Science*, 2009, 87(1): 24–36
7. Kneeland J, Li C, Basarab J, Snelling W M, Benkel B, Murdoch B, Hansen C, Moore S S. Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of *Bos taurus*. *Journal of Animal Science*, 2004, 82(12): 3405–3414
8. Bolormaa S, Pryce J E, Hayes B J, Goddard M E. Multivariate analysis of a genome-wide association study in dairy cattle. *Journal of Dairy Science*, 2010, 93(8): 3818–3833
9. Mai M D, Sahana G, Christiansen F B, Guldbrandtsen B. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *Journal of Animal Science*, 2010, 88(11): 3522–3528
10. Hampe J, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, De La Vega F M, Briggs J, Günther S, Prescott N J, Onnie C M, Häsler R, Sipos B, Fölsch U R, Lengauer T, Platzer M, Mathew C G, Krawczak M, Schreiber S. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature Genetics*, 2007, 39 (2): 207–211
11. Sun L D, Xiao F L, Li Y, Zhou W M, Tang H Y, Tang X F, Zhang H, Schaarschmidt H, Zuo X B, Foelster-Holst R, He S M, Shi M, Liu Q, Lv Y M, Chen X L, Zhu K J, Guo Y F, Hu D Y, Li M, Li M, Zhang Y H, Zhang X, Tang J P, Guo B R, Wang H, Liu Y, Zou X Y, Zhou F S, Liu X Y, Chen G, Ma L, Zhang S M, Jiang A P, Zheng X D, Gao X H, Li P, Tu C X, Yin X Y, Han X P, Ren Y Q, Song S P, Lu Z Y, Zhang X L, Cui Y, Chang J, Gao M, Luo X Y, Wang P G, Dai X, Su W, Li H, Shen C P, Liu S X, Feng X B, Yang C J, Lin G S, Wang Z X, Huang J Q, Fan X, Wang Y, Bao Y X, Yang S, Liu J J, Franke A, Weidinger S, Yao Z R, Zhang X J. Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. *Nature Genetics*, 2011, 43(7): 690–694
12. García-Gámez E, Gutiérrez-Gil B, Sahana G, Sánchez J P, Bayón Y, Arranz J J. GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the *LALBA* gene. *PLoS ONE*, 2012, 7(10): e47782
13. Magwire M M, Fabian D K, Schweyen H, Cao C, Longdon B, Bayer F, Jiggins F M. Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila melanogaster*. *PLOS Genetics*, 2012, 8(11): e1003057
14. Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, Zhang Q. Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS ONE*, 2010, 5(10): e13661
15. Hirschhorn J N. Genomewide association studies—illuminating biologic pathways. *The New England Journal of Medicine*, 2009, 360(17): 1699–1701
16. Hardy J, Singleton A. Genomewide association studies and human disease. *The New England Journal of Medicine*, 2009, 360(17): 1759–1768
17. Grisart B, Farnir F, Karim L, Cambisano N, Kim J J, Kvasz A, Mni M, Simon P, Frère J M, Coppieters W, Georges M. Genetic and functional confirmation of the causality of the *DGAT1 K232A* quantitative trait nucleotide in affecting milk yield and composition. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101(8): 2398–2403
18. Blott S, Kim J J, Moisio S, Schmidt-Küntzel A, Cornet A, Berzi P, Cambisano N, Ford C, Grisart B, Johnson D, Karim L, Simon P, Snell R, Spelman R, Wong J, Vilkki J, Georges M, Farnir F, Coppieters W. Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics*, 2003, 163(1): 253–266
19. Coppieters W, Riquet J, Arranz J J, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Nezer C, Simon P, Vanmanshoven P, Wagenaar D, Georges M. A QTL with major effect on milk yield and composition maps to bovine chromosome 14. *Mammalian Genome*, 1998, 9(7): 540–544
20. Sun D, Jia J, Ma Y, Zhang Y, Wang Y, Yu Y, Zhang Y. Effects of

DGAT1 and *GHR* on milk yield and milk composition in the Chinese dairy population. *Animal Genetics*, 2009, 40(6): 997–1000

21. Riquet J, Coppieters W, Cambisano N, Arranz J J, Berzi P, Davis S K, Grisart B, Farnir F, Karim L, Mni M, Simon P, Taylor J F, Vanmanshoven P, Wagenaar D, Womack J E, Georges M. Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 1999, 96(16): 9252–9257
22. Looft C, Reinsch N, Karall-Albrecht C, Paul S, Brink M, Thomsen H, Brockmann G, Kühn C, Schwerin M, Kalm E. A mammary gland EST showing linkage disequilibrium to a milk production QTL on bovine Chromosome 14. *Mammalian Genome*, 2001, 12(8): 646–650
23. Wang H, Jiang L, Liu X, Yang J, Wei J, Xu J, Zhang Q, Liu J F. A post-GWAS replication study confirming the *PTK2* gene associated with milk production traits in Chinese Holstein. *PLoS ONE*, 2013, 8 (12): e83625
24. Cases S, Smith S J, Zheng Y W, Myers H M, Lear S R, Sande E, Novak S, Collins C, Welch C B, Lusis A J, Erickson S K, Farese R V Jr. Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 1998, 95(22): 13018–13023
25. Bionaz M, Loor J J. Gene networks driving bovine milk fat synthesis during the lactation cycle. *BMC Genomics*, 2008, 9(1): 366
26. Lengi A J, Corl B A. Identification and characterization of a novel bovine stearoyl-CoA desaturase isoform with homology to human SCD5. *Lipids*, 2007, 42(6): 499–508