

RESEARCH ARTICLE

SNV discovery and functional candidate gene identification for milk composition based on whole genome resequencing of Holstein bulls with extremely high and low breeding values

Shan Lin¹, Hongyan Zhang¹, Yali Hou², Lin Liu³, Wenhui Li¹, Jianping Jiang¹, Bo Han¹, Shengli Zhang¹, Dongxiao Sun^{1*}

1 Department of Animal Genetics and Breeding, College of Animal Science and Technology, Key Laboratory of Animal Genetics and Breeding of the Ministry of Agriculture, National Engineering Laboratory for Animal Breeding, China Agricultural University, Beijing, China, **2** Laboratory of Disease Genomics and Individualized Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China, **3** Beijing Dairy Cattle Center, Beijing, China

* sundx@cau.edu.cn



OPEN ACCESS

Citation: Lin S, Zhang H, Hou Y, Liu L, Li W, Jiang J, et al. (2019) SNV discovery and functional candidate gene identification for milk composition based on whole genome resequencing of Holstein bulls with extremely high and low breeding values. PLoS ONE 14(8): e0220629. <https://doi.org/10.1371/journal.pone.0220629>

Editor: Sujan Mamidi, HudsonAlpha Institute for Biotechnology, UNITED STATES

Received: January 4, 2019

Accepted: July 19, 2019

Published: August 1, 2019

Copyright: © 2019 Lin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: This work was financially supported by the National Natural Science Foundation of China (31872330, 31802041), Beijing Dairy Industry Innovation Team (BAIC06-2017/2018), Beijing Science and Technology Program (D171100002417001), National Science and Technology Programs of China (2013AA102504),

Abstract

We have sequenced the whole genomes of eight proven Holstein bulls from the four half-sib or full-sib families with extremely high and low estimated breeding values (EBV) for milk protein percentage (PP) and fat percentage (FP) using Illumina re-sequencing technology. Consequently, 2.3 billion raw reads were obtained with an average effective depth of 8.1×. After single nucleotide variant (SNV) calling, total 10,961,243 SNVs were identified, and 57,451 of them showed opposite fixed sites between the bulls with high and low EBVs within each family (called as common differential SNVs). Next, we annotated the common differential SNVs based on the bovine reference genome, and observed that 45,188 SNVs (78.70%) were located in the intergenic region of genes and merely 11,871 SNVs (20.67%) located within the protein-coding genes. Of them, 13,099 common differential SNVs that were within or close to protein-coding genes with less than 5 kb were chosen for identification of candidate genes for milk compositions in dairy cattle. By integrated analysis of the 2,657 genes with the GO terms and pathways related to protein and fat metabolism, and the known quantitative trait loci (QTLs) for milk protein and fat traits, we identified 17 promising candidate genes: *ALG14*, *ATP2C1*, *PLD1*, *C3H1orf85*, *SNX7*, *MTHFD2L*, *CDKN2D*, *COL5A3*, *FDX1L*, *PIN1*, *FIG4*, *EXOC7*, *LASP1*, *PGS1*, *SAO*, *GPLD1* and *MGEA5*. Our findings provided an important foundation for further study and a prompt for molecular breeding of dairy cattle.

Introduction

Milk yield, milk protein and fat traits are main economic traits and important breeding goals of dairy industry. Compared to the standard phenotypic data based methods, marker-assisted

earmarked fund for Modern Agro-industry Technology Research System (CARS-36), and the Program for Changjiang Scholar and Innovation Research Team in University (IRT_15R62).

Competing interests: The authors have declared that no competing interests exist.

selection is expected to lead faster genetic progress by using information at the DNA level. Of note, genomic selection (GS) with the application of high density SNP chips has become the most popular and efficient technology in dairy cattle breeding since the first report of GS in 2001 by Meuwissen et al [1, 2]. Using publicly available quantitative trait loci (QTLs) and genome-wide association study (GWAS) data can improve the accuracy of whole genome prediction (WGP) compared to the chip-based GBLUP and BayesB methods et al[3]. Over the last few decades, with linkage (LA) or linkage and linkage disequilibrium (LA/LD) analysis, candidate genes approach and genome-wide association analysis (GWAS)[4], a great amount of QTLs and genetic associations for milk yield and milk composition have been identified in dairy cattle since the first report of QTL mapping in Holstein by Georges et al [5]. So far, the Cattle QTL database contains 3,996, 17,677, and 19,895 loci for milk yield, milk protein and fat, respectively (December 23, 2018, <http://www.animalgenome.org/cgi-bin/QTLdb/>). Nonetheless, merely *DGATI*, *GHR*, and *ABCG2* gene have been validated to be true major genes for milk composition traits until now [6–11].

In recent years, the development of bioinformatics software and cost reduction of next generation sequencing (NGS) has opened a new era for genomics and molecular biology[12]. Compared to the traditional Sanger capillary electrophoresis sequencing method [13, 14], NGS technologies that is massively parallel DNA sequencing methods, provide higher throughput data with lower cost and make population-scale genome research possible[15–17]. Moreover, NGS can detect rare mutations, solve the disequilibrium between the rare causal mutations, genotype SNPs and distinguish structural variants [15, 18]. As for all kinds of variants, single nucleotide polymorphisms (SNPs) are the most widespread and wide-used in identification of genes for complex traits [19, 20]. Some whole genome resequencing studies in cattle have been reported on SNPs and copy number variations (CNVs) for genetic differences between the Black Angus and Holstein[21], Hanwoo-specific structural variations and selection signatures for meat quality and disease resistance traits in Hanwoo [22], haplotype under selection in USA Holstein [23] and evolutionary analysis in Japanese Kuchinoshima-Ushi [24]. In our previous studies, we detected some CNVs and insertions and deletions (indels) associated with milk protein and fat in Chinese Holstein [25, 26]. In the present study, we searched for differential SNVs between the Holstein bulls with extremely high and low estimated breeding values (EBVs) for milk protein percentage (PP) and fat percentage (FP) traits based on whole genome sequencing data, and identified candidate genes for milk compositions by integrating biological functions and the known QTL data.

Materials and methods

Sample selection and resequencing

Eight Holstein bulls were selected from the Beijing Dairy Cattle Center (<http://www.bdcc.com.cn/>) that consisted of four full-sib and/or half-sib families, and each family contains two bulls who have extremely high and low EBVs for milk protein percentage (PP) and fat percentage (FP) with reliabilities of more than 0.85. The detailed information of the 8 bulls were described previously [25, 26].

The frozen semen samples were used for genomic DNA collection with the standard phenol/chloroform extraction method. 1% agarose gels and Nano Drop 2000 (Thermo Scientific Inc. Waltham, DE, USA) were performed for the DNA concentration and purity control. The purified DNAs were then used for library construction. Eight paired-end libraries (read length = 2×100 bp) with one library for each bull were constructed, and subsequently sequenced on Illumina HiSeq2000 instruments (Illumina Inc., San Diego, CA, USA).

Read mapping and SNV calling

By using the Burrows–Wheeler Alignment tool (BWA ver. 0.6.2)[27], the sequenced reads were aligned to the bovine reference genome assembly UMD3.1.69 (ftp://ftp.ensembl.org/pub/release-69/fasta/bos_taurus/dna/) with the default parameters. NGS QC Toolkit with default parameters was applied to reduce mapping error rate [28]. By comparing 8 individual sequence to the bovine reference genome respectively, we called SNVs for each bull based on SAM tools (ver. 0.1.19)[29] with following criteria: base quality score ≥ 20 ; read depth < 100 for each individual; and non-reference allele supporting reads > 3 . Based on this, 8 sets of SNV data for 8 bulls could be obtained.

Functional annotation and SNV filtering

After SNV calling, the SNVs were annotated by ANNOVAR[30] using the RefSeq gene sets (14,912 genes; the gene sets is available from the UCSC download site <http://hgdownload.cse.ucsc.edu/goldenPath/bosTau6/database/>). The region that was close to a gene with less than 1kb was defined as upstream/downstream and that with more than 1kb was defined as inter-genic region.

Afterwards, every single nucleotide which was polymorphic between the two bulls with high and low EBVs within each family was preserved. Then the SNVs with opposite fixed sites across four families were chosen and defined as ‘common differential SNVs’. Fixed sites, which were SNVs with opposite fixed alleles in the high and low group were used for identification of candidate genes.

Functional enrichment analysis

After annotation, we selected the genes that included or were closed to the common differential SNVs with less than 5 kb. Then, we performed Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Medical Subject Headings (MeSH) enrichment for these genes. KOBAS tool (<http://kobas.cbi.pku.edu.cn/>) was used for GO and KEGG pathway enrichment and MeSH ORA was applied for MeSH enrichment[31–33]. All packages used in MeSH analysis are available in the releases of BIOCONDUCTOR (<http://bioconductor.org/>). P value of < 0.05 determined by Fisher’s exact test was set as the criteria for significance.

Apart from the genes that were referred in the significantly enriched pathways, we also remained the genes that were not significantly enriched but involved in the 8 well-known pathways related to protein, fat, and fatty acid metabolisms based on the KEGG pathway website (<http://www.kegg.jp/>), including mTOR, insulin, AMPK, PPAR, Jak-STAT, PI3K-Akt, MAPK, and TGF- β .

Positions comparison with known QTL database

Afterwards, we obtained the genetic position of each gene based on its physical position and compared with the confidence intervals and the peak positions of the previously reported QTLs that have been shown to be associated with milk composition traits (<http://www.animalgenome.org/cgi-bin/QTLdb>). The genes that were close to the peaks of QTLs with less than 1 cM were remained.

Results

Read mapping and SNV detection

With Illumina HiSeq 2000, we sequenced the genomic DNA samples of the eight Holstein bulls with extremely high and low EBVs for milk protein percentage and milk fat percentage

Table 1. Summary of the sequencing result and SNV counts for 8 extreme Holstein bulls.

Sib-family	Sample	Raw reads	Mapped reads	Mapped reads (%)	Uniquely mapped reads (%)	Genome coverage (%)	Sequencing depth(X)	SNV
full-sib1	high 1	289,952,310	261,783,075	92.01	82.9	98.58	8	4,925,685
	low 1	286,870,238	252,201,294	90.24	80.88	98.55	8	4,636,009
full-sib2	high 2	292,878,886	257,840,281	91.28	82.97	98.59	8	5,075,588
	low 2	272,948,496	241,748,531	91.48	81.69	98.52	8	4,166,954
half-sib1	high 3	251,953,446	218,677,291	89.03	81.15	98.34	7	4,306,882
	low 3	337,815,303	314,651,325	96.73	83.75	98.4	10	5,403,284
half-sib2	high 4	288,003,254	253,311,627	91.01	82.45	98.61	8	4,572,191
	low 4	283,359,516	255,124,411	91.89	85.16	98.57	8	3,399,107
average		287,972,681	256,917,229	91.71	82.62	98.52	8.1	4,560,713

<https://doi.org/10.1371/journal.pone.0220629.t001>

[25, 26]. As a result, a total of 2,303,781,449 raw reads were obtained. Of these, 2,055,337,835 reads (91.71%) were finally mapped to the reference genome (UMD 3.1.69) while the average proportion of uniquely mapped reads was 82.62% with an average depth of 8.1×, and the genome coverage was approximately 98% in each individual.

By using SAM tools with variant filtration process, 10,961,243 SNVs were identified in total after removing duplicates among 8 bulls at an average of 4,560,713 within each individual (Table 1).

SNV annotation and genomic distribution

The 10,945,507 SNVs were annotated based on the bovine gene set in RefSeq database (including 14,912 genes) (S1 Table). Most of the SNVs (79.63%) were detected in intergenic region, and 2,156,190 (19.70%) SNVs were located within genes including introns (19.13%), exons (0.32%) and untranslated regions (UTR) (0.24%), and other region (0.68%) (ncRNA_exonic, ncRNA_intronic and up/downstream) (Table 2 and Fig 1). Of the total 35,505 exonic SNVs, 11,843 nonsynonymous nucleotide substitutions were included.

Identification of common differential SNVs

Out of the 10,945,507 annotated SNVs, 57,451 that were fixed sites between the bulls with extremely high and low EBVs across four families were chosen for the further analysis, which number across chromosomes ranged from 761 to 5,044 (Fig 2). As a result of annotation, 57,419 common differential SNVs were successfully classified into 9 functional categories: the majority was found in intergenic and intronic regions (78.70% and 20.10%, respectively), whereas fewer SNVs were located in exon (0.31%), exonic ncRNA (0.003%), UTR (0.26%) and up/downstream (0.63%) (Table 3 and Fig 3).

Subsequently, we further identified 2,657 protein-coding genes that included or were nearby the common differential SNVs with less than 5 kb. Of these, 13,099 SNVs were remained, including 11,498 (87.78%) in intron, 176 (1.34%) in exon, 154 (1.18%) in UTR, and 355 (2.71%) in upstream/downstream while only 6.99% were detected in intergenic region (Table 4 and Fig 4).

Genes Ontology and pathway analyses

To further identify candidate genes for milk protein and fat traits, we performed functional analysis on the above-mentioned 2,657 genes with KOBAS online tool and MeSH ORA. A total of 6,819 GO terms and 286 KEGG pathways were observed, among them with 1,011 terms and 73 pathways were significantly enriched ($P < 0.05$; S2 Table). 23 significant MeSH

Table 2. Annotation of 10,945,507 SNVs across 8 bulls.

Category	Number of SNP	Percentage of SNP % ¹
intergenic	8,715,765	79.63
upstream ^a	33,515	0.31
downstream ^b	36,840	0.34
upstream and downstream ^c	624	0.01
3' UTR	22,495	0.21
5' UTR	3,832	0.04
ncRNA__exonic ^d	216	0.002
ncRNA__intronic ^e	2357	0.02
intronic	2,094,358	19.13
exonic		0.32
nonsynonymous	11,843	
synonymous	19,617	
stop gain	75	
stop loss	5	
unknown	3,965	

¹Percentage was calculated based on total annotated SNVs.

^aupstream from the nearest gene (<1kb).

^bdownstream away from the nearest gene (<1kb).

^cvariant located in both upstream and downstream regions for two different genes (<1kb).

^dnon-coding RNA expressed within exon of a gene

^enon-coding RNA expressed within intron of a gene

<https://doi.org/10.1371/journal.pone.0220629.t002>

terms were detected by MeSH ORA ($P < 0.05$; [S3 Table](#)). Of these, 29 genes were enriched in Mesh term of Amino Acids (MeSH:D000596) in the Chemicals and Drugs category which was associated with protein synthesis and metabolism. Thereby, we identified 1,354 genes that were involved in 133 significant GO terms, pathways and Mesh terms relevant to protein, lipid, and fatty acid synthesis and metabolism such as protein metabolic, cellular protein modification, lipid modification, phospholipid metabolic, glycerophospholipid metabolic, sphingolipid metabolism, glycerolipid metabolic, fat cell differentiation, insulin resistance, insulin secretion and MAPK signaling pathways.

Besides the genes which were significantly enriched in pathways associated with protein and fat synthesis and metabolism, we selected 23 additional genes that were not significantly enriched but participated in six well-known pathways such as mTOR, AMPK, Jak-STAT, PI3K-Akt, PPAR and TGF- β . Thus, 1,377 candidate genes were obtained for milk protein and fat traits.

Position comparison with known QTLs and identification of promising candidates associated with milk protein and fat traits

We further compared the physical positions of the 1,377 candidate genes with the previously reported QTLs for milk fat and protein in dairy cattle (<http://www.animalgenome.org/cgi-bin/QTLdb>). Consequently, 94 genes were found to be adjacent to the peak positions of QTLs with less than 1.0 cM. Of these, 17 genes with 21 common differential SNVs in exon, UTR, upstream and downstream were identified as promising candidates affecting milk protein and fat traits. They included UDP-N-acetylglucosaminyltransferase subunit (*ALG14*), ATPase secretory pathway Ca²⁺ transporting 1 (*ATP2C1*), phosphatidylcholine-specific (*PLD1*),

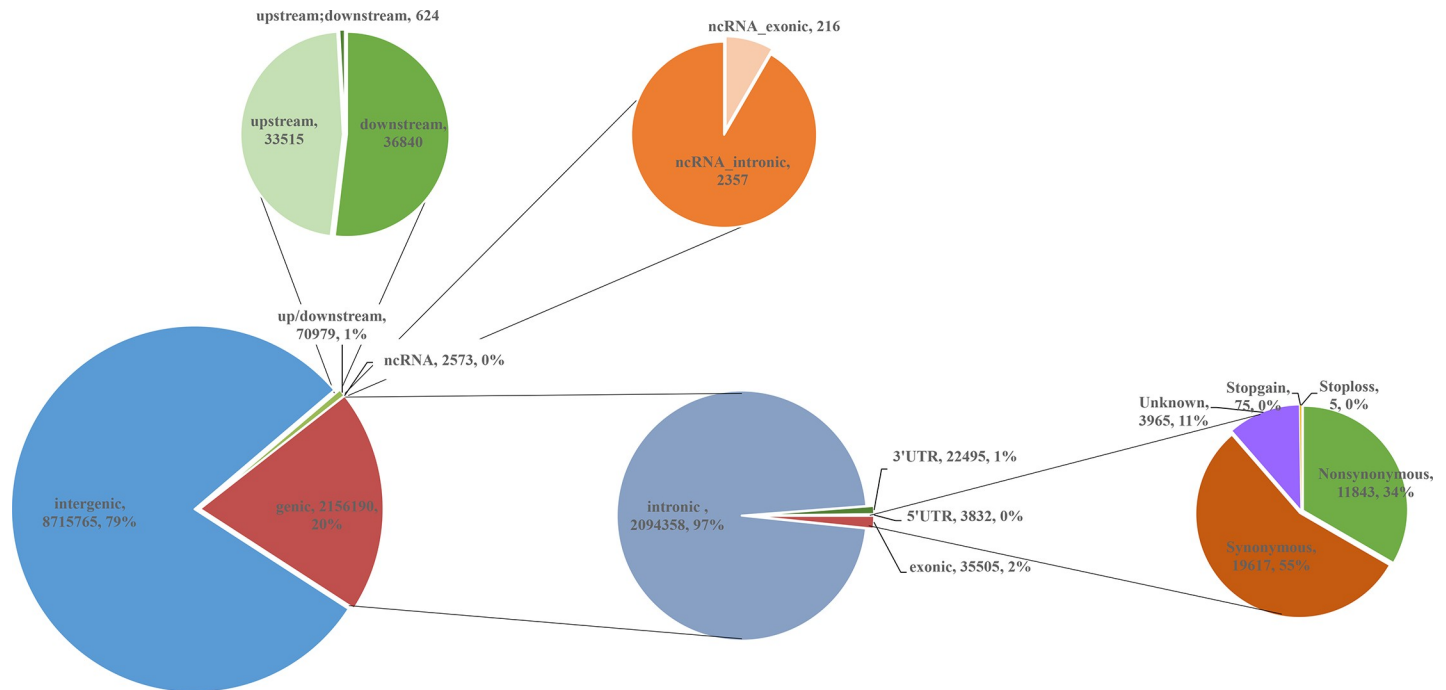


Fig 1. Distribution of the 10,945,507 annotated SNVs according to the functional category. After annotation of all SNVs among eight bulls, we found 8,715,765 intergenic SNVs (away from protein-coding genes more than 1 kb), 70,979 up/downstream SNVs, 2,573 ncRNA SNVs, 2,094,358 intronic SNVs, 26,327 untranslated regions (UTRs), 19,617 synonymous SNVs, 11,843 nonsynonymous substitutions, 75 stop gain SNVs, 5 stop loss SNVs and 3,965 unknown.

<https://doi.org/10.1371/journal.pone.0220629.g001>

glycosylated lysosomal membrane protein (*C3H1orf85*), sorting nexin 7 (*SNX7*), methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 2 like (*MTHFD2L*), cyclin dependent kinase inhibitor 2D (*CDKN2D*), alpha3 (V) collagen chain (*COL5A3*), ferredoxin 1-like (*FDX1L*), NIMA-interacting 1 (*PINI*), phosphoinositide 5-phosphatase (*FIG4*), exocyst complex component 7 (*EXOC7*), LIM and SH3 protein 1 (*LASP1*), phosphatidylglycerophosphate synthase 1 (*PGS1*), primary amine oxidase, liver isozyme (*SAO*), glycosylphosphatidylinositol specific phospholipase D1 (*GPLD1*) and OGA O-GlcNAcase (*MGEA5*). The alleles of the common differential SNV in high and low groups and the adjacent QTLs of 17 candidate genes were shown in Tables 5 and 6, respectively.

Discussion

In this study, based on the whole genome resequencing data of 8 proven Holstein bulls from the four half-sib or full-sib families with extremely high and low EBVs for milk protein and fat percentages, we obtained 57,419 common differential SNVs between high and low groups, and further identified 17 promising candidate genes for milk composition traits by integrating the positions of SNVs in gene regions, the known QTLs and the biological functions of genes.

Since the first report of SNV detection by the whole genome resequencing in cattle, a number of SNVs have been detected in different cattle breeds. In this study, a total of 10,961,243 SNVs were identified in 8 Holstein bulls (average 4,560,713 for each), which was much more than those in Holstein bulls reported by Paul et al. (SNP = 3,755,663) [21], but fewer than other Holstein bulls studies (SNP = 12,434,860 and 26.7 million, respectively) [23, 47]. This was probably due to the different sequencing depth and coverage.

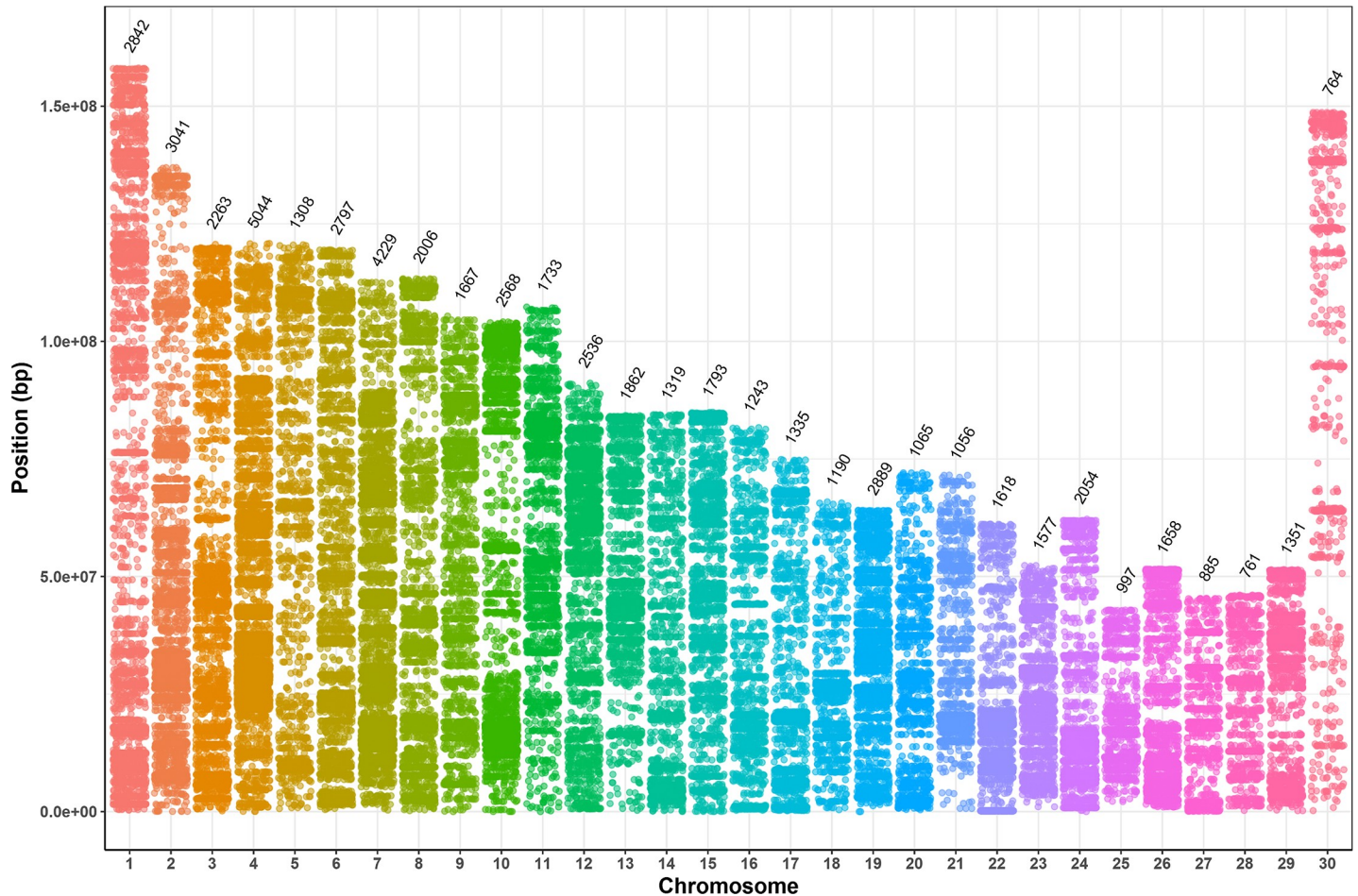


Fig 2. The number of common differential SNVs in each chromosome. Each point represents the location of a SNV on chromosome and the number above every chromosome represents the counts of SNV in this chromosome.

<https://doi.org/10.1371/journal.pone.0220629.g002>

Candidate genes

The 17 identified promising candidate genes that contained one or two common differential SNVs were specifically illustrated follow.

Candidate genes with SNV in exon. SNV in exon of a gene, especially nonsynonymous variants, potentially had a bigger influence on gene function. *COL5A3* with a nonsynonymous SNV in exon 2 encodes collagen type V alpha 3 belong to a superfamily of proteins. *COL5A3* takes part in protein digestion, absorption and PI3K-Akt signaling pathway. Previous study found that obese black women exhibited higher expression of *COL5A1* (collagen Valpha1), and *COL6A1* (collagen VIalpha1) than obese white women in gluteal [48]. In our previous RNA sequencing study among 3 milking period Holstein cattle, Collagen VI was found involved in regulating fat metabolism[49]. *COL5A3* is an important element of the microenvironment of certain highly specialized cell types in white adipose tissue and have profound effects on function of such cells [50]. Actually, nonsynonymous and synonymous coding SNPs show similar likelihood and effect size of traits [51–54]. *ATP2C1* with 2 synonymous SNVs encodes a protein belongs to the family of P-type cation transport ATPases which catalyzes the hydrolysis of ATP coupled with the transport of calcium ions. *ATP2C1* activity is associated with the sphingomyelin content of the trans-Golgi network membrane and it regulates

Table 3. Annotation of 57,419 common differential SNVs across 8 bulls.

Category	Number of SNVs	Percentage of SNVs % ¹
intergenic	45,188	78.70
upstream ^a	161	0.28
downstream ^b	193	0.34
upstream;downstream ^c	4	0.01
3'UTR	140	0.24
5'UTR	14	0.02
ncRNA_exonic ^d	2	0.003
intronic	11,541	20.10
exonic		0.31
nonsynonymous	45	
synonymous	116	
unknown	15	

¹Percentage was calculated based on annotated 57,419 common differential SNVs.

^aupstream from the nearest gene (<1kb).

^bdownstream away from the nearest gene (<1kb).

^cvariant located in both upstream and downstream regions for two different genes (<1kb).

^dnon-coding RNA expressed within exon of a gene

<https://doi.org/10.1371/journal.pone.0220629.t003>

proteases within the trans-Golgi network that require for virus glycoprotein maturation [55, 56]. The study of rat demonstrated that *ATP2C1* played a role in the control of beta-cell Ca (2+) homeostasis and insulin secretion [57]. In addition, Golgi Ca²⁺/H⁺ antiporter as a contributor to mammary Golgi calcium transport needs was related to the role of *ATP2C1* and *ATP2C2* [58]. *PLD1* encodes a phosphatidylcholine-specific phospholipase which catalyzes the hydrolysis of phosphatidylcholine in order to yield phosphatidic acid and choline. The deficiency of *PLD1* or *PLD2* activity promotes elevated free fatty acids (FFA) levels and are insulin as well as glucose intolerant [59]. Besides, *PLD1* regulates COPII vesicle transport from the endoplasmic reticulum (ER) to the Golgi apparatus by regulating Sec13/31 recruitment from the cytosol to the ER membrane during COPII vesicle formation [60]. *EXOC7* encodes a protein which is a component of the exocyst complex that plays a critical role in vesicular trafficking and the secretory pathway by targeting post-Golgi vesicles to the plasma membrane. *EXOC7* is a direct substrate of the extracellular signal-regulated kinases 1/2, their phosphorylation enhances the binding of *EXOC7* to other exocyst components and promotes the assembly of the exocyst complex [61, 62]. PIPKI γ and phosphatidyl inositol phosphate pools at nascent E-cadherin contacts cue *EXOC7* targeting and orient the tethering of exocyst-associated E-cadherin [62]. The protein encoded by *FIG4* belongs to the SAC domain-containing protein gene family. *FIG4* binds to hepatitis C virus and modulates particle formation in a cholesterol ester-related manner [63].

Candidate genes with SNV in regulatory regions. SNV in regulatory regions probably regulates the translation processes of a gene. *ALG14* with a SNV in 5'UTR is a member of the glycosyltransferase 1 family. The protein encoded by *ALG14* and *ALG13* are thought to be subunits of UDP-GlcNAc transferase, which catalyzes the first two committed steps in endoplasmic reticulum N-linked glycosylation. *ALG14* coordinate recruitment of catalytic *ALG7* and *ALG13* to the endoplasmic reticulum membrane for initiating lipid-linked oligosaccharide biosynthesis at the N- and C-termini and interacted formation of the active UDP-N-acetylglucosamine transferase complex at the C terminus mediates [64, 65]. *CDKN2D*, *LASP1* and *PIN1* respectively contained 1, 1 and 2 SNVs in 3'UTR. *CDKN2D* encoded a protein which is a

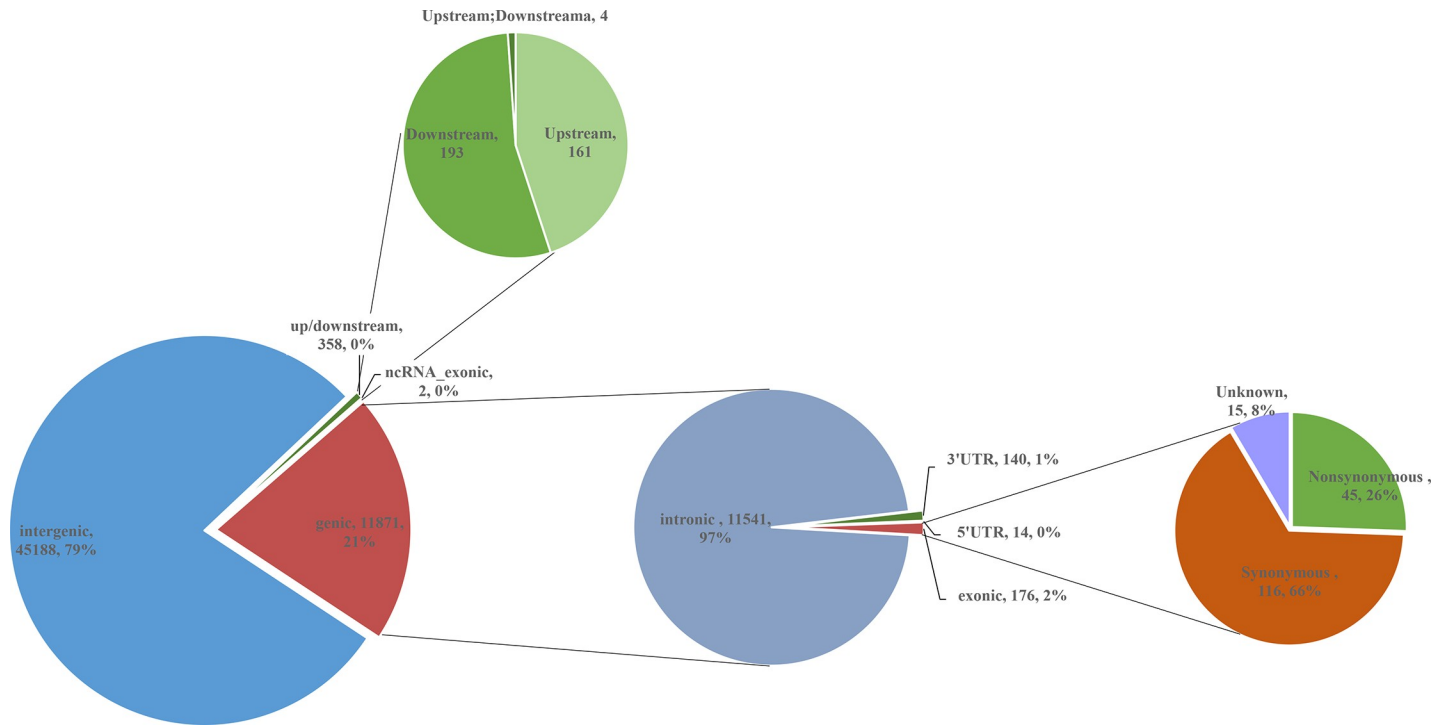


Fig 3. Distribution of the common differential SNVs in nine functional categories. After annotation, we found 45,188 intergenic SNVs (away from protein coding genes more than 1kb), 358 up/downstream SNVs, 2 ncRNA SNVs, 11,541 intronic SNVs, 154 untranslated regions (UTRs), 116 synonymous SNVs, 45 nonsynonymous substitutions and 15 unknown.

<https://doi.org/10.1371/journal.pone.0220629.g003>

member of the INK4 family of cyclin-dependent kinase inhibitors that form a stable complex with *CDK4* or *CDK6*, and prevent the activation of the CDK kinases, thus function as a cell growth regulator that controls cell cycle G1 progression. *FDX1L* encodes a member of the

Table 4. Annotation of 13,099 common differential SNVs that were included or nearby 2,657 protein-coding genes.

Category	Number of SNVs	Percentage of SNVs % ¹
intregenic	916	6.99
upstream ^a	161	1.23
downstream ^b	190	1.45
upstream;downstream ^c	4	0.03
3'UTR	140	0.11
5'UTR	14	1.07
intronic	11,498	87.78
exonic		1.34
nonsynonymous	45	
synonymous	116	
unknown	15	

¹Percentage was calculated based on 13,099 SNVs.

^aupstream from the nearest gene (<1kb).

^bdownstream away from the nearest gene (<1kb).

^cvariant located in both upstream and downstream regions for two different genes (<1kb).

<https://doi.org/10.1371/journal.pone.0220629.t004>

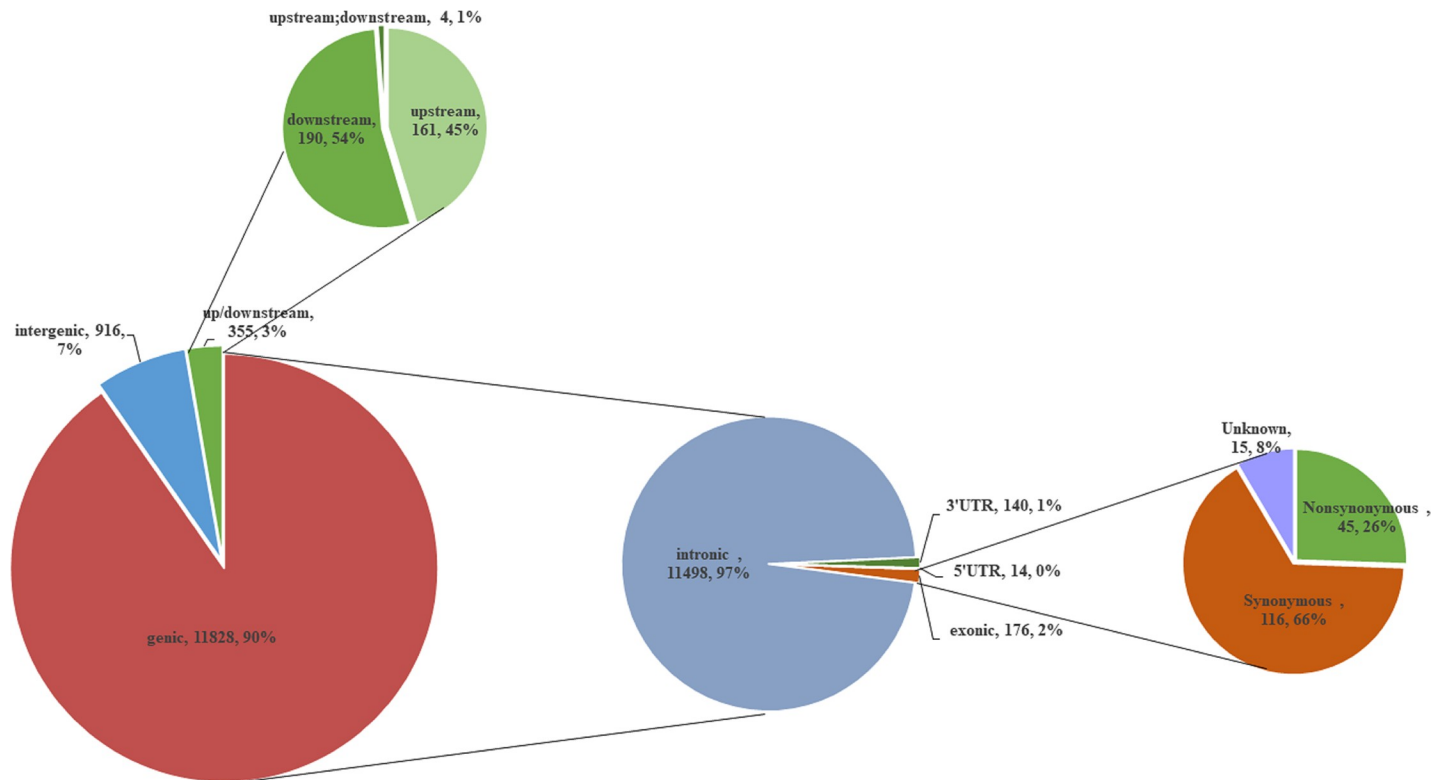


Fig 4. Distribution of the 13,099 SNVs in eight functional categories. A total of 13,099 SNVs was included or nearby the protein-coding genes with less than 5 kb. Of these, 916 SNVs were located in intergenic region (away from protein coding genes more than 1kb), 355 in up/downstream, 11,498 in intron, 154 untranslated regions (UTRs), 176 in exon (116 synonymous SNVs, 45 nonsynonymous substitutions and 15 unknown).

<https://doi.org/10.1371/journal.pone.0220629.g004>

ferredoxin family. The mutation of genes that encoded proteins involved in either the lipoic acid (*LIPT1* and *LIPT2*) or mitochondrial ISC biogenesis (*FDX1L*, *ISCA2*, *IBA57*, *NFU1*, *BOLA3*) pathway led a heterogeneous group of diseases with a wide variety of clinical symptoms and combined enzymatic defects [66]. The protein encoded by *LASPI* is a subfamily of LIM proteins and also a member of the nebulin family of actin-binding proteins. *LASPI* activates the PI3K/AKT signaling pathway which is well-known pathways for protein and fat synthesis and metabolism [67]. *LASPI* was significantly upregulated in breast cancer tissues and cell lines and identified as a target gene of miR-133a [68]. Comparing gene expression profiles of lactating bovine mammary tissue against nonlactating tissue on the BMAM microarray, *LASPI* exhibited differential expression [69]. *PINI* encodes one of the PPIases, which specifically binds to phosphorylated ser/thr-pro motifs to catalytically regulate the post-phosphorylation conformation of its substrates and involved in the regulation of cell growth. Besides, *PINI* can enhance adipocyte differentiation by regulating the function of PPAR gamma [70]. Another study suggested that *PINI* expression in pancreatic beta-cells was obviously changed in obese knockout mice from diet high in fat or sucrose [71].

Candidate genes with SNV in upstream and downstream. Transcription factors interact with specific nucleotide sequences known as transcription factor binding site and these interactions are implicated in regulation of the gene expression. The upstream and downstream regions of genes contain variety of elements/binding sites, which apparently infer on a particular gene the inducibility. *C3H1orf85*, *GPLD1*, *MTHFD2L*, *LASPI*, *MGEA5*, *PGS1*, *SAO*, *SNX7* contained at least one SNV located in upstream and downstream (<1000 bp) of these gene.

Table 5. The information about the SNV allele in high and low groups of 17 candidate genes.

Gene	Gene start ¹	Gene end	Chromosome	Location of SNV	Physical position of SNV	SNV allele in high group	SNV allele in low group
ALG14	48600669	48699217	1	UTR5	48600677	T	C
ATP2C1	140368052	140522627	1	exonic	140375966	A	G
			1	exonic	140388073	A	G
PLD1	96517508	96676253	1	exonic	96594836	C	T
C3H1orf85	14558546	14561400	3	upstream	14558465	T	C
SNX7	44657521	44775833	3	downstream	44656702	A	G
MTHFD2L	90842884	90986715	6	downstream	90987596	A	G
CDKN2D	16298106	16300588	7	UTR3	16298320	G	T
COL5A3	15768627	15813561	7	exonic	15769886	G	A
FDX1L	16068661	16073072	7	UTR3	16068671	C	G
PIN1	15532997	15545250	7	UTR3	15544947	C	T
			7	UTR3	15544948	C	T
FIG4	40873044	41046003	9	exonic	40957415	G	A
EXOC7	56190845	56208380	19	exonic	56206567	G	A
LASP1	40090994	40131374	19	UTR3	40129782	C	G
			19	downstream	40131920	A	G
PGS1	54404339	54441590	19	downstream	54404267	T	C
SAO	43555807	43559785	19	downstream	43560755	A	G
			19	downstream	43560756	A	C
GPLD1	32984455	33036038	23	upstream	32984009	C	T
MGEA5	22390769	22417775	26	upstream	22418458	A	G

¹genomic coordinates of genes were based on Bos_taurus_UMD_3.1

<https://doi.org/10.1371/journal.pone.0220629.t005>

C3H1orf85 encodes glycosylated lysosomal membrane protein and was also known as *GLMP*. Data indicated that increased flux of glucose, increased de novo lipogenesis and lipid accumulation were detected in lysosomal protein NCU-G1 (*GLMP*) gt/gt primary hepatocytes [72]. Compared with the wild-type myotubes, myotubes from *GLMP* (gt/gt) mice metabolized glucose faster and had a larger pool of intracellular glycogen, while oleic acid uptake, storage and oxidation were significantly reduced [73]. The nuclear proteins by O-linked N-acetylglucosamine (*MGEA5*) addition and removal on serine and threonine residues is catalyzed by OGT (MIM 300255), which adds *O-GlcNAc*, and *MGEA5*, a glycosidase that removes *O-GlcNAc* modifications [74]. *PGS1* encodes a phosphatidylglycero-phosphate synthase. In cancer cachexia, TNFalpha induces a higher energy wasting in liver mitochondria by increasing cardiolipin content via upregulation of phosphatidylglycerophosphate synthase (*PGPS*) expression [75]. *PGS1* gene in *Saccharomyces cerevisiae* played a vital role in cells impaired in the mitochondrial DNA, is localized in the mitochondria and expressed in response to inositol and choline [76]. The protein encoded by *SAO* is part of the anion exchanger (AE) family and is expressed in the erythrocyte plasma membrane. *MTHFD2L* encodes a mitochondrial methylenetetrahydrofolate dehydrogenase isozyme expressed in adult tissues. *SNX7* encodes a member of the sorting nexin family that contain a phox (PX) domain, which is a phosphoinositide binding domain, and are involved in intracellular trafficking. In zebrafish, *SNX7* is a liver-enriched anti-apoptotic protein and indispensable for the liver development [77]. The protein encoded by *GPLD1* is a GPI degrading enzyme. *GPLD1* hydrolyzes the inositol phosphate linkage in proteins anchored by phosphatidylinositol glycans, thereby releasing the attached protein from the plasma membrane. AMPK suppresses *PLD* activity, and *PLD* suppresses AMPK via mTOR [78]. Additionally, *GPLD1* influences triglyceride-rich lipoprotein metabolism [79].

Table 6. The detailed information of 17 candidate genes and the related QTLs.

Gene	Position(bp)	Position (cM) ¹	Previously reported QTL			
			Distance to QTL peak (cM)	CI and peak location(cM)	Trait	Reference
<i>ALG14</i>	Chr1:48600669–48699217	Chr3:48.4	0	48.37–48.37(peak:48.37)	PY	Marete et al., <i>Frontiers in genetics</i> , 2018[34]
<i>ATP2C1</i>	Chr1:140368052–140522627	Chr1:134.9	0.3	134.24–135.04 (peak:134.64)	PP	Russo et al., <i>Animal genetics</i> , 2012[35]
<i>PLD1</i>	Chr1:96517508–96676253	Chr1:93.7	0.2	77.7–122.3(peak:93.5)	PY	Nadesalingam et al., <i>Mammalian genome</i> ,2001 [36]
<i>C3H1orf85</i>	Chr3:14558546–14561400	Chr3:24.3	0.7	22.6–27.4(peak:25)	FY, PP	Ashwell et al., <i>J Dairy Sci</i> ,2004[37]
<i>SNX7</i>	Chr3:44657521–44775833	Chr3:46.7	0.8	45.93–45.93(peak:45.95)	FY	Cole JB et al., <i>BMC Genomics</i> , 2011[38]
				45.93–45.93(peak:45.95)	PY	Cole JB et al., <i>BMC Genomics</i> , 2011[38]
<i>MTHFD2L</i>	Chr6:90842884–90986715	Chr6:100	0.9	100.09–100.09 (peak:100.9)	PP	Olsen HG et al., <i>Genetics, selection, evolution : GSE</i> , 2016[39]
			0.4	100.28–100.49 (peak:100.38)	PP	Zhou Y et al., <i>BMC genomics</i> , 2018[40]
<i>CDKN2D</i>	Chr7:16298106–16300588	Chr7:18.4	0.6	15.2–38.5(peak:17.8,15.9)	PP	Ron et al., <i>Journal of dairy science</i> , 2004[41]
<i>COL5A3</i>	Chr7:15768627–15813561	Chr7:17.9	0.1	15.2–38.5(peak:17.8,15.9)	PP	Ron et al., <i>Journal of dairy science</i> , 2004[41]
<i>FDX1L</i>	Chr7:16068661–16073072	Chr7:18.2	0.4	15.2–38.5(peak:15.9,17.8)	PP	Ron et al., <i>Journal of dairy science</i> ,2004[41]
<i>PIN1</i>	Chr7:15532997–15545250	Chr7:17.5	0.3	15.2–38.5(peak:17.8,15.9)	PP	Ron et al., <i>Journal of dairy science</i> ,2004[41]
<i>FIG4</i>	Chr9:40873044–41046003	Chr9:45	0.6	42.5–50(peak:44.4,49.1)	FY, PY	Schnabel et al., <i>Animal Genetics</i> , 2005[42]
<i>EXOC7</i>	Chr19:56190845–56208380	Chr19:96.1	0.9	82.8–101.4(peak:95.2)	PY	Boichard et al., <i>Genetics,selection,evolution:GSE</i> , 2003[43]
<i>LASPI</i>	Chr19:40090994–40131374	Chr19:51.3	0.7	2.4–91(peak:60.4)	FP	Bennewitz et al., <i>Genetics</i> , 2004[44]
<i>PGS1</i>	Chr19:54404339–54441590	Chr19:93.1	0.7	60.7–106.2(PEAK:92.4)	FY	Boichard et al., <i>Genetics,selection,evolution:GSE</i> , 2003[43]
<i>SAO</i>	Chr19:43555807–43559785	Chr19:74.5	0.5	70.24–77.386 (peak:75.0)	FP	Viitala SM et al., <i>J Dairy Sci</i> , 2003[45]
<i>GPLD1</i>	Chr23:32984455–33036038	Chr23:44.2	0.6	42.4–58.2(peak:43.6)	FY	Plante et al., <i>J Dairy Sci</i> ,2001 [46]
<i>MGEA5</i>	Chr26:22390769–22417775	Chr26:32.5	0.8	31.72–31.72(peak:31.72)	PP	Cole JB, et al, <i>BMC Genomics</i> , 2011[38]

¹The linkage position was estimated relative to UMD3.1 and based on the QTL mapper v.2.019 at www.animalgenome.org/cgi-bin/QTLdb/.

PP: protein percentage; PY: protein yield; FP: fat percentage; FY: fat yield.

<https://doi.org/10.1371/journal.pone.0220629.t006>

Overexpressing *GPLD1* in an insulinoma cell line enhanced glucose-stimulated insulin secretion [79].

The 17 candidate genes and 21 SNVs identified in this study still need further in vivo and in vitro experiments to validate their biological function and to explore molecular mechanisms for formation of milk protein and fat traits.

The interpretation of the findings from the present study still has limitations. When performed function enrichment for genes that included or were closed to the common differential SNVs with less than 5 kb, non-coding RNAs and genes could be disregarded because the current software and tools can only annotate limited protein-coding genes. Therefore, the omission of genes that haven not been studied yet is a general problem in present function study.

Conclusions

In this study, by resequencing the whole genome of eight proven Holstein bulls with extremely high and low EBVs of milk protein percentage and fat percentage, we successfully identified 10,961,243 SNVs and detected 57,451 common differential SNVs with opposite fixed sites between high and low groups. Subsequently, 2,657 genes that included or were nearby the common differential SNVs were obtained. Further, through integrating GO, KEGG pathways and Mesh enrichment results, the known QTLs for milk composition and common differential SNVs located in exon and flanking regions, we identified 17 promising candidate genes for milk protein and fat, including *ALG14*, *ATP2C1*, *PLD1*, *C3H1orf85*, *SNX7*, *MTHFD2L*, *CDKN2D*, *COL5A3*, *FDX1L*, *PIN1*, *FIG4*, *EXOC7*, *LASPI*, *PGS1*, *SAO*, *GPLD1* and *MGEA5*. And the 17 genes identified in this study will provide a useful resource for future genomic selection (GS) in dairy cattle.

Ethics statement

All protocols for semen samples of China Holstein bulls were reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) at China Agricultural University. Semen samples were collected specifically for this study following standard procedures with the full agreement of the Beijing Dairy Cattle Center who owned the animals.

Supporting information

S1 Table. Functional annotation of each identified SNVs.
(ZIP)

S2 Table. GO and KEGG pathways enrichment results for 2,657 genes by KOBAS tool.
(XLSX)

S3 Table. Mesh enrichment results for 2,657 genes by MeSH ORA.
(XLSX)

Author Contributions

Conceptualization: Shengli Zhang, Dongxiao Sun.

Data curation: Shan Lin, Lin Liu.

Formal analysis: Shan Lin, Lin Liu.

Funding acquisition: Dongxiao Sun.

Investigation: Jianping Jiang.

Methodology: Shan Lin, Yali Hou, Jianping Jiang.

Project administration: Shengli Zhang, Dongxiao Sun.

Resources: Lin Liu, Wenhui Li, Bo Han.

Software: Shan Lin, Hongyan Zhang, Yali Hou, Wenhui Li.

Validation: Shan Lin, Jianping Jiang, Bo Han.

Visualization: Hongyan Zhang.

Writing – original draft: Shan Lin.

Writing – review & editing: Dongxiao Sun.

References

1. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001; 157(4):1819–29. WOS:000168223400036. PMID: [11290733](#)
2. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*. 2009; 4(4):e5350. <https://doi.org/10.1371/journal.pone.0005350> PMID: [19390634](#); PubMed Central PMCID: PMCPMC2669730.
3. Zhang Z, Ober U, Erbe M, Zhang H, Gao N, He JL, et al. Improving the Accuracy of Whole Genome Prediction for Complex Traits Using the Results of Genome Wide Association Studies. *Plos One*. 2014; 9(3). ARTN e93017 <https://doi.org/10.1371/journal.pone.0093017> WOS:000333459900143. PMID: [24663104](#)
4. Andersson L. Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica*. 2009; 136(2):341–9. <https://doi.org/10.1007/s10709-008-9312-4> PMID: [18704695](#).
5. Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, et al. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*. 1995; 139(2):907–20. PMID: [7713441](#); PubMed Central PMCID: PMCPMC1206390.
6. Jiang L, Liu JF, Sun DX, Ma PP, Ding XD, Yu Y, et al. Genome Wide Association Studies for Milk Production Traits in Chinese Holstein Population. *Plos One*. 2010; 5(10). ARTN e13661 <https://doi.org/10.1371/journal.pone.0013661> WOS:000283537000030. PMID: [21048968](#)
7. Schopen GCB, Visker MHPW, Koks PD, Mullaart E, van Arendonk JAM, Bovenhuis H. Whole-genome association study for milk protein composition in dairy cattle. *Journal of Dairy Science*. 2011; 94(6):3148–58. <https://doi.org/10.3168/jds.2010-4030> WOS:000290777800049. PMID: [21605784](#)
8. Mai MD, Sahana G, Christiansen FB, Guldbrandtsen B. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *J Anim Sci*. 2010; 88(11):3522–8. <https://doi.org/10.2527/jas.2009-2713> PMID: [20656975](#).
9. Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, et al. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res*. 2002; 12(2):222–31. <https://doi.org/10.1101/gr.224202> PMID: [11827942](#).
10. Minozzi G, Nicolazzi EL, Stella A, Biffani S, Negrini R, Lazzari B, et al. Genome wide analysis of fertility and production traits in Italian Holstein cattle. *PLoS One*. 2013; 8(11):e80219. <https://doi.org/10.1371/journal.pone.0080219> PMID: [24265800](#); PubMed Central PMCID: PMCPMC3827211.
11. Blott S, Kim JJ, Moio S, Schmidt-Kuntzel A, Cornet A, Berzi P, et al. 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):253–66. WOS:000181039700023. PMID: [12586713](#)
12. Park ST, Kim J. Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. *International neurourology journal*. 2016; 20(Suppl 2):S76–83. <https://doi.org/10.5213/inj.1632742.371> PMID: [27915479](#); PubMed Central PMCID: PMCPMC5169091.
13. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977; 74(12):5463–7. <https://doi.org/10.1073/pnas.74.12.5463> PMID: [271968](#); PubMed Central PMCID: PMCPMC431765.
14. Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A*. 1977; 74(2):560–4. <https://doi.org/10.1073/pnas.74.2.560> PMID: [265521](#); PubMed Central PMCID: PMCPMC392330.
15. Clark AG HM, Bustamante CD, Williamson SH, Nielsen R. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res*. 2005; 15(11).
16. Shendure J, Ji H. Next-generation DNA sequencing. *Nature biotechnology*. 2008; 26(10):1135–45. <https://doi.org/10.1038/nbt1486> PMID: [18846087](#).
17. Miller JR, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. *Genomics*. 2010; 95(6):315–27. <https://doi.org/10.1016/j.ygeno.2010.03.001> PMID: [20211242](#); PubMed Central PMCID: PMCPMC2874646.
18. Nielsen R, Hubisz MJ, Clark AG. Reconstituting the frequency spectrum of ascertained single-nucleotide polymorphism data. *Genetics*. 2004; 168(4):2373–82. <https://doi.org/10.1534/genetics.104.031039> PMID: [15371362](#); PubMed Central PMCID: PMCPMC1448751.
19. Lu D MS, Sargolzaei M Kelly M, Vander Voort G, Caldwell T, Wang Z, Plastow G, and Moore S. Genome-wide association analyses for growth and feed efficiency traits in beef cattle. *J Anim Sci*. 2013;91.

20. Utsunomiya YT, O'Brien A.M.P., Sonstegard T.S., Van Tassell C.P., do Carmo A.S., Meszaros G., Soelkner J., and Garcia J.F. Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan method. *Plos One*. 2013;8.
21. Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y, Liao X, et al. Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC genomics*. 2011; 12:559. <https://doi.org/10.1186/1471-2164-12-559> PMID: 22085807; PubMed Central PMCID: PMC3229636.
22. Choi JW, Choi BH, Lee SH, Lee SS, Kim HC, Yu D, et al. Whole-Genome Resequencing Analysis of Hanwoo and Yanbian Cattle to Identify Genome-Wide SNPs and Signatures of Selection. *Mol Cells*. 2015; 38(5):466–73. <https://doi.org/10.14348/molcells.2015.0019> PMID: 26018558; PubMed Central PMCID: PMC4443289.
23. Larkin DM, Daetwyler HD, Hernandez AG, Wright CL, Hetrick LA, Boucek L, et al. Whole-genome resequencing of two elite sires for the detection of haplotypes under selection in dairy cattle. *P Natl Acad Sci USA*. 2012; 109(20):7693–8. <https://doi.org/10.1073/pnas.1114546109> WOS:000304369800034. PMID: 22529356
24. Kawahara-Miki R, Tsuda K, Shiwa Y, Arai-Kichise Y, Matsumoto T, Kanesaki Y, et al. Whole-genome resequencing shows numerous genes with nonsynonymous SNPs in the Japanese native cattle Kuchinoshima-Ushi. *BMC Genomics*. 2011; 12:103. <https://doi.org/10.1186/1471-2164-12-103> PMID: 21310019; PubMed Central PMCID: PMC3048544.
25. Gao YH, Jiang JP, Yang SH, Hou YL, Liu GE, Zhang SG, et al. CNV discovery for milk composition traits in dairy cattle using whole genome resequencing. *Bmc Genomics*. 2017;18. <https://doi.org/10.1186/s12864-016-3397-4> 26510.1186/s12864-017-3636-3. WOS:000397755800001.
26. Jiang JP, Gao YH, Hou YL, Li WH, Zhang SL, Zhang Q, et al. Whole-Genome Resequencing of Holstein Bulls for Indel Discovery and Identification of Genes Associated with Milk Composition Traits in Dairy Cattle. *Plos One*. 2016; 11(12). ARTN e0168946 <https://doi.org/10.1371/journal.pone.0168946> WOS:000391222000088. PMID: 28030618
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168; PubMed Central PMCID: PMC2705234.
28. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 2012; 7(2):e30619. <https://doi.org/10.1371/journal.pone.0030619> PMID: 22312429; PubMed Central PMCID: PMC3270013.
29. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352> PMID: 19505943; PubMed Central PMCID: PMC2723002.
30. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38(16):e164. <https://doi.org/10.1093/nar/gkq603> PMID: 20601685; PubMed Central PMCID: PMC2938201.
31. Xie C, Mao XZ, Huang JJ, Ding Y, Wu JM, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res*. 2011; 39:W316–W22. <https://doi.org/10.1093/nar/gkr483> WOS:000292325300051. PMID: 21715386
32. Morota G, Penagaricano F, Petersen JL, Ciobanu DC, Tsuyuzaki K, Nikaido I. An application of MeSH enrichment analysis in livestock. *Anim Genet*. 2015; 46(4):381–7. <https://doi.org/10.1111/age.12307> PMID: 26036323; PubMed Central PMCID: PMC45032990.
33. Morota G, Beissinger TM, Penagaricano F. MeSH-Informed Enrichment Analysis and MeSH-Guided Semantic Similarity Among Functional Terms and Gene Products in Chicken. *G3-Genes Genom Genet*. 2016; 6(8):2447–53. <https://doi.org/10.1534/g3.116.031096/-/DC1> WOS:000381282300019.
34. Marete AG, Gulbrandsen B, Lund MS, Fritz S, Sahana G, Boichard D. A Meta-Analysis Including Pre-selected Sequence Variants Associated With Seven Traits in Three French Dairy Cattle Populations. *Front Genet*. 2018; 9:522. <https://doi.org/10.3389/fgene.2018.00522> PMID: 30459810; PubMed Central PMCID: PMC6232291.
35. Russo V, Fontanesi L, Dolezal M, Lipkin E, Scotti E, Zambonelli P, et al. A whole genome scan for QTL affecting milk protein percentage in Italian Holstein cattle, applying selective milk DNA pooling and multiple marker mapping in a daughter design. *Animal Genetics*. 2012; 43:72–86. <https://doi.org/10.1111/j.1365-2052.2012.02353.x> WOS:000305788600009. PMID: 22742505
36. Nadesalingam J, Plante Y, Gibson JP. Detection of QTL for milk production on Chromosomes 1 and 6 of Holstein cattle. *Mamm Genome*. 2001; 12(1):27–31. PMID: 11178740.
37. Ashwell MS, Heyen DW, Sonstegard TS, Van Tassell CP, Da Y, VanRaden PM, et al. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J Dairy Sci*. 2004; 87(2):468–75. [https://doi.org/10.3168/jds.S0022-0302\(04\)73186-0](https://doi.org/10.3168/jds.S0022-0302(04)73186-0) PMID: 14762090.

38. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ Jr., Crooker BA, et al. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *Bmc Genomics*. 2011; 12:408. <https://doi.org/10.1186/1471-2164-12-408> PMID: 21831322; PubMed Central PMCID: PMCPMC3176260.
39. Olsen HG, Knutsen TM, Lewandowska-Sabat AM, Grove H, Nome T, Svendsen M, et al. Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production. *Genet Sel Evol*. 2016; 48(1):79. <https://doi.org/10.1186/s12711-016-0257-2> PMID: 27760518; PubMed Central PMCID: PMCPMC5072345.
40. Zhou Y, Connor EE, Wiggans GR, Lu Y, Tempelman RJ, Schroeder SG, et al. Genome-wide copy number variant analysis reveals variants associated with 10 diverse production traits in Holstein cattle. *Bmc Genomics*. 2018; 19(1):314. <https://doi.org/10.1186/s12864-018-4699-5> PMID: 29716533; PubMed Central PMCID: PMCPMC5930521.
41. Ron M, Feldmesser E, Golik M, Tager-Cohen I, Kliger D, Reiss V, et al. A complete genome scan of the Israeli Holstein population for quantitative trait loci by a daughter design. *Journal of Dairy Science*. 2004; 87(2):476–90. [https://doi.org/10.3168/jds.S0022-0302\(04\)73187-2](https://doi.org/10.3168/jds.S0022-0302(04)73187-2) WOS:000189076500025. PMID: 14762091
42. Schnabel RD, Sonstegard TS, Taylor JF, Ashwell MS. Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. *Anim Genet*. 2005; 36(5):408–16. <https://doi.org/10.1111/j.1365-2052.2005.01337.x> PMID: 16167984.
43. Boichard D, Grohs C, Bourgeois F, Cerqueira F, Faugeras R, Neau A, et al. Detection of genes influencing economic traits in three French dairy cattle breeds. *Genet Sel Evol*. 2003; 35(1):77–101. <https://doi.org/10.1186/1297-9686-35-1-77:2002037> PMID: 12605852; PubMed Central PMCID: PMCPMC2732691.
44. Bennewitz J, Reinsch N, Guiard V, Fritz S, Thomsen H, Looft C, et al. Multiple quantitative trait loci mapping with cofactors and application of alternative variants of the false discovery rate in an enlarged granddaughter design. *Genetics*. 2004; 168(2):1019–27. <https://doi.org/10.1534/genetics.104.030296> PMID: 15514072; PubMed Central PMCID: PMCPMC1448815.
45. Viitala SM, Schulman NF, de Koning DJ, Elo K, Kinos R, Virta A, et al. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *Journal of Dairy Science*. 2003; 86(5):1828–36. [https://doi.org/10.3168/jds.S0022-0302\(03\)73769-2](https://doi.org/10.3168/jds.S0022-0302(03)73769-2) WOS:000182985600032. PMID: 12778594
46. Plante Y, Gibson JP, Nadesalingam J, Mehrabani-Yeganeh H, Lefebvre S, Vandervoort G, et al. Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. *J Dairy Sci*. 2001; 84(6):1516–24. [https://doi.org/10.3168/jds.S0022-0302\(01\)70185-3](https://doi.org/10.3168/jds.S0022-0302(01)70185-3) PMID: 11417712.
47. Daetwyler HD, Capitan A, Pausch H, Stothard P, van Binsbergen R, Brondum RF, et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat Genet*. 2014; 46(8):858–65. <https://doi.org/10.1038/ng.3034> PMID: 25017103.
48. Kotze-Horstmann LM, Keswell D, Adams K, Dlamini T, Goedecke JH. Hypoxia and extra-cellular matrix gene expression in adipose tissue associates with reduced insulin sensitivity in black South African women. *Endocrine*. 2017; 55(1):144–52. <https://doi.org/10.1007/s12020-016-1089-0> PMID: 27628582.
49. Liang RB, Han B, Li Q, Yuan YW, Li JG, Sun DX. Using RNA sequencing to identify putative competing endogenous RNAs (ceRNAs) potentially regulating fat metabolism in bovine liver. *Sci Rep-Uk*. 2017; 7. ARTN 6396 <https://doi.org/10.1038/s41598-017-06634-w> WOS:000406279200003. PMID: 28743867
50. Huang GR, Ge GX, Wang DY, Gopalakrishnan B, Butz DH, Colman RJ, et al. alpha 3(V) Collagen is critical for glucose homeostasis in mice due to effects in pancreatic islets and peripheral tissues. *J Clin Invest*. 2011; 121(2):769–83. <https://doi.org/10.1172/JCI45096> WOS:000286913800038. PMID: 21293061
51. Chen R, Davydov EV, Sirota M, Butte AJ. Non-Synonymous and Synonymous Coding SNPs Show Similar Likelihood and Effect Size of Human Disease Association. *Plos One*. 2010;5(10). ARTN e13574 <https://doi.org/10.1371/journal.pone.0013574> WOS:000283419100014. PMID: 21042586
52. Chamary JV, Parmley JL, Hurst LD. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat Rev Genet*. 2006; 7(2):98–108. <https://doi.org/10.1038/nrg1770> WOS:000234714000012. PMID: 16418745
53. Duret L. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev*. 2002; 12(6):640–9. [https://doi.org/10.1016/S0959-437x\(02\)00353-2](https://doi.org/10.1016/S0959-437x(02)00353-2) WOS:000179003900003. PMID: 12433576
54. Comeron JM. Selective and mutational patterns associated with gene expression in humans: Influences on synonymous composition and intron presence. *Genetics*. 2004; 167(3):1293–304. <https://doi.org/10.1534/genetics.104.026351> WOS:000223109300023. PMID: 15280243

55. Hoffmann HH, Schneider WM, Blumen VA, Scull MA, Hovnanian A, Brummelkamp TR, et al. Diverse Viruses Require the Calcium Transporter SPCA1 for Maturation and Spread. *Cell Host Microbe*. 2017; 22(4):460–+. <https://doi.org/10.1016/j.chom.2017.09.002> WOS:000412768900009. PMID: 29024641
56. Deng YQ, Pakdel M, Blank B, Sundberg EL, Burd CG, von Blume J. Activity of the SPCA1 Calcium Pump Couples Sphingomyelin Synthesis to Sorting of Secretory Proteins in the Trans-Golgi Network. *Dev Cell*. 2018; 47(4):464–+. <https://doi.org/10.1016/j.devcel.2018.10.012> WOS:000450620800012. PMID: 30393074
57. Mitchell KJ, Tsuboi T, Rutter GA. Role for plasma membrane-related Ca²⁺-ATPase-1 (ATP2C1) in pancreatic beta-cell Ca²⁺ homeostasis revealed by RNA silencing. *Diabetes*. 2004; 53(2):393–400. <https://doi.org/10.2337/diabetes.53.2.393> WOS:000188739600016. PMID: 14747290
58. Reinhardt TA, Lippolis JD, Sacco RE. The Ca²⁺/H⁺ antiporter TMEM165 expression, localization in the developing, lactating and involuting mammary gland parallels the secretory pathway Ca²⁺ ATPase (SPCA1). *Biochem Biophys Res Commun*. 2014; 445(2):417–21. <https://doi.org/10.1016/j.bbrc.2014.02.020> WOS:000333228700026. PMID: 24530912
59. Viera JT, El-Merahbi R, Nieswandt B, Stegner D, Sumara G. Phospholipases D1 and D2 Suppress Appetite and Protect against Overweight. *Plos One*. 2016; 11(6). ARTN e0157607 <https://doi.org/10.1371/journal.pone.0157607> WOS:000377822200066. PMID: 27299737
60. Nakagawa H, Hazama K, Ishida K, Komori M, Nishimura K, Matsuo S. Inhibition of PLD1 activity causes ER stress via regulation of COPII vesicle formation. *Biochem Biophys Res Commun*. 2017; 490(3):895–900. <https://doi.org/10.1016/j.bbrc.2017.06.137> WOS:000406820600047. PMID: 28648601
61. Ren JQ, Guo W. ERK1/2 Regulate Exocytosis through Direct Phosphorylation of the Exocyst Component Exo70. *Dev Cell*. 2012; 22(5):967–78. <https://doi.org/10.1016/j.devcel.2012.03.005> WOS:000304291700010. PMID: 22595671
62. Xiong XH, Xu QW, Huang Y, Singh RD, Anderson R, Leof E, et al. An association between type I gamma PI4P 5-kinase and Exo70 directs E-cadherin clustering and epithelial polarization. *Mol Biol Cell*. 2012; 23(1):87–98. <https://doi.org/10.1091/mbc.E11-05-0449> WOS:000298661200009. PMID: 22049025
63. Cottarel J, Plissonnier ML, Kullolli M, Pitteri S, Clement S, Millarte V, et al. FIG4 is a hepatitis C virus particle-bound protein implicated in virion morphogenesis and infectivity with cholesterol ester modulation potential. *J Gen Virol*. 2016; 97:69–81. <https://doi.org/10.1099/jgv.0.000331> WOS:000372062500008. PMID: 26519381
64. Lu JS, Takahashi T, Ohoka A, Nakajima K, Hashimoto R, Miura N, et al. Alg14 organizes the formation of a multiglycosyltransferase complex involved in initiation of lipid-linked oligosaccharide biosynthesis. *Glycobiology*. 2012; 22(4):504–16. <https://doi.org/10.1093/glycob/cwr162> WOS:000301005600006. PMID: 22061998
65. Gao XD, Moriyama S, Miura N, Dean N, Nishimura SI. Interaction between the C Termini of Alg13 and Alg14 Mediates Formation of the Active UDP-N-acetylglucosamine Transferase Complex. *J Biol Chem*. 2008; 283(47):32534–41. <https://doi.org/10.1074/jbc.M804060200> WOS:000260893700042. PMID: 18809682
66. Lebigot E, Gaignard P, Dorboz I, Slama A, Rio M, de Lonlay P, et al. Impact of mutations within the [Fe-S] cluster or the lipoic acid biosynthesis pathways on mitochondrial protein expression profiles in fibroblasts from patients. *Mol Genet Metab*. 2017; 122(3):85–94. <https://doi.org/10.1016/j.ymgme.2017.08.001> WOS:000418395100011. PMID: 28803783
67. Liu Y, Gao Y, Li DH, He LY, Lao IW, Hao B, et al. LASP1 promotes glioma cell proliferation and migration and is negatively regulated by miR-377-3p. *Biomed Pharmacother*. 2018; 108:845–51. <https://doi.org/10.1016/j.biopha.2018.09.068> WOS:000450101800097. PMID: 30372896
68. Sui YM, Zhang XL, Yang HL, Wei W, Wang ML. MicroRNA-133a acts as a tumour suppressor in breast cancer through targeting LASP1. *Oncol Rep*. 2018; 39(2):473–82. <https://doi.org/10.3892/or.2017.6114> WOS:000422855800003. PMID: 29207145
69. Suchyta SP, Sipkovsky S, Halgren RG, Kruska R, Elftman M, Weber-Nielsen M, et al. Bovine mammary gene expression profiling using a cDNA microarray enhanced for mammary-specific transcripts. *Physiol Genomics*. 2003; 16(1):8–18. <https://doi.org/10.1152/physiolgenomics.00028.2003> WOS:000187327300002. PMID: 14559974
70. Han Y, Lee SH, Bahn M, Yeo CY, Lee KY. Pin1 enhances adipocyte differentiation by positively regulating the transcriptional activity of PPARgamma. *Mol Cell Endocrinol*. 2016; 436:150–8. <https://doi.org/10.1016/j.mce.2016.07.030> PMID: 27475846
71. Nakatsu Y, Mori K, Matsunaga Y, Yamamotoya T, Ueda K, Inoue Y, et al. The prolyl isomerase Pin1 increases beta-cell proliferation and enhances insulin secretion. *J Biol Chem*. 2017; 292(28):11886–95. <https://doi.org/10.1074/jbc.M117.780726> WOS:000405485600025. PMID: 28566287

72. Kong XY, Kase ET, Herskedal A, Schjalm C, Damme M, Nasset CK, et al. Lack of the Lysosomal Membrane Protein, GLMP, in Mice Results in Metabolic Dysregulation in Liver. *Plos One*. 2015; 10(6). UNSP e0129402 <https://doi.org/10.1371/journal.pone.0129402> WOS:000355652200160. PMID: 26047317
73. Kong XY, Feng YZ, Eftestol E, Kase ET, Haugum H, Eskild W, et al. Increased glucose utilization and decreased fatty acid metabolism in myotubes from *Glmp(gt/gt)* mice. *Arch Physiol Biochem*. 2016; 122(1):36–45. <https://doi.org/10.3109/13813455.2015.1120752> WOS:000370643400006. PMID: 26707125
74. Gao Y, Wells L, Comer FI, Parker GJ, Hart GW. Dynamic O-glycosylation of nuclear and cytosolic proteins—Cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *J Biol Chem*. 2001; 276(13):9838–45. <https://doi.org/10.1074/jbc.M010420200> WOS:000167996400036. PMID: 11148210
75. Peyta L, Jarnouen K, Pinault M, Coulouarn C, Guimaraes C, Goupille C, et al. Regulation of hepatic cardiolipin metabolism by TNFalpha: Implication in cancer cachexia. *Biochim Biophys Acta*. 2015; 1851(11):1490–500. <https://doi.org/10.1016/j.bbali.2015.08.008> PMID: 26327596.
76. Dzugasova V, Obernauerova M, Horvathova K, Vachova M, Zakova M, Subik J. Phosphatidylglycerol-phosphate synthase encoded by the *PEL1/PGS1* gene in *Saccharomyces cerevisiae* is localized in mitochondria and its expression is regulated by phospholipid precursors. *Curr Genet*. 1998; 34(4):297–302. <https://doi.org/10.1007/s002940050399> WOS:000077150600007. PMID: 9799363
77. Feng Z, Xu T, Xu J. [Expression, purification and phosphoinositide binding specificity of recombinant human SNX7 expressed in *Escherichia coli*]. *Sheng Wu Gong Cheng Xue Bao*. 2014; 30(9):1436–45. PMID: 25720158.
78. Mukhopadhyay S, Saqcena M, Chatterjee A, Garcia A, Frias MA, Foster DA. Reciprocal regulation of AMP-activated protein kinase and phospholipase D. *J Biol Chem*. 2015; 290(11):6986–93. <https://doi.org/10.1074/jbc.M114.622571> PMID: 25632961; PubMed Central PMCID: PMC4358122.
79. Raikwar NS, Cho WK, Bowen RF, Deeg MA. Glycosylphosphatidylinositol-specific phospholipase D influences triglyceride-rich lipoprotein metabolism. *Am J Physiol-Endoc M*. 2006; 290(3):463–E70. <https://doi.org/10.1152/ajpendo.00593.2004> WOS:000235210000009. PMID: 16219662