

Article

Genome-Wide Identification of Discriminative Genetic Variations in Beef and Dairy Cattle via an Information-Theoretic Approach

Soo-Jin Kim ¹ , Jung-Woo Ha ² and Heebal Kim ^{1,3,4,*}

¹ Department of Agricultural Biotechnology and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea; kim.soojin@outlook.com

² Clova AI Research, NAVER Corp., Seongnam 13561, Korea; jungwoo.ha@navercorp.com

³ Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 08826, Korea

⁴ C&K Genomics, Seoul 05836, Korea

* Correspondence: heebal@snu.ac.kr; Tel.: +82-2880-4803

Received: 26 May 2020; Accepted: 19 June 2020; Published: 22 June 2020



Abstract: Analyzing the associations between genotypic changes and phenotypic traits on a genome-wide scale can contribute to understanding the functional roles of distinct genetic variations during breed development. We performed a whole-genome analysis of Angus and Jersey cattle breeds using conditional mutual information, which is an information-theoretic method estimating the conditional independency among multiple factor variables. The proposed conditional mutual information-based approach allows breed-discriminative genetic variations to be explicitly identified from tens of millions of SNP (single nucleotide polymorphism) positions on a genome-wide scale while minimizing the usage of prior knowledge. Using this data-driven approach, we identified biologically relevant functional genes, including breed-specific variants for cattle traits such as beef and dairy production. The identified lipid-related genes were shown to be significantly associated with lipid and triglyceride metabolism, fat cell differentiation, and muscle development. In addition, we confirmed that milk-related genes are involved in mammary gland development, lactation, and mastitis-associated processes. Our results provide the distinct properties of Angus and Jersey cattle at a genome-wide level. Moreover, this study offers important insights into discovering unrevealed genetic variants for breed-specific traits and the identification of genetic signatures of diverse cattle breeds with respect to target breed-specific properties.

Keywords: cattle genome-wide analysis; conditional mutual information; Angus and Jersey cattle; genetic variations; single nucleotide polymorphisms

1. Introduction

Manipulating domesticated animals by inbreeding and artificial selection has led to the development of a multitude of individual cattle breeds. As a result, many cattle breeds have become highly specialized for meat or milk production subsequent to strong genetic selection for these traits. In this context, investigating the associations between genetic variations and phenotypes has significant potential for understanding the heritability of complex traits in cattle. Moreover, such a study will identify distinct genetic factors that are likely to relate to breed-specific characteristics.

Meat and milk yield are important economic factors for cattle production, and Angus and Jersey cattle are representative breeds for beef and dairy traits, respectively. The Angus breed has been intensively selected over the last few decades to reduce several recessive genetic disorders [1]. Jersey cattle were originally bred on the British Channel Island of Jersey, and a number of Jersey breeds have become highly specialized for milk production factors, such as high butterfat content [2].

Thus, analyzing the genetic profiles of Angus and Jersey breeds can aid in investigating important economic traits such as meat and milk production. Recently, the accumulation of massive genotype information on numerous cattle breeds has facilitated detailed studies of cattle genetic variants for the design and development of livestock. A large-scale analysis of the bovine genome may also have an impact on cattle farming by providing new insights for cattle breeding and production programs.

Several studies have been performed to obtain evidence of selection on a genome-wide level in cattle [3–8], and various statistical methods have been successfully proposed to detect selection signatures from genetic polymorphism data. The allele frequency spectrum and haplotype segregation are key concepts for inferring the signatures of selection in populations. The fixation index (F_{ST}) [9] and the cross-population composite likelihood ratio (XP-CLR) [10] are based on variations in the allele frequencies in populations for detecting genomic regions under selection. Linkage information is also employed to identify selection signatures in populations by investigating long-range haplotypes. Long-range haplotype methods use the integrated haplotype score (iHS) [11] and the across-population extended haplotype homozygosity (XP-EHH) [12] to identify alleles segregating in a population based on haplotype length. These methods have relied on variation patterns (e.g., allele frequencies or long haplotypes) as constraints for efficient measurements; however, they may cause SNPs to dominate which are indirectly or implicitly related to these constraints, because they are very sensitive to SNP ascertainment bias [10,13].

In this study, we performed a comparative genome analysis of two cattle breeds, Angus and Jersey, using an information-theoretic approach. The proposed mutual information extractor based on conditional mutual information (CMI) is a principled data-driven method and minimizes the use of prior knowledge to identify discriminative SNPs determining breed-specific traits, such as meat and milk yield. Thus, our analysis focuses on systematically detecting the distinct SNPs that discriminate cattle breeds on a genome-wide scale. Together with detecting new genetic patterns from cattle genome, we found putative genes showing genetic signatures that may have contributed to the development of Angus- and Jersey-specific phenotypes, such as beef and dairy production.

2. Materials and Methods

2.1. Sequencing, Quality Control, and Variant Calling

We collected whole-blood samples (10 mL) from 10 Angus cattle and 10 Jersey cattle. The Angus and Jersey samples originated from Chamtwoo (Seoul, Korea), the Seoul Milk Cooperative (Yangpyeong, Gyeonggi, Korea), and the Korea Federation of Livestock Cooperatives (Dangjin, Chungnam, Korea). The blood samples were obtained from jugular veins. The DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega, Seoul, Korea). The collection of blood samples was performed in accordance with the guidelines given by the relevant agricultural institutions. All methods involving animal works were approved by the Institutional Animal Care and Use Committee of the National Institute of Animal Science (NIAS) in Korea under approval numbers NIAS-2014-093.

We produced pair-end reads using an Illumina HiSeq 2000 and isolated DNA from whole blood using a G-DEXTMIIb Genomic DNA Extraction Kit (iNtRoN Biotechnology, Seongnam, Korea) following the instructions of the manufacturer. We used the Covaris System to generate 3 µg of genomic DNA for generating the ~300 bp inserts. The fragments of the shared DNA were end-repaired, A tailed, adaptor-ligated, and amplified using the TruSeq DNA Sample Prep. Kit (Illumina, San Diego, CA, USA). Paired-end sequencing was performed on the Illumina HiSeq 2000 platform using the TruSeq SBS Kit v3-HS (Illumina, San Diego, CA, USA) with NICEM (National Instrumentation Center for Environmental Management, Seoul, Korea). Raw Sequence data are available from the National Centre for Biotechnology Information (NCBI) with the Bioproject accession numbers PRJNA318087 (Angus) and PRJNA318089 (Jersey).

A per-base sequence was quality checked with fastQC, which calculates various quality metrics for raw reads. Next, the pair-end sequence reads were aligned to a bovine reference genome (UMD 3.1)

using Bowtie2 [14]. We used default parameters (except the “-no-mixed” option) to inhibit unpaired alignments for paired reads.

We used open-source software packages for downstream processing and variant calling. Potential PCR duplicates were filtered using the “REMOVE_DUPLICATEDS = true” option in “MarkDuplicates” of Picard. We then used SAMtools [15] to generate index files for the reference and bam files. GATK [16] was used to correct misalignments due to the presence of indels by performing a local realignment of reads with “RealignerTargetCreator” and “IndelRealigner” modules. The “UnifiedGenotyper” and “SelectVariants” modules of GATK were used for calling candidate SNPs. Next, we used the “VariantFiltration” of GATK to filter variants and avoid possible false positives with the following options: SNPs with QUAL (Phred-scaled quality score) < 30 were filtered; SNPs with MQ0 (the number of reads with a mapping quality of zero across all samples) > 4 and QD (variants confidence/quality by depth; low scores are indicative of false positives and artifacts) < 5 were filtered, and SNPs with FS (Phred-scaled p -value using Fisher’s exact test) > 200 were filtered. We used BEAGLE [17] to impute missing genotypes and infer haplotype phases for the whole set of cattle populations simultaneously.

2.2. Conditional Mutual Information

Information theory has provided a theoretical basis in many data analysis and machine learning tasks since it was proposed for communication and compression perspectives [18,19]. In particular, mutual information (MI) has been a metric widely applied for extracting significant variables from high dimensional data, including gene expression and sequencing data [20,21]. Since it is defined with the entropy of random variables similar to the other information-theoretical methods, such as Kullback–Leibler divergence (KL-divergence), MI formulates the conditional independency between two random variables.

It is straightforward to apply MI to select significant genetic factors associated with cattle breed, and MI can be used to calculate the associations between the SNP positions and breeds. Assuming that genetic factors (e.g., SNP) and class variables (e.g., cattle breed) are random variables, MI can formulate the dependency of each genetic factor on cattle trait. In formal, let x and y denote an SNP position and breed variables. Given an SNP position and breed variables, MI between the SNP position and cattle breed, $I(x;y)$, is defined with the entropy of x and the conditional entropy of x and y as follows:

$$I(x;y) = H(x) - H(x|y),$$

$$I(x;y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)},$$

$$\text{s.t. } H(x) = - \sum_{x \in X} p(x) \log p(x), \text{ and } H(x|y) = H(x,y) - H(y),$$

where X and Y denote the SNP position and the breed variable sets, respectively. When $I(x;y)$ is equal to 0, the SNP corresponding to x is independent of breed, and it means that the SNP does not influence breed in the aspect of a pairwise relationship.

Since MI is a metric for representing the independency between two variables, it is not trivial to characterize the effects of multiple factors on determining cattle breed. To solve this issue, in this study, we employed conditional mutual information (CMI) among two SNP variables and a breed variable as the criterion for extracting important SNP positions [22,23]:

$$CI(y;x_1|x_2) = CI(y;x_1, x_2) - CI(y;x_2),$$

where x_1 and x_2 are two SNP variables.

We are interested in $CI(y; x_1, x_2)$ and call it the mutual information extractor in the rest of this paper. Mutual information extractors can be defined from the chain rule for mutual information:

$$CI(y; x_1, x_2) = CI(y; x_1|x_2) + CI(y; x_2),$$

$$\text{s.t. } CI(y; x_1|x_2) = \sum_{s_1 \in x_1} \sum_{s_2 \in x_2} \sum_{y \in Y} p_{Y, x_1, x_2}(y, s_1, s_2) \log \frac{p_{x_2}(s_2) p_{Y, x_1, x_2}(y, s_1, s_2)}{p_{Y, x_2}(c, s_2) p_{x_1, x_2}(s_1, s_2)},$$

where s_n is the allele value of the n -th SNP x_n .

A mutual information extractor quantifies the associations between SNPs at two loci and breeds. Since $CI(y; x_1, x_2)$ is nonnegative, the same for mutual information, the mutual information extractor is equal to 0 when all three variables are conditionally independent. This property is suitable for detecting discriminative two-locus haplotypes influencing cattle breed. Thus, our method is effective in detecting the discriminative SNPs showing a high dependence between the haplotypes of two adjacent loci and breed.

2.3. XP-CLR and XP-EHH Tests

We performed cross-population composite likelihood ratio (XP-CLR) and cross-population extended haplotype homozygosity (XP-EHH) tests for detecting the selective signatures in Angus and Jersey cattle. These two statistics are representative methods that use different criteria to detect the genomic regions under selection in populations. The XP-CLR statistic is able to detect regions of rapid changes in allele frequency at a locus with random drift [10]. In contrast, the XP-EHH statistic is designed to identify nearly fixed selective signatures by comparing the haplotypes of two populations by measuring linkage disequilibrium [12].

The XP-CLR test is based on the detection of multi-locus allele frequency differentiation across populations, which is not as affected by ascertainment bias [10]. We used the following parameters: non-overlapping sliding windows of 50 kb, a maximum number of SNPs within each window of 400, and the correlation level of the SNPs' contribution to the XP-CLR results down-weighted to 0.95. The regions with XP-CLR values in the top 1% of the empirical distributions in the Angus and Jersey samples were designated as candidate sweeps.

In addition, the XP-EHH is designed to find alleles with an increase in frequency to the point of fixation or near-fixation in one of the populations by comparing haplotypes from two populations [12]. It means that it detects SNPs which are under selection in one population but not in others. So, the extreme XP-EHH scores potentially describe the selection of a particular population. We computed the EHH and the log-ratio values of the iHH (integrated EHH) for the pairwise test of the Angus and Jersey populations. The log ratios were normalized to have a mean of 0 and a variance of 1. In addition, XP-EHH scores are directional. A positive score indicates that selection is likely to have happened in population A, while a negative score means the selection probably occurs in population B. In this study, an XP-EHH value indicating a positive score suggests selection in Angus cattle, whereas a negative score signifies selection in Jersey cattle. We selected the regions with XP-EHH scores in the top and bottom 1% of the empirical distributions (empirical p -value $< 1.0 \times 10^{-3}$), and the selected genomic regions were annotated to the closest genes.

3. Results

We performed a comparative genome-wide analysis for identifying discriminative genetic variations between Angus and Jersey cattle using enhanced methods based on the information-theoretic approach (Figure 1).

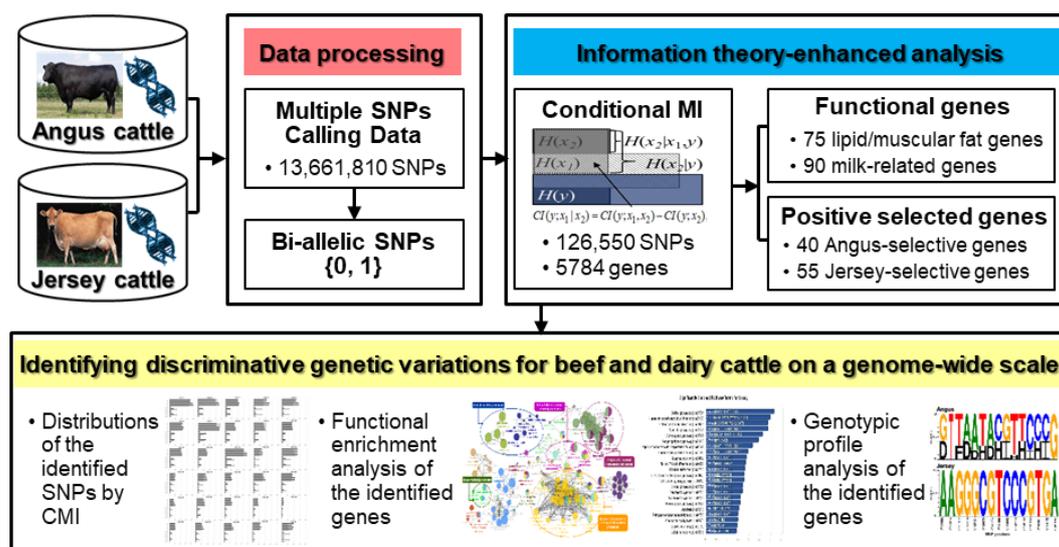


Figure 1. Schematic overview of information theory-enhanced analysis on cattle genome for identifying discriminative genetic variations between Angus and Jersey.

3.1. SNP Detection

The genomes of the Angus and Jersey cattle were sequenced to approximately $15.79\times$ coverage on average, with a total of 840,132,997,679 bp in 8,401,720,919 reads. The pair-end sequence reads were aligned with an average alignment rate of 97.54%, and the reads covered 98.82% of the genome across all of the samples on average (Table S1). A total of ~13 million SNPs were obtained after filtering the potential PCR duplicates and correcting misalignments (Table S2).

3.2. Population Structures

Principal component analysis (PCA), a linear dimensionality reduction method, is broadly used to extract the fundamental structure of a dataset via the projection of individuals into a subspace spanned by the largest principal components [24,25]. In genetics, given that there are a large number of SNPs for many individuals, PCA can be applied to infer the patterns of population structure. To detect the genetic structure of populations, we conducted a PCA on SNP genotype data extracted from Angus and Jersey breed samples via genome-wide complex trait analysis (GCTA) [26] as implemented in EIGENSTRAT [24]. The analysis disregards breed membership but, nevertheless, displays clear breed structures as samples from the same breed cluster together. The Angus samples are separated from the Jersey samples in the projection subspace (the largest PC was 19.06% of the total variation), as shown in Figure S1. This separation indicates that these two breeds show no evidence of admixture with each other.

3.3. Extraction of Discriminative SNPs Based on the Information-Theoretic Method

We used a mutual information extractor for the Angus and Jersey breeds to identify the candidate SNPs with discriminative potential. We extracted 126,550 SNPs annotated to as 5874 genes (Figure S2 and Table S2). Table S3 shows the detailed distribution of the number of the extracted SNPs on each chromosome. The extracted SNPs have high CMI values ($\theta = 0.693$) at a significant level (p -value equal to 2.98×10^{-5}). To overcome any bias caused by the small number of samples, we set a strict p -value for estimating statistical significance (p -value less than 1.0×10^{-3}) compared with those used in other studies [27].

Figure 2 shows the distributions of the SNPs identified by CMI distinguishing between two breeds on each chromosome, excluding the mitochondrial genome. It presents the distributions of the identified SNPs on each chromosome for AA, TT, GG, CC, AT/TA, AG/GA, AC/CA, TG/GT, TC/CT,

and GC/CG genotypes, respectively, in Angus and Jersey. Figure S3 is the graph of the ratio of the SNPs identified by CMI with a significant p -value to total the SNPs across all chromosomes. The result shows that the distribution of SNP extracted by mutual information extractor was different in Angus versus Jersey cattle. These distributions of the SNPs on each chromosome can provide the information on genomic locations that are likely to have received selection pressure and possess the ability to distinguish Angus and Jersey breeds. Hence, the regions containing the extracted SNPs can offer specific candidate areas for a fine-grained mapping of the genes that are important for discriminating between the two breeds.

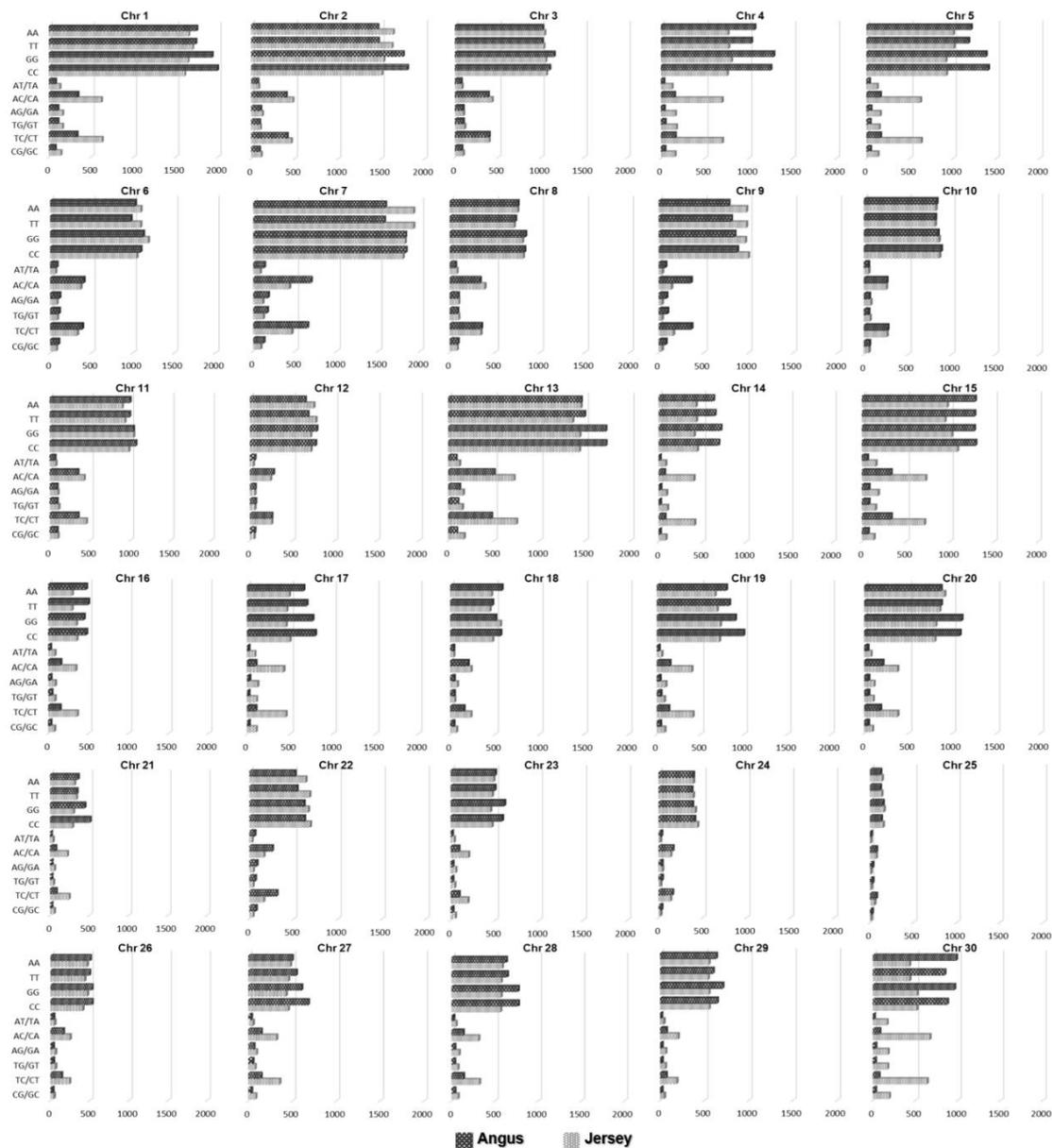


Figure 2. Distributions of the SNPs identified by conditional mutual information (CMI) distinguishing between Angus and Jersey on each chromosome. Patterned black and grey bars indicate the numbers of SNPs identified by CMI in Angus and Jersey according to the genotypes of SNPs on each chromosome. For all graphs, the x -axis is the number of SNPs, and the y -axis represents the genotypes of SNPs (AA, TT, GG, CC, AT/TA, AG/GA, AC/CA, TG/GT, TC/CT, and GC/CG).

Many of the genes mapped in the regions of the extracted SNPs were highly associated with functional genes for beef and dairy traits. To evaluate this finding, we collected a list of 185 lipid and intramuscular fat-related genes and 256 mammary gland/milk-related genes from the literature [28–34]. We performed an evaluation of the functional gene enrichment in the identified genes using a hypergeometric test. Figure 3 shows that the identified genes were statistically overrepresented in the compiled list of literature-reviewed genes with significant p -values. We found 75 genes that were enriched in a catalog of genes involved in lipid and intramuscular fat-related functions (p -value equal to 3.28×10^{-4}), and 90 genes that were overrepresented in a list of mammary gland/milk-related genes (p -value equal to 1.51×10^{-2}). These overrepresented genes are listed in Tables S4 and S5. This result indicates that the regions involved in the SNPs identified by the mutual information extractors include many functional genes that are closely associated with breed-specific characteristics. The functional analysis of these genes is detailed in the next section.

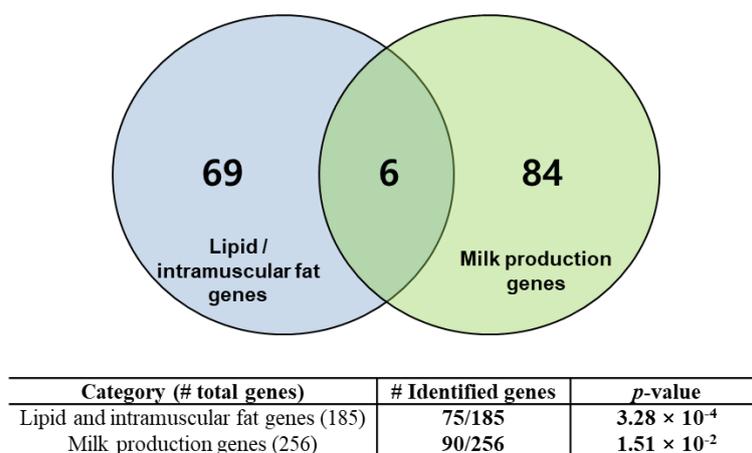


Figure 3. Functional genes identified based on conditional mutual information. The identified genes were enriched in lipid/intramuscular fat genes or mammary gland/milk production-related genes with significant p -value levels.

The identified SNPs that overlap the lipid/intramuscular fat or milk-related genes had dissimilar patterns in terms of heterozygosity in the Angus versus the Jersey breeds (Figures S4 and S5). Table 1 shows the frequencies of heterozygosity for the lipid/intramuscular fat and milk-related SNPs for the two breeds. The total average frequency of heterozygosity of the lipid/intramuscular fat-related SNPs in the Angus breed was 0.391. Interestingly, when one or more Jersey individuals exhibited heterozygous alleles at an SNP locus of these lipid/intramuscular fat genes, all the Angus individuals' alleles at this SNP locus were homozygous (the number of lipid-intramuscular fat genes “with heterozygosity in Jersey” was 0 in the Angus breed). In contrast, if all the alleles of the lipid/intramuscular fat-related genes were homozygous in the Jersey breed, the frequency of heterozygosity at the same SNP position in the Angus breed reached 2.813. We also observed that the frequency of heterozygosity was 0.431 for the milk-related SNPs in Jersey. Similarly, if the alleles of the milk-related genes at an SNP locus were heterozygous in one or more Angus individuals, this heterozygosity did not occur at the same SNP locus in Jersey individuals (the number of milk-related genes in Jersey “with heterozygosity in Angus” was 0). Moreover, if all the alleles of the milk-related SNPs appeared to be homozygous in the Angus breed, the heterozygosity frequency at the same SNP position was 4.046 in the Jersey breed. As shown in Table 1, if the allele genotype of the identified SNPs was heterozygous in any individual of a breed, that heterozygosity did not occur in the same SNP of the other breed. We found large differences in the patterns of heterozygosity between the Angus and Jersey breeds, particularly for SNPs involved in breed-specific genes.

Table 1. The average of the heterozygosity frequency for the lipid-intramuscular fat-related genes in Angus and milk-related genes in Jersey.

Average of Heterozygosity Frequency of Lipid-Related Genes in Angus	
Total heterozygosity frequency in Angus	0.391
With homozygosity in Jersey	2.813
With heterozygosity in Jersey	0
Average of heterozygosity frequency of milk-related genes in Jersey	
Total heterozygosity frequency in Jersey	0.431
With homozygosity in Angus	4.046
With heterozygosity in Angus	0

“Total average of heterozygosity frequency” is the average of the occurrence of heterozygosity per locus of all the SNPs in Angus and Jersey breed, respectively. “With homozygosity in Jersey (or Angus)” denotes the SNPs where all the individuals in a specific breed had homozygous alleles. In contrast, “With heterozygosity in Jersey (or Angus)” indicates the cases that the SNPs had at least one individual with the heterozygous allele in a breed.

3.4. Identification of Breed-Specific Genes

We found discriminative SNPs based on conditional mutual information from the genomes of Angus and Jersey cattle with significant *p*-value levels. The candidate regions indicating strong associations between SNPs and phenotypic traits can be genetically important sites for Angus and Jersey selection. Moreover, these regions contain key genes associated with functional roles for beef and dairy production, and the identified genes were validated with a literature review and gene ontology (GO) analysis.

Several of the identified genes were strongly associated with lipid metabolism for meat and milk production. FASN, LPL, and SCD, important lipogenic enzymes, have been reported to have an influence on lipid deposition, metabolism, and synthesis and are involved in the mammary regulation of milk fat synthesis [35–37]. In particular, SCD is a lipogenic enzyme responsible for influencing the fatty acid composition of muscle and adipose tissue, and the SCD genotype may be a marker for enhancing the nutritional quality of milk [38,39]. Genetic variations in INSIG1 are also related to the ratio of saturated to unsaturated fatty acids in milk, and the activity of INSIG1 affects cholesterol metabolism, lipogenesis, and glucose homeostasis in adipose tissue [39,40]. In addition, several studies have reported that GHR is a key gene influencing milk composition and yield and that polymorphisms in GHR are related to beef marbling [41,42]. Moreover, PPARGC1A is known to be a regulator of energy metabolism and controls the proliferation and differentiation of brown adipocytes [43]. The PPARGC1A gene has also been observed to play a role in the regulation of milk fat synthesis in dairy cattle [44].

In addition, we found genes involved in adipogenesis and adipose cellular functions. EBF1 has been reported to inhibit the differentiation of intramuscular adipocytes by increasing anti-adipogenic factors [43]. Recent studies indicated that FGFs (including FGF1 and FGF2) play a positive role in adipogenesis [43]. Specifically, FGF1 has pro-adipogenic activity on preadipocytes [45], and FGF2 induces the development and growth of adipose-tissue in muscles [46]. IGF1 also has distinct effects on preadipocytes and potentially on mature adipocytes [43,47]. In addition, MYOG regulates the formation of muscle myofibers, which are associated with meat production capacity and harbor several QTL for weight and marbling in cattle [37,48]. Moreover, TTN is one of the marker genes for marbling in beef, and polymorphisms in TTN are closely associated with myofibrillogenesis, which increases marbling levels [49].

Furthermore, we identified genes highly specialized for milk production and mammary gland-related processes. CSN1S1 and CSN3 are known to be key milk protein genes. These genes are closely related to milk yield parameters and milk quality, and many studies have reported that polymorphisms of casein genes influence milk composition and milk protein synthesis [50–52]. MFGE8 is specifically observed in the mammary glands of lactating mice and is overexpressed during lactation and associated with an increase in milk fat content [53]. In addition, GLYCAM1, a member of

the mucin family, is a milk protein synthesized in the mammary gland that encodes a milk fat globule glycoprotein [54]. In addition to the genes responsible for dairy yield traits, we identified the gene for the KIT ligand, KITLG. Missense variations in KITLG influence the roan/white coat color in cattle [55]. KITLG is also known to be an attractive candidate gene for moderating coat color in pigs [56].

In domesticated animals, research has supported the importance of the conservation of specific alleles or genotypes [57]. In particular, the widely conserved casein loci affect milk production and quality. Thus, we analyzed the genotype profiles of the found SNPs by our method in the identified casein genes, CSN1S1 and CSN3, in Angus and Jersey cattle. Interestingly, the result showed that the analyzed sequence logos clearly revealed different genotype profiles for each breed from the identified SNPs (Figure 4). The genome regions containing these genes were highly conserved, with less genetic variation observed in Jersey versus the Angus breed.

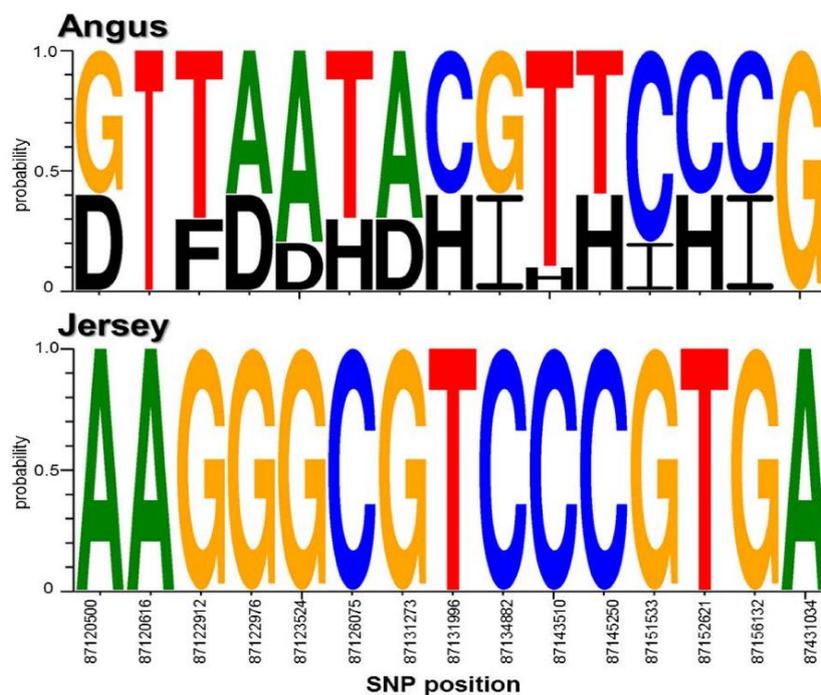


Figure 4. Genotypic profiles of the identified casein genes in Angus and Jersey. This is the genotype profiles of the SNPs in CSN1S1 and CSN3 in Angus and Jersey breeds. A, T, G, C, B, D, E, F, H, and I in the figure indicate AA, TT, GG, CC, AT/TA, AG/GA, AC/CA, TG/GT, TC/CT, and GC/CG genotypes, respectively.

3.5. Functional Enrichment Analysis of the Identified Genes

To obtain insights into the biological processes involving the genes identified by conditional mutual information, we performed a functional analysis using ClueGO [58]. We identified functional genes that are useful for estimating the economic value of cattle, including 75 lipid and intramuscular fat-related genes and 90 milk production-related genes. Figures 5 and 6 show the functional effects of these genes on biological processes. As shown in Figure 5, a large majority of the terms obtained by analyzing the 75 lipid and intramuscular fat-related genes were significantly associated with lipid and triglyceride metabolism, brown fat cell/fat cell differentiation, and energy metabolism, including the regulation of glucose, glycogen, and fatty acid metabolic processes. In particular, muscular and bone development-related terms were significantly enriched in our gene list and specifically included muscle adaptation, activity and hypertrophy, and bone remodeling. The core GO terms obtained by analyzing the 90 milk production-related genes were also related to mammary gland development, lactation, lipid metabolism, and mastitis-related processes, including JUN kinase activity regulation,

MAPK kinase cascades, and the WNT signaling pathway (Figure 6). The details of the analyzed GO terms are described in Tables S6 and S7.

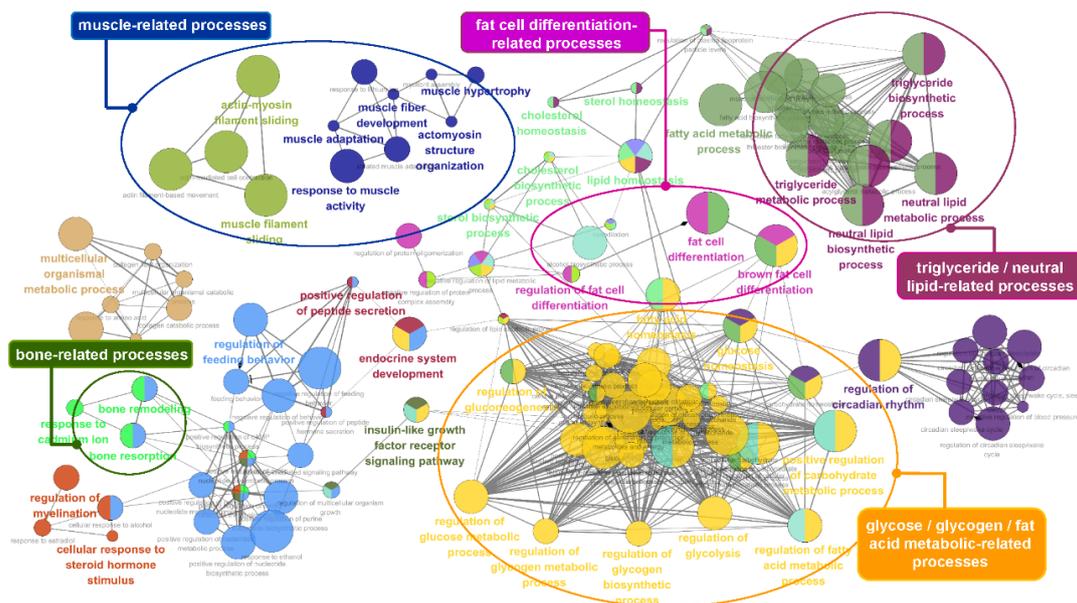


Figure 5. Gene ontology (GO) functional enrichment analysis of the 75 identified genes that overlapped with the lipid metabolism and intramuscular fat genes. The analyzed GO network consisted of distinct functional groups that are associated with energy- and lipid-related processes, such as triglyceride, fatty acid, and glucose metabolism as well as fat cell differentiation. The development of muscle- and bone-related processes is also annotated. The GO functionally grouped networks use terms as nodes (Bonferroni p -value < 0.05) and are linked according to their kappa score level (≥ 0.4). The size of the nodes corresponds to the statistical significance of the terms.

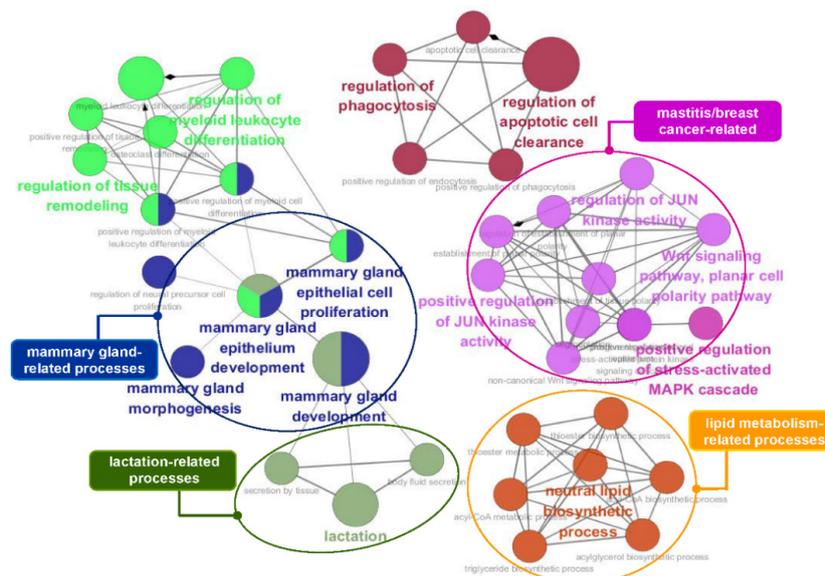


Figure 6. GO functional enrichment analysis of the 90 identified genes that overlapped with the mammary gland/milk production-related genes. This GO network comprised distinct functional groups that are implicated in mammary gland- and lactation-related processes. Moreover, several functional groups, such as Wnt signaling, the MAPK cascade, and JUN kinase activity, are associated with mastitis and breast cancer. GO functionally grouped networks use terms as nodes (Bonferroni p -value < 0.05) and are linked according to their kappa score level (≥ 0.4). The size of the nodes corresponds to the statistical significance of the terms.

3.6. Distinct Genetic Variation on the Mitochondrial Genome

In this study, we identified eight SNPs (24.2% of the total SNPs on the mitochondrial genome) based on our mutual information extractors that act as genetic markers on the mitochondrial genome, discriminating between Angus and Jersey cattle. These SNPs included five genes in a total of 33 SNPs marked with a significant p -value ($< 1.0 \times 10^{-4}$) (Figure S6), and many of the included genes are implicated in energy mechanisms in cattle.

Three of the five genes (ND1, ND2, and COX1) are associated with the development of intramuscular fat content in muscles. In detail, ND1 is highly expressed in oxidative muscles with higher intramuscular fat content [59], and ND2 is significantly correlated with marbling fat content in loin muscles [60]. COX is also strongly associated with triacylglycerol, a chief component of fat in muscles [61]. In addition, we analyzed the functional coherence of ND1, ND2, and COX1 with gene ontology analysis (Figure S7) [62]. Many of the overrepresented GO terms are closely associated with energy metabolism. The abundant GO terms include “generating energy for ATP synthesis”, “energy derivation through oxidation and respiration”, “oxidative phosphorylation”, and “phosphorus metabolic processes.” The genotypes of the identified SNPs also showed clearly dissimilar patterns in the Angus versus the Jersey cattle (Figure S8).

3.7. Analysis of the Overlapped Genetic Signatures Using Diverse Statistics

Combining different statistical methods can be more powerful than a single test to localize a source of selection if each statistic provides distinct information about the selective signatures [13]. We found putative genes showing genetic signatures that may have contributed to the development of Angus and Jersey-specific phenotypes by combining diverse statistics. We applied XP-CLR and XP-EHH tests to detect the putative selection genes by measuring changes in the allele frequency spectrum and the characteristics of extended haplotype homozygosity. In the results of the XP-CLR analysis, 230 and 203 putative selection genes were detected in Angus and Jersey, respectively, with the top 1% of their empirical distributions (empirical p -value $< 1.0 \times 10^{-3}$) (Table S8). Of these genes, 157 genes and 181 genes were shared among Angus and Jersey from MI and XP-CLR. The 226 Angus-selective genes and 253 Jersey-selective genes were detected with p -value $< 1.0 \times 10^{-3}$ using XP-EHH tests (Table S9). Of these 226 and 253 genes, 131 and 192 genes were found in common between MI and XP-EHH for Angus and Jersey, respectively. Finally, we observed 40 Angus-selective genes and 55 Jersey-selective genes at the intersection of MI, XP-CLR, and XP-EHH selection candidates, with the exception of various types of RNA, including 5S_rRNA, 7SK, U6, and so on (Tables S10 and S11).

KEGG pathway analysis was performed by KOBAS 3.0, which is the latest web server for functional sets enrichment of genes [63]. The pathway analysis of 40 genes extracted in common among XP-CLR, XP-EHH and MI for Angus-selective genes showed significantly enriched terms “Glycosylphosphatidylinositol (GPI)-anchor biosynthesis”, “Hippo signaling pathway”, “Fatty acid elongation”, and “Base excision repair” with a p -value < 0.05 (Figure 7). PIGC, which exhibits fat depot-specific mRNA expression, is known to associate with lipid metabolism and obesity [64]. Moreover, ACAA2 is essential for de novo fatty acid synthesis and the activation of long-chain fatty acids and is expressed in the subcutaneous fat tissue of beef cattle involved in adipogenesis [65,66]. TEAD1 is well known as a mediator of skeletal muscle development, and transcriptional regulation of TEAD1 to muscle-specific genes is implemented in cooperation with numerous cofactors such as FoxO3a, which plays a key role in the muscle fiber types affecting meat color, meat tenderness and intramuscular fat content [67].

The 55 positively selected genes in Jersey compared to Angus were mainly involved in the nervous systems, immune systems, infectious diseases, signal transductions, environmental adaptation, endocrine systems, cell growth and death, and lipid metabolisms with a significant p -value < 0.05 (Figure 8). GNG11 and GNGT1 were significantly over-represented in several annotated pathways. GNG11 and GNGT1 code for G proteins, which function as key attributes of innate immune responses, and these are involved in functions relating to mastitis resistance [68]. In particular, PLCL1 encodes a

protein that is involved in a component in the phospho-dependent endocytosis process of the GABA-A receptor. It is also located in CHR2: 86831095–87004473, of which the region is included in a specific trait of milk association QTL relating to milk fat percentage [69,70].

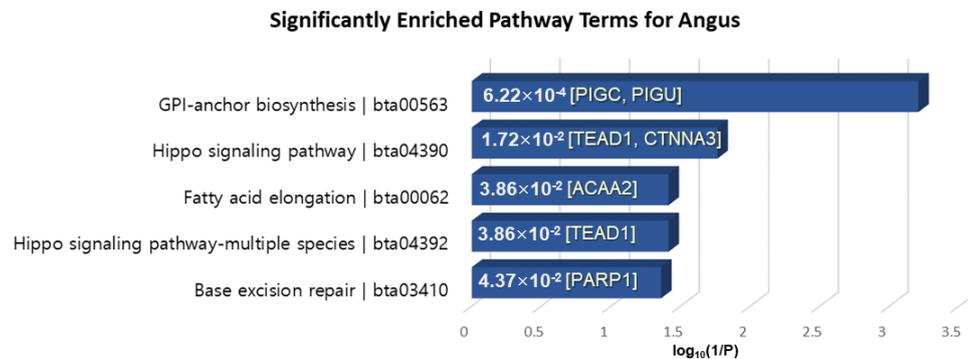


Figure 7. KEGG pathway enrichment analysis of the 40 overlapped genes in the Angus breed. The x-axis is the $\log_{10}(1/p\text{-value})$, and the annotated genes of the enriched pathways with significant p -values shown in bar graphs.

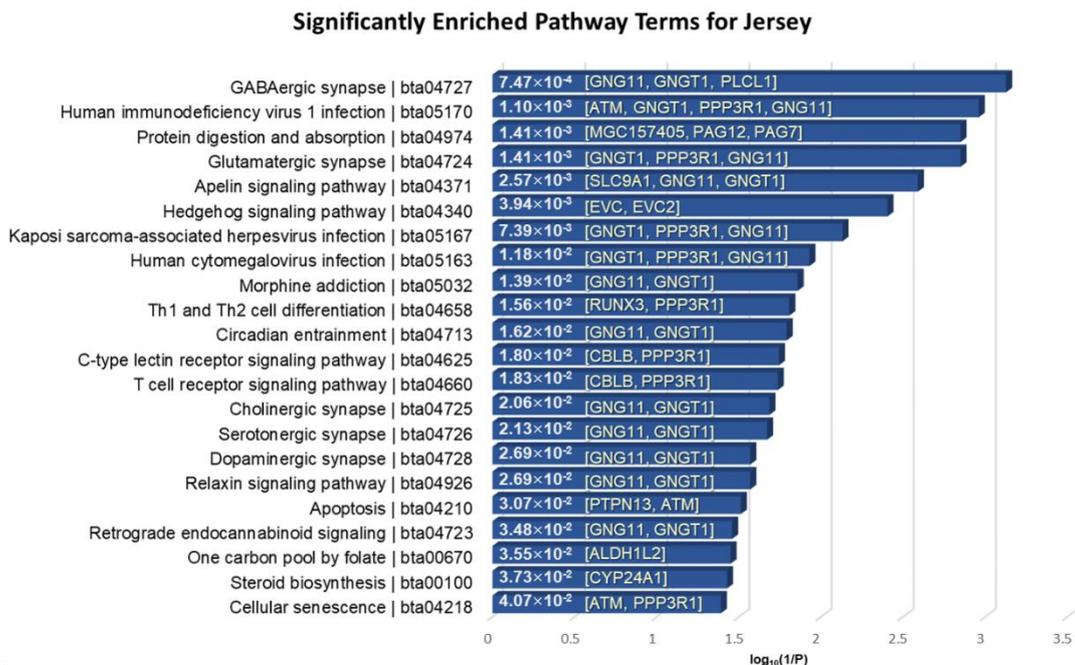


Figure 8. KEGG pathway enrichment analysis of the 55 overlapped genes in Jersey breed. The x-axis indicates the $\log_{10}(1/p\text{-value})$, and the annotated genes of the enriched pathways with significant p -values shown in bar graphs.

4. Discussion

Our study provides insights supporting the identification of discriminative SNPs with breed-specific genetic variations from the whole cattle genome. The used mutual information extractor explicitly found several genetic variations influencing beef and dairy traits in Angus and Jersey cattle from large-scale genome data. Genotype profiles using these phenotypic traits from the cattle genome can be analyzed to identify the key functional genes involved in the formation of breed characteristics.

In this study, we identified discriminative SNPs between Angus and Jersey breeds, and several analysis results based on the identified SNPs confirmed distinct differences between two cattle. The identified genes, including distinct SNPs, are associated with breed-specific functions, such as

meat and milk production, respectively, and the contained SNPs showed clearly dissimilar genotypic patterns. Furthermore, several functional enrichment analyses revealed that distinct functional terms were enriched in the identified genes.

Interestingly, the enriched GO terms in the fat-related genes were associated with lipid and energy metabolism, fat cell differentiation, and muscular/bone development-related terms, whereas the overrepresented terms in the milk-related genes were primarily involved in mammary gland development, lactation, and lipid metabolism. The lipid metabolism terms were enriched in both the fat- and milk-related genes because the traits for both meat and milk are closely associated with the lipid activity of cattle adipocytes [71]. These findings present that our approach can offer a new source of genetic variations influencing the breed-specific traits and show promise for advancing cattle genome research.

Moreover, we investigated the allele patterns of the identified SNPs associated with major functional genes. In particular, we found that casein genes clearly exhibited different breed-specific genotype profiles. The frequency of heterozygosity per SNP locus also showed clearly distinguishable patterns in Angus versus Jersey cattle. Interestingly, if the alleles of a candidate SNP are heterozygous in one specific breed, this heterozygosity does not occur in the same SNP locus of the other breed. This pattern indicates that genetic variation is associated with different properties depending on specific cattle breeds, which can influence the distinct traits of each individual. In addition, the distinct breed-specific allelic patterns of the candidate functional SNPs can provide insights for the discovery of new breed-specific hallmarks for discriminating between beef and milk cattle breeds. Furthermore, this approach allows us to understand the distinct genetic mechanisms underlying the formation of breed characteristics in domestic animals.

Despite numerous studies on the cattle genome, the details of the genetic variations in the mitochondrial genome of cattle have not been much explored relatively. However, mitochondria are important for metabolism, nutrition, and health in humans and animals. Several mitochondrial DNA mutants are reported in connection with a variety of complex traits, such as human disease, longevity, and so on [72,73]. Moreover, several studies have presented that mitochondrial genome polymorphisms in livestock are associated with economic traits, including meat quality, milk-yield, production, and reproduction [74–77]. In particular, considerable mitochondrial DNA diversity has been detected in dairy cattle, and the differences in mitochondrial DNA have been significantly related to milk-yield traits [78,79]. In this study, we found functional genomic regions with discriminative genetic variations between two cattle breeds on the mitochondrial genome. More interestingly, the contained SNP genotypes in the identified genes also showed explicitly different patterns in Angus versus Jersey cattle. Thus, the SNP positions identified on the mitochondrial genome are distinct regions with high discriminative capability, and their variation can be used to recognize genetic features for classifying the two breeds. Moreover, this finding can provide new clues for mitochondrial genome studies in cattle for economic traits.

Finally, our analysis focused on identifying the genetic variations distinguishing each cattle breed and representing the functional traits of cattle from genome sequence data, minimizing the use of genetic assumptions. To compare our method with other statistical methods, we performed an analysis of the selection signatures in the Angus and Jersey cattle using two approaches with different theoretical bases. Many statistical approaches, including F_{ST} , iHS , XP-CLR, and XP-EHH, have been developed to detect the footprints left by selection in genomes. These methods rely on patterns of variation caused by the changes arising quickly in a population, such as allele frequency and haplotype length, to efficiently detect the genomic regions under selection. In addition, these methods use different time frames. In particular, XP-CLR, which utilizes changes in the phase of the allele frequency distribution between populations, has the power to identify older signatures compared with those based on extended linkage disequilibrium, such as XP-EHH [80].

In this study, we analyzed using XP-CLR and XP-EHH with conditional mutual information for providing meaningful results. The selection signature results are not completely consistent,

but we found several candidate genomic regions. Frequently, combining different approaches can be more powerful than a single test if each statistic provides distinct information about the selective signatures [13]. Genetic regions revealed by XP-CLR and XP-EHH were putatively under positive selection, some of which could be crucial for understanding their unique properties. This is possible to produce larger lists of likely selective sweeps, and it may allow us to better understand how selection has effected the variations of a specific breed.

These statistical approaches are useful for detecting the genomic features that accompany the introduction of evolutionarily selective alleles in genome-wide studies, whereas our proposed method has the advantage of identifying potentially discriminative genetic variations in genome sequences. Our method can also assist with hypothesis formulation for genetic mechanisms in cattle, and thus, it provides a new approach for studying the distinct genomic regions related to breed-specific characteristics. Moreover, conditional mutual information can contribute to investigating the associations between distinct SNPs relevant to traits of interest and can considerably aid in understanding the evolution of cattle.

5. Conclusions

Our results described that beef and dairy cattle clearly show genetic differences at a genome-wide level. These implicate that the identified genes based on the extracted SNPs using conditional mutual information can contribute to discriminating the phenotypes of Angus and Jersey cattle, including beef and milk-yield traits. Moreover, the found SNPs showed that they can be involved in different molecular functions and mechanisms influencing the phenotypic differences between cattle breeds of distinct economic significance. Our analysis may provide potential genetic markers for the improvement in livestock productivity, and show the value of comparative genome study in cattle breeds based on an information-theoretic approach.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/11/6/678/s1>. Figure S1: Projection of 20 Angus and Jersey individuals plotted on the top two principal components, Figure S2: Ratio of the identified SNPs to all the SNPs on each chromosome and the number of genes, including the identified SNPs, Figure S3: Distribution of 858 SNPs for heterozygosity in the identified lipid/intramuscular fat-related genes in Angus versus Jersey cattle, Figure S4: Distribution of 852 SNPs for heterozygosity in the identified milk production-related genes in Angus versus Jersey cattle, Figure S5: Thirty-three SNPs on the mitochondrial genome plotted with negative log-scaled *p*-values, Figure S6: Significant GO terms for ND1, ND2, and COX1 on the mitochondrial genome, Figure S7: Genotype profiles for each SNP locus on the mitochondrial genome of Angus and Jersey breeds, Table S1: Summary of sequencing data, Table S2: Chromosomal distributions of the number of SNPs, Table S3: Chromosomal distributions of the number of SNPs identified by CMI, Table S4: The 75 identified lipid/intramuscular fat-related genes and the number of identified SNPs, Table S5: The 90 identified mammary gland/milk production-related genes and the number of identified SNPs, Table S6: Significant GO terms in the 75 identified lipid/intramuscular fat-related genes, Table S7: Significant GO terms in the 90 identified mammary gland/milk production-related genes, Table S8: Summary of XP-CLR, Table S9: Summary of XP-EHH, Table S10: Forty genes in the intersection of XP-CLR, XP-EHH, and conditional mutual information for Angus, Table S11: Fifty-five genes in the intersection of XP-CLR, XP-EHH, and conditional mutual information for Jersey.

Author Contributions: S.-J.K. designed the experiment and method, performed the analysis of genome data and biological interpretation from the results, and wrote the manuscript. J.-W.H. designed and implemented the method and wrote the methods in the manuscript. H.K. supervised and managed the whole study. The final manuscript was reviewed and approved by all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the Next-Generation BioGreen 21 Program (Project No. PJ01323701), Rural Development Administration, Republic of Korea. Also, it was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Republic of Korea (NRF-2018R1D1A1B07050393).

Conflicts of Interest: The authors declare that they have no competing interests.

References

- Ciepluch, A.; Rutkowska, K.; Oprządek, J.; Poławska, E. Genetic disorders in beef cattle: A review. *Genes Genom.* **2017**, *39*, 461–471. [[CrossRef](#)] [[PubMed](#)]
- Mai, M.D.; Sahana, G.; Christiansen, F.B.; Guldbrandtsen, B. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *J. Anim. Sci.* **2010**, *88*, 3522–3528. [[CrossRef](#)] [[PubMed](#)]
- Hayes, B.J.; Chamberlain, A.J.; Maceachern, S.; Savin, K.; McPartlan, H.; MacLeod, I.; Sethuraman, L.; Goddard, M.E. A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Anim. Genet.* **2009**, *40*, 176–184. [[CrossRef](#)] [[PubMed](#)]
- Rothammer, S.; Seichter, D.; Förster, M.; Medugorac, I. A genome-wide scan for signatures of differential artificial selection in ten cattle breeds. *BMC Genom.* **2013**, *14*, 908. [[CrossRef](#)] [[PubMed](#)]
- Utsunomiya, Y.T.; Pérez O'Brien, A.M.; Sonstegard, T.S.; Van Tassell, C.P.; do Carmo, A.S.; Mészáros, G.; Sölkner, J.; Garcia, J.F. Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. *PLoS ONE* **2013**, *8*, e64280. [[CrossRef](#)] [[PubMed](#)]
- Lee, H.J.; Kim, J.; Lee, T.; Son, J.K.; Yoon, H.B.; Baek, K.S.; Jeong, J.Y.; Cho, Y.M.; Lee, K.T.; Yang, B.C.; et al. Deciphering the genetic blueprint behind Holstein milk proteins and production. *Genome Biol. Evol.* **2014**, *6*, 1366–1374. [[CrossRef](#)]
- Taye, M.; Lee, W.; Jeon, S.; Yoon, J.; Dessie, T.; Hanotte, O.; Mwai, O.A.; Kemp, S.; Cho, S.; Oh, S.J.; et al. Exploring evidence of positive selection signatures in cattle breeds selected for different traits. *Mamm. Genome* **2017**, *28*, 528–541. [[CrossRef](#)]
- Xu, L.; Yang, L.; Zhu, B.; Zhang, W.; Wang, Z.; Chen, Y.; Zhang, L.; Gao, X.; Gao, H.; Liu, G.E.; et al. Genome-wide scan reveals genetic divergence and diverse adaptive selection in Chinese local cattle. *BMC Genom.* **2019**, *20*, 494. [[CrossRef](#)]
- Weir, B.S.; Cockerham, C.C. Estimating F-statistics for the analysis of population structure. *Evolution* **1984**, *38*, 1358–1370.
- Chen, H.; Patterson, N.; Reich, D. Population differentiation as a test for selective sweeps. *Genome Res.* **2010**, *20*, 393–402. [[CrossRef](#)]
- Voight, B.F.; Kudaravalli, S.; Wen, X.; Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol.* **2006**, *4*, e72.
- Sabeti, P.C.; Varilly, P.; Fry, B.; Lohmueller, J.; Hostetter, E.; Cotsapas, C.; Xie, X.; Byrne, E.H.; McCarroll, S.A.; Gaudet, R.; et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* **2007**, *449*, 913–918. [[CrossRef](#)] [[PubMed](#)]
- Grossman, S.R.; Shlyakhter, I.; Karlsson, E.K.; Byrne, E.H.; Morales, S.; Frieden, G.; Hostetter, E.; Angelino, E.; Garber, M.; Zuk, O.; et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **2010**, *327*, 883–886. [[CrossRef](#)]
- Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
- Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
- Nekrutenko, A.; Taylor, J. Next-generation sequencing data interpretation: Enhancing reproducibility and accessibility. *Nat. Rev. Genet.* **2012**, *13*, 667–672. [[CrossRef](#)]
- Browning, S.R.; Browning, B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **2007**, *81*, 1084–1097. [[CrossRef](#)]
- Kingma, D.P.; Welling, M. Auto-encoding variational Bayes. In Proceedings of the Second International Conference on Learning Representations, Banff, AB, Canada, 14–16 April 2014.
- Bassily, R.; Nissim, K.; Smith, A.; Steinke, T.; Stemmer, U.; Ullman, J. Algorithmic stability for adaptive data analysis. In Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing, Cambridge, MA, USA, 19–21 June 2016; pp. 1046–1059.
- Na, Y.J.; Sohn, K.A.; Kim, J.H. Interpretation of personal genome sequencing data in terms of disease ranks based on mutual information. *BMC Med. Genom.* **2015**, *8*, S4. [[CrossRef](#)] [[PubMed](#)]

21. Roche, K.; Feltus, F.A.; Park, J.P.; Coissieux, M.M.; Chang, C.; Chan, V.B.S.; Bentires-Alj, M.; Booth, B.W. Cancer cell redirection biomarker discovery using a mutual information approach. *PLoS ONE* **2017**, *12*, e0179265. [[CrossRef](#)]
22. Wyner, A.D. A definition of conditional mutual information for arbitrary ensembles. *Inf. Control* **1978**, *38*, 51–59. [[CrossRef](#)]
23. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: New York, NY, USA, 2006.
24. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **2006**, *38*, 904–909. [[CrossRef](#)] [[PubMed](#)]
25. Abraham, G.; Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS ONE* **2014**, *9*, e93766. [[CrossRef](#)] [[PubMed](#)]
26. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76. [[CrossRef](#)]
27. Pickrell, J.K.; Coop, G.; Novembre, J.; Kudaravalli, S.; Li, J.Z.; Absher, D.; Srinivasan, B.S.; Barsh, G.; Myers, R.M.; Feldman, M.W.; et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **2009**, *19*, 826–837. [[CrossRef](#)] [[PubMed](#)]
28. Ogorevc, J.; Kunej, T.; Razpet, A.; Dovc, P. Database of cattle candidate genes and genetic markers for milk production and mastitis. *Anim. Genet.* **2009**, *40*, 832–851. [[CrossRef](#)]
29. Ron, M.; Israeli, G.; Seroussi, E.; Weller, J.I.; Gregg, J.P.; Shani, M.; Medrano, J.F. Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. *BMC Genom.* **2007**, *8*, 183. [[CrossRef](#)]
30. Clark, D.L.; Boler, D.D.; Kutzler, L.W.; Jones, K.A.; McKeith, F.K.; Killefer, J.; Carr, T.R.; Dilger, A.C. Muscle gene expression associated with increased marbling in beef cattle. *Anim. Biotechnol.* **2011**, *22*, 51–63. [[CrossRef](#)] [[PubMed](#)]
31. De Jager, N.; Hudson, N.J.; Reverter, A.; Barnard, R.; Café, L.M.; Greenwood, P.L.; Dalrymple, B.P. Gene expression phenotypes for lipid metabolism and intramuscular fat in skeletal muscle of cattle. *J. Anim. Sci.* **2013**, *91*, 1112–1128. [[CrossRef](#)]
32. Lim, D.; Kim, N.K.; Park, H.S.; Lee, S.H.; Cho, Y.M.; Oh, S.J.; Kim, T.H.; Kim, H. Identification of candidate genes related to bovine marbling using protein-protein interaction networks. *Int. J. Biol. Sci.* **2011**, *7*, 992–1002. [[CrossRef](#)]
33. Wang, Y.H.; Bower, N.I.; Reverter, A.; Tan, S.H.; De Jager, N.; Wang, R.; McWilliam, S.M.; Café, L.M.; Greenwood, P.L.; Lehnert, S.A. Gene expression patterns during intramuscular fat development in cattle. *J. Anim. Sci.* **2009**, *87*, 119–130. [[CrossRef](#)]
34. Zhao, F.; McParland, S.; Kearney, F.; Du, L.; Berry, D.P. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet. Sel. Evol.* **2015**, *47*, 49. [[CrossRef](#)] [[PubMed](#)]
35. Harvatine, K.J.; Perfield, J.W.; Bauman, D.E. Expression of enzymes and key regulators of lipid synthesis is upregulated in adipose tissue during CLA-induced milk fat depression in dairy cows. *J. Nutr.* **2009**, *139*, 849–854. [[CrossRef](#)] [[PubMed](#)]
36. Jeong, J.; Kwon, E.G.; Im, S.K.; Seo, K.S.; Baik, M. Expression of fat deposition and fat removal genes is associated with intramuscular fat content in longissimus dorsi muscle of Korean cattle steers. *J. Anim. Sci.* **2012**, *90*, 2044–2053. [[CrossRef](#)] [[PubMed](#)]
37. Smith, K.R.; Duckett, S.K.; Azain, M.J.; Sonon, R.N.; Pringle, T.D., Jr. The effect of anabolic implants on intramuscular lipid deposition in finished beef cattle. *J. Anim. Sci.* **2007**, *85*, 430–440. [[CrossRef](#)]
38. Estany, J.; Ros-Freixedes, R.; Tor, M.; Pena, R.N. A functional variant in the stearoyl-coA desaturase gene promoter enhances fatty acid desaturation in pork. *PLoS ONE* **2014**, *9*, e86177. [[CrossRef](#)]
39. Rincon, G.; Islas-Trejo, A.; Castillo, A.R.; Bauman, D.E.; German, B.J.; Medrano, J.F. Polymorphisms in genes in the SREBP1 signaling pathway and SCD are associated with milk fatty acid composition in Holstein cattle. *J. Dairy Res.* **2012**, *79*, 66–75. [[CrossRef](#)]
40. Dong, X.Y.; Tang, S.Q. Insulin-induced gene: A new regulator in lipid metabolism. *Peptides* **2010**, *31*, 2145–2150. [[CrossRef](#)]
41. Komisarek, J.; Michalak, A.; Walendowska, A. The effects of polymorphisms in DGAT1, GH and GHR genes on reproduction and production traits in Jersey cows. *Anim. Sci.* **2011**, *29*, 29–36.

42. Yamada, T. Genetic dissection of marbling trait through integration of mapping and expression profiling. *Anim. Sci. J.* **2014**, *85*, 349–355. [[CrossRef](#)]
43. Rosen, E.D.; MacDougald, O.A. Adipocyte differentiation from the inside out. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 885–896. [[CrossRef](#)]
44. Weikard, R.; Kühn, C.; Goldammer, T.; Freyer, G.; Schwerin, M. The bovine PPARGC1A gene: Molecular characterization and association of an SNP with variation of milk fat synthesis. *Physiol. Genom.* **2005**, *21*, 1–13. [[CrossRef](#)] [[PubMed](#)]
45. Hutley, L.; Shurety, W.; Newell, F.; McGeary, R.; Pelton, N.; Grant, J.; Herington, A.; Cameron, D.; Whitehead, J.; Prins, J. Fibroblast growth factor 1: A key regulator of human adipogenesis. *Diabetes* **2004**, *53*, 3097–3106. [[CrossRef](#)] [[PubMed](#)]
46. Kim, S.; Ahn, C.; Bong, N.; Choe, S.; Lee, D.K. Biphasic effects of FGF2 on adipogenesis. *PLoS ONE* **2015**, *10*, e0120073. [[CrossRef](#)] [[PubMed](#)]
47. Holly, J.; Sabin, M.; Perks, C.; Shield, J. Adipogenesis and IGF-1. *Metab. Syndr. Relat. Disord.* **2006**, *4*, 43–50. [[CrossRef](#)] [[PubMed](#)]
48. Casas, E.; Keele, J.W.; Shackelford, S.D.; Koohmaraie, M.; Stone, R.T. Identification of quantitative trait loci for growth and carcass composition in cattle. *Anim. Genet.* **2004**, *35*, 2–6. [[CrossRef](#)] [[PubMed](#)]
49. Yamada, T.; Sasaki, S.; Sukegawa, S.; Yoshioka, S.; Takahagi, Y. Association of a single nucleotide polymorphism in titin gene with marbling in Japanese Black beef cattle. *BMC Res. Notes* **2009**, *2*, 78. [[CrossRef](#)]
50. Kuceriva, J.; Matejcek, A.; Jandurova, O.M.; Sørensen, P.; Nemcova, E.; Stipkova, M.; Kott, T.; Bouska, J.; Frelich, J. Milk protein genes CSN1S1, CSN2, CSN3, LGB and their relation to genetic values of milk production parameters in Czech Fleckvieh. *Czech J. Anim. Sci.* **2006**, *6*, 241–247.
51. Lemay, D.G.; Lynn, D.J.; Martin, W.F.; Neville, M.C.; Casey, T.M.; Rincon, G.; Kriventseva, E.V.; Barris, W.C.; Hinrichs, A.S.; Molenaar, A.J.; et al. The bovine lactation genome: Insights into the evolution of mammalian milk. *Genome Biol.* **2009**, *10*, R43. [[CrossRef](#)]
52. Bonfatti, V.; Giannini, M.; Gervaso, M.; Coletta, A.; Dacasto, M.; Carnier, P. Effect of CSN1S1-CSN3 (α (S1)- κ -casein) composite genotype on milk production traits and milk coagulation properties in Mediterranean water buffalo. *J. Dairy Sci.* **2012**, *95*, 3435–3443. [[CrossRef](#)]
53. Aoki, N.; Ishii, T.; Ohira, S.; Yamaguchi, Y.; Negi, M.; Adachi, T.; Nakamura, R.; Matsuda, T. Stage specific expression of milk fat globule membrane glycoproteins in mouse mammary gland: Comparison of MFG-E8, butyrophilin, and CD36 with a major milk protein, beta-casein. *Biochim. Biophys. Acta* **1997**, *1334*, 182–190. [[CrossRef](#)]
54. Dowbenko, D.; Kikuta, A.; Fennie, C.; Gillett, N.; Lasky, L.A. Glycosylation-dependent cell adhesion molecule 1 (GlyCAM 1) mucin is expressed by lactating mammary gland epithelial cells and is present in milk. *J. Clin. Investig.* **1993**, *92*, 952–960. [[CrossRef](#)] [[PubMed](#)]
55. Andersson, L.; Georges, M. Domestic-animal genomics: Deciphering the genetics of complex traits. *Nat. Rev. Genet.* **2004**, *5*, 202–212. [[CrossRef](#)] [[PubMed](#)]
56. Hadjiconstantouras, C.; Sargent, C.A.; Skinner, T.M.; Archibald, A.L.; Haley, C.S.; Plastow, G.S. Characterization of the porcine KIT ligand gene: Expression analysis, genomic structure, polymorphism detection and association with coat colour traits. *Anim. Genet.* **2008**, *39*, 217–224. [[CrossRef](#)]
57. Petit, R.J.; Mousadik, A.; Pons, O. Identifying populations for conservation on the basis of genetic markers. *Cons. Biol.* **1997**, *12*, 844–855. [[CrossRef](#)]
58. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W.H.; Pages, F.; Trajanoski, Z.; Galon, J. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25*, 1091–1093. [[CrossRef](#)] [[PubMed](#)]
59. Kim, N.K.; Lim, J.H.; Song, M.J.; Kim, O.H.; Park, B.Y.; Kim, M.J.; Hwang, I.H.; Lee, S.C. Comparisons of longissimus muscle metabolic enzymes and muscle fiber types in Korean and western pig breeds. *Meat Sci.* **2008**, *78*, 455–460. [[CrossRef](#)] [[PubMed](#)]
60. Kim, N.K.; Cho, Y.M.; Jung, Y.S.; Kim, G.S.; Heo, K.N.; Lee, S.H.; Lim, D.; Cho, S.; Park, E.W.; Yoon, D. Gene expression profiling of metabolism-related genes between top round and loin muscle of Korean cattle (Hanwoo). *J. Agric. Food Chem.* **2009**, *57*, 10898–10903. [[CrossRef](#)]

61. Jurie, C.; Cassar-Malek, I.; Bonnet, M.; Leroux, C.; Bauchart, D.; Boulesteix, P.; Pethick, D.W.; Hocquette, J.F. Adipocyte fatty acid binding protein and mitochondrial enzyme activities in muscles as relevant indicators of marbling in cattle. *J. Anim. Sci.* **2007**, *85*, 2660–2669. [[CrossRef](#)]
62. Huang, W.; Sherman, B.T.; Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2008**, *4*, 44–57. [[CrossRef](#)]
63. Xie, C.; Mao, X.; Huang, J.; Ding, Y.; Wu, J.; Dong, S.; Kong, L.; Gao, G.; Li, C.Y.; Wei, L. KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **2011**, *39*, 316–322. [[CrossRef](#)]
64. Schleinitz, D.; Klötting, N.; Lindgren, C.M.; Breitfeld, J.; Dietrich, A.; Schön, M.R.; Lohmann, T.; Dreßler, M.; Stumvoll, M.; McCarthy, M.I.; et al. Fat depot-specific mRNA expression of novel loci associated with waist-hip ratio. *Int. J. Obes.* **2014**, *38*, 120–125. [[CrossRef](#)] [[PubMed](#)]
65. Wang, H.; Zheng, Y.; Wang, G.; Li, H. Identification of microRNA and bioinformatics target gene analysis in beef cattle intramuscular fat and subcutaneous fat. *Mol. Biosyst.* **2013**, *9*, 2154–2162. [[CrossRef](#)] [[PubMed](#)]
66. Taye, M.; Yoon, J.; Dessie, T.; Cho, S.; Oh, S.J.; Lee, H.K.; Kim, H. Deciphering signature of selection affecting beef quality traits in Angus cattle. *Genes Genom.* **2018**, *40*, 63–75. [[CrossRef](#)] [[PubMed](#)]
67. Qiu, H.; Wang, F.; Liu, C.; Xu, X.; Liu, B. TEAD1-dependent expression of the FoxO3a gene in mouse skeletal muscle. *BMC Mol. Biol.* **2011**, *12*, 1. [[CrossRef](#)]
68. Wijga, S.; Bastiaansen, J.W.; Wall, E.; Strandberg, E.; de Haas, Y.; Giblin, L.; Bovenhuis, H. Genomic associations with somatic cell score in first-lactation Holstein cows. *J. Dairy Sci.* **2012**, *95*, 899–908. [[CrossRef](#)]
69. Raven, L.A.; Cocks, B.G.; Hayes, B.J. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genom.* **2014**, *15*, 62. [[CrossRef](#)]
70. Shin, D.; Lee, C.; Park, K.D.; Kim, H.; Cho, K.H. Genome-association analysis of Korean Holstein milk traits using genomic estimated breeding value. *Asian-Australas. J. Anim. Sci.* **2017**, *30*, 309–319.
71. Patton, S.; Kesler, E.M. Saturation in milk and meat fats. *Science* **1967**, *156*, 1365–1366. [[CrossRef](#)]
72. Taylor, R.W.; Turnbull, D.M. Mitochondrial DNA mutations in human disease. *Nat. Rev. Genet.* **2005**, *6*, 389–402. [[CrossRef](#)]
73. Shadyab, A.H.; LaCroix, A.Z. Genetic factors associated with longevity: A review of recent findings. *Ageing Res. Rev.* **2015**, *19*, 1–7. [[CrossRef](#)]
74. Fernández, A.I.; Alves, E.; Fernández, A.; de Pedro, E.; López-García, M.A.; Ovilo, C.; Rodríguez, M.C.; Sillio, L. Mitochondrial genome polymorphisms associated with longissimus muscle composition in Iberian pigs. *J. Anim. Sci.* **2008**, *86*, 1283–1290.
75. Wang, J.; Xiang, H.; Liu, L.; Kong, M.; Yin, T.; Zhao, X. Mitochondrial haplotypes influence metabolic traits across bovine inter- and intra-species cybrids. *Sci. Rep.* **2017**, *7*, 4179. [[CrossRef](#)] [[PubMed](#)]
76. Hauswirth, W.W.; Laipis, P.J. Mitochondrial DNA polymorphism in maternal lineages of Holstein cows. *Proc. Nat. Acad. Sci. USA* **1982**, *79*, 4686–4690. [[CrossRef](#)] [[PubMed](#)]
77. Sutarno; Cummins, J.M.; Greeff, J.; Lymbery, A.J. Mitochondrial DNA polymorphisms and fertility in beef cattle. *Theriogenology* **2002**, *57*, 1603–1610. [[CrossRef](#)]
78. Ron, M.; Genis, I.; Ezra, E.; Yoffe, O.; Weller, J.I.; Shani, M. Mitochondrial DNA polymorphism and determination of effects on economic traits in dairy cattle. *Anim. Biotechnol.* **1992**, *3*, 201–219. [[CrossRef](#)]
79. Schutza, M.M.; Freeman, A.E.; Lindberg, G.L.; Koehler, C.M.; Beitz, D.C. The effect of mitochondrial DNA on milk production and health of dairy cattle. *Livest. Prod. Sci.* **1994**, *37*, 283–295. [[CrossRef](#)]
80. Oleksyk, T.K.; Smith, M.W.; O'Brien, S.J. Genome-wide scans for footprints of natural selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2010**, *365*, 185–205. [[CrossRef](#)]

