

## Article

# Genome-Wide Signal Selection Analysis Revealing Genes Potentially Related to Sheep-Milk-Production Traits

Ruonan Li <sup>1,2,3</sup>, Yuhetian Zhao <sup>1,2</sup>, Benmeng Liang <sup>1,2</sup>, Yabin Pu <sup>1,2</sup>, Lin Jiang <sup>1,2</sup>  and Yuehui Ma <sup>1,2,\*</sup> 

<sup>1</sup> State Key Laboratory of Animal Biotech Breeding, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China; liruonan1998424@163.com (R.L.); jianglin@caas.cn (L.J.)

<sup>2</sup> Key Laboratory of Animal Genetics Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, Chinese Academy of Agricultural Sciences (CAAS), Beijing 100193, China

<sup>3</sup> Teaching and Research Centre (TERRA), Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium

\* Correspondence: mayuehui@caas.cn

**Simple Summary:** In our research, candidate genes related to sheep-milk production were revealed by a genome-resequencing analysis and a genome-signal-selection analysis, and a RT-qPCR experiment was performed to prove the expression levels of these candidate genes, the results showed that the *FCGR3A* gene's expression level had a significant negative relationship with sheep-milk production.

**Abstract:** Natural selection and domestication have shaped modern sheep populations into a vast range of phenotypically diverse breeds. Among these breeds, dairy sheep have a smaller population than meat sheep and wool sheep, and less research is performed on them, but the lactation mechanism in dairy sheep is critically important for improving animal-production methods. In this study, whole-genome sequences were generated from 10 sheep breeds, including 57 high-milk-yield sheep and 44 low-milk-yield sheep, to investigate the genetic signatures of milk production in dairy sheep, and 59,864,820 valid SNPs (Single Nucleotide Polymorphisms) were kept after quality control to perform population-genetic-structure analyses, gene-detection analyses, and gene-function-validation analyses. For the population-genetic-structure analyses, we carried out PCA (Principal Component Analysis), as well as neighbor-joining tree and structure analyses to classify different sheep populations. The sheep used in our study were well distributed in ten groups, with the high-milk-yield-group populations close to each other and the low-milk-yield-group populations showing similar classifications. To perform an exact signal-selection analysis, we used three different methods to find SNPs to perform gene-annotation analyses within the 995 common regions derived from the fixation index ( $F_{ST}$ ), nucleotide diversity ( $\theta\pi$ ), and heterozygosity rate (ZHp) results. In total, we found 553 genes that were located in these regions. These genes mainly participate in the protein-binding pathway and the nucleoplasm-interaction pathway, as revealed by the GO- and KEGG-function-enrichment analyses. After the gene selection and function analyses, we found that *FCGR3A*, *CTSK*, *CTSS*, *ARNT*, *GHR*, *SLC29A4*, *ROR1*, and *TNRC18* were potentially related to sheep-milk-production traits. We chose the strongly selected genes, *FCGR3A*, *CTSK*, *CTSS*, and *ARNT* during the signal-selection analysis to perform a RT-qPCR (Real time Quantitative Polymerase Chain Reaction) experiment to validate their expression-level relationship with milk production, and the results showed that *FCGR3A* has a significant negative relationship with sheep-milk production, while other three genes did not show any positive or negative relations. In this study, it was discovered and proven that the candidate gene *FCGR3A* potentially contributes to the milk production of dairy sheep and a basis was laid for the further study of the genetic mechanism underlying the strong milk-production traits of sheep.

**Keywords:** dairy sheep; whole-genome resequencing; RT-qPCR; milk production; gene expression



**Citation:** Li, R.; Zhao, Y.; Liang, B.; Pu, Y.; Jiang, L.; Ma, Y. Genome-Wide Signal Selection Analysis Revealing Genes Potentially Related to Sheep-Milk-Production Traits. *Animals* **2023**, *13*, 1654. <https://doi.org/10.3390/ani13101654>

Academic Editors: Marios Mataragas, Loulouda Bosnea and Maria Luisa Dettori

Received: 17 April 2023

Revised: 10 May 2023

Accepted: 11 May 2023

Published: 16 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The sheep is one of the earliest domesticated-farm-animal species and has experienced evolution and domestication over thousands of years [1]. Dairy sheep are traditionally farmed in southern Europe (France, Italy, Spain, Greece), central Europe (Hungary and the Czech and Slovak Republics), eastern Europe (Romania and Ukraine), and countries in the Middle East, such as Turkey and Iran [2]. Organized sheep-breeding programs were developed from at least the 1960s [3]. The most efficient selection scheme for local dairy sheep is based on the pyramidal management of the population, with the breeders of the nucleus flocks at the top, and pedigree, official milk recording, AI (artificial insemination) breeding, controlled natural mating, and breeding-value estimation (i.e., BLUP) are carried out to make genetic progress. In 2013, dairy small ruminants accounted for a minor part of the total agricultural output in France, Italy, and Spain (0.9 to 1.8%) and a larger part in Greece (8.8%) [4]. In these European countries, the dairy-sheep industry is based on local breeds and crossbreeds raised under semi-intensive and intensive systems, and it is concentrated in a few regions. While with the development of dairy-sheep breeding and the emergence of dairy products emerging, dairy ruminants have become a major part of European agricultural income [5,6]. The average flock size varies from small to medium (140 to 333 ewes/farm), and the average milk yield ranges from low to middle (170 to 500 L/ewe) [7], showing substantial space for improvement in relation to cows' milk [8]. Furthermore, sheep milk has higher protein, fat, lactose, total non-fat solids, and ash contents and a higher nutritional value than cows' and goats' milk [9], which makes it suitable for processing into various types of dairy products. Most sheep milk is sold to industries and then processed into traditional cheese products, most of which are made into Protected Denomination of Origin (PDO) cheeses for gourmets. The animals' udder health is also critical in sheep-milk quantity and quality. Mastitis is an inflammation of the mammary gland that is usually caused by pathogens, mainly bacteria, which develop in the udder tissue after infections in the teat canal. It is one of the most prevalent and costly diseases in the dairy industry due to the significant reductions in milk production and physical harm that it causes [10–12]. In previous studies, some quantitative trait loci (QTLs) and SNPs associated with SCC (somatic cell counts) that serve as markers of mastitis were identified based on linkage-disequilibrium analyses of different dairy-sheep breeds [10,13,14], leading to significant improvements in ewes' ability to resist mastitis [15,16].

The dramatic decreases in the costs of whole-genome sequencing (WGS) and RT-qPCR experiments on animals have made it possible to scan the complete genomes of thousands of animals. Using genome information makes it possible to explain parts of the total genetic variance that are difficult to measure, such as low-heritability, sex-limited, and postmortem traits. To help clarify the link between the genome sequence and real data of important traits, this research was designed to gain a better understanding of sheep lactating genes by comparing the SNP-variant information on different sheep breeds with significantly different dairy-production characteristics. These breeds significantly differ at two dairy levels, since there are high-milk-yield (East Friesian sheep, 700 kg/lactation, Dairy Meade sheep, 500 kg/lactation and Awassi sheep) and low-milk-yield sheep breeds (Hu sheep, small-tailed Han sheep, and Churra sheep) [17]. To this end, we performed a whole-genome sequencing (WGS) analysis of these sheep breeds, as we found that in previous studies, little research was performed using whole-genome sequencing on these dairy sheep populations. First, we re-sequenced the crossed breed of Dairy Meade sheep and small-tailed Han sheep. The information on this sheep breed will be uploaded to a public genome-sequence database for researchers to use freely (NCBI). In our research, a genetic-structure analysis was performed to establish the relationships and geographical distances between 10 sheep breeds, and then genome-wide-signal scans were carried out to identify significant genes associated with sheep-milk production. The genes validated by the RT-qPCR results and their relationship with milk production represent significant

progress in our knowledge of the gene-expression levels of lactating sheep and, thus, provide potentially valuable information for future dairy-sheep studies.

## 2. Material and Methods

### 2.1. Animals

One hundred and one sheep samples were sequenced in this study. This step was approved by the Animal Ethics Committee of China. In total, 41 sheep-ear samples were provided by the M-Natural Animal Husbandry Technology Company, including 36 sheep with high levels of milk production and 5 sheep with low levels of milk production. The high-milk-yield populations included Dairy Meade (DM) sheep (11), the crossbreed of Dairy Meade sheep and small-tailed Han sheep DM (F1) (15), and the crossbreed of DM (F1) and small-tailed Han sheep, DM (F2) (10). The low-milk-yield population consisted of small-tailed Han sheep (STHS) (5). The sequence data of the 60 remaining sheep were downloaded from (<https://www.ncbi.nlm.nih.gov/sra>, accessed on 15 March 2022). These sheep were from three high-milk-yield breeds, East Friesian (EFR) (10), Awassi (AWS) (2), and Dairy Meade (9), and four low-milk-yield breeds: Hu (HS) (14), Fin (FS) (9), Suffolk (SFK) (10), and Churra (CS) (6) (Table 1).

**Table 1.** 101 sheep information.

Breed	Abbreviation	Number	Milk Yield per Lactation/kg	Data Source	Classification	Reference
Dairy Meade sheep	DM	11	350–600	M-Natural farm	High yield	[18,19]
Dairy Meade (F1) sheep	DMF1	15	350	M-Natural farm	High yield	[18,20]
Dairy Meade (F2) sheep	DMF2	10	500	M-Natural farm	High yield	[18,20]
East Friesian sheep	EFR	10	500–700	NCBI/PRJNA624020	High yield	[21,22]
Dairy Meade sheep	DM	9	500	NCBI/PRJNA624020	High yield	[19]
Awassi sheep	AWS	2	300–500	NCBI/MW260509	High yield	[22]
Small-tailed Han sheep	STHS	5	100	M-Natural farm	Low yield	[23]
Churra sheep	CS	6	100–200	NCBI/PRJNA395499	Low yield	[24]
Hu sheep	HS	14	100–240	NCBI/PRJNA624020	Low yield	[25,26]
Suffolk sheep	SFK	10	–	NCBI/PRJNA624020	Low yield	[27]
Finland sheep	FS	9	–	NCBI/PRJNA624020	Low yield	[28]

### 2.2. Whole-Genome-Resequencing Analysis

We sequenced 41 sheep genomes at an average depth-of-coverage of 10X, which contained 4 sheep breeds. To this end, genomic DNA was extracted from ear tissue using the Wizard<sup>®</sup> Genomic DNA purification kit (Promega, A1125, Madison, WI, USA). The A260/A280 ratio using Nano-Drop ND-2000 (Thermo Fisher Scientific, MA, USA) was used to check the quality and integrity of the extracted DNA and agarose-gel electrophoresis was used to check its quality. Good-quality DNA from each collected sample was sequenced on the MGI-SEQ2000 platform from the Beijing Compass Biotechnology Company (Beijing, China). The satisfactory sequence data obtained were used to characterize individual genomes at a minimum of 10X depth of coverage, with more than 30 G of raw data from each sample. After quality control, data were used to perform genome mapping and SNP calling. Resequencing data from a further 60 sheep, downloaded from NCBI, were analyzed by Plink parameters on Linux (San Francisco, CA, USA) (Version V1.90). The details are listed in Section 2.3. All high-quality double-trimmed read pairs were aligned against the reference assembly Oar v.4.0 genome [29] ([https://www.ncbi.nlm.nih.gov/assembly/GC\\_A\\_000298735.2](https://www.ncbi.nlm.nih.gov/assembly/GC_A_000298735.2), accessed on 15 March 2022) using BWA software (<https://sourceforge.net/projects/bio-bwa/files/>, accessed on 15 March 2022) (version: 0.7.12). Paired-end reads that mapped to exactly the same position on the reference genome were deleted by the Mark Duplicates in Picard (picard-tools-1.56, at <http://picard.sourceforge.net>, accessed on 15 March 2022). Additional realignment of indels and SNPs was performed by using the Genome Analysis Toolkit (GATKV4.0) (<https://gatk.broadinstitute.org/hc/en-us>, accessed on 15 March 2022) [30] and sequence alignment (SAM tools) [31]. The GATK was used to identify SNP variation in each individual sample. The SNPs that did not meet the

following criteria were excluded: (1) SNP call rate > 99.6%; (2) minor-allele frequency > 0.01; (3) missing rate is lower than 0.1; and (4) all loci followed the Hardy–Weinberg rule; (5) linkage-disequilibrium loci were excluded ( $R^2 < 2$ ). All SNPs were annotated using the Bio Mart 4.0 (<https://asia.ensembl.org/info/data/biomart/index.html>, accessed on 15 March 2022) [32] based on the gene-reference genome provided by the Oar v.4.0 genome from NCBI.

### 2.3. Population Structures and Phylogenetic Analysis

Population structures among all sheep breeds were investigated using a total of 59,864,820 high-quality SNPs. The VCF (Visual Component Framework) files were transformed in the PLINK format by using VCF Tools (<https://vcftools.sourceforge.net/>, accessed on 15 March 2022) [33], and the SNP filtering was undertaken under the following conditions: minor-allele frequency (MAF) greater than 0.05, call-out rate higher than 0.9, missing-genotype rate > 0.05, and Hardy–Weinberg rules >  $1 \times 10^6$ . Pruning was carried out within windows of 50 SNPs and in 5 steps, using the indep-pairwise 50 5 2 parameters in PLINK [34] (`plink -file qc -indep 50 5 2 -chr-set 27 -recode -out qc_prune`) to obtain linkage-disequilibrium SNPs. This provided a total of 4,751,642 unlinked SNPs for PCA, ADMIXTURE, and neighbor-joining-tree-construction analysis. The PCA was conducted using the GCTA software (<https://yanglab.westlake.edu.cn/software/gcta/>, accessed on 15 March 2022) [35]; the first two components were plotted using with the R program gg plot package (<https://www.r-project.org/>, accessed on 15 April 2022) (version 4.1.0). Individual ancestry origins were predicted using replicates in the ADMIXTURE (<http://software.genetics.ucla.edu/admixture/download.html>, accessed on 15 April 2022) (version 1.3.0) [36], and assuming 2 to 4 ancestral populations (K), with the smallest CV error as the correct result. Neighbor-joining tree was constructed using the PHYLIP v3.69 (<https://evolution.genetics.washington.edu/phylip.html>, accessed on 15 April 2022) [37] based on the pairwise genetic distance matrix revealed by the PLINKv1.9 (`plink1 -sheep -file prune -cluster -distance-matrix -out`), and then MEGA7 (<https://www.megasoftware.net/>, accessed on 15 April 2022) [38] was used to construct the phylogenetic trees.

### 2.4. Genome-Wide Signal-Selection Scan and Gene Annotation

We used all SNPs that passed quality control to detect signatures of selection in high- and low-milk-yield sheep within 101 sheep genomes. Genome-wide signal-selection analysis was performed by using  $F_{ST}$  fixation indices (population-differentiation value), ( $\theta\pi$ ), the nucleotide-diversity ratio ( $\theta\pi\text{-HMY}/\theta\pi\text{-LMY}$ ), HMY (high milk yield), LMY (low milk yield), and the transformed heterozygosity score ( $ZH_P$ ). The window-based  $ZH_P$  method was calculated by the formula  $ZH_P = (H_p - \mu H_p) / \sigma H_p$ , where  $\mu$  is the overall average signal value of heterozygosity and  $\sigma$  is the standard deviation of all windows of each group. High-milk-yield group included 57 samples from five breeds (Dairy Meade sheep = 20, DM (F1) = 15, DM (F2) = 10, East Friesian sheep = 10, and Awassi sheep = 2), while low-milk-yield group included 44 individuals across five breeds (STHS = 5, Hu sheep = 14, Fin sheep = 9, Churra sheep = 6 and Suffolk sheep = 10). The  $F_{ST}$  fixation,  $ZH_P$  [39], and  $\log_2\theta\pi$  (high/low), were calculated within 100-kb sliding windows and 10-kb steps to obtain overlapping regions. The variation regions within the highest 5% of all three statistics were considered as candidate selections, and then candidate genes were annotated by a genomic-database search and annotating engine, Bio Mart (<https://asia.ensembl.org/info/data/biomart/index.html>, accessed on 15 March 2022) [40].

### 2.5. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes

Functional enrichments for Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using g: Profiler (<https://biit.cs.ut.ee/gprofiler/gost>, accessed on 1 June 2022) for all selected genes. The enrichment

significance was assessed using a condition for *g*: SCS ‘Set Counts and Sizes’, and the *p* value was set at <0.05.

### 2.6. Validation of RT-qPCR Experiment

We randomly selected 11 Lacaune sheep in from a farm in Belgium (<https://lesfauve.slaineux.wordpress.com/le-troupeau/>, accessed on 1 September 2022) to determine their genes’ expression levels. These were frequently selected in our research, since Lacaune sheep are among the most famous and heavily produced dairy sheep in France, and imported into Belgium. This farm efficiently recorded the milk-production data from different milk lactations and information on newly lambing ewes. Fresh milk was sampled from 11 newly lambed ewes. From each ewe, 50 mL of milk was sampled in non-RNA se tubs. All these steps complied with the requirements of European Animal Welfare Committee. The milk was stored at 4 °C during transport from farm to laboratory and processed immediately for somatic cell isolation. Subsequently, 50 µL of 0.5 M EDTA was added for each 50-mL milk sample. The samples were then centrifuged for 10 min at 2000× *g*, after which the supernatant containing cream and skin milk was removed, followed by washing in 10 mL PBS. Samples were then centrifuged a second time, the maximum amount of supernatant was removed, and the lysed samples were stored at −80 °C. The RNA was extracted according to the procedure of the isolation of RNA from non-fibrous tissue (Promega Inc., Madison, WI, USA), after which concentration and A260/A280 of RNA were tested; the standards were concentration >10 ng/µL and 2.2 > A260/A280 > 1.8. Satisfactory RNA was used to perform reverse transcription in accordance with Protocole reverse transcription (Promega Inc., Madison, WI, USA). The conditions of reverse transcription were as follows: 25 °C, 5 min; 42 °C, 60 min; 70 °C, 15 min, 4 °C, 20 min; 4 cycles in total. The cDNA was stored at −20 °C. Quantitative real-time PCR was performed using DNA, primers were designed to amplify target genes and reference genes, reference genes were selected from NCBI database, which were similarly expressed in all tissues, the genes’ premier pairs for RT-qPCR were designed by Eurogentec CO., (Brussel, Belgium), and RT-qPCR analyses were performed using Rotor Gene 6000 system (Corbett) and SYBR green (Gembloux, Belgium). Each amplification reaction contained 2 µL of cDNA and 18 µL of SYBR master mix (Thermo Fisher Scientific, MA, USA), including 500 nM of primers at a final volume of 20 µL. Reactions were performed as follows: 95 °C for 5 s, 60 °C for 30 s, 72 °C for 45 s, for 40 cycles in total, with data collection at the annealing step.

### 2.7. The Relationship between Gene Expression and Milk Production

Candidate genes from RT-qPCR experiment were used to perform a linear correlation analysis between genes’ relative expression values and milk production. The milk-production recordings were from Belgian farm with the same Lacaune sheep as those in 2022 (Data S4), as well as the production data from 2015 to 2021 (Data S5). We used the BLUPF90+ (<http://nce.ads.uga.edu/html/projects/programs/>, accessed on 15 March 2023) program to obtain estimated values (similar to estimated breeding values) of milk-production traits of each ewe from 2015 to 2021. Usually, higher estimated values meant the animal had stronger production traits, so we used these values to conduct a linear analysis with  $\Delta$ CT gene values and actual milk production from 2022, as well as the estimated values from period of 2015–2021 to show their relationships. Animal model is needed in BLUPF90+. In our research, the animal model was as follows:  $y_{ijk} = A_i + S_j + \beta x_{ijk} + e_{ijk}$ , where  $y_{ijk}$  is observed milk production,  $A_i$  is fixed effect of lactation days, lactation number, birth year, and milking times,  $S_j$  is random effect of animals,  $\beta x_{ijk}$  is regression coefficient of variance of 1.0, and  $e_{ijk}$  is residual effect. The estimations were given after data analysis using BLUPF90+ program.

### 3. Results

#### 3.1. Genomic Variants and Principal Component Analysis

We combined the genome sequences of 101 high- and low-production dairy sheep. The genomes were mapped and the SNPs were called using PLINK. After the joint calling and quality control, we detected 4,751,642 SNPs in total. To study the population stratification, three different approaches were employed, of which PCA is the first to be discussed here. The intention behind the use of PCA was to reduce the dimensionality of the genomic relationship matrix so that individuals (and breeds) were separated along different principal components (PCs). The first three principal components were able to explain most of the variation: PC1 explained about 50.26%, PC2 explained about 23.51%, and PC3 explained about 14.95%. Principal component 1 (PC1) differentiated the breeds in this study into three main clusters (Figure 1A): Cluster1 (high-yield breeds, EFR, DM, DMF1, and DMF2); Cluster2 (Mongolian low-yield breeds, HS, STHS, CS, and AWS) and Cluster3 (low-yield breeds, FS and SFK). However, overlap was observed among the DM sheep, DMF1 and DMF2, within Cluster 1. The reason for this may have been that these sheep have extremely close genetic relations. The PCA separated the different sheep breeds in this study, with SFK the most clearly separated due to its longer distance from and non-blood relations with the other sheep. In the PCA plots, the high-yield breeds were closer to the Mongolian sheep than to the low-yield sheep, and in the high-yield cluster, the EFR sheep were clearly separated from the DM sheep. The separation of the clusters corresponded to the geographical origins and blood mixes of these sheep breeds, and may have also revealed their genetic distance.

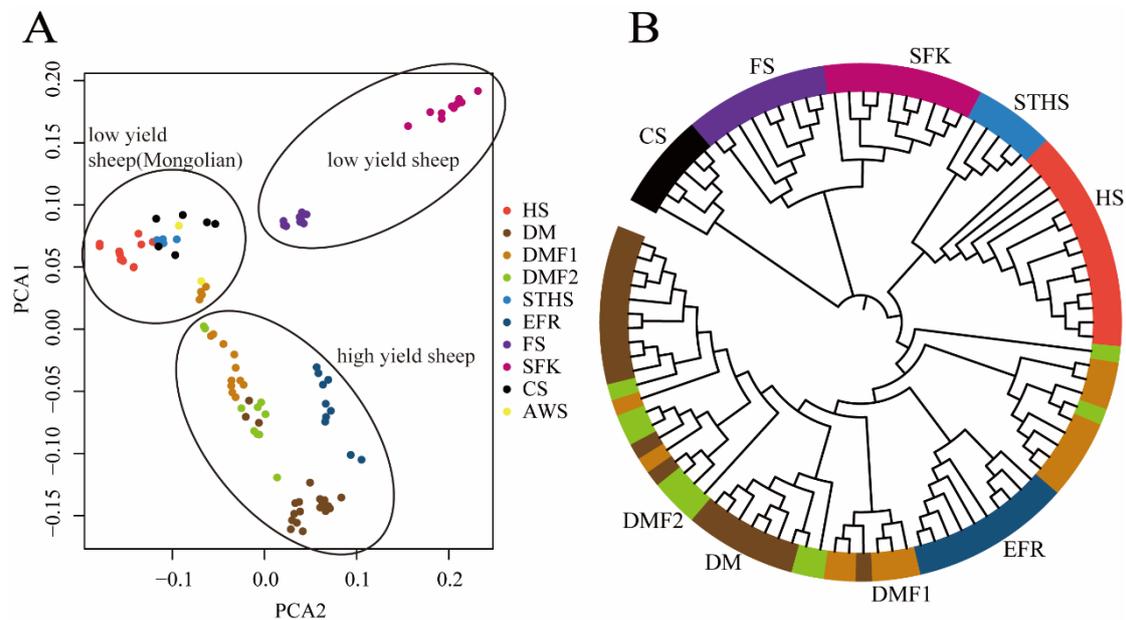
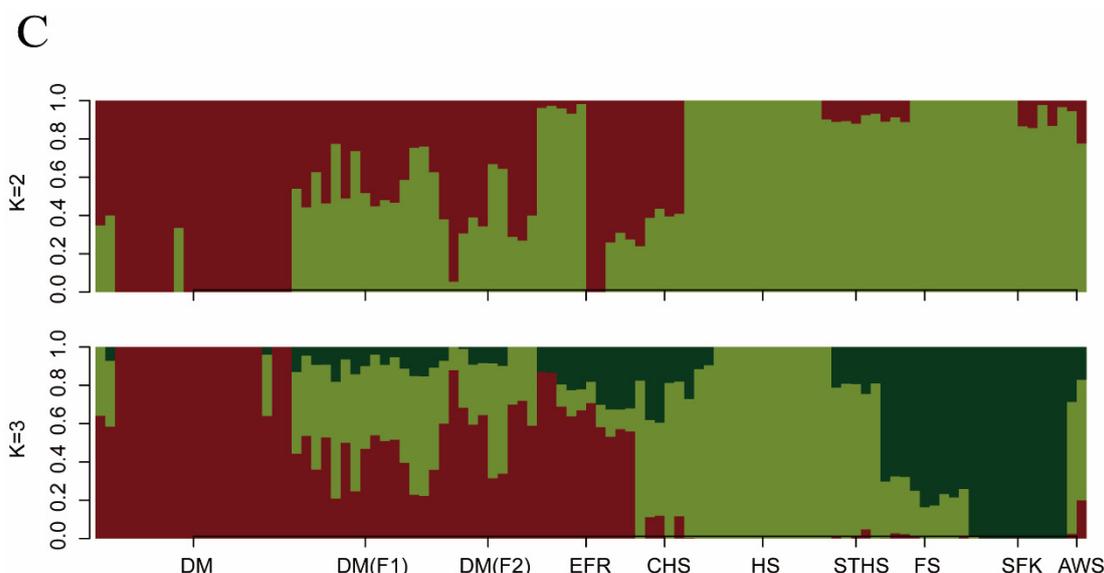


Figure 1. Cont.



**Figure 1.** (A) PCA analysis of ten sheep breeds, which were divided into three clusters: Cluster1, high-yield breeds (East Friesian sheep (EFR), Dairy Meade sheep (DM), Dairy Meade (DM)F1, Dairy Meade (DM)F2), Cluster2, Mongolian low-yield breeds (Hu sheep (HS), Small-Tail Han Sheep (STHS), Churra sheep (CS) and Awassi sheep (AWS)), and Cluster3, low-yield breeds (Finland sheep and Suffolk sheep). (B) The neighbor-joining tree divided one hundred and one sheep into nine different groups. The nine main branches with closest relationships, DM, DMF1, DMF2, EFR, CS, FS, HS, STHS, and SFK, breeds, were clustered into closer bunches. (C) Ancestor analysis was performed by ADMIXTURE; when  $K = 2$ , all sheep were divided into two groups (high- and low-yield sheep), and when  $K = 3$ , all sheep were separated into three groups (high-yield, low-yield, and Mongolian low-yield sheep).

### 3.2. Neighbor-Joining-Tree Analysis

In this study, the phylogenetic neighbor-joining tree divided the high- and low-production groups into nine different clusters (Figure 1B). The nine main branches were as follows: DM, DMF1, DMF2, EFR, CS, FS, HS, STHS, and SFK. According to the tree, we observed that the DM and DMF2 sheep branches were close to each other and that some of the DM, DMF1, and DMF2 sheep were mixed together, indicating their genetic similarity and probable sharing of a single ancestral line. The DM sheep is the hybrid offspring of EFR breeds and New Zealand sheep [41], so it was the second-nearest to the EFR sheep. Our phylogenetic results showed two breeds (DMF1 and DMF2) in the different sub-branches, which might have been due to the hybridization between the DM breeds and the STHS breeds. For the HS and STHS sheep, all the Mongolian sheep had similar genetic information, and their branches were close to each other. Similar to the PCA, there was a close genetic distance between the FS and SFK sheep, since there are shorter geographical distances between them. In the NJ-tree picture, the Awassi breed was not clearly separated, possibly due to its small numbers. It should be noted that the results of the phylogenetic analysis were reliable due to the large number of SNPs from the 10 sheep breeds used in this study.

### 3.3. Ancestor Analysis

To further explore the population structure among individuals from the different breeds in this study, a model-based hierarchical clustering analysis was undertaken for  $K$ -values of 2–7 (with  $K$  the user-defined number of biological ancestral populations). The cross-validation (CV) estimates revealed the  $K$ -value of 3 to be the best fit for the nine populations; it demonstrated the lowest CV error among all the  $K$  values. The bar plot of the ADMIXTURE results from the  $K$ -values of 2 and 3 is presented in Figure 1C. The analysis with  $K = 2$  separated the high-milk-yield sheep breeds from the low-yield breeds. When

$K = 3$ , the FS and SFK were separated from the other breeds in the low-yield group. The DM and EFR sheep had the highest milk yield among all the populations, the Mongolian CHS, HS, and STHS sheep had average milk production [26], and the FS and SFK sheep produced the least milk. It is known that FS and SFK sheep are clearly separated from the other two sheep populations, and this was validated by the results obtained from the PCA analysis and the admixture analysis.

### 3.4. Signal-Selection Analysis and Gene Ontology

We scanned the genomes of 101 high- and low-milk-yield sheep for signals of positive selection with different traits. To achieve this, we calculated three complementary statistics along the sheep reference genome Oar v.4.0 ([https://www.ncbi.nlm.nih.gov/assembly/GCA\\_000298735.2](https://www.ncbi.nlm.nih.gov/assembly/GCA_000298735.2), accessed on 15 March 2022) using 100-kb-long sliding windows and 15-kb step sizes. The first statistic was the population-differentiation index  $F_{ST}$ , for identifying genomic regions with different allelic frequencies between high and low groups. The second statistic measured the differences in nucleotide diversity between the two groups ( $\theta\pi(\text{high/low})$ ), with high or low  $\theta\pi$  values indicating positive selection of different milk-production sheep, respectively. The third measure was also used to calculate the selected candidate SNPs using the index of the transformed heterozygosity score ( $ZH_P$ ) (Figure 2). With regard to the limiting of the false-positive identifications, we considered the 5% most heavily overlapping regions from all three scans, which provided a total of 995 overlapping SNP regions and 553 protein-coding genes and small RNA. The selection candidates identified in the high- and low-yield groups are provided in Data S1. Most of the protein-coding genes were significantly enriched using the functional gene ontology and KEGG categories related to the protein binding, RNA binding, molecular transport, and cytokine-receptor activity ( $p = 9.23 \times 10^{-3}$ ) (Figure 3, Data S2). Importantly, through the gene annotation, the three highest 5% genomic regions providing the strongest signatures of selection containing *FCGR3A*, *CTSK*, *CTSS*, *ARNT*, *GHR*, and *SLC29A4* were selected as the strongest annotation genes by comparing the high- and low-milk-yield groups.

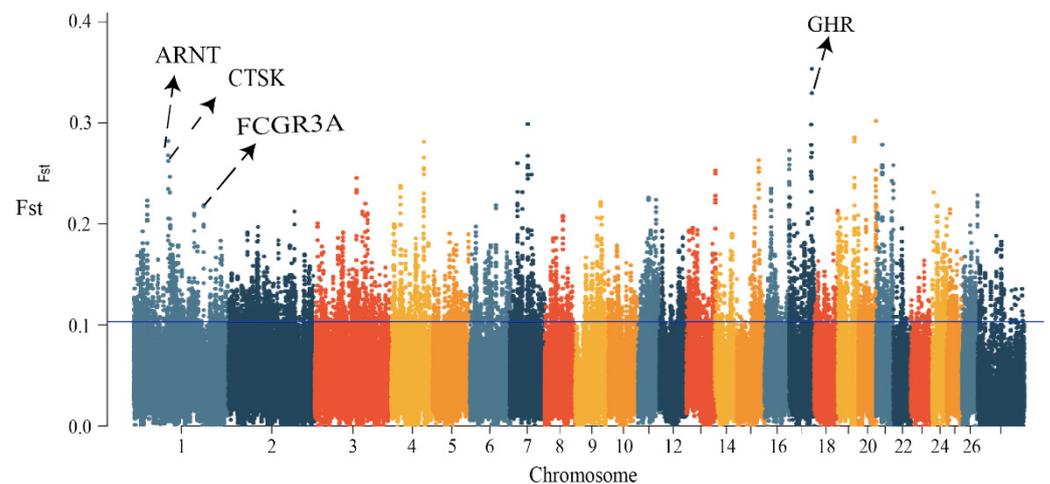
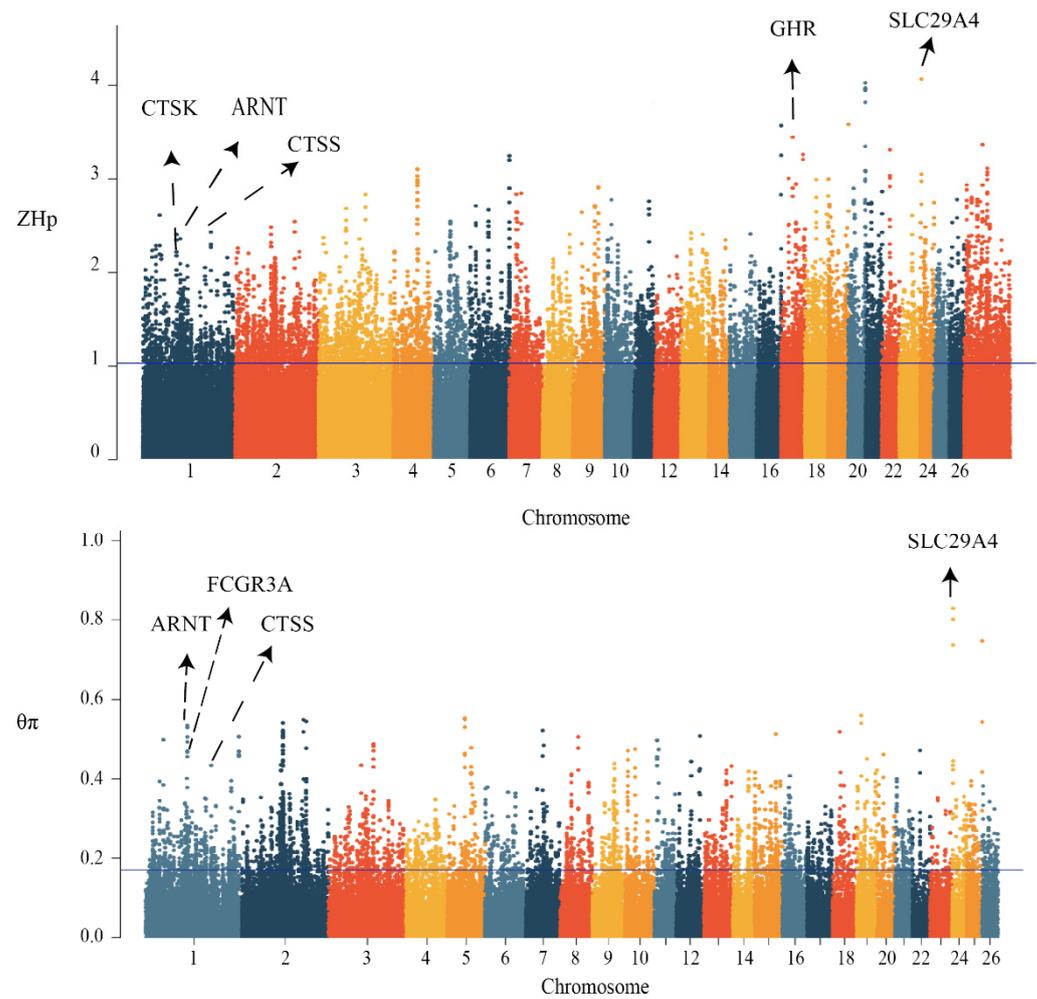
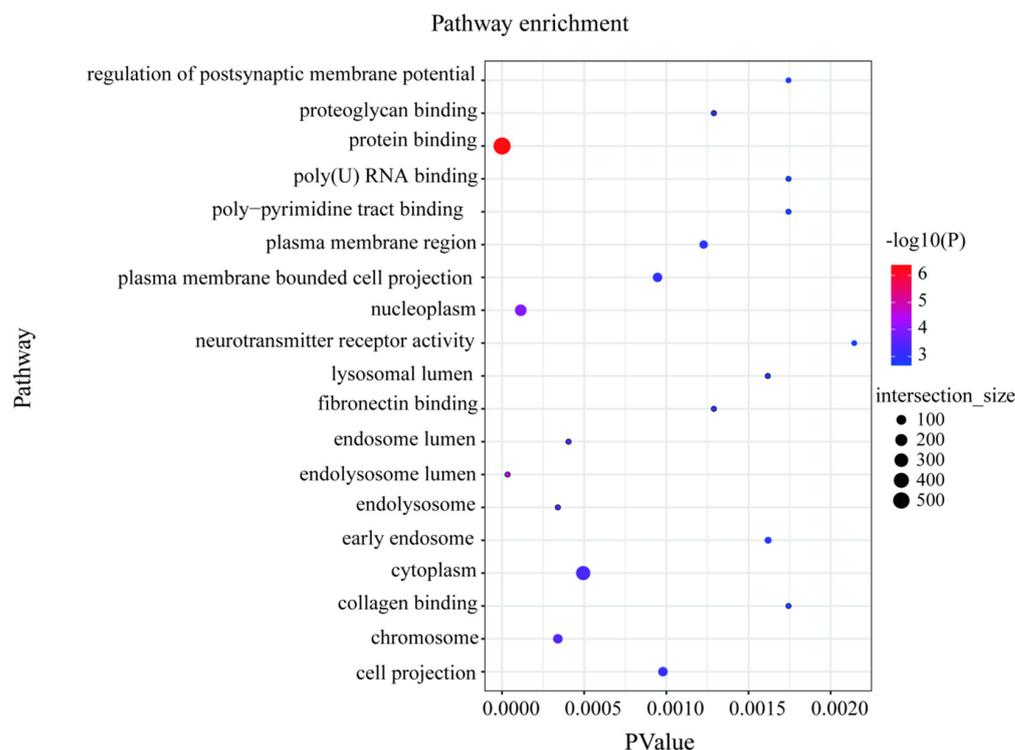


Figure 2. Cont.



**Figure 2.** Manhattan plots of genome-signature-selection analysis of total 101 sheep. Three statistical methods were applied to obtain positive selection scans for strong milk-production traits. High-yield sheep (HY) were compared with low-yield (LY) controls. The population genetic differentiation for  $F_{ST}$  values, the transformed heterozygosity score  $ZH_p$ , and the nucleotide-diversity  $\theta\pi$  ratios ( $\theta\pi_{HY}/\theta\pi_{LY}$ ) were calculated within 100-kb sliding windows. The significance threshold of selection signature was set to top 5% percentile outliers for each individual test and is indicated with horizontal dashed lines.



**Figure 3.** Positive selection of total overlapping genes was performed using functional analysis of Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathways. The dots colored blue to red means more genes participated in the functional pathway.

### 3.5. RT-qPCR Results

We chose the four genes that were most frequently selected by the signature-selection analysis shown in Figure 2, which were *FCGR3A*, *CTSK*, *CTSS*, and *ARNT*, according to the NCBI database and previous studies. We chose *GAPDH* gene as reference gene to perform RT-qPCR validation. The primer sequence is listed in Table 2. All the gene-expression values were derived from 11 ewes; the CT and  $\Delta$ CT values of each gene are listed in Data S3, with  $\Delta$ CT values meaning the relative expression of each gene.

**Table 2.** Premier sequences of target genes and reference genes.

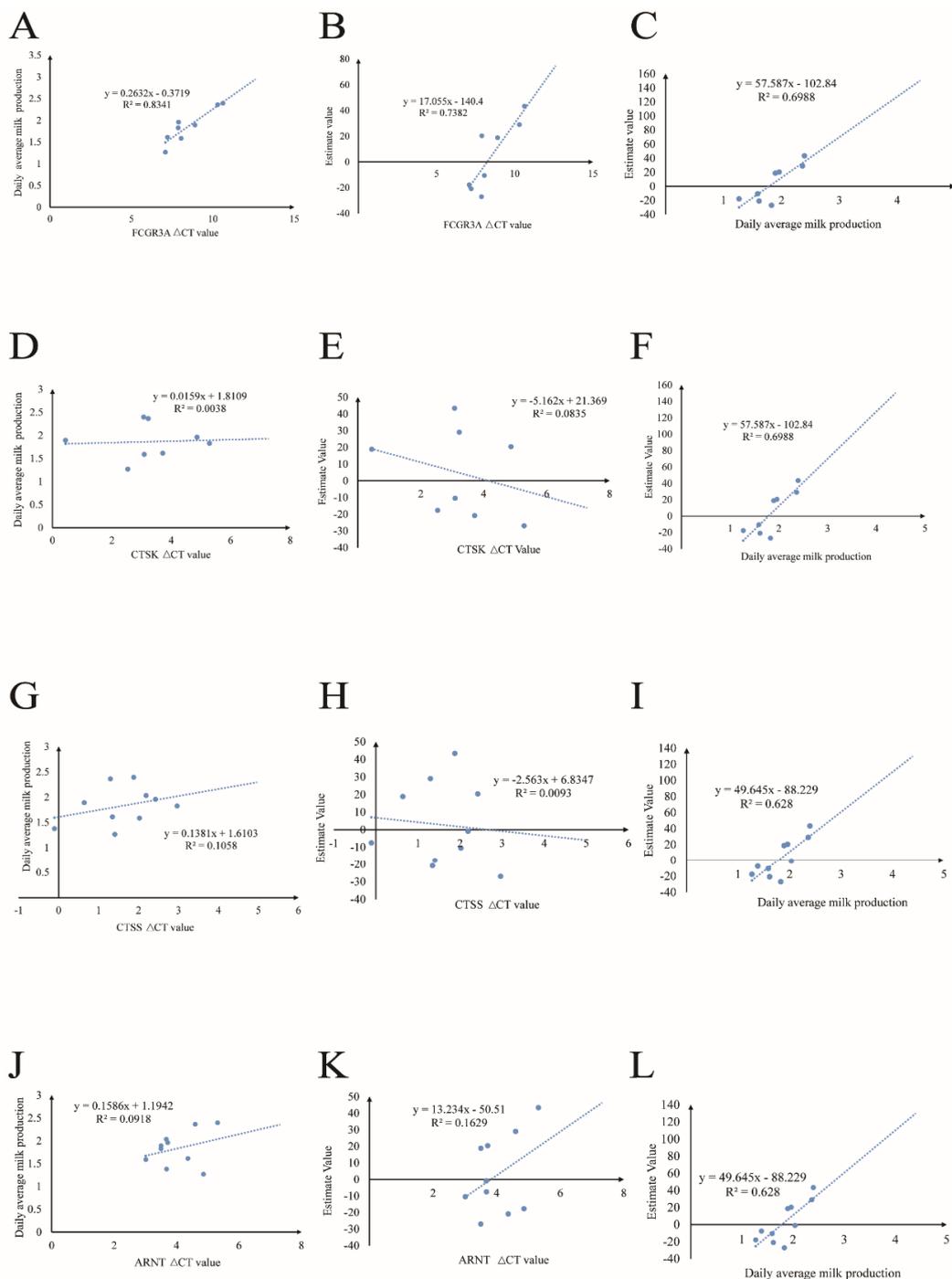
Gene	Primer Sequence	Length	Synthesis Scale	Format
FCGR3A-F *	CTTAGGACAAATGAAGGCTCTGA	23	10 nmol	TE-100 um
FCGR3A-R *	CTGCCTCTCCACCACGAAT	19	10 nmol	TE-100 um
ARNT-F	CAGGCCGGGTGGTATATGTC	20	10 nmol	TE-100 um
ARNT-R	TGGAAAGCTGCTCACGAAGT	20	10 nmol	TE-100 um
CTSS-F	AAGCTGGTGTCTCTGAGTGC	20	10 nmol	TE-100 um
CTSS-R	CGCCATTGCAGCCCTTATTC	20	10 nmol	TE-100 um
CTSK-F	ATGCAAGCCTGACCTCCTTC	20	10 nmol	TE-100 um
CTSK-R	CCAGTTTCTCCCCAGCTGT	21	10 nmol	TE-100 um
GAPDH-F	TCGGAGTGAACGGATTGGC	20	10 nmol	TE-100 um
GAPDH-R	CCGTTCTCTGCCTTGACTGT	20	10 nmol	TE-100 um

\*-F means the forward primers of genes; -R means the reverse primers of genes.

### 3.6. The Relationship between Gene Expression and Milk Production

The milk-production data were sourced from the Belgian farm. They covered the average daily milk production from four months in 2022: September, October, November, and December (Data S4). A linear analysis of the  $\Delta$ CT values of each gene was performed using the daily average milk production from 2022 and values estimated using the program of BLUPf90+ (Data S5), and the correlations between daily milk production and

the estimated values are also shown in Figure 4. The results showed that the *FCGR3A* gene had significant correlations with milk production and estimate values. The higher  $\Delta$ CT values meant lower expression values during the lactation period. It was revealed that the *FCGR3A* gene had a significantly negative relationship with milk production and the estimate values ( $R^2 = 0.8452$ ,  $R^2 = 0.7382$ ); furthermore, as predicted, the relationship between actual milk production in 2022 and the estimate value was positive ( $R^2 = 0.6988$ ). The *CTSK*, *CTSS*, and *ARNT* did not show significant differences between their expression values, the milk production values, and the estimate values, but the correlation coefficients between actual milk production in 2022 and the estimate values for the 2015–2021 milk production were positive ( $R^2 = 0.6988$ ,  $R^2 = 0.628$ ,  $R^2 = 0.628$ ).



**Figure 4.** (A–C) The correlations between *FCGR3A* gene's relative expression value and the daily

average milk production and estimated values, as well as the relevant coefficient of daily average milk production and estimated values. (D–F) The correlations between *CTSK* gene's relative expression value and the daily average milk production and estimated values, as well as the relevant coefficient of daily average milk production and estimate values. (G–I) The correlations between *CTSS* gene's relative expression value and the daily average milk production and estimated values, as well as the relevant coefficient of daily average milk production and estimate value. (J–L) The correlations between *ARNT* gene's relative expression value and the daily average milk production and estimated values, as well as the relevant coefficient of daily average milk production and estimated values.

#### 4. Discussion

In this study, we sequenced the genomes of 101 sheep with high and low levels of milk production. Using a neighbor-joining-tree analysis, we divided these sheep into 10 groups according to their genetic relationships and background, and then, using a principal component analysis and a structure analysis, these 10 groups were divided into three clusters: high-milk-yield sheep (EFR, DM, DMF1, DMF2), Mongolian low-milk-yield sheep (HS, CS, AWS, STHS), and low-milk-yield sheep (SFK and FS). The scanning of the genomes of high- and low-yield sheep breeds revealed that the *FCGR3A*, *ARNT*, *CTSK*, *CTSS*, and *GHR* genes were the strongest candidates. According to the RT-qPCR experiment and the analysis of the relationship between these candidate genes'  $\Delta$ CT values and milk production, the *FCGR3A* gene had a significant relation with Lacune sheep milk production. All of these genes have also been reported to be highly expressed during the lactation period in cattle [42] and buffalo, reflecting the similar biological functions of these genes when they are expressed in lactating mammary glands. The GO and KEGG analyses showed that most of these genes were significantly enriched for mammary-gland-specific GO terms (cytokine-receptor activity and protein binding), as well as establishing the cellular functions and protease bindings. These genes were found to have high and low levels of milk production, milk protein, milk fat, milk lactose, cheese traits, somatic cell counts, etc. For some of these traits, it is not easy to find a direct relationship with milk quantity. This was probably the reason why we did not find a significant correlation coefficient between the *CTSK*, *CTSS*, or *ARNT* expression values and milk production. Furthermore, as we found in many publications, some of the most significant genes in our research play a crucial role in sheep-mastitis traits and in the immune-response process, which improves milk production. Sheep-mammary-gland mastitis is also a common source of economic losses on sheep farms, so it is also important for us to focus on some of the genes related to mammary-disease resistance.

##### *Candidate Genes Potentially Associated with Some Milk Traits*

The *FCGR3A* (Fc fragment of IgG, low-affinity IIIa, receptor) gene is considered a novel and promising candidate for relieving stress, inflammation, and disease [43], as well as dairy-cattle mastitis. The binding of FCGRs to the Fc region of immunoglobulins mediates a variety of immune functions, such as antigen presentation, the clearance of immune complexes, the phagocytosis of pathogens, and cytokine production [44], they work together to resist viral injection to improve udder health and, thus, increase the production of milk. It is reasonable to suggest that lactation periods cause the increase in expression of a large number of genes, resulting in improvements in performance, as mastitis or *Staphylococcus aureus* infections often occur during lactation [45,46]. With regard to *CTSS* (Cathepsin S)-encoding protease, Sodhi et al. (2021) found a higher expression of most of the cathepsin genes (*CTSS*, *CTSD*, and *CTSK*) during the mid-to-late lactation stages, emphasizing their potential roles in milk synthesis, since the expressions of most of the proteases were higher during peak lactation [47]. The higher expression of cathepsin-encoding genes during late lactation stages could be attributed to the fact that *CTSS* plays a crucial role in mammary-gland involution [48]. Previous studies found the role of the *CTSK* gene in influencing cheese traits, as it is a protease-encoding casein that increases milk protein [49].

Similar results were found when evaluating the expression patterns of important protease-pathway-associated *CTSK* genes derived from different lactation stages in Sahiwal cows and Murrah buffalo. In both breeds, the RNA-expression levels of these genes were higher in the late lactation stages than in the early lactation stages. Lactation induces bone loss in order to provide sufficient calcium in milk, and to prevent this from taking place, the *CTSK* gene elevates its expression to increase osteocyte numbers, in order to maintain the balance of bone and milk calcium [50,51]. This proves that the *CTSK* gene plays an important role in milk-calcium traits. Furthermore, a differential expression analysis of Churra and Assaf sheep allowed us to notice some genes that were significantly and differentially expressed between the two breeds. These genes were mainly associated with protein-protease activity. Furthermore *CTSK* was differentially expressed and was selected as a candidate gene associated with cheese traits in this study [49]. These findings confirmed previous results, which highlighted the importance of the expression of genes encoding for some proteases in sheep milk. The aryl hydrocarbon receptor nuclear translocator (*ARNT*) can interact with the AHR aryl hydrocarbon receptor). The AHR is restricted to the cytoplasm in its unbound state [52,53]. Once activated, the AHR is combined with the nucleus of the *ARNT* and forms an active complex with the AHR nuclear translocator (*ARNT*) to alter the expressions of target genes [54]. It was reported that an association between AHR activation and *ARNT* causes changes in milk production [55]. More specifically, pregnant mice exposed to *TCDD* (an *ARNT* agonist) in vivo produced lower levels of the milk proteins,  $\beta$ -casein and whey protein [56]. The *GHR* is a regulator of developing growth and, as a growth hormone, it has important effects on carbohydrate, protein, and lipid metabolism. In cattle, mutations in *GHR* have been associated with milk yield and composition in Ayrshire, Holstein, and Jersey cattle. Dettori et al. (2018) reported that variations of the ovine *GHR* gene might affect milk-quality traits in Sarda sheep [57]. The rs55631463 *GHR* SNP genotype affects milk fat and protein yield, and the rs411154235 SNP is associated with lactose content in milk, as it encourages the transformation of glucose into glycogen to help glycogen cellular deposits [58].

In our study, except for *FCGR3A*, the strongest gene candidates, *CTSK*, *CTSS*, *ARNT*, were not found to be significantly relevant to milk-production quantity. This was potentially due to the small number of Lacune sheep selected to validate their function. Since our objective was to detect milk-yield genes, not milk-composition genes, it was difficult to directly find the relationships between these genes, which may participate in milk-protein or milk-fat synthesis. Thus, further measures need to be applied in the future to prove that these genes are correlated with other sheep-milk traits. In particular, Dairy Meade should be tested, since this breed, or crosses in which it is involved, comprised a high proportion of the high-milk-yield group.

## 5. Conclusions

In conclusion, our research is the first attempt to report a genome-signature-selection scan revealing genes that are associated with high- and low-milk-yield sheep breeds by using whole genome sequencing. Some of the most significant candidate genes associated with milk yield were identified through a combination of genome-signature-selection scanning and RT-qPCR experimentation. In the high- and low-yield groups, 553 genes were detected that were enriched with significant protein-receptors-combing pathways. Under selective pressure, we selected the four genes, *FCGR3A*, *CTSS*, *CTSK*, *ARNT*, that were most likely to be related to milk-production traits, in order to perform a correlations analysis, in which *FCGR3A* was found to have a significant relationship with daily milk yield, as it participates in the process of immune reaction. Furthermore, since higher gene-expression values mean severer mastitis, they are negatively correlated with milk production. These results were supported by the correlation between the *FCGR3A*  $\Delta$ CT values and the estimated values of the corresponding ewes. Therefore, these results could provide a better genetic perspective on the phenotypic differences between different-milk-yield groups for similar studies. Our findings provide an insight into the dynamic characterization of

sheep-mammary-gland gene expression, and the identified candidate genes can provide valuable information for future functional characterization, as well as contributing to a better understanding of the genetic mechanisms underlying the milk-production traits in sheep.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ani13101654/s1>, Data S1: Selection candidates by Overlapping Fst, ZHp and  $a^{+a-}$  analysis, Data S2: GO and KEGG pathway enriched by detecting genes, Data S3: CT values of target gene and reference gene, Data S4: Milk production of ewes in 2022, Data S5: Estimate value of experiment ewes from BLUPf90.

**Author Contributions:** Investigation, software, data analysis, and writing—original draft, R.L.; validation, resources, and picture draft, Y.Z.; methodology, B.L.; tissue sampling, Y.P.; writing—review and editing, and supervision, L.J. and Y.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by National Key Research and Development Program of China (SQ2022YFD1300062), the earmarked fund for Modern Agro-Industry Technology Research System (CARS-39-01), and the Agricultural Science and Technology Innovation Program of China (ASTIP-IAS01).

**Institutional Review Board Statement:** All experimental procedures were approved by the Animal Ethics Committee of the Chinese Academy of Agricultural Sciences (AEC-CAAS-20191106, Beijing, China).

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article's additional files; further inquiries can be directed to the corresponding author. The raw sequence reads of sheep were deposited in NCBI's Sequence Read Archive (SRA) database and accessible, through accession no. PRJNA914094, for genomic sequences of 41 dairy sheep.

**Acknowledgments:** Ruonan Li gratefully acknowledges financial support from the China Scholarship Council (no. 202003250150).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Ahmad, S.F.; Mehrotra, A.; Charles, S.; Ganai, N.A. Analysis of selection signatures reveals important insights into the adaptability of high-altitude Indian sheep breed Changthangi. *Gene* **2021**, *799*, 145809. [[CrossRef](#)] [[PubMed](#)]
2. Zygoyiannis, D. Sheep production in the world and in Greece. *Small Rumin. Res.* **2006**, *62*, 143–147. [[CrossRef](#)]
3. Seykora, A.J.; McDaniel, B.T. Genetics statistics and relationships of teat and udder traits, somatic cell counts, and milk production. *J. Dairy Sci.* **1986**, *69*, 2395–2407. [[CrossRef](#)] [[PubMed](#)]
4. Champion, F.; Dockes, A.; Lagriffoul, G.; Mottet, A.; Morin, E.; Neumeistier, D.; Perrot, C. *Bergers Demain en Brebis Laitières; études sur la production ovine laitière à l'horizon 2020; Éléments de diagnostic et propositions d'actions; Collection Résultats*, Institut de l'Élevage: Paris, France, 2013; pp. 7–47.
5. Correddu, F.; Gaspa, G.; Cesarani, A.; Macciotta, N.P.P. Phenotypic and genetic characterization of the occurrence of noncoagulating milk in dairy sheep. *J. Dairy Sci.* **2022**, *105*, 6773–6782. [[CrossRef](#)] [[PubMed](#)]
6. Martínez-Valladares, M.; Martín-Ramos, E.; Esteban-Ballesteros, M.; Balaña-Fouce, R.; Rojo-Vázquez, F.A. Effect of level of infection by gastrointestinal nematodes and anthelmintic treatment on milk yield in dairy sheep. *Parasite* **2021**, *28*, 71. [[CrossRef](#)]
7. Pulina, G.; Milán, M.J.; Lavín, M.P.; Theodoridis, A.; Morin, E.; Capote, J.; Thomas, D.L.; Francesconi, A.H.D.; Caja, G. Invited review: Current production trends, farm structures, and economics of the dairy sheep and goat sectors. *J. Dairy Sci.* **2018**, *101*, 6715–6729. [[CrossRef](#)]
8. Jandal, J.M. Comparative aspects of goat and sheep milk. *Small Rumin. Res.* **1996**, *22*, 177–185. [[CrossRef](#)]
9. Davies, G.; Genini, S.; Bishop, S.C.; Giuffra, E. An assessment of opportunities to dissect host genetic variation in resistance to infectious diseases in livestock. *Animals* **2009**, *3*, 415–436. [[CrossRef](#)]
10. Merz, A.; Stephan, R.; Jöhler, S. Staphylococcus aureus Isolates from Goat and Sheep Milk Seem to Be Closely Related and Differ from Isolates Detected from Bovine Milk. *Front. Microbiol.* **2016**, *7*, 319. [[CrossRef](#)]
11. Sutura, A.M.; Portolano, B.; Gerlando, R.D.; Sardina, M.T.; Mastrangelo, S.; Tolone, M. Determination of milk production losses and variations of fat and protein percentages according to different levels of somatic cell count in Valle del Belice dairy sheep. *Small Rumin. Res.* **2018**, *162*, 39–42. [[CrossRef](#)]

12. Banos, G.; Bramis, G.; Bush, S.J.; Clark, E.L.; McCulloch, M.E.B.; Smith, J.; Schulze, G.; Arsenos, G.; Hume, D.A.; Psifidi, A. The genomic architecture of mastitis resistance in dairy sheep. *BMC Genom.* **2017**, *18*, 624. [[CrossRef](#)] [[PubMed](#)]
13. Sutera, A.M.; Moscarelli, A.; Mastrangelo, S.; Sardina, M.T.; Tolone, M. Genome-Wide Association Study Identifies New Candidate Markers for Somatic Cells Score in a Local Dairy Sheep. *Front. Genet.* **2021**, *12*, 643531. [[CrossRef](#)]
14. Sutera, R.; Mastrangelo, D. Gerlando Genome-wide association studies for milk production traits in Valle del Belice sheep using repeated measures. *Anim. Genet.* **2019**, *50*, 311–314. [[CrossRef](#)]
15. Di Gerlando, R.; Sutera, A.M.; Mastrangelo, S.; Tolone, M.; Portolano, B.; Sottile, G.; Bagnato, A.; Strillacci, M.G.; Sardina, M.T.; Davoli, R. Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. *PLoS ONE* **2019**, *14*, e0215204. [[CrossRef](#)] [[PubMed](#)]
16. Marina, H.; Gutiérrez-Gil, B.; Esteban-Blanco, C.; Suárez-Vega, A.; Pelayo, R.; Arranz, J.J. Analysis of Whole Genome Resequencing Datasets from a Worldwide Sample of Sheep Breeds to Identify Potential Causal Mutations Influencing Milk Composition Traits. *Animals* **2020**, *10*, 1542. [[CrossRef](#)] [[PubMed](#)]
17. Bai, U.; Su, X.; Zheng, Z.; Zhang, L.; Ma, Y.; Dou, Y.; Zhang, X.; Su, G.; Li, G.; Zhang, L. Comparative metabolomics analysis of Small-Tailed Han and DairyMeade ovine milk. *Eur. Food Res. Technol.* **2021**, *247*, 2869–2876. [[CrossRef](#)]
18. Ma, Y.; Zheng, Z.; Zhang, L.; Dou, Y.; Bai, U.; Liu, X.; Zhang, X.; Su, X.; Zhang, L. Effects Study of Early Weaning Time on the Growth and Development of Dairy Sheep Lamb. *Int. J. Food Sci. Agric.* **2021**, *247*, 2869–2876. [[CrossRef](#)]
19. Su, X.; Zheng, Z.; Zhang, L.; Bai, U.; Ma, Y.; Dou, Y.; Zhang, X.; Su, G.; Zhou, N.; Li, G.; et al. Proteomic profiling of ovine milk after grading up. *J. Dairy Res.* **2021**, *88*, 201–204. [[CrossRef](#)] [[PubMed](#)]
20. Li, X.; Yuan, L.; Wang, W.; Zhang, D.; Zhao, Y.; Chen, J.; Xu, D.; Zhao, L.; Li, F.; Zhang, X. Whole genome re-sequencing reveals artificial and natural selection for milk traits in East Friesian sheep. *Front. Vet. Sci.* **2022**, *9*, 1034211. [[CrossRef](#)]
21. McKusick, B.; Thomas, D.; Berger, Y. Effect of weaning system on commercial milk production and lamb growth of East Friesian dairy sheep. *J. Dairy Sci.* **2001**, *84*, 1660–1668. [[CrossRef](#)]
22. Talafha, A.Q.; Ababneh, M.M. Awassi sheep reproduction and milk production. *Trop. Anim. Health Prod.* **2011**, *43*, 1319–1326. [[CrossRef](#)] [[PubMed](#)]
23. Wang, J.; Huitong, Z.; Hickford, J.G.H.; Hao, Z.; Shen, J.; Luo, Y.; Hu, J.; Liu, X.; Li, S. Comparison of the transcriptome of the ovine mammary gland in lactating and non-lactating small-tailed han sheep. *Front. Genet.* **2020**, *11*, 472. [[CrossRef](#)] [[PubMed](#)]
24. Gutiérrez-Gil, B.; El-Zarei, M.F.; Alvarez, L.; Bayón, Y.; De La Fuente, L.F.; San Primitivo, F.; Arranz, J.-J. Quantitative trait loci underlying milk production traits in sheep. *Anim. Gen.* **2009**, *40*, 423–434. [[CrossRef](#)] [[PubMed](#)]
25. Yue, G. Reproductive characteristics of Chinese Hu sheep. *Anim. Reprod. Sci.* **1996**, *44*, 223–230. [[CrossRef](#)]
26. Hao, Z.; Zhou, H.; Hickford, J.G.H.; Gong, H.; Wang, J.; Hu, J.; Liu, X.; Li, S.; Zhao, M.; Luo, Y. Identification and characterization of circular RNA in lactating mammary glands from two breeds of sheep with different milk production profiles using RNA-Seq. *Genomics* **2020**, *112*, 2186–2193. [[CrossRef](#)]
27. Slen, S.; Clark, R.; Hironaka, R. A comparison of milk production and its relation to lamb growth in five breeds of sheep. *Can. J. Anim. Sci.* **1963**, *43*, 16–21. [[CrossRef](#)]
28. Majjala, K. Productivity of pure Finnsheep in Finland and abroad. *Livest. Prod. Sci.* **1977**, *4*, 355–377. [[CrossRef](#)]
29. Li, X.; Yang, J.; Shen, M.; Xie, X.L.; Liu, G.J.; Xu, Y.X.; Lv, F.H.; Yang, H.; Yang, Y.L.; Liu, C.B.; et al. Whole-genome resequencing of wild and domestic sheep identifies genes associated with morphological and agronomic traits. *Nat. Commun.* **2020**, *11*, 2815. [[CrossRef](#)] [[PubMed](#)]
30. McCormick, R.F.; Truong, S.K.; Mullet, J.E. RIG: Recalibration and interrelation of genomic sequence data with the GATK. *G3 (Bethesda Md.)* **2015**, *5*, 655–665. [[CrossRef](#)]
31. Zhou, Z.; Jiang, Y.; Wang, Z.; Gou, Z.; Lyu, J.; Li, W.; Yu, Y.; Shu, L.; Zhao, Y.; Ma, Y.; et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* **2015**, *33*, 408–414. [[CrossRef](#)] [[PubMed](#)]
32. Morton, L.M.; Karyadi, D.M.; Hartley, S.W.; Frone, M.N.; Sampson, J.N.; Howell, R.M.; Neglia, J.P.; Arnold, M.A.; Hicks, B.D.; Jones, K.; et al. Subsequent Neoplasm Risk Associated With Rare Variants in DNA Damage Response and Clinical Radiation Sensitivity Syndrome Genes in the Childhood Cancer Survivor Study. *JCO Precis Oncol.* **2020**, *4*, 926–936. [[CrossRef](#)] [[PubMed](#)]
33. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [[CrossRef](#)] [[PubMed](#)]
34. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)]
35. 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*, 76–82. [[CrossRef](#)]
36. Shriner, D. Overview of Admixture Mapping. *Curr. Protoc. Hum. Genet.* **2017**, *94*, 1.23.21–21.23.28. [[CrossRef](#)]
37. Felsenstein, J. *PHYLIP: Phylogeny Inference Package Version 3.68*; Department of Genome Sciences and Department of Biology, University of Washington: Seattle, WA, USA, 2008.
38. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]

39. GX, E.; Basang, W.D.; Zhu, Y.B. Whole-genome analysis identifying candidate genes of altitude adaptive ecological thresholds in yak populations. *J. Anim. Breed Genet.* **2019**, *136*, 371–377. [[CrossRef](#)]
40. Frankish, A.; Diekhans, M.; Ferreira, A.M.; Johnson, R.; Jungreis, I.; Loveland, J.; Mudge, J.M.; Sisu, C.; Wright, J.; Armstrong, J.; et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **2019**, *47*, D766–D773. [[CrossRef](#)]
41. Luo, R.; Zhang, X.; Wang, L.; Zhang, L.; Li, G.; Zheng, Z. GLIS1, a potential candidate gene affect fat deposition in sheep tail. *Mol. Biol. Rep.* **2021**, *48*, 4925–4931. [[CrossRef](#)]
42. Suárez-Vega, A.; Gutiérrez-Gil, B.; Klopp, C.; Robert-Granie, C.; Tosser-Klopp, G.; Arranz, J.J. Characterization and Comparative Analysis of the Milk Transcriptome in Two Dairy Sheep Breeds using RNA Sequencing. *Sci. Rep.* **2015**, *5*, 18399. [[CrossRef](#)]
43. Li, C.; Cai, W.; Zhou, C.; Yin, H.; Zhang, Z.; Loo, J.J.; Sun, D.; Zhang, Q.; Liu, J.; Zhang, S. RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population. *Sci. Rep.* **2016**, *6*, 26813. [[CrossRef](#)]
44. Clynes, R.A.; Towers, T.L.; Presta, L.G.; Ravetch, J.V. Inhibitory Fc receptors modulate in vivo cytotoxicity against tumor targets. *Nat. Med.* **2000**, *6*, 443–446. [[CrossRef](#)] [[PubMed](#)]
45. Wang, D.; Liu, L.; Augustino, S.M.; Duan, T.; Hall, T.J.; MacHugh, D.E.; Dou, J.; Zhang, Y.; Wang, Y.; Yu, Y. Identification of novel molecular markers of mastitis caused by *Staphylococcus aureus* using gene expression profiling in two consecutive generations of Chinese Holstein dairy cattle. *J. Anim. Sci. Biotechnol.* **2020**, *11*, 98. [[CrossRef](#)]
46. Seroussi, E.; Blum, S.E.; Krifucks, O.; Shirak, A.; Jacoby, S.; Leitner, G. Basal levels of CD18 antigen presenting cells in cow milk associate with copy number variation of Fc gamma receptors. *Genes* **2020**, *11*, 952. [[CrossRef](#)]
47. Sodhi, M.; Sharma, M.; Sharma, A.; Verma, P.; Mohanty, A.; Kataria, R.S.; Shandilya, U.K.; Kumari, P.; Mukesh, M. Expression profile of different classes of proteases in milk derived somatic cells across different lactation stages of indigenous cows (*Bos indicus*) and riverine buffaloes (*Bubalus bubalis*). *Anim. Biotechnol.* **2023**, *34*, 15–24. [[CrossRef](#)] [[PubMed](#)]
48. Seol, M.B.; Bong, J.J.; Baik, M. Involvement of Cathepsin D in Apoptosis of Mammary Epithelial Cells. *Asian-Australas. J. Anim. Sci.* **2006**, *19*, 1100–1105. [[CrossRef](#)]
49. Suárez-Vega, A.; Gutiérrez-Gil, B.; Arranz, J.J. Transcriptome expression analysis of candidate milk genes affecting cheese-related traits in 2 sheep breeds. *J. Dairy Sci.* **2016**, *99*, 6381–6390. [[CrossRef](#)]
50. Wysolmerski, J.J. Osteocytes remove and replace perilacunar mineral during reproductive cycles. *Bone* **2013**, *54*, 230–236. [[CrossRef](#)]
51. Lotinun, S.; Ishihara, Y.; Nagano, K.; Kiviranta, R.; Carpentier, V.T.; Neff, L.; Parkman, V.; Ide, N.; Hu, D.; Dann, P. Cathepsin K-deficient osteocytes prevent lactation-induced bone loss and parathyroid hormone suppression. *J. Clin. Investig.* **2019**, *129*, 3058–3071. [[CrossRef](#)] [[PubMed](#)]
52. Tijet, N.; Boutros, P.C.; Moffat, I.D.; Okey, A.B.; Tuomisto, J.; Pohjanvirta, R. Aryl hydrocarbon receptor regulates distinct dioxin-dependent and dioxin-independent gene batteries. *Mol. Pharm.* **2006**, *69*, 140–153. [[CrossRef](#)]
53. Girolami, F.; Spalenza, V.; Carletti, M.; Sacchi, P.; Rasero, R.; Nebbia, C. Modulation of aryl hydrocarbon receptor target genes in circulating lymphocytes from dairy cows bred in a dioxin-like PCB contaminated area. *Sci. Total Environ.* **2013**, *450*, 7–12. [[CrossRef](#)] [[PubMed](#)]
54. Basham, K.J.; Leonard, C.J.; Kieffer, C.; Shelton, D.N.; McDowell, M.E.; Bhonde, V.R.; Looper, R.E.; Welm, B.E. Dioxin exposure blocks lactation through a direct effect on mammary epithelial cells mediated by the aryl hydrocarbon receptor repressor. *Toxicol. Sci. Off. J. Soc. Toxicol.* **2015**, *143*, 36–45. [[CrossRef](#)] [[PubMed](#)]
55. Collins, L.L.; Lew, B.J.; Lawrence, B.P. TCDD exposure disrupts mammary epithelial cell differentiation and function. *Reprod. Toxicol.* **2009**, *28*, 11–17. [[CrossRef](#)] [[PubMed](#)]
56. Vorderstrasse, B.A.; Fenton, S.E.; Bohn, A.A.; Cundiff, J.A.; Lawrence, B.P. A novel effect of dioxin: Exposure during pregnancy severely impairs mammary gland differentiation. *Toxicol. Sci. Off. J. Soc. Toxicol.* **2004**, *78*, 248–257. [[CrossRef](#)] [[PubMed](#)]
57. Dettori, M.L.; Pazzola, M.; Paschino, P.; Amills, M.; Vacca, G.M. Association between the GHR, GHRHR, and IGF1 gene polymorphisms and milk yield and quality traits in Sarda sheep. *J. Dairy Sci.* **2018**, *101*, 9978–9986. [[CrossRef](#)] [[PubMed](#)]
58. Kim, S.-H.; Park, M.-J. Effects of growth hormone on glucose metabolism and insulin resistance in human. *Ann. Pediatr. Endocrinol. Metab.* **2017**, *22*, 145. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.