

RESEARCH ARTICLE

# Whole-Genome Resequencing of Holstein Bulls for Indel Discovery and Identification of Genes Associated with Milk Composition Traits in Dairy Cattle

Jianping Jiang<sup>1</sup>, Yahui Gao<sup>1</sup>, Yali Hou<sup>2</sup>, Wenhui Li<sup>1</sup>, Shengli Zhang<sup>1</sup>, Qin Zhang<sup>1</sup>, Dongxiao Sun<sup>1\*</sup>

**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

\* [sundx@cau.edu.cn](mailto:sundx@cau.edu.cn)



**OPEN ACCESS**

**Citation:** Jiang J, Gao Y, Hou Y, Li W, Zhang S, Zhang Q, et al. (2016) Whole-Genome Resequencing of Holstein Bulls for Indel Discovery and Identification of Genes Associated with Milk Composition Traits in Dairy Cattle. PLoS ONE 11(12): e0168946. doi:10.1371/journal.pone.0168946

**Editor:** Sudhindra R. Gadagkar, Midwestern University, UNITED STATES

**Received:** August 12, 2016

**Accepted:** December 8, 2016

**Published:** December 28, 2016

**Copyright:** © 2016 Jiang 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 paper and its Supporting Information files.

**Funding:** This work was financially supported by the National Science and Technology Programs of China (2013AA102504, 2011BAD28B02, 2014ZX08009-053B), Beijing Natural Science Foundation (6152013), the Beijing Dairy Industry Innovation Team (BAIC06-2016), the earmarked fund for Modern Agro-industry Technology

## Abstract

The use of whole-genome resequencing to obtain more information on genetic variation could produce a range of benefits for the dairy cattle industry, especially with regard to increasing milk production and improving milk composition. In this study, we sequenced the genomes of eight Holstein bulls from four half- or full-sib families, with high and low estimated breeding values (EBVs) of milk protein percentage and fat percentage at an average effective depth of 10×, using Illumina sequencing. Over 0.9 million nonredundant short insertions and deletions (indels) [1–49 base pairs (bp)] were obtained. Among them, 3,625 indels that were polymorphic between the high and low groups of bulls were revealed and subjected to further analysis. The vast majority (76.67%) of these indels were novel. Follow-up validation assays confirmed that most (70%) of the randomly selected indels represented true variations. The indels that were polymorphic between the two groups were annotated based on the cattle genome sequence assembly (UMD3.1.69); as a result, nearly 1,137 of them were found to be located within 767 annotated genes, only 5 (0.138%) of which were located in exons. Then, by integrated analysis of the 767 genes with known quantitative trait loci (QTL); significant single-nucleotide polymorphisms (SNPs) previously identified by genome-wide association studies (GWAS) to be associated with bovine milk protein and fat traits; and the well-known pathways involved in protein, fat synthesis, and metabolism, we identified a total of 11 promising candidate genes potentially affecting milk composition traits. These were *FCGR2B*, *CENPE*, *RETSAT*, *ACSBG2*, *NFKB2*, *TBC1D1*, *NLK*, *MAP3K1*, *SLC30A2*, *ANGPT1* and *UGDH*. Our findings provide a basis for further study and reveal key genes for milk composition traits in dairy cattle.

Research System (CARS-37) and the Program for Changjiang Scholar and Innovation Research Team in University (IRT1191). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

## Introduction

In recent years, the establishment of next-generation sequencing (NGS) technology has led to substantial progress in terms of the discovery of large numbers of single-nucleotide polymorphisms (SNPs), insertions and deletions (indels), and large structural variations (SVs). Short indels are now recognized as the second most common form of genomic variation [1], which contribute to phenotypic diversity in many species and human diseases [2], such as cystic fibrosis [3] and fragile X syndrome [4]. In cattle, indels have been shown to contribute to complex traits, such as the double-musled phenotype [5]. Since Mills et al. constructed the initial map of human indel variations [6], and with the increased availability and advances of high-throughput sequencing technology, indels have been discovered in individual genomes across different species [1, 7–12]. At the time of writing, more than 8 million indels are registered in the dbSNP bovine database ([ftp://ftp.ncbi.nih.gov/snp/organisms/cow\\_9913/VCF/](ftp://ftp.ncbi.nih.gov/snp/organisms/cow_9913/VCF/), updated in July 11, 2016). Therefore, it is now feasible to identify and characterize the molecular basis of complex traits using indels.

Milk production in dairy cattle is an economically important variable, especially the traits of milk yield and milk composition. Numerous studies have focused on the detection of quantitative trait loci (QTL) for milk production traits [13–19], since the first report on QTL mapping in dairy cattle by Georges et al. [20]. In recent decades, based on QTL mapping and genome-wide association studies (GWASs), numerous candidate genes and mutations associated with milk production traits have been identified, including two confirmed causal mutations: *DGAT1* p.Lys232Ala and *GHR* p.Phe279Tyr, which have large effects on milk yield and composition [21, 22]. However, to date, only limited investigations to identify genes associated with milk production traits using whole-genome resequencing have been reported [23–25].

The aims of this study were thus to detect indels across the whole genome in eight Holstein bulls with extremely high and low estimated breeding values (EBVs) of milk protein percentage and fat percentage using NGS, and to explore the short indels [1–49 base pairs (bp)] that were polymorphic between these two groups and to identify functional genes that are important for milk protein and fat traits.

## Materials and Methods

Frozen semen samples were collected along with the regular quarantine inspection of the farms and in strict accordance with the protocol approved by the Institutional Animal Care and Use Committee (IACUC) at China Agricultural University.

### Sample collection

Eight bulls from two half-sib families and two full-sib families ( $n = 2$  each) were selected from the Beijing Dairy Cattle Center (<http://www.bdcc.com.cn/>), according to their EBVs of milk protein percentage (PP) and fat percentage (FP), which were calculated based on a multiple trait random regression test-day model using the software RUNGE by the Dairy Data Center of China (<http://www.holstein.org.cn/>). The reliability of EBVs for milk protein percentage and fat percentage of each bull was more than 0.80. The two bulls in each group showed high and low EBVs for milk protein percentage and fat percentage. Detailed information on the eight bulls is presented in [Table 1](#).

### DNA library construction and sequencing

Genomic DNA was extracted from frozen semen samples using the standard phenol/chloroform extraction method. DNA degradation and contamination were monitored on 1% agarose

**Table 1. Descriptive statistics of PP and FP EBVs for eight bulls.**

Sib-family	Sample	EBV for PP	EBV for FP	Reliability for PP	Reliability for FP
Full-sib1	1	0.03	0.10	0.99	0.99
	2	-0.13	-0.31	0.97	0.87
Full-sib2	3	0.08	0.56	0.99	0.98
	4	-0.03	0.27	0.98	0.98
Half-sib1	5	0.22	0.09	0.91	0.8
	6	0.01	-0.26	0.99	0.99
Half-sib2	7	0.07	-0.14	0.98	0.99
	8	-0.06	-0.26	0.99	0.99

doi:10.1371/journal.pone.0168946.t001

gels and the concentration and purity were assessed on NanoDrop 2000 (Thermo Scientific Inc. Waltham, DE, USA); the high-quality DNAs were then used for library construction. Two paired-end libraries were constructed for each individual, the read length was 2×100 bp, and sequencing was performed using Illumina HiSeq2000 instruments (Illumina Inc., San Diego, CA, USA).

### Read mapping and variant calling

The cattle reference genome assembly (UMD3.1.69) was downloaded from the Ensembl Genome Browser website (<http://asia.ensembl.org>). We applied the NGS QC Toolkit [26] with default parameters to minimize mapping errors. To map reads to the reference genome, we used the Burrows–Wheeler Alignment tool (BWA ver. 0.6.2) [27], mainly with the default parameters. To call indels, SAMtools (ver. 0.1.19) [28] was applied. The criteria were as follows: overall base quality score ≥20; read depth <100 for each individual; and alternative allele on either forward or reverse supporting reads >3.

### Indel validation

To evaluate the reliability of the resequencing data, 20 randomly selected indels were validated using a PCR assay and direct sequencing. PCR primers were designed using PRIMER3 (<http://frodo.wi.mit.edu/primer3>) and Oligo 6.0 to amplify the 150–300 bp fragments containing the indels (S1 Table). The PCR reaction was performed in a volume of 25 µl containing 50–100 ng of genomic DNA, 10 µl of premix, 1 µl of each primer, and 11 µl of PCR buffer, using the following PCR program: 94°C for 10 min; 35 cycles of 94°C for 40 s, 55–60°C for 40 s, and 72°C for 40 s; and then 72°C for 7 min at the end of the final cycle. The PCR products were purified and then subjected to Sanger sequencing.

### Detection of differential variants and functional annotation

Among all of the indels from the eight bulls retained after filtering, we selected those that were polymorphic between the two groups with high and low EBVs, which were called the ‘common differential variants’. The pipeline in this study was as follows: we first collected all the variants that were polymorphic between the two individuals within each family, and then retained the common differential variants with the same allelic distribution directions between the high and low individuals across the four families. Subsequently, these common differential indels were annotated by ANNOVAR [29] based on the cattle gene set downloaded from the Ensembl website ([ftp://ftp.ensembl.org/pub/release-79/gtf/bos\\_taurus](ftp://ftp.ensembl.org/pub/release-79/gtf/bos_taurus)) to identify the indels within functional genes. Moreover, we compared the differential indels mentioned above with the variants in cattle SNP database (<http://www.ncbi.nlm.nih.gov>).

## Integrated analysis of resequencing, QTL, and GWAS data and known pathways

We specifically focused on those common differential indels between the high and low groups located in genic regions or flanking regions of genes to identify candidate genes and causal mutations related to milk composition traits. On the one hand, we estimated the positions of genes containing these indels using QTL mapper v.2.019 ([www.animalgenome.org/cgi-bin/QTLdb/](http://www.animalgenome.org/cgi-bin/QTLdb/)), and then compared the genetic positions of the genes with the confidence intervals and peak positions of previously reported QTL (<http://www.animalgenome.org/cgi-bin/QTLdb/>, updated in April 29, 2016). We also compared the physical chromosomal positions of the above genes with the SNPs determined by previous GWASs [13, 15] to be significantly associated with milk protein and fat traits of dairy cattle. Then, we performed Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the genes containing the identified indels using KOBAS tool [30]. During the functional annotation of genes, a *P* value of <0.05 determined by Fisher's exact test was set as the criteria for significance. In addition, we downloaded the genes involved in well-known pathways for protein, fat, and fatty acid metabolism from the KEGG pathway website (<http://www.kegg.jp/>), consisting the mTOR, insulin, AMPK, PPAR, Jak-STAT, PI3K-Akt, MAPK, and TGF-β signaling pathways, and further determined whether these genes were included in the gene list which were identified by integrated analysis of QTL and GWAS data.

## Results

### Data production and read mapping

After resequencing of the genomic DNA samples for eight Holstein bulls, a total of 2,303,781,448 short reads were obtained, and the average proportion of uniquely mapped reads was as high as 82.62%. The sequencing depth ranged from 7× to 10×, with an average of 8.1×, and the genome coverage was approximately 98% in each individual (Table 2).

### Indel detection

In total, 912,302 nonredundant short indels (1–49 bp) were obtained from the eight bulls after filtering using SAMtools. The number of indels detected in each bull varied from 323,490 to 411,070, with an average of 365,169 (Table 2). Among them, we focused on the unique indels that were polymorphic between the bulls with high and low EBVs of milk protein percentage and fat percentage, which were inferred to be relevant to the milk protein and fat trait. As a result, after filtering, we obtained 3,625 common differential indels that were polymorphic

**Table 2. Summary of sequencing, mapping statistics, and indel count for individuals.**

Sample	Raw reads	Mapped reads	Uniquely mapped reads (%)	Genome coverage (%)	Sequencing depth (X)	Indel
1	289,952,310	261,783,075	83.75	98.58	8	363,757
2	286,870,238	252,201,294	81.15	98.55	8	346,453
3	292,878,886	257,840,281	81.69	98.59	8	366,402
4	272,948,496	241,748,531	82.97	98.52	8	360,827
5	251,953,446	218,677,291	82.90	98.34	7	323,490
6	337,815,303	314,651,325	80.88	98.40	10	411,070
7	288,003,254	253,311,627	82.45	98.61	8	368,232
8	283,359,516	255,124,411	85.16	98.57	8	381,119
Average	287,972,681	256,917,229	82.62	98.52	8.1	365,169

doi:10.1371/journal.pone.0168946.t002



**Fig 1. Distribution of allele frequencies for the common differential indels between high and low groups.** The first line represents insertion frequency for each common differential indel locus in low group, and the second line represents insertion frequency for each common differential indel locus in high group.

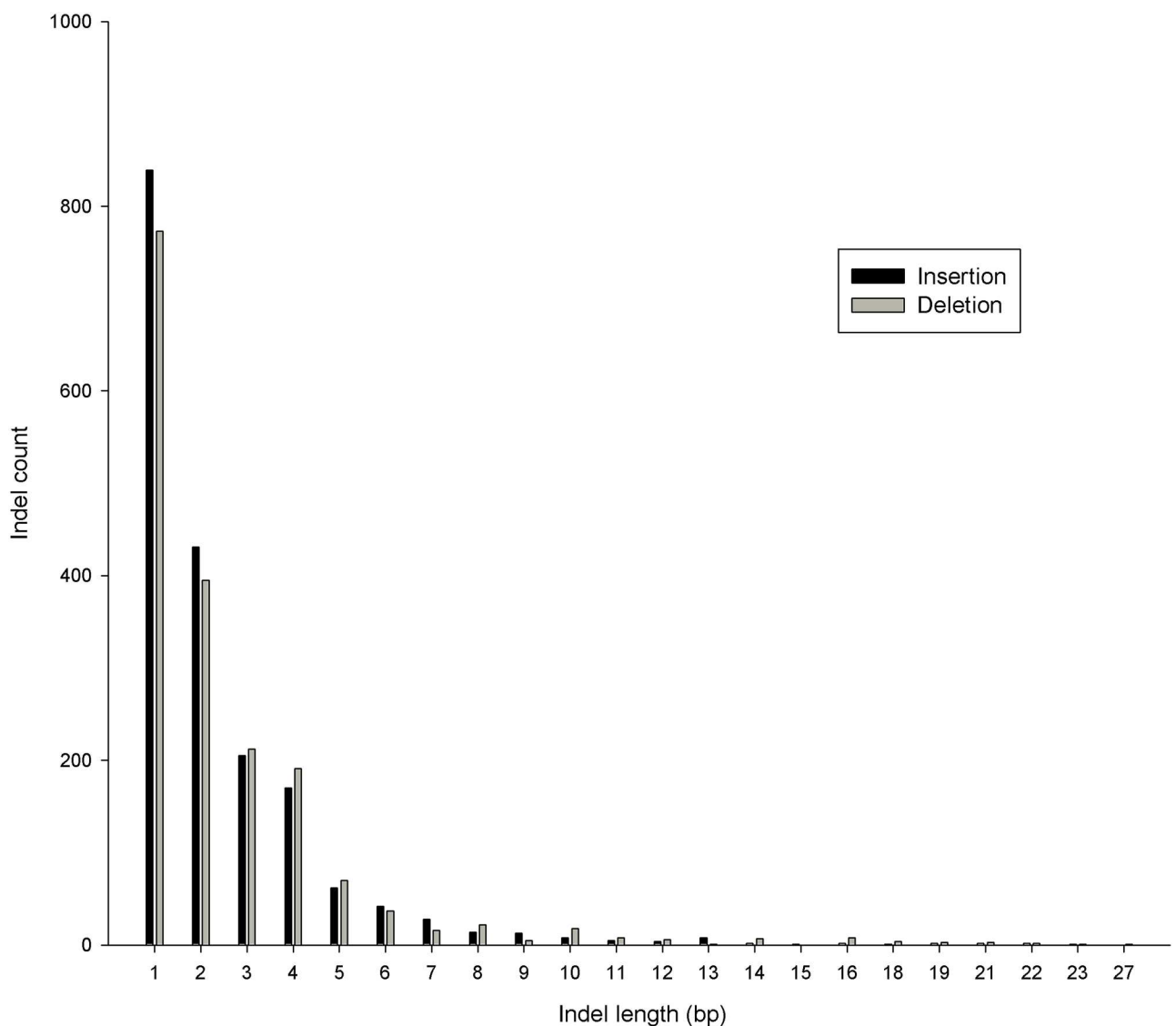
doi:10.1371/journal.pone.0168946.g001

between the high and low groups of bulls. The distribution of allele frequencies of these common differential indels was calculated and presented in Fig 1. We further compared our results with the variants in the SNP database (NCBI dbSNP, updated in July 11, 2016) and found that the majority of the common differential indels (76.67%) were novel.

To estimate the accuracy and reliability of our resequencing data, PCR amplification and Sanger sequencing were applied to validate 20 randomly selected indels with a range of lengths from 1 to 21 bp with the same genomic DNA samples of the eight bulls as used for resequencing. We found that the sequences of the 14 indels were consistent with those from our whole-genome resequencing data, indicating that the accuracy of our results was 70% (S1 Table). The primer sequences are shown in the supplemental material (S1 Table).

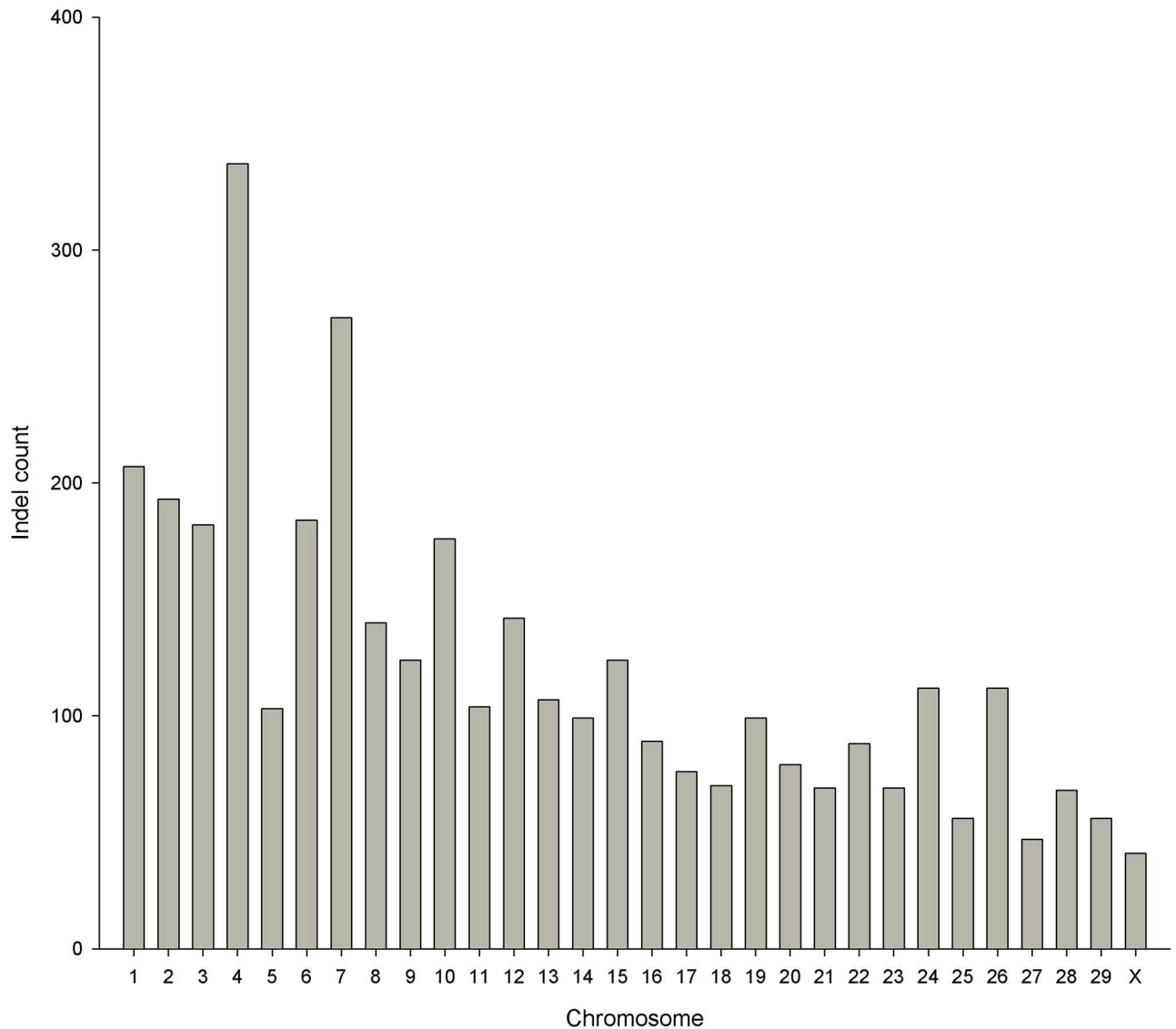
### Functional annotation and genomic distribution

Among the 3,625 common indels that were polymorphic between the high and low bull groups, the largest indel was 27 bp, but most (97.96%) of them were less than 10 bp. Single-bp indels (44.47%) were the most common, and the number of insertions (n = 1,842) was slightly more than the number of deletions (n = 1,783). The cumulative length of the indels reached 9 kb. The distribution of indel length is shown in Fig 2. The indels were proportionally



**Fig 2. Distribution of indel length (bp).** The indel represents the common differential one between the high and low groups.

doi:10.1371/journal.pone.0168946.g002



**Fig 3. The number of indels in each chromosome.** The indel represents the common differential one between the high and low groups.

doi:10.1371/journal.pone.0168946.g003

distributed among autosomes corresponding to chromosome length; the number of indels was lower on X chromosome than on autosomes (Fig 3).

Functional annotation of the 3,625 common differential indels was performed against known genes from the Ensembl website using ANNOVAR software (Table 3). As a result of this, the majority (68.634%) of indels was found to be located in intergenic regions, and the remaining 1,137 indels were within 767 genes, including in introns (29.379%), exons (0.138%), 3' untranslated region (3' UTR) (0.276%), and other regions (ncRNAs, splicing site, etc.). Regarding the exonic indels that were identified, there were 2 frameshift indels with lengths not divisible by 3, and 3 nonframeshift indels, called 3n indels [31], with lengths divisible by 3.

**Table 3. Annotation of 3,625 common differential indels for eight bulls.**

Category	Indel	Indel % <sup>1</sup>
All	3625	100
Intergenic	2488	68.634
Upstream	30	0.828
Downstream	24	0.662
Upstream;downstream <sup>a</sup>	1	0.028
3' UTR	10	0.276
Splicing	1	0.028
ncRNA_exonic	1	0.028
Intronic	1065	29.379
Exonic		0.138
Non-frameshift deletion	1	
Non-frameshift insertion	2	
Frameshift deletion	1	
Frameshift insertion	1	

<sup>1</sup>Percentage was calculated based on total annotated indels.

<sup>a</sup>Variant located in both upstream and downstream regions (possibly for two different genes).

doi:10.1371/journal.pone.0168946.t003

## Integrated analysis with known QTL and significant SNPs identified by GWASs

In this part of the study, we first compared the genetic positions of the 1,137 indels located in 767 genes with QTL known to be associated with milk protein and fat traits in the Cattle QTL database (<http://www.animalgenome.org/cgi-bin/QTLdb/BT/index>). We found that 869 indels, corresponding to 605 unique genes, overlapped with the confidence intervals of 538 known QTL. Notably, 542 of them were located in 372 functional genes that were close to the peak position of the known QTL with a genetic distance of less than 5 cM. Furthermore, of the 1,137 indels, 926 located in 635 genes were close (< 5 Mb) to significant SNPs identified by previous GWASs. Hence, a total of 695 of the genes in these two groups were retained for further analyses.

## Gene Ontology and pathway analyses

We performed GO enrichment and KEGG pathway analyses using the KOBAS tool based on the 695 genes mentioned above. In total, 306 GO terms and 29 KEGG pathways were significantly enriched ( $P < 0.05$ ; S2 Table), mainly including 'inorganic cation transmembrane transport', metabolic process related to protein and lipid metabolism such as 'lipid biosynthetic process', 'phosphatidylglycerol biosynthetic process', 'glycosaminoglycan biosynthetic process', 'protein kinase A binding', 'protein kinase C activity' and 'protein phosphatase inhibitor activity' *etc.*, and terms associated with the cell cycle and cell adhesion such as 'regulation of cell cycle G<sub>1</sub>/S phase transition' and 'protein complex involved in cell adhesion'. In addition, significant enriched KEGG pathways consisted of 'thyroid hormone synthesis', 'aldosterone synthesis and secretion', 'GnRH signaling pathway', MAPK signaling pathway, 'gap junction', and 'phosphatidylinositol signaling system' *et al.*

We also found that 34 out of the 695 genes with an indel polymorphic between the high and low groups of bulls were involved in several well-known pathways for protein and fat synthesis and metabolism. These included the mTOR and Jak-STAT signaling pathways, both of



which play pivotal roles in the synthesis of milk proteins; the AMPK signaling pathway as an energy supplement system; the insulin and PPAR signaling pathway for fat synthesis; the PI3K-Akt signaling pathway together with the MAPK signaling pathway, which play roles in cellular functions, such as transcription, translation, and cell proliferation; and the TGF- $\beta$  signaling pathway, which is essential for mammary epithelial cell proliferation.

### Identification of candidate genes associated with milk protein and fat traits

Based on the known QTL and significant SNPs identified by previous GWASs, along with biological functions of genes, 11 genes that containing at least one common differential indel locus were considered as promising candidates for an association with milk protein and fat traits. Of them, four genes, including Acyl-CoA synthetase bubblegum family member 2 (*ACSBG2*), nuclear factor kappa B subunit 2 (*NFKB2*), TBC1 domain family member 1 (*TBC1D1*) and mitogen-activated protein kinase kinase kinase 1 (*MAP3K1*), were located near to both the peak locations of the reported QTL and significant SNPs for milk protein and fat traits, and were involved in the well-known protein and lipid metabolic pathways. The Fc fragment of IgG receptor IIb (*FCGR2B*) and centromere protein E (*CENPE*) genes were located in QTL regions and near to the significant SNPs reported to be associated with milk protein and fat traits. Two genes, solute carrier family 30 member 2 (*SLC30A2*) and UDP-glucose 6-dehydrogenase (*UGDH*) were very close to the peak position of QTL and significant SNPs for milk protein and fat traits. Two genes, nemo like kinase (*NLK*) and angiotensin 1 (*ANGPT1*), were near to the significant SNPs for milk protein percentage, and participated in the MAPK and PI3K-Akt signaling pathways. In addition, retinol saturase (*RETSAT*) was also near to significant SNPs for milk protein and fat traits. The chromosomal positions of these 11 genes and detailed information about their nearest QTL and significant SNPs were shown in Tables A and B in [S1 File](#). The indel genotypes in these 11 genes between high and low groups were shown in [Table 4](#).

### Discussion

With the rapid development of sequencing technologies and bioinformatics tools, it is now feasible to sequence the complete genome of many species, such as human, mouse and *Arabidopsis* [32–34]. Compared with traditional technologies, NGS has numerous major advantages, including enabling sequencing that is cheaper and can be performed on a larger scale. Moreover, unlike GWASs, it can identify numerous rare and common variants as genetic causes of complex traits, especially in human Mendelian disorders [35, 36]. In this paper, we presented the indels in eight Holstein bulls determined using Illumina sequencing, with the aim of detecting the distribution of variants across the cattle genome and identifying the candidate genes for milk protein and fat traits. To our knowledge, this is one of the first studies involving whole-genome resequencing for milk protein and fat traits in dairy cattle; based on the indels that were polymorphic between the high and low groups, we identified a few candidate genes potentially associated with milk protein and fat traits.

### Indel discovery

Since the first report of the construction of a human indel map [6], numerous indels have been found in different species. In our research, we identified approximately 365,000 short indels for each bull, which was much more than for the Fleckvieh bull ( $n = 115,000$ ) [37], but slightly fewer than for the Japanese native cattle Kuchinoshima-Ushi ( $n = 629,256$ ) [38], Chikso (Korean brindle cattle) ( $n = 551,363$ ) [24], Korean brown cattle ( $n = 697,048$ ) [39], Jeju black

**Table 4. The indel genotypes in 11 genes between high and low groups.**

Gene name	Indel	Location	Indel sequence	Indel genotypes in high group				Indel genotypes in low group			
				sample1	sample3	sample5	sample7	sample2	sample4	sample6	sample8
<i>FCGR2B</i>	1N ins	exon	G	ins/del	del/del	ins/del	del/del	ins/ins	ins/del	ins/ins	ins/del
<i>CENPE</i>	3N ins	exon	AGA	del/del	del/del	del/del	del/del	ins/del	ins/del	ins/ins	ins/ins
	3N del	exon	TAG	ins/ins	ins/ins	del/ins	ins/ins	del/ins	del/ins	del/del	del/del
	3N ins	intron	GTT	del/del	del/del	del/del	ins/del	ins/del	ins/del	ins/ins	ins/ins
	1N ins	intron	T	del/del	del/del	ins/del	ins/del	ins/del	ins/del	ins/ins	ins/ins
	1N del	intron	A	ins/ins	ins/ins	ins/ins	del/ins	del/del	del/ins	del/del	del/del
	21N ins	intron	ACTTAAGTATATAACCTTAAC	del/del	del/del	del/del	del/del	ins/del	ins/del	ins/ins	ins/ins
	2N del	intron	CC	ins/ins	ins/ins	del/ins	del/ins	del/ins	del/ins	del/del	del/del
	1N del	intron	C	ins/ins	ins/ins	ins/ins	del/ins	del/ins	del/ins	del/del	del/del
<i>RETSAT</i>	4N ins	intron	ACAC	ins/del	ins/ins	ins/ins	ins/ins	del/del	ins/del	del/del	del/del
	2N del	3'UTR	AA	del/del	del/ins	del/del	del/ins	ins/ins	ins/ins	del/ins	ins/ins
<i>ACSBG2</i>	9N ins	intron	ATTCTGGGG	ins/ins	ins/del	ins/ins	ins/del	del/del	del/del	del/del	del/del
	3N ins	upstream	GGC	ins/ins	ins/ins	ins/ins	ins/ins	del/del	del/del	ins/del	ins/del
<i>NFKB2</i>	1N ins	intron	C	ins/ins	ins/ins	ins/ins	ins/ins	del/del	del/del	ins/del	ins/del
	2N ins	upstream	GG	ins/ins	ins/ins	ins/ins	ins/ins	del/del	del/del	del/del	del/del
<i>TBC1D1</i>	1N del	intron	T	ins/ins	ins/ins	ins/ins	ins/ins	del/ins	del/ins	del/ins	del/ins
<i>NLK</i>	1N ins	intron	T	ins/ins	ins/del	ins/ins	ins/ins	del/del	del/del	ins/del	ins/del
	2N del	intron	AT	del/ins	del/ins	del/del	del/del	ins/ins	ins/ins	del/ins	del/ins
	1N del	intron	A	del/ins	del/ins	del/del	del/del	ins/ins	ins/ins	del/ins	del/ins
	4N del	intron	AAAA	ins/ins	del/ins	ins/ins	ins/ins	del/ins	del/del	del/ins	del/del
<i>MAP3K1</i>	5N del	intron	CATTT	ins/ins	ins/ins	ins/ins	del/ins	del/del	del/del	del/ins	del/del
<i>SLC30A2</i>	6N del	intron	TTTTTG	del/ins	del/ins	del/ins	del/ins	ins/ins	ins/ins	ins/ins	ins/ins
<i>ANGPT1</i>	2N ins	intron	AT	del/del	del/del	del/del	del/del	ins/ins	ins/ins	ins/ins	ins/del
<i>UGDH</i>	1N ins	intron	T	del/del	del/del	del/del	del/del	ins/ins	ins/del	ins/ins	ins/del
	1N ins	intron	G	del/del	del/del	del/del	del/del	ins/ins	ins/del	ins/ins	ins/del

N: nucleotide; ins: insertion; del: deletion.

doi:10.1371/journal.pone.0168946.t004

cattle (n = 702,965) [39], Korean Holstein (n = 631,332) [39] and French dairy breeds (n = 873,372) [25], which was probably due to the sequencing fold coverage and different genetic backgrounds. In addition, in this study, the indels were validated by sequencing at a rate as high as 70%, which was the same as in a previous study (70%) [24]. In the current research, approximately 89% of insertions and 88% of deletions were less than 4 bp, and the number of indels in each chromosome declined with decreasing chromosome length. All of these findings were similar to those in previous studies [24, 37, 38]. The distributions in each chromosome and the lengths of the indels determined here were in line with recent findings in other species, such as chicken, Tibetan macaque, Quarter Horse, and sorghum [7–9, 11, 40].

### Candidate genes for milk composition

In the present study, we performed integrated analysis of previously reported QTL, significant SNPs identified by previous GWASs, and the biological functions of genes, which enabled us to identify 11 candidate genes potentially associated with milk protein and fat traits. These genes contained at least one indel that was polymorphic between the high and low groups as described below.

**Candidate genes containing exonic indel.** Indels in a gene coding region, especially frameshift indels, had a major influence on gene function [41, 42]. Among the 11 candidate genes, the *FCGR2B* gene with one frameshift indel in exon 7 encodes a transmembrane glycoprotein that is a member of the low-affinity immunoglobulin gamma Fc receptors and plays a key role in phagocytosis and the clearing of immune complexes. Bovine mammary gland, which is a product of the innate immune system, is active during lactation. *FCGR2B* might affect the milk protein and fat traits through playing an indispensable role in preventing the individuals from disease infection during lactation. *CENPE* with two exonic indels (exons 58 and 68) and seven intronic indels was a component of the kinetochore complex and part of the spindle assembly checkpoint those were responsible for maintain chromosome congression and mitotic checkpoint [43]. It was a candidate gene for triple-negative breast cancer, and played a crucial role in suppressing tumorigenesis [44] Depletion of *CENPE* resulted in promoting the Drosophila epithelial tissues overgrowth [45].

**Candidate genes with indel in regulatory regions.** *RETSAT* also known as *PPSIG* with one indel in 3' UTR and one indel in intron, was found to be a target gene of PPAR $\alpha$  which played an essential role in lipid metabolism and energy balance [46], and responsible for adipogenesis promotion and lipid deposition [47]. *ACSBG2* with one indel in upstream region and one indel in intron encodes a protein that belongs to the acyl-CoA synthetase family, which plays a pivotal role in lipid metabolism and lipid droplet formation. Claire et al. identified *ACSBG2* as a candidate gene that was significantly related to abdominal fat deposition in chicken [48]. In addition, Sun et al. demonstrated that one SNP located in an intron of *ACSBG2* was also significantly associated with yolk development in chicken [49]. *NFKB2* with one indel in upstream region encodes a protein subunit of the transcription factor complex nuclear factor-kappa-B (NFkB), which regulates the transcription of genes related to cell differentiation and proliferation. It was indispensable for normal development of the bovine mammary gland [50]. In our previous study, we identified that a SNP (ARS-BFGL-NGS-107403) close to *NFKB2* gene was significant associated with C14:1 and C14 indices of milk fatty acids in Chinese Holstein population, and the indel identified in current study was only 0.27 Mb away from the significant SNP mentioned above [51].

**Candidate genes with intronic indel.** *TBC1D1* with an intronic indel is a Rab GTPase activating protein, which was involved in AMPK signaling pathway and was important to maintain the glucose and energy homeostasis [52]. Dysfunction of *TBC1D1* in mice induced that lipogenic gene expressions were up-regulated through IGF1R–PKB–mTOR signaling pathway, including *FASN*, *ACL*, *ACCI* and *SCD* [53]. Polymorphism of *TBC1D1* was found linked to the severe obesity in human [54]. *NLK* with four intronic indels is a serine/threonine-protein kinase, belonging to the MAP kinase (MAPK) subfamily, and is involved in MAPK signaling pathway. Yuan et al. identified the NLK as a negative regulator of mTORC1 complex which was important to the milk protein synthesis [55]. Polymorphisms of *NLK* were strongly associated with obesity in human and fatty acid composition in pig [56, 57].

*MAP3K1* also named as *MEKK1* with one intronic indel, encodes a serine/threonine kinase and is involved in MAPK signaling pathway. Also, it was demonstrated that *MAP3K1* was related to human breast cancer progression [58, 59], thus, *MAP3K1* may play a vital role in keeping bovine mammary gland function well. *SLC30A2* was also known as the zinc transporter ZnT2 with one intronic indel, and played an important role in transporting of Zn into mammary epithelial cells and secreting into the milk [60]. It was found that loss of *ZnT2* can impair mammary gland development and reduce milk composition in mouse [61, 62], and polymorphisms of *ZnT2* were associated with human mammary gland dysfunction [63, 64].

*ANGPT1* also named as *Ang1* with one indel in intron, was one of the ligands for TEK Receptor Tyrosine Kinase (*TEK*) and involved in the PI3K-Akt signaling pathway that is

correlated to milk protein synthesis [55]. Mice deficient with *ANGPT1* were shown poor vascular network [65]. Bovine mammary gland was with an advanced microvasculature which were requisite for mammary microcirculation of nutrients transport and supply to maintain milk synthesis and secretory [66]. *UGDH* with two intronic indels also located within a QTL region on chromosome 6 linked to milk protein and fat traits, and its biological function and genetic effects revealed that it was a key candidate gene associated with such traits [67].

Although we identified 11 candidate genes based on the indels in genic and gene flanking regions that were polymorphic between high and low groups, further association analysis should be performed to validate functionally important indels and to characterize more important molecular markers of milk protein and fat traits. Along with the development of the human ENCODE project (2011), similar analyses could be performed for the intergenic variants. This would make it possible to identify more important potential functional variants within the intergenic regions of the genome that play major roles in regulating genes affecting bovine milk composition traits. Therefore, the complete analysis based on a large number of indels in our study should constitute a valuable resource for improving the molecular breeding of dairy cattle.

## Conclusion

In this study, we resequenced eight Holstein bulls and detected 912,302 indels through cattle genome. We also successfully identified 3,625 indels that were polymorphic between high and low groups throughout the genome. Based on the integrated analysis of the differential indels located in genic and flanking regions, previously reported QTL, and significant SNPs identified by GWASs, along with well-known pathways and biological functions of genes, we identified 11 promising candidate genes that may affect milk protein and fat traits of dairy cattle, including *FCGR2B*, *CENPE*, *RETSAT*, *ACSBG2*, *NFKB2*, *TBC1D1*, *NLK*, *MAP3K1*, *SLC30A2*, *ANGPT1* and *UGDH*. And the indels of 11 genes identified in this study will provide a useful resource for future association studies to discovery the major alleles affecting milk composition traits in dairy cattle.

## Supporting Information

**S1 File. This file contains the detailed information of QTL, significant SNPs and well-known pathways of 11 candidate genes. Table A.** Detailed information on the reported QTL containing the 8 genes for the high and low groups regarding EBV for milk protein and fat traits. **Table B.** Detailed information of the nearest and most significant SNPs from previous GWASs and well-known pathways to the 11 genes between the high and low groups for milk protein and fat traits.

(DOCX)

**S1 Table. Summary of validated indels.**

(XLSX)

**S2 Table. Functional enrichment of 695 genes containing the indels that were polymorphic between the high and low groups by KOBAS tool.**

(XLSX)

## Acknowledgments

The authors would also like to appreciate the kind help from official Dairy Data Center of China in providing the official EBVs data.

## Author Contributions

**Conceptualization:** DS QZ.

**Data curation:** JJ YG YH.

**Formal analysis:** JJ YG YH.

**Funding acquisition:** DS.

**Investigation:** JJ YG.

**Methodology:** DS YH.

**Project administration:** DS SZ QZ.

**Resources:** JJ DS.

**Software:** YH.

**Supervision:** DS.

**Validation:** YG WL.

**Writing – original draft:** JJ.

**Writing – review & editing:** DS.

## References

1. Montgomery SB, Goode DL, Kvikstad E, Albers CA, Zhang ZD, Mu XJ, et al. The origin, evolution, and functional impact of short insertion-deletion variants identified in 179 human genomes. *Genome research*. 2013; 23(5):749–61. doi: [10.1101/gr.148718.112](https://doi.org/10.1101/gr.148718.112) PMID: [23478400](https://pubmed.ncbi.nlm.nih.gov/23478400/)
2. Weischenfeldt J, Symmons O, Spitz F, Korbelt JO. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature reviews Genetics*. 2013; 14(2):125–38. doi: [10.1038/nrg3373](https://doi.org/10.1038/nrg3373) PMID: [23329113](https://pubmed.ncbi.nlm.nih.gov/23329113/)
3. Collins FS, Drumm ML, Cole JL, Lockwood WK, Vande Woude GF, Lannuzzi MC. Construction of a general human chromosome jumping library, with application to cystic fibrosis. *Science*. 1987; 235(4792):1046–9. PMID: [2950591](https://pubmed.ncbi.nlm.nih.gov/2950591/)
4. Warren ST, Zhang F, Licameli GR, Peters JF. The fragile X site in somatic cell hybrids: an approach for molecular cloning of fragile sites. *Science*. 1987; 237(4813):420–3. PMID: [3603029](https://pubmed.ncbi.nlm.nih.gov/3603029/)
5. Grobet L, Martin LR, Poncelet D, Pirottin D, Brouwers B, Riquet J, et al. A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nature genetics*. 1997; 17(1):71–4. doi: [10.1038/ng0997-71](https://doi.org/10.1038/ng0997-71) PMID: [9288100](https://pubmed.ncbi.nlm.nih.gov/9288100/)
6. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, et al. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome research*. 2006; 16(9):1182–90. doi: [10.1101/gr.4565806](https://doi.org/10.1101/gr.4565806) PMID: [16902084](https://pubmed.ncbi.nlm.nih.gov/16902084/)
7. Doan R, Cohen ND, Sawyer J, Ghaffari N, Johnson CD, Dindot SV. Whole-genome sequencing and genetic variant analysis of a Quarter Horse mare. *BMC genomics*. 2012; 13(1):1.
8. Fan W-L, Ng CS, Chen C-F, Lu M-YJ, Chen Y-H, Liu C-J, et al. Genome-wide patterns of genetic variation in two domestic chickens. *Genome biology and evolution*. 2013; 5(7):1376–92. doi: [10.1093/gbe/evt097](https://doi.org/10.1093/gbe/evt097) PMID: [23814129](https://pubmed.ncbi.nlm.nih.gov/23814129/)
9. Yan Y, Yi G, Sun C, Qu L, Yang N. Genome-wide characterization of insertion and deletion variation in chicken using next generation sequencing. *PloS one*. 2014; 9(8):e104652. doi: [10.1371/journal.pone.0104652](https://doi.org/10.1371/journal.pone.0104652) PMID: [25133774](https://pubmed.ncbi.nlm.nih.gov/25133774/)
10. Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, et al. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nature genetics*. 2010; 42(11):1027–30. doi: [10.1038/ng.684](https://doi.org/10.1038/ng.684) PMID: [20972441](https://pubmed.ncbi.nlm.nih.gov/20972441/)
11. Zheng L-Y, Guo X-S, He B, Sun L-J, Peng Y, Dong S-S, et al. Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). *Genome biology*. 2011; 12(11):1.

12. Lam H-M, Xu X, Liu X, Chen W, Yang G, Wong F-L, et al. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature genetics*. 2010; 42(12):1053–9. doi: [10.1038/ng.715](https://doi.org/10.1038/ng.715) PMID: [21076406](https://pubmed.ncbi.nlm.nih.gov/21076406/)
13. Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, et al. Genome wide association studies for milk production traits in Chinese Holstein population. *PloS one*. 2010; 5(10):e13661. doi: [10.1371/journal.pone.0013661](https://doi.org/10.1371/journal.pone.0013661) PMID: [21048968](https://pubmed.ncbi.nlm.nih.gov/21048968/)
14. Mai M, Sahana G, Christiansen F, Guldbbrandtsen 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):3522–8. doi: [10.2527/jas.2009-2713](https://doi.org/10.2527/jas.2009-2713) PMID: [20656975](https://pubmed.ncbi.nlm.nih.gov/20656975/)
15. 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. doi: [10.1186/1471-2164-12-408](https://doi.org/10.1186/1471-2164-12-408) PMID: [21831322](https://pubmed.ncbi.nlm.nih.gov/21831322/)
16. Schopen G, Visker M, Koks P, Mullaart E, van Arendonk J, Bovenhuis H. Whole-genome association study for milk protein composition in dairy cattle. *Journal of dairy science*. 2011; 94(6):3148–58. doi: [10.3168/jds.2010-4030](https://doi.org/10.3168/jds.2010-4030) PMID: [21605784](https://pubmed.ncbi.nlm.nih.gov/21605784/)
17. Meredith BK, Kearney FJ, Finlay EK, Bradley DG, Fahey AG, Berry DP, et al. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC genetics*. 2012; 13(1):1.
18. Guo J, Jorjani H, Carlborg Ö. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. *BMC genetics*. 2012; 13(1):1.
19. 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. doi: [10.1371/journal.pone.0080219](https://doi.org/10.1371/journal.pone.0080219) PMID: [24265800](https://pubmed.ncbi.nlm.nih.gov/24265800/)
20. 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](https://pubmed.ncbi.nlm.nih.gov/7713441/)
21. Grisart B, Farnir F, Karim L, Cambisano N, Kim J-J, Kvasz A, et al. 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):2398–403. doi: [10.1073/pnas.0308518100](https://doi.org/10.1073/pnas.0308518100) PMID: [14983021](https://pubmed.ncbi.nlm.nih.gov/14983021/)
22. Blott S, Kim J-J, Moio S, Schmidt-Küntzel 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. PMID: [12586713](https://pubmed.ncbi.nlm.nih.gov/12586713/)
23. Stothard P, Choi J-W, 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(1):1.
24. Choi J-W, Liao X, Park S, Jeon H-J, Chung W-H, Stothard P, et al. Massively parallel sequencing of Chikso (Korean brindle cattle) to discover genome-wide SNPs and InDels. *Molecules and cells*. 2013; 36(3):203–11. doi: [10.1007/s10059-013-2347-0](https://doi.org/10.1007/s10059-013-2347-0) PMID: [23912596](https://pubmed.ncbi.nlm.nih.gov/23912596/)
25. Boussaha M, Esquerré D, Barbieri J, Djari A, Pinton A, Letaief R, et al. Genome-wide study of structural variants in bovine Holstein, Montbéliarde and Normande dairy breeds. *PloS one*. 2015; 10(8):e0135931. doi: [10.1371/journal.pone.0135931](https://doi.org/10.1371/journal.pone.0135931) PMID: [26317361](https://pubmed.ncbi.nlm.nih.gov/26317361/)
26. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PloS one*. 2012; 7(2):e30619. doi: [10.1371/journal.pone.0030619](https://doi.org/10.1371/journal.pone.0030619) PMID: [22312429](https://pubmed.ncbi.nlm.nih.gov/22312429/)
27. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
28. 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. doi: [10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352) PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/)
29. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*. 2010; 38(16):e164–e. doi: [10.1093/nar/gkq603](https://doi.org/10.1093/nar/gkq603) PMID: [20601685](https://pubmed.ncbi.nlm.nih.gov/20601685/)
30. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic acids research*. 2011; 39(suppl 2):W316–W22.
31. Hu J, Ng PC. SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. *PLoS One*. 2013; 8(10):e77940. doi: [10.1371/journal.pone.0077940](https://doi.org/10.1371/journal.pone.0077940) PMID: [24194902](https://pubmed.ncbi.nlm.nih.gov/24194902/)

32. Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ, et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature*. 2007; 448(7157):1050–3. doi: [10.1038/nature06067](https://doi.org/10.1038/nature06067) PMID: [17660834](https://pubmed.ncbi.nlm.nih.gov/17660834/)
33. Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, et al. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nature genetics*. 2007; 39(9):1151–5. doi: [10.1038/ng2115](https://doi.org/10.1038/ng2115) PMID: [17676040](https://pubmed.ncbi.nlm.nih.gov/17676040/)
34. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, et al. Whole-genome patterns of common DNA variation in three human populations. *Science*. 2005; 307(5712):1072–9. doi: [10.1126/science.1105436](https://doi.org/10.1126/science.1105436) PMID: [15718463](https://pubmed.ncbi.nlm.nih.gov/15718463/)
35. Metzker ML. Sequencing technologies—the next generation. *Nature reviews genetics*. 2010; 11(1):31–46. doi: [10.1038/nrg2626](https://doi.org/10.1038/nrg2626) PMID: [19997069](https://pubmed.ncbi.nlm.nih.gov/19997069/)
36. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. The next-generation sequencing revolution and its impact on genomics. *Cell*. 2013; 155(1):27–38. doi: [10.1016/j.cell.2013.09.006](https://doi.org/10.1016/j.cell.2013.09.006) PMID: [24074859](https://pubmed.ncbi.nlm.nih.gov/24074859/)
37. Eck SH, Benet-Pagès A, Flisikowski K, Meitinger T, Fries R, Strom TM. Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. *Genome biology*. 2009; 10(8):1.
38. 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(1):1.
39. Choi J-W, Liao X, Stothard P, Chung W-H, Jeon H-J, Miller SP, et al. Whole-genome analyses of Korean native and Holstein cattle breeds by massively parallel sequencing. *PloS one*. 2014; 9(7): e101127. doi: [10.1371/journal.pone.0101127](https://doi.org/10.1371/journal.pone.0101127) PMID: [24992012](https://pubmed.ncbi.nlm.nih.gov/24992012/)
40. Fan Z, Zhao G, Li P, Osada N, Xing J, Yi Y, et al. Whole-genome sequencing of Tibetan macaque (*Macaca thibetana*) provides new insight into the macaque evolutionary history. *Molecular biology and evolution*. 2014:msu104.
41. Ng PC, Levy S, Huang J, Stockwell TB, Walenz BP, Li K, et al. Genetic variation in an individual human exome. *PLoS Genet*. 2008; 4(8):e1000160. doi: [10.1371/journal.pgen.1000160](https://doi.org/10.1371/journal.pgen.1000160) PMID: [18704161](https://pubmed.ncbi.nlm.nih.gov/18704161/)
42. Pelak K, Shianna KV, Ge D, Maia JM, Zhu M, Smith JP, et al. The characterization of twenty sequenced human genomes. *PLoS Genet*. 2010; 6(9):e1001111. doi: [10.1371/journal.pgen.1001111](https://doi.org/10.1371/journal.pgen.1001111) PMID: [20838461](https://pubmed.ncbi.nlm.nih.gov/20838461/)
43. Weaver BAA, Silk AD, Montagna C, Verdier-Pinard P, Cleveland DW. Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell*. 2007; 11(1):25–36. doi: [10.1016/j.ccr.2006.12.003](https://doi.org/10.1016/j.ccr.2006.12.003) PMID: [17189716](https://pubmed.ncbi.nlm.nih.gov/17189716/)
44. Kung PP, Martinez R, Zhu Z, Zager M, Blasina A, Rymer I, et al. Chemogenetic evaluation of the mitotic kinesin CENP-E reveals a critical role in triple-negative breast cancer. *Mol Cancer Ther*. 2014; 13(8):2104–15. doi: [10.1158/1535-7163.MCT-14-0083-T](https://doi.org/10.1158/1535-7163.MCT-14-0083-T) PMID: [24928852](https://pubmed.ncbi.nlm.nih.gov/24928852/)
45. Clemente-Ruiz M, Muzzopappa M, Milan M. Tumor suppressor roles of CENP-E and Nsl1 in *Drosophila* epithelial tissues. *Cell cycle*. 2014; 13(9):1450–5. doi: [10.4161/cc.28417](https://doi.org/10.4161/cc.28417) PMID: [24626182](https://pubmed.ncbi.nlm.nih.gov/24626182/)
46. Lee SS, Chan W-Y, Lo CK, Wan DC, Tsang DS, Cheung W-T. Requirement of PPAR $\alpha$  in maintaining phospholipid and triacylglycerol homeostasis during energy deprivation. *Journal of lipid research*. 2004; 45(11):2025–37. doi: [10.1194/jlr.M400078-JLR200](https://doi.org/10.1194/jlr.M400078-JLR200) PMID: [15342691](https://pubmed.ncbi.nlm.nih.gov/15342691/)
47. Moise AR, Lobo GP, Erokwu B, Wilson DL, Peck D, Alvarez S, et al. Increased adiposity in the retinol saturase-knockout mouse. *FASEB J*. 2010; 24(4):1261–70. doi: [10.1096/fj.09-147207](https://doi.org/10.1096/fj.09-147207) PMID: [19940255](https://pubmed.ncbi.nlm.nih.gov/19940255/)
48. D'Andre HC, Paul W, Shen X, Jia X, Zhang R, Sun L, et al. Identification and characterization of genes that control fat deposition in chickens. *Journal of animal science and biotechnology*. 2013; 4(1):1.
49. Sun C, Lu J, Yi G, Yuan J, Duan Z, Qu L, et al. Promising Loci and Genes for Yolk and Ovary Weight in Chickens Revealed by a Genome-Wide Association Study. *PloS one*. 2015; 10(9):e0137145. doi: [10.1371/journal.pone.0137145](https://doi.org/10.1371/journal.pone.0137145) PMID: [26332579](https://pubmed.ncbi.nlm.nih.gov/26332579/)
50. Sun X-F, Zhang H. NFKB and NFKBI polymorphisms in relation to susceptibility of tumour and other diseases. 2007.
51. Li C, Sun D, Zhang S, Wang S, Wu X, Zhang Q, et al. Genome wide association study identifies 20 novel promising genes associated with milk fatty acid traits in Chinese Holstein. *PloS one*. 2014; 9(5): e96186. doi: [10.1371/journal.pone.0096186](https://doi.org/10.1371/journal.pone.0096186) PMID: [24858810](https://pubmed.ncbi.nlm.nih.gov/24858810/)
52. Chen S, Murphy J, Toth R, Campbell DG, Morrice NA, Mackintosh C. Complementary regulation of TBC1D1 and AS160 by growth factors, insulin and AMPK activators. *Biochem J*. 2008; 409(2):449–59. doi: [10.1042/BJ20071114](https://doi.org/10.1042/BJ20071114) PMID: [17995453](https://pubmed.ncbi.nlm.nih.gov/17995453/)

53. Chen L, Chen Q, Xie B, Quan C, Sheng Y, Zhu S, et al. Disruption of the AMPK–TBC1D1 nexus increases lipogenic gene expression and causes obesity in mice via promoting IGF1 secretion. *Proceedings of the National Academy of Sciences*. 2016; 113(26):7219–24.
54. Stone S, Abkevich V, Russell DL, Riley R, Timms K, Tran T, et al. TBC1D1 is a candidate for a severe obesity gene and evidence for a gene/gene interaction in obesity predisposition. *Human molecular genetics*. 2006; 15(18):2709–20. doi: [10.1093/hmg/ddl204](https://doi.org/10.1093/hmg/ddl204) PMID: [16893906](https://pubmed.ncbi.nlm.nih.gov/16893906/)
55. Bionaz M, Looor JJ. Gene networks driving bovine mammary protein synthesis during the lactation cycle. *Bioinformatics and biology insights*. 2011; 5:83–98. doi: [10.4137/BBI.S7003](https://doi.org/10.4137/BBI.S7003) PMID: [21698073](https://pubmed.ncbi.nlm.nih.gov/21698073/)
56. Pei YF, Zhang L, Liu Y, Li J, Shen H, Liu YZ, et al. Meta-analysis of genome-wide association data identifies novel susceptibility loci for obesity. *Hum Mol Genet*. 2014; 23(3):820–30. doi: [10.1093/hmg/ddt464](https://doi.org/10.1093/hmg/ddt464) PMID: [24064335](https://pubmed.ncbi.nlm.nih.gov/24064335/)
57. Supakankul P, Mekchay S. Association of NLK polymorphisms with intramuscular fat content and fatty acid composition traits in pigs. *Meat science*. 2016; 118:61–5. doi: [10.1016/j.meatsci.2016.03.025](https://doi.org/10.1016/j.meatsci.2016.03.025) PMID: [27050409](https://pubmed.ncbi.nlm.nih.gov/27050409/)
58. Cuevas BD, Winter-Vann AM, Johnson NL, Johnson GL. MEK1 controls matrix degradation and tumor cell dissemination during metastasis of polyoma middle-T driven mammary cancer. *Oncogene*. 2006; 25(36):4998–5010. doi: [10.1038/sj.onc.1209507](https://doi.org/10.1038/sj.onc.1209507) PMID: [16568086](https://pubmed.ncbi.nlm.nih.gov/16568086/)
59. Glubb DM, Maranian MJ, Michailidou K, Pooley KA, Meyer KB, Kar S, et al. Fine-scale mapping of the 5q11.2 breast cancer locus reveals at least three independent risk variants regulating MAP3K1. *The American Journal of Human Genetics*. 2015; 96(1):5–20. doi: [10.1016/j.ajhg.2014.11.009](https://doi.org/10.1016/j.ajhg.2014.11.009) PMID: [25529635](https://pubmed.ncbi.nlm.nih.gov/25529635/)
60. Lopez V, Kelleher SL. Zinc transporter-2 (ZnT2) variants are localized to distinct subcellular compartments and functionally transport zinc. *Biochem J*. 2009; 422(1):43–52. doi: [10.1042/BJ20081189](https://doi.org/10.1042/BJ20081189) PMID: [19496757](https://pubmed.ncbi.nlm.nih.gov/19496757/)
61. Lee S, Hennigar SR, Alam S, Nishida K, Kelleher SL. Essential Role for Zinc Transporter 2 (ZnT2)-mediated Zinc Transport in Mammary Gland Development and Function during Lactation. *J Biol Chem*. 2015; 290(21):13064–78. doi: [10.1074/jbc.M115.637439](https://doi.org/10.1074/jbc.M115.637439) PMID: [25851903](https://pubmed.ncbi.nlm.nih.gov/25851903/)
62. Dempsey C, McCormick NH, Croxford TP, Seo YA, Grider A, Kelleher SL. Marginal maternal zinc deficiency in lactating mice reduces secretory capacity and alters milk composition. *J Nutr*. 2012; 142(4):655–60. doi: [10.3945/jn.111.150623](https://doi.org/10.3945/jn.111.150623) PMID: [22357740](https://pubmed.ncbi.nlm.nih.gov/22357740/)
63. Seo YA, Kelleher SL. Functional analysis of two single nucleotide polymorphisms in SLC30A2 (ZnT2): implications for mammary gland function and breast disease in women. *Physiol Genomics*. 2010; 42A(4):219–27. doi: [10.1152/physiolgenomics.00137.2010](https://doi.org/10.1152/physiolgenomics.00137.2010) PMID: [20858712](https://pubmed.ncbi.nlm.nih.gov/20858712/)
64. Alam S, Hennigar SR, Gallagher C, Soybel DI, Kelleher SL. Exome Sequencing of SLC30A2 Identifies Novel Loss- and Gain-of-Function Variants Associated with Breast Cell Dysfunction. *J Mammary Gland Biol Neoplasia*. 2015; 20(3–4):159–72. doi: [10.1007/s10911-015-9338-z](https://doi.org/10.1007/s10911-015-9338-z) PMID: [26293594](https://pubmed.ncbi.nlm.nih.gov/26293594/)
65. Sato TN, Tozawa Y, Deutsch U, Wolburg-Buchholz K, Fujiwara Y, Gendron-Maguire M, et al. Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. *Nature*. 1995; 376(6535):70–4. doi: [10.1038/376070a0](https://doi.org/10.1038/376070a0) PMID: [7596437](https://pubmed.ncbi.nlm.nih.gov/7596437/)
66. Prosser CG, Davis SR, Farr VC, Lacasse P. Regulation of blood flow in the mammary microvasculature. *J Dairy Sci*. 1996; 79(7):1184–97. doi: [10.3168/jds.S0022-0302\(96\)76472-X](https://doi.org/10.3168/jds.S0022-0302(96)76472-X) PMID: [8872713](https://pubmed.ncbi.nlm.nih.gov/8872713/)
67. Xu Q, Mei G, Sun D, Zhang Q, Zhang Y, Yin C, et al. Detection of genetic association and functional polymorphisms of UGDH affecting milk production trait in Chinese Holstein cattle. *BMC genomics*. 2012; 13(1):1.