

Bayesian-genome-wide association study and post-GWAS on reproductive traits of Holstein dairy cattle

Jeyran Jabbari Tourchi¹, Sadegh Alijani¹, Seyed Abbas Rafat¹ and Mokhtar Ali Abbasi²

Research Article

Cite this article: Jabbari Tourchi J, Alijani S, Rafat SA and Abbasi MA (2024). Bayesian-genome-wide association study and post-GWAS on reproductive traits of Holstein dairy cattle. *Journal of Dairy Research* **91**, 146–155. <https://doi.org/10.1017/S0022029924000414>

Received: 7 April 2023

Revised: 4 April 2024

Accepted: 23 April 2024

Keywords:

Bayesian method; candidate genes; gene set enrichment analysis; reproductive traits; windows

Corresponding author:

Sadegh Alijani;

Email: sad-ali@tabrizu.ac.ir

¹Department of Animal Science, Faculty of Agricultural Sciences, University of Tabriz, Tabriz, Iran and

²Animal Science Research Institute of Iran, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran

Abstract

Bayesian multiple regression models are often used for genomic selection, where all markers are adjusted simultaneously as random effects to reduce the false-discovery rate. The purpose of this research was to identify individual candidate genes for some reproductive traits in Holstein cattle through genome wide association studies (GWAS) using a Bayesian method in combination with post-GWAS analysis. Reproductive traits included: days open (DO), pregnancy rate (PR), calving interval (CI) and age at first calving (AFC). The animals were genotyped using single-nucleotide polymorphism (SNP) panels of different densities imputed to a 50 K SNP density. After quality control, we included 2400 genotyped animals. According to the Bayesian analysis, there were 19 windows with an explained additive genetic variance of >0.1 percent for CI, DO, AFC and PR in Holstein cattle, which were 3, 3, 6 and 7, respectively. Using Bayesian analysis, 79 genes were located within or nearby (250-kb) 19 significant SNPs/windows in the *Bos taurus* autosomes. Among these genes, we identified 25 candidate genes for reproductive traits, namely CHD7, CLVS1, EVX2, MAT2B, NUDCD2, GPR39, NCKAP5, LYPD1, HOXD13, SEMA5B, CCNG1, SEMA5A, BRF1, PSEN2, CACHD1, SUGTA, ELF1, SNORA70, AKT1, TM2D1, SLF1, MCTPA, PAB2A, MTRF1 and ADCY2. Additionally, another nine candidate genes (CLVS1, GPR39, CENPF, AMOT, ARF1, CCDC186, ADCY2, BDP1 and AMOTL1) were identified in the network cluster analysis as hub genes for reproductive traits. The results of gene set enrichment analysis (GSEA) and pathway analysis, suggest that the most important gene ontology term involving cellular metabolic process was related to the AFC trait. To summarize, Bayesian methods were used to identify SNPs and candidate genes that could be useful in genomic selection to improve reproductive traits of Holstein dairy cattle.

In recent times, genetic progress in production animal breeding has increasingly become the result of selection based on genomic information of important economic traits (Miglior *et al.*, 2017). Genome-wide association studies (GWAS) can be used to identify genomic regions or causative mutations that account for a significant proportion of the genetic variance associated with a trait. The GWAS approach is based on the assessment of correlations between genetic markers and the trait of interest in a large number of samples (Jiang *et al.*, 2019). The ability to identify single-nucleotide polymorphisms (SNPs) that are associated with economically important traits is the main advantage of GWAS in dairy cattle breeding. Various statistical models are used in GWAS, including Bayesian multiple regression and generalized linear mixed models. Since most economic traits are controlled by multiple loci, estimates of marker effects by linear mixed models have low accuracy. Another challenge of GWAS is the multiple testing correction. Many loci may fail to overcome the strict criterion of significance test because the normal Bonferroni correction is too cautious (Wang *et al.*, 2016). When the goal is reduction of experimental power to achieve high stringency, traditional multiple testing corrections such as Bonferroni do not take into account the genotype dependencies among SNPs in linkage disequilibrium (Fernando *et al.*, 2004). An alternative approach is to use Bayesian multiple regression models for genomic selection, where all markers are adjusted simultaneously as random effects to reduce the false-discovery rate (Kemper *et al.*, 2015). A clear advantage of multi-locus models is that they do not require Bonferroni correction, making them more robust approaches for GWAS. Although Bayesian methods use shrinkage approaches to handle a large sample size of markers, they find difficulty when the number of markers increases due to memory allocation or computation time limitations. Furthermore, these models suffer from multicollinearity problems in situations where marker density is unusually high (Segura *et al.*, 2012). Some of these problems can be solved using the Bayesian multiple regression model, which was initially designed for genome prediction.

© The Author(s), 2024. Published by Cambridge University Press on behalf of Hannah Dairy Research Foundation



Bayesian methods allow the effects of all genotyped markers to be adjusted simultaneously by accounting for different prior distributions of marker effects. Furthermore, additional sources of information such as omics data or annotation information can be incorporated (Sollero *et al.*, 2017). The purpose of the research described here was to identify individual candidate genes for some reproductive traits in Holstein cattle through a GWAS using a Bayesian method in combination with post-GWAS analysis.

Materials and methods

Phenotypic data

Pedigree and phenotypic records for insemination and calving dates of Iranian Holstein dairy cattle were obtained from the Animal Breeding Center of Iran. The records of Holstein heifers and cows (lactation 1–3) were collected between 2000 and 2021. Reproductive traits included were days open (DO), pregnancy rate (PR), calving interval (CI) and age at first calving (AFC). A summary of the pedigree file that was used in genetic analysis is presented in online Supplementary Table S1. Descriptive statistics for AFC and reproductive traits are presented in online Supplementary Table S2.

Genotypic data

The animals were genotyped using SNP panels of different densities (Illumina and GeenSeek Genomic companies) imputed to a 50 K SNP density. The following quality control (QC) criteria were used to evaluate the processes: deviations from Hardy–Weinberg equilibrium (P -value $< 10^{-6}$), animal call rate ($> 90\%$), SNP call rate ($> 95\%$) and minor allele frequency (> 0.01). We performed quality control using the PLINK software. Additionally, duplicate, twin and animals who failed parentage tests were eliminated from the analysis. QC was initially carried out on the 54 001 SNP array and SNPs on chromosome X or with an unclear physical location were eliminated. Thus, 51 185 markers were found on the 54 001 SNP chip. The density of each SNP chip after imputation was 51 185. The 2400 animals genotyped with 41 099 SNP markers demonstrated success with QC filtering for association analysis. The SNP number and animal number of the final analysis are presented in online Supplementary Table S3.

Statistical model

To estimate genetic parameters for the AFC trait, the following sire model was fitted:

$$y_{ik} = \mu + HYS_i + \alpha_k + e_{ik} \quad (1)$$

Where: y_{ik} is the k th observation; μ is the population mean; fixed effects include: HYS_i : herd-years-season; α_k is additive genetic $\alpha_k \sim N(0, A\alpha_a^2)$, and e_{ik} : is the random residual error $e_{ik} \sim N(0, I\alpha_e^2)$.

Using the following repeatability single-trait model, the genetic variance components, breeding values, and corresponding reliability reproductive traits were estimated:

$$y_{ijklmk} = \mu + HYS_i + b(\text{Age}_j - \overline{\text{Age}\dots}) + P_l + Pe_m + \alpha_k + e_{ijklmk} \quad (2)$$

Where Age_j : is the k th covariate of age at calving; Pe_m : is the random effect of the permanent environment; P_l : is the fixed

effect of parity (3 levels); other terms are defined above. Variance components and heritability were estimated by Bayesian method and using Gibbs2f90 software from the BLUPF90 program family. We used the deregressed proofs (DRP) as pseudo phenotypes for DGV and GWAS prediction in the next multiple-step method. To calculate DRP, DRP packages from the R programming language were utilized.

Bayesian genome-wide association analysis

Using the Bayes- CP_i method, GWAS were performed with GenSel software, according to the presented model:

$$y_i = \mu + \sum_{j=1}^k Z_j U_j \delta_j + e_i \quad (3)$$

where y_i is DRP on animal i for the respective trait; μ is the population mean; k is the number of markers; Z_j is the vector includes genotypes for the SNP $_j$ (coded as: -10 homozygous dominant, -10 homozygous recessive and 0 heterozygous); U_j is the random substitution effect for marker j ; δ_j is a random 0/1 variable showing the absence or presence of marker j , and e_i is the random residual effects assumed normally distributed $N(0, \alpha_e^2)$. We used the Markov Chain Monte Carlo (MCMC) method with the GenSel implementation to obtain the posterior distributions of the SNP effects. After a burn-in period of 1000 iterations in which data were discarded, 51 000 iterations were performed to determine the posterior mean effect of each SNP. Multiple regression models with $\pi = 0.9$ were fitted simultaneously in each MCMC iteration of the Bayesian variable selection process. Finally, we displayed the Manhattan plot of genetic variance using the R program.

Genome-wide association study and functional enrichment analyses

Following GWAS analysis, gene set enrichment analysis (GSEA) was used to identify molecular pathways and putative functional categories associated with reproductive traits in Holstein cattle. Information from the NCBI platform (National Science for Biotechnology Information) was utilized in this context. The Ensembl (<https://www.ensembl.org>; NCBI, <https://www.ncbi.nlm.nih.gov/genome/>; genome-browser <https://genome.ucsc.edu>) platform was used to assess the biological functions of the genes and genetic ontology. STRING (<https://string-db.org>) databases were employed for GSEA. The BiomaRt package was used for positional genes were located within or nearby 250 kb of the significant SNPs.

Results and discussion

Genetic parameter estimation

Since reproductive traits are known to have low heritability, achieving increased genetic progress compared to higher heritable variables is more challenging. Estimates of heritability for reproductive traits (AFC, CI, DO, and PR) in Iranian Holstein cattle are presented in Table 1. Compared to other reports on Holstein cows the heritability value for the AFC trait was higher, however, other traits showed comparable heritability. Previous studies have estimated heritability similarly for the AFC trait (0.19) in dairy cattle (Ali *et al.*, 2019), and for reproductive traits

Table 1. Variance components, and heritability estimates for fertility traits in Iranian Holstein cattle

Trait	Additive genetic variance	Permanent variance	Residual variance	Heritability
AFC	1.20	–	5.300	0.18
CI	316.67	192.34	5986.3	0.05
DO	210.21	115.83	4518.0	0.04
PR	12.57	7.2222	276.21	0.04

in Iranian Holstein cows (Faraji *et al.*, 2011). The heritability of the AFC trait varies between 0.05 and 0.74 in different reports (Pirlo *et al.*, 2000). Moreover, heritability is a specific property that is unique for population, environmental conditions and imposed management decisions.

Genome-wide association analysis

Using GWAS based on Bayesian methods, we were able to determine which SNP marker regions were most significant in the Illumina BovineSNP50 panel. The relationship between SNPs and reproductive traits in Holstein cattle is represented by a Manhattan plot in Fig. 1, and Table 2 shows the results of the GWAS for reproductive traits. GWAS results include chromosomal location and proportion of variance of 1-Mb genome windows by informative regions (more than 1.0%). According to the Bayesian analysis, there were 19 windows with an explained additive genetic variance of >0.1 percent for CI, DO, AFC and PR in Holstein cattle, which were 3, 3, 6, and 7, respectively. Quantitative trait loci (QTL) associated with the target traits were defined as windows that account for $\geq 1.0\%$ of the total variance.

Using Bayesian analysis for reproductive traits in Holstein cows

Using Bayesian analysis, 79 genes were located within or nearby (250-kb) to 19 significant SNPs/windows in the *Bos taurus* autosomes. The windows with the most informative located within or nearby (250-kb) detected CI, DO, AFC and PR traits on (BTA7) at 75 Mb 0.18%, (BTA3) at 84 Mb 0.23%, (BTA3) at 17 Mb 0.21% and (BTA21) at 69 Mb 0.28%, respectively. Significant SNPs reproductive traits in Iranian Holstein cattle are shown in Table 2.

Due to the high frequency of QTL controlling reproductive traits in this region, we identified 132 important SNPs through Bayesian analysis. There is particular interest in the AFC trait on BTA3, as shown in Fig. 1b. The region of BTA3 at 1-Mb was the most significant window region. However, significant SNPs were also associated with heifer reproductive trait within the SPRR4, CLVS1, SNORA70, CCNG1, NUDCD2 and MAT2B genes. A total of 23 genes were found within or near the 250 kb of these SNPs.

The association analysis detected significant SNPs related to the PR trait, and we identified 152 significant SNP, whilst BTA21 at 1-Mb was the most significant window region as shown in Fig. 1d. Among the significant SNPs associated with the PR trait, SNP/windows on chromosome 21 were located within the BRF1 gene. BRF1 gene is one of the important genes affecting fertility traits (Butler *et al.*, 2020).

BTA7 at 1-Mb was the most significant window region, with 70 significant SNPs identified for the CI trait, as shown in Fig. 1c. We identified 57 significant SNP for the DO trait, and BTA3 at 1-Mb was the most significant window region (Fig. 1a). Our results for the 1-Mb and significant regions are in agreement with previously found regions (BTA3, BTA14, BTA1, BTA21, BTA17 and BTA12) in the animal QTL database that may have an impact on reproductive traits. With Bayesian-GWAS, a substantially higher level of genetic variance allowed for the identification of the more significant essential regions.

Finding the genomic regions in GWAS that influence reproductive traits is interesting because dairy cow fertility is a complex trait with low heritability that involves multiple biological pathways. However, this study provided new information on many genes and molecular networks affecting reproductive traits in Holstein cattle, as shown in Table 2. The CHD7 gene, which encodes the protein known as chromodomain helicase DNA binding protein 7, was one of the important positional candidate genes within 1-Mb upstream and downstream of this overlapping window. In addition, the genes on BTA14 (AFC) have previously been reported by Mohammadi *et al.* (2020). In our research, the CHD7 gene was identified in the AFC (BTA14) trait, consistent with the findings of Mota *et al.* (2017), which plays an important role in reproductive traits. In addition, this gene influences the growth trait and fertility of Simmental cattle (Fonseca *et al.*, 2018), meat quality traits in Brahman cattle (Braz *et al.*, 2021) and lactation persistency in Canadian Holsteins (Do *et al.*, 2017). In Nellore cattle, the CLVS1 gene plays an essential role in determining the age at first calving trait (Mota *et al.*, 2017). The SEMA5A and MCTP1 genes play a significant antibacterial role in mammary gland infection by controlling neutrophil migration to the infection site and also affect carcass traits, beef tenderness and oviduct-embryo interactions, once again in Nellore cattle (Santana *et al.*, 2015). The ELF1 gene has been reported as a candidate gene related to feed efficiency traits that help to reduce environmental consequences from livestock production, enhancing beef cattle profitability (De Lima *et al.*, 2020). The small proline-rich proteins (SPRR4) gene affects reproductive traits, and 175-kb flanking regions of BTB-00661215 SNP were detected in Iranian Holstein cows as a potential candidate gene for the AFC trait (Mohammadi *et al.*, 2022). Additionally, in Brahman cattle, the PAB2A gene is associated with scrotal growth and male fertility (Soares *et al.*, 2017). The genes HOXD, EVX2 and KIAA1715 were found to play important roles in limb function, digit number and lumbar vertebrae in the potential genome region of sheep (Ren *et al.*, 2016). In the peri-implantation phase of pregnancy, the MAT2B gene regulates the expression of critical genes required for cell-specific metabolism and transport of polyamines in the sheep endometrium, affecting fertility and offspring health, growth rates and changes in the embryonic stage of bovine fibroblast cells with DNA methylation, but also has relationships with milk production characteristics of Chinese Holstein cows (Ye *et al.*, 2022). The expression of cow genes related to the cell cycle, membrane stability and apoptosis, as well as other embryonic genes, is influenced by the NudC domain containing 2 (PDCD2) gene (Hoelker *et al.*, 2012). The G protein-coupled receptor 39 gene (GPR39), as in other mammals, is crucial for digestive and metabolic processes in cows (Yamamoto *et al.*, 2008). The NCK associated protein 5 (NCKAP5) gene was studied in various cattle breeds and crosses to assess its influence on temperamental traits. This gene encodes a transcription factor

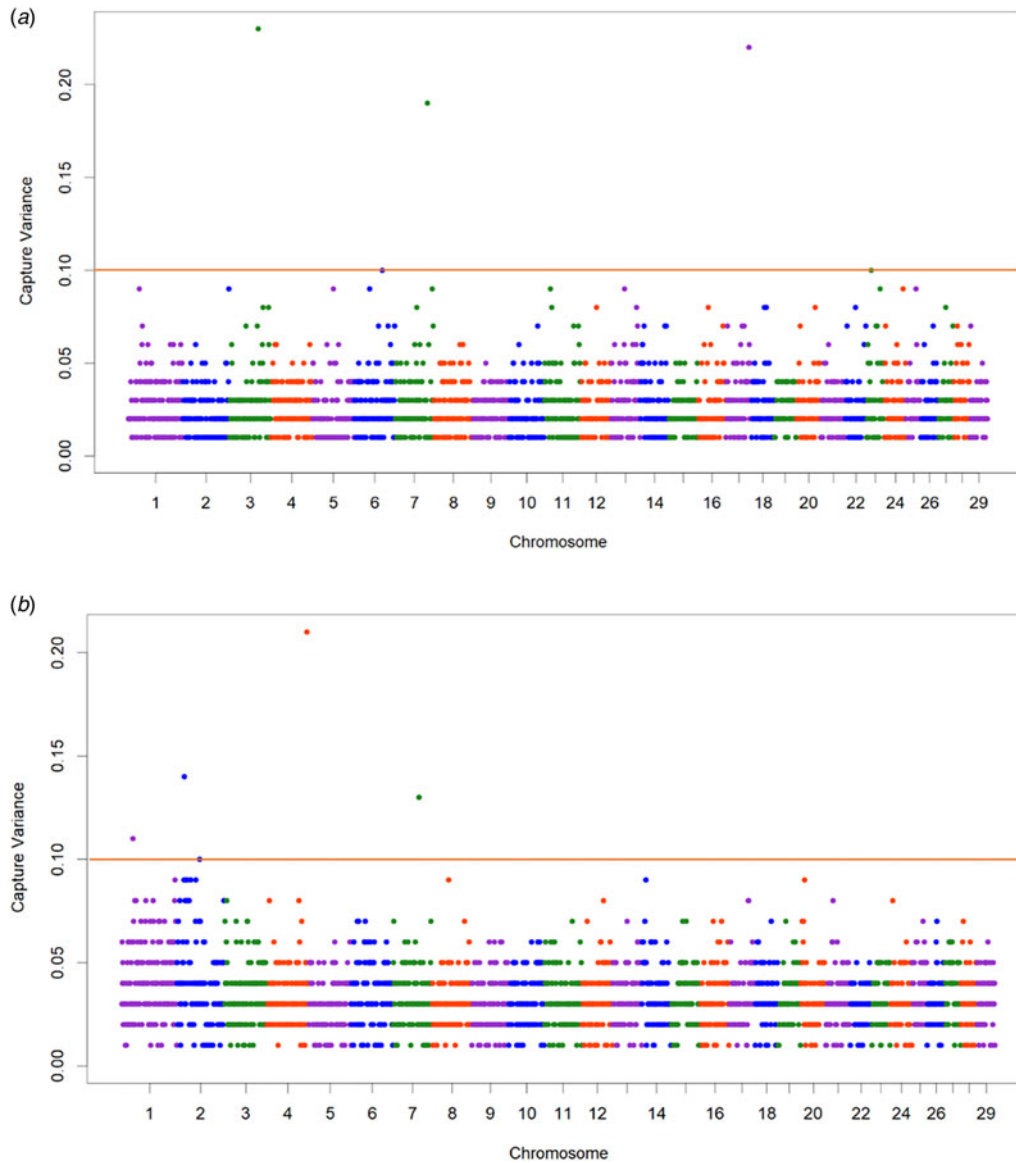


Figure 1. Manhattan plots of the genome-wide association studies for reproductive traits. The x-axis and the y-axis represent the chromosome number and the observed windows, respectively. The SNPs were plotted against their genomic positions. (a) DO, (b) AFC, (c) CI, (d) PR.

essential for osteoblast development and chondrocyte maturation, processes that may influence crucial phenotypes of the Limousin breed (Mariadassou *et al.*, 2020). In addition, HOXD13 and SLF1 genes were found to affect sperm morphometry through inbreeding in a cattle breed and related to sperm DNA methylation and reproductive traits in cattle (Liu *et al.*, 2019). We found a significant SNP on BTA20 for PR traits in Holstein cows. This SNP is located in the semaphorin5A (SEMA5A) gene. The SEMA5B gene is an essential paralog of SEMA5A and is crucial for fertility pathways in cows (Cai *et al.*, 2019) as well as improvements in mastitis resistance and there are indications that the FEZL-SEMA5A pathway may regulate innate immunity and neuronal development in cattle. According to Han and Peñagaricano (2016), the BRF1 gene has been associated with bull fertility in Holstein cattle. The CACHD1 and CDC42BPA genes reportedly affect the quality of meat produced beef cattle (Ryu and Lee, 2016). According to Li *et al.* (2020), the PSEN2 and SUGT1 genes are linked to a haplotype that in dairy cows affects fertility

and leads to altered conception, early embryonic losses, pregnancy losses and mutations. This haplotype has also been linked to heritable abnormalities called BMS (bovine male subfertility) in the Russian Simmental population. The biological mechanisms underlying a number of economically important traits in cattle breeds are linked to the SNORA70 gene (Taye *et al.*, 2017). Furthermore, activation of the AKT1 gene in the epithelial cells of the mammary gland plays a role in the transmission of lactation signals (Wan *et al.*, 2011). Research has shown that the TM2D1 gene is associated with adaptation and fertility traits in Colombian creole cattle breeds, as well as with apoptosis of ovarian follicles in cattle and buffalo (Zoheir *et al.*, 2021). The KIAA0825 gene was found to have an impact on endoparasite infection, general health and fertility in German Holsteins (Wolf *et al.*, 2021). The adenylate cyclase 2 (ADCY2) gene is one of the primary candidates for reproductive traits, having effects on age at first calving and first calving interval in water dairy buffaloes and connected to several biological processes,

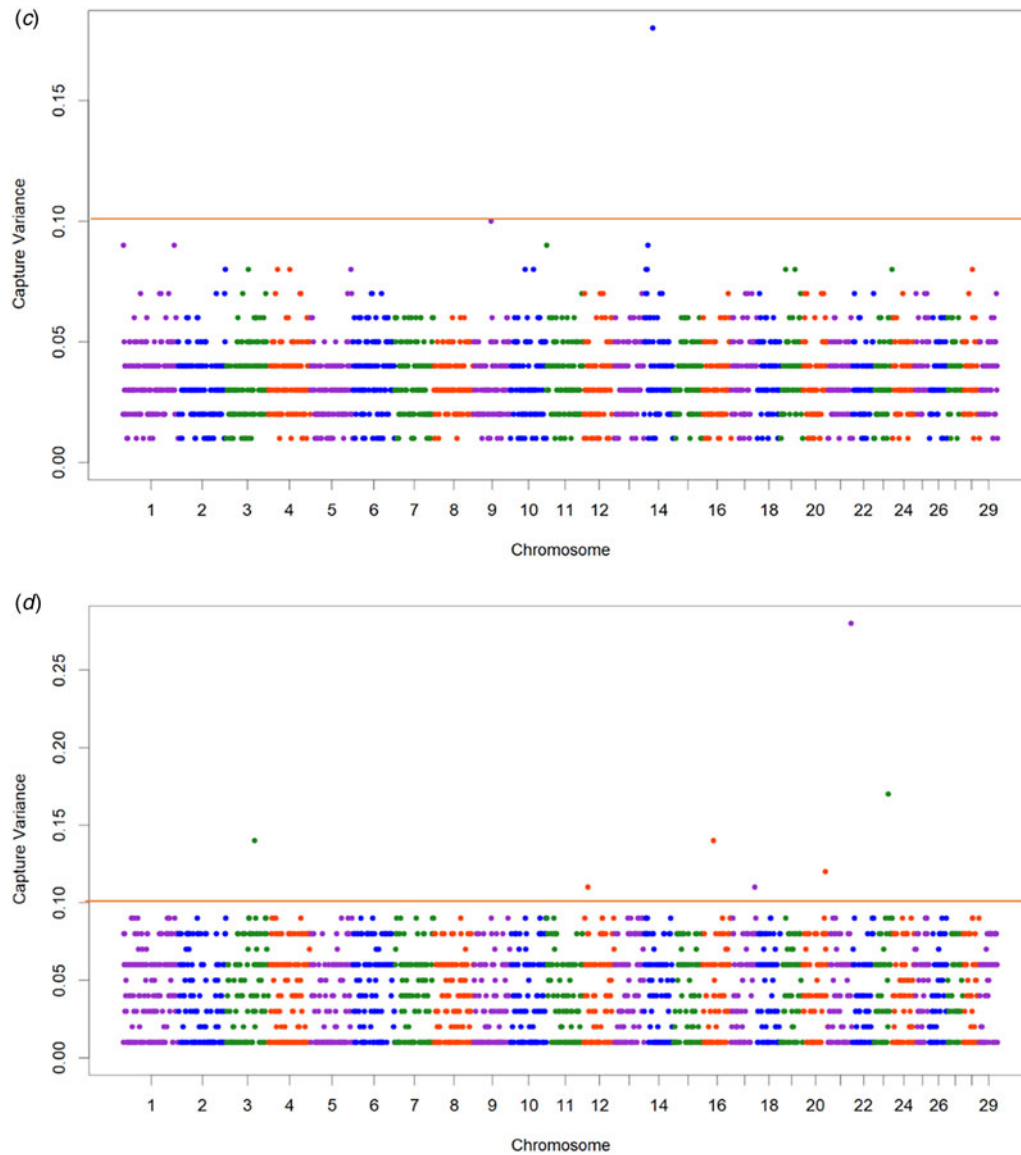


Figure 1. Continued.

and the Yanbian cattle have previously shown an association with meat tenderness (Silva *et al.*, 2020). Mutations in the MTRF1 gene have related to a significant effect on body weight and other quantitative traits, as reported by (Reynolds *et al.*, 2022).

Post-GWAS (GSEA and network cluster analysis)

Gene enrichment analysis was used in GWAS to identify potential functional categories and molecular pathways associated with reproductive traits in Iranian Holstein cattle, and these data are presented in online Supplementary Table S4. A total of 43 biological terms were significantly associated with reproductive traits (FDR < 0.05). Given that many gene ontology (GO) terms are correlated, the data need to be interpreted with caution. However, after the Bonferroni correction for post-GWAS, the top 43 terms were highly enriched in the overexpressed-up gene set. As concerns significant GO terms for biological processes for reproductive traits in our Iranian Holstein cows, the most significant terms were 'GO:0006412' Translation involved the cellular

metabolic process related to the AFC trait. Network clusters, and interactions between potential genes for reproductive traits are shown in Fig. 2. The genes were classified into three clusters by the network analysis using the K-means algorithm technique (red, green and blue). In the first (red) cluster CLVS1 and GPR39 (in relation to AFC), AMOT and AMOTL1 (in relation to DO), ARF1 and CCDC186 (in relation to CI), ADCY2 and BDP1 (in relation to PR) genes were the hub genes of the network. Moreover, the hub genes of the second cluster (green) were CHD7 and ICT1 (in relation to AFC), ANKRD32 and ATM (in relation to DO), GHRL and GPR39 (in relation to CI), C14orf79 and CEPT1 (in relation to PR). Also, BHMT and BHMT2 hub genes of the third cluster were related to AFC, FAM81B and PITPNB (in relation to DO), CA8 and ITPR1 (in relation to CI) and FASTKD3 and JAG2 (in relation to PR). The first cluster's annotated genes were significantly involved in the regulation of body weight, gastrointestinal mobility, hormone secretion and cell death (in relation to AFC), and may play a role in the polarity, proliferation and migration of endothelial cells (in

Table 2. Significant SNPs identified for reproductive traits in Iranian Holstein cows using Bayes-GWAS approach

Trait	Start	End	No. of SNPs	% Var	BTA	bp	Symbols of nearby genes
AFC	ARS-BFGL-NGS-31442	INRA-198	23	0.21	3	17547502	SPRR4
	Hapmap25446-BTC-054694	Hapmap25761-BTC-065280	23	0.16	14	26728705	CHD7; CLVS1
	BTB-00083524	ARS-BFGL-NGS-32709	25	0.14	2	21003570	LNPk, EVX2, HOXD13, bta-mir-10b, HOXD3, HOXD4, HOXD8, HOXD9, HOXD10, HOXD11, HOXD12, HOXD13
	BTB-01557864	BTB-00328082	18	0.13	7	75017155	5S_Rrna, CCNG1, NUDCD2, HMMR, MAT2B
	Hapmap43017-BTA-69173	BTA-69212-no-rs	20	0.11	1	32036293	-
	BFGL-NGS-111934	BFGL-NGS-110609	23	0.1	2	65004490	NCKAP5, LYPD1, GPR39
	BFGL-NGS-111934	BFGL-NGS-110609	23	0.1	2	65004490	NCKAP5, LYPD1, GPR39
CI	Hapmap25446-BTC-054694	Hapmap25761-BTC-065280	23	0.11	14	26003598	CA8, RAB2A
	ARS-BFGL-NGS-51043	BTB-00397064	24	0.1	9	50027556	-
	ARS-BFGL-NGS-1567	BTA-68573-no-rs	19	0.23	3	84094824	PATJ, TM2D1, NFIA,
DO	BTA-41804-no-rs	Hapmap41535-BTA-41786	24	0.22	17	67131762	MN1, PITPNB
	Hapmap51706-BTA-25479	Hapmap43946-BTA-80204	14	0.19	7	94105221	KIAA0825, SLF1, MCTP1, TTC37, FAM81B
	ARS-BFGL-NGS-107488	BFGL-NGS-119656	21	0.28	21	69484017	BRF1, NUDT14, JAG2, BTBD6, PACS2, GPR132, CDCA4, CLBA1, AHNAK2, PLD4, CEP170B, AKT1, ZBTB42, JAG2
PR	BFGL-NGS-115157	ARS-BFGL-NGS-69151	22	0.17	23	40154308	CAP2, RBM24, bta-mir-2325b, STMND1, U6, ATXN1, GMPR
	ARS-BFGL-BAC-33363	BTB-01274755	22	0.14	16	30042285	PSEN2, COQ8A, CDC42BPA
	BTB-00139181	BTB-00139874	23	0.14	3	81047780	CACHD1, UBE2U, PGM1, EFCAB7
	ARS-BFGL-NGS-22583	ARS-BFGL-BAC-36834	22	0.12	20	64002006	SEMA5A, MTRR, ADCY2, FASTKD3
	BTA-41804-no-rs	Hapmap41535-BTA-41786	24	0.11	17	67131762	MN1, PITPNB
	ARS-BFGL-NGS-43211	BTA-31554-no-rs	18	0.11	12	11179949	CNMD, SUGT1, ELF1, SNORA70, MTRF1

relation to DO), affect the cell cycle and differentiation of various tissues (in relation to CI), and be involved for instance in cell cycle control and signal transduction (in relation to PR).

Gene ontology (GO) terms for biological processes

The biological process ('GO:0006412') associated with the regulation of gene expression, protein translation and indeed any process related to the AFC trait are presented in online Supplementary Table S4 and the important role of this terms in feed efficiency in beef cattle and bovine oocyte quality and female fertility (Macaulay *et al.*, 2016) has already been reported. The association of embryonic limb morphogenesis ('GO:0030326') term for the DO trait is a biological process that occurs in the embryo by which the anatomical structures of the limb are generated and organized and the biosynthetic process ('GO:0044272') terms related with mastitis induced by *E. coli* was demonstrated by Darang (Darang *et al.*, 2022). Additionally, the role of this GO:0051649 term in meat performance in pigs has been reported by Wimmers *et al.* (2010). The genes included in ('GO:1901566') affect organonitrogen compound biosynthetic processes, the decline in follicle-stimulating hormone, the bovine granulosa

cell pathway and the mitochondrial translational termination pathway, whilst ('GO:0070126') affects pre-implantation embryo development in cattle (Clark, 2021).

Network cluster analysis

Network analysis of the first cluster CLVS1 and GPR39 hub genes associated with the AFC trait is shown in Fig. 2a. The GPR39 gene is known for its important role in gastrointestinal and metabolic functions in cattle and other mammals. Additionally, the CLVS1 gene is a potential candidate gene for carcass traits due to its biological functions. Regarding the gene hub-related PR trait (Fig. 2d), the AMOT gene has a significant impact on regulating gene expression in mouse, cow and human morula embryos. The role of the hub gene, ADP ribosylation factor 1 (ARF1) in cell proliferation and lactation of bovine mammary epithelial cells has been investigated. Shown in Fig. 2c, the CCDC186 hub gene (first cluster involved in CI) helps in milk production traits in water buffalo, as well as affecting of feed efficiency in mid-lactation Holstein dairy cows. In addition, the role of the BDP1 gene associated with white blood cell count in Chinese Holstein cows has been reported by Chen *et al.* (2022). The CHD7 gene

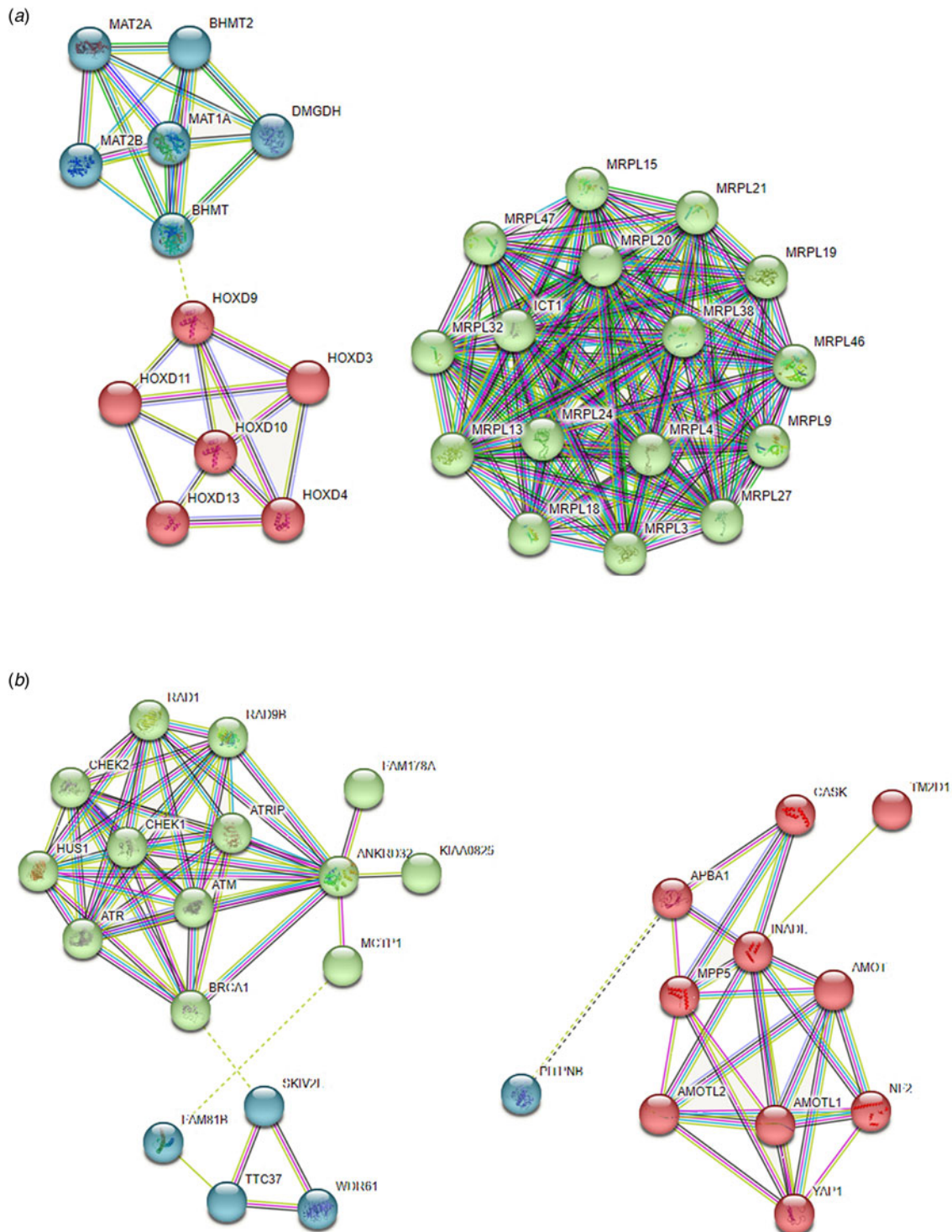


Figure 2. Network clusters between the identified genes for (a): age at first calving (AFC), (b) days open (DO), (c) calving interval (CI), and (d) pregnancy rate (PR) traits in Iranian Holstein cattle.

(second cluster, associated with AFC) is involved in the organization of chromatin and has associated GO terms related to biological processes including in-utero embryonic development and cow genitalia development and may play an important role in regulating puberty and reproduction as well as having effects on in vitro fertilization, embryo transfer and superovulation in Holstein cattle. Additionally, in beef cattle, the CEPT1 gene

may have an important regulatory function in lipid metabolism. Finally, the ITPR1 gene, a member of the third cluster, was associated with both male and female fertility and also with embryo size and morphology in both mice and humans (Cole *et al.*, 2018).

In conclusion, using Bayesian analysis, 79 genes were located within or near to 19 significant SNPs/windows in the *Bos taurus* autosomes. Among these genes, 25 candidate genes for

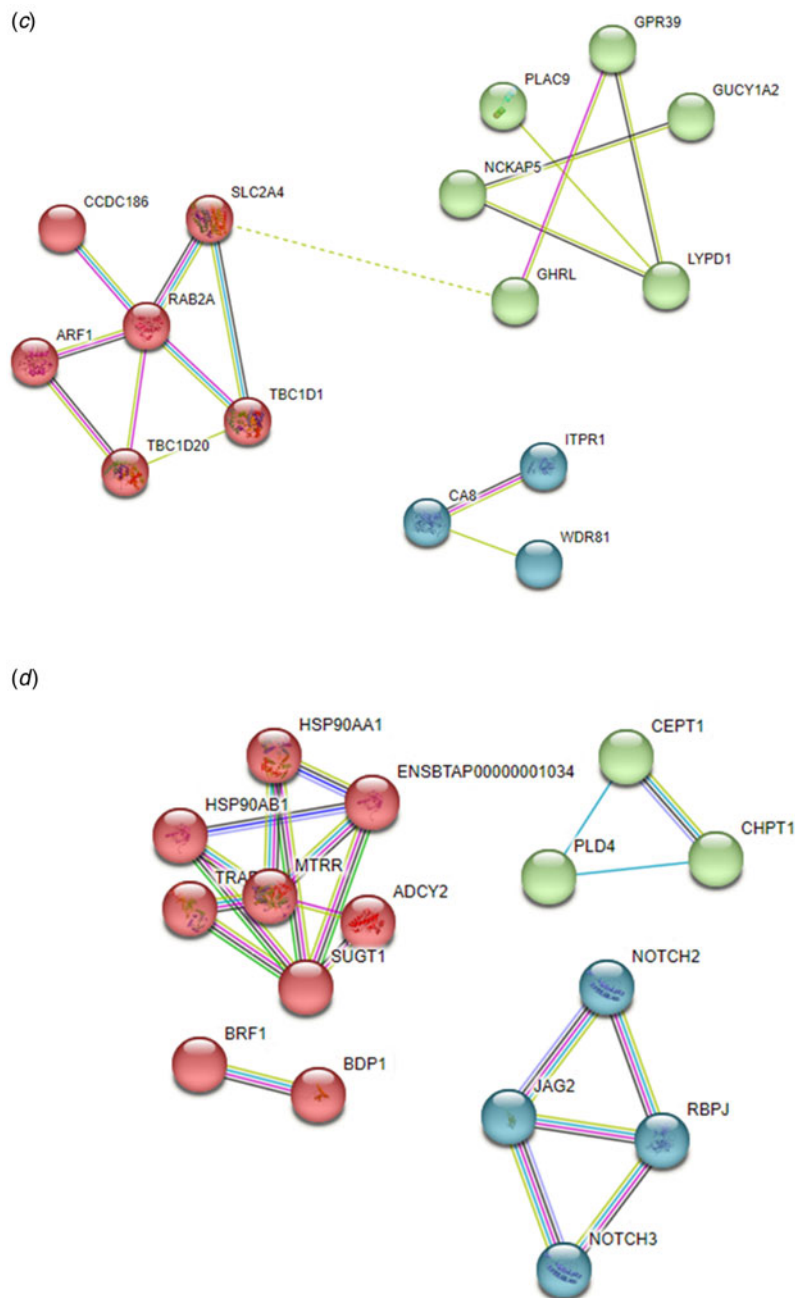


Figure 2. Continued.

reproductive traits were identified, namely CHD7, CLVS1, EVX2, MAT2B, NUDCD2, GPR39, NCKAP5, LYPD1, HOXD13, SEMA5B, CCNG1, SEMA5A, BRF1, PSEN2, CACHD1, SUGTA, ELF1, SNORA70, AKT1, TM2D1, SLF1, MCTPA, PAB2A, MTRF1 and ADCY2. Additionally, the candidate genes CLVS1, GPR39, CENPF, AMOT, ARF1, CCDC186, ADCY2, BDP1 and AMOTL1 were identified in the network cluster analysis as hub genes for reproductive traits. The results of gene set enrichment analysis and pathway analysis suggest that the most important gene ontology term involving cellular metabolic process was related to the AFC trait. A combination of Bayesian GWAS and post-GWAS analysis can be effectively used to identify potential candidate genes and causative mutations that could improve reproductive traits in dairy cattle.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0022029924000414>

Acknowledgements. The Iranian Animal Breeding Center is acknowledged by the authors for supplying the genotypes and information.

References

- Ali I, Muhammad Suhail S and Shafiq M (2019) Heritability estimates and genetic correlations of various production and reproductive traits of different grades of dairy cattle reared under subtropical condition. *Reproduction in Domestic Animals* **54**, 1026–1033.
- Braz CU, Rowan TN, Schnabel RD and Decker JE (2021) Genome-wide association analyses identify genotype-by-environment interactions of growth traits in Simmental cattle. *Scientific Reports* **11**, 13335.
- Butler ML, Bormann JM, Weaber RL, Grieger DM and Rolf MM (2020) Selection for bull fertility: a review. *Translational Animal Science* **4**, 423–441.

- Cai Z, Gulbrandsen B, Lund MS and Sahana G (2019) Prioritizing candidate genes for fertility in dairy cows using gene-based analysis, functional annotation and differential gene expression. *BMC Genomics* **20**, 1–9.
- Chen Q, Yang M, Liu X, Zhang J, Mi S, Wang Y, Xiao W and Yu Y (2022) Blood transcriptome analysis and identification of genes associated with supernumerary teats in Chinese Holstein cows. *Journal of Dairy Science* **105**, 9837–9852.
- Clark K (2021) Paternal effects on pre-implantation embryo development in cattle. In Book Paternal effects on pre-implantation embryo development in cattle, Vol., pp. (Ed. Eds. Editor). Series Paternal effects on pre-implantation embryo development in cattle. University of Missouri–Columbia.
- Cole J, Gaddis KP, Null D, Maltecca C and Clay J (2018) Genome-wide association study and gene network analysis of fertility, retained placenta, and metritis in US Holstein cattle. *World (Oakland, Calif)* **1**, 171.
- Darang E, Pezeshkian Z, Mirhoseini SZ and Ghovvati S (2022) Identification of key genes and potential pathways associated with mastitis induced by *E. coli*. *Biochemical Genetics* **61**, 202–220.
- De Lima AO, Koltjes JE, Diniz WJ, De Oliveira PS, Cesar AS, Tizioto PC, Afonso J, de Souza MM, Petrini J and Rocha MI (2020) Potential biomarkers for feed efficiency-related traits in Nelore cattle identified by co-expression network and integrative genomics analyses. *Frontiers in Genetics* **11**, 189.
- Do D, Bissonnette N, Lacasse P, Miglior F, Sargolzaei M, Zhao X and Ibeagha-Awemu E (2017) Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle. *Journal of Dairy Science* **100**, 1955–1970.
- Faraji AH, Aslaminejad AA and Farhangfar H (2011) Estimation of genetic parameters and trends for age at first calving and calving interval in Iranian Holstein cows.
- Fernando R, Nettleton D, Southey B, Dekkers J, Rothschild M and Soller M (2004) Controlling the proportion of false positives in multiple dependent tests. *Genetics* **166**, 611–619.
- Fonseca P, Id-Lahoucine S, Reverter A, Medrano JF, Fortes MS, Casellas J, Miglior F, Brito L, Carvalho MRS and Schenkel FS (2018) Combining multi-OMICs information to identify key-regulator genes for pleiotropic effect on fertility and production traits in beef cattle. *PLoS One* **13**, e0205295.
- Han Y and Peñagaricano F (2016) Unravelling the genomic architecture of bull fertility in Holstein cattle. *BMC Genetics* **17**, 1–11.
- Hoelker M, Salilew-Wondim D, Drillich M, Christine G-B, Ghanem N, Goetze L, Tesfaye D, Schellander K and Heuwieser W (2012) Transcriptional response of the bovine endometrium and embryo to endometrial polymorphonuclear neutrophil infiltration as an indicator of sub-clinical inflammation of the uterine environment. *Reproduction, Fertility and Development* **24**, 778–793.
- Jiang J, Ma L, Prakapenka D, VanRaden PM, Cole JB and Da Y (2019) A large-scale genome-wide association study in US Holstein cattle. *Frontiers in Genetics* **10**, 412.
- Kemper KE, Reich CM, Bowman PJ, Vander Jagt CJ, Chamberlain AJ, Mason BA, Hayes BJ and Goddard ME (2015) Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genetics Selection Evolution* **47**, 1–17.
- Li Y, Yuan J, Chen F, Zhang S, Zhao Y, Chen X, Lu L, Zhou L, Chu CY and Sun H (2020) Long noncoding RNA SAM promotes myoblast proliferation through stabilizing Sugt1 and facilitating kinetochore assembly. *Nature Communications* **11**, 2725.
- Liu S, Fang L, Zhou Y, Santos DJ, Xiang R, Daetwyler HD, Chamberlain AJ, Cole JB, Li C-J and Yu Y (2019) Analyses of inter-individual variations of sperm DNA methylation and their potential implications in cattle. *BMC Genomics* **20**, 1–14.
- Macaulay AD, Gilbert I, Scantland S, Fournier E, Ashkar F, Bastien A, Saadi HAS, Gagné D, Sirard M-A and Khandjian EW (2016) Cumulus cell transcripts transit to the bovine oocyte in preparation for maturation. *Biology of Reproduction* **94**, 16, 11–11.
- Mariadassou M, Ramayo-Caldas Y, Charles M, Féménia M, Renand G and Rocha D (2020) Detection of selection signatures in Limousin cattle using whole-genome resequencing. *Animal Genetics* **51**, 815–819.
- Miglior F, Fleming A, Malchiodi F, Brito LF, Martin P and Baes CF (2017) A 100-year review: identification and genetic selection of economically important traits in dairy cattle. *Journal of Dairy Science* **100**, 10251–10271.
- Mohammadi A, Alijani S, Rafat SA and Abdollahi-Arpanahi R (2020) Estimation of genetic and phenotypic parameters and trends of female fertility traits of cows and sire conception rate in Iranian Holstein. *Journal of Animal Science Research* **30**, 37–53.
- Mohammadi A, Alijani S, Rafat S and Abdollahi-Arpanahi R (2022) Single-step genome-wide association study and candidate genes networks affecting reproductive traits in Iranian Holstein cattle. *Livestock Science* **262**, 104971.
- Mota R, Guimarães S, Fortes M, Hayes B, Silva F, Verardo L, Kelly M, de Campos C, Guimarães J and Wenceslau R (2017) Genome-wide association study and annotating candidate gene networks affecting age at first calving in Nelore cattle. *Journal of Animal Breeding and Genetics* **134**, 484–492.
- Pirlo G, Miglior F and Speroni M (2000) Effect of age at first calving on production traits and on difference between milk yield returns and rearing costs in Italian Holsteins. *Journal of Dairy Science* **83**, 603–608.
- Ren X, Yang G-L, Peng W-F, Zhao Y-X, Zhang M, Chen Z-H, Wu F-A, Kantanen J, Shen M and Li M-H (2016) A genome-wide association study identifies a genomic region for the polycerate phenotype in sheep (*Ovis aries*). *Scientific Reports* **6**, 21111.
- Reynolds EG, Lopdell T, Wang Y, Tiplady KM, Harland CS, Johnson TJ, Neeley C, Carnie K, Sherlock RG and Couldrey C (2022) Non-additive QTL mapping of lactation traits in 124,000 cattle reveals novel recessive loci. *Genetics Selection Evolution* **54**, 5.
- Ryu J and Lee C (2016) Genetic association of marbling score with intragenic nucleotide variants at selection signals of the bovine genome. *Animal: An International Journal of Animal Bioscience* **10**, 566–570.
- Santana M, Ventura RV, Utsunomiya YT, Neves H, Alexandre PA, Oliveira Junior GA, Gomes R, Bonin M, Coutinho LL and Garcia JF (2015) A genomewide association mapping study using ultrasound-scanned information identifies potential genomic regions and candidate genes affecting carcass traits in Nelore cattle. *Journal of Animal Breeding and Genetics* **132**, 420–427.
- Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q and Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44**, 825–830.
- Silva DB, Fonseca LF, Pinheiro DG, Magalhães AF, Muniz MM, Ferro JA, Baldi F, Chardulo LA, Schnabel RD and Taylor JF (2020) Spliced genes in muscle from Nelore Cattle and their association with carcass and meat quality. *Scientific Reports* **10**, 14701.
- Soares A, Guimarães S, Kelly M, Fortes M, e Silva F, Verardo L, Mota R and Moore S (2017) Multiple-trait genomewide mapping and gene network analysis for scrotal circumference growth curves in Brahman cattle. *Journal of Animal Science* **95**, 3331–3345.
- Sollero BP, Junqueira VS, Gomes CC, Caetano AR and Cardoso FF (2017) Tag SNP selection for prediction of tick resistance in Brazilian Braford and Hereford cattle breeds using Bayesian methods. *Genetics Selection Evolution* **49**, 1–15.
- Taye M, Lee W, Jeon S, Yoon J, Dessie T, Hanotte O, Mwai OA, Kemp S, Cho S and Oh SJ (2017) Exploring evidence of positive selection signatures in cattle breeds selected for different traits. *Mammalian Genome* **28**, 528–541.
- Wan Z-Y, Tong H-L, Li Q-Z and Gao X-J (2011) Influence on cellular signal transduction pathway in dairy cow mammary gland epithelial cells by galactopoietic compound isolated from *Vaccariae segetalis*. *Agricultural Sciences in China* **10**, 619–630.
- Wang S-B, Feng J-Y, Ren W-L, Huang B, Zhou L, Wen Y-J, Zhang J, Dunwell JM, Xu S and Zhang Y-M (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific Reports* **6**, 1–10.
- Wimmers K, Murani E and Ponsuksili S (2010) Functional genomics and genetical genomics approaches towards elucidating networks of genes affecting meat performance in pigs. *Briefings in Functional Genomics* **9**, 251–258.

- Wolf MJ, Yin T, Neumann GB, Korkuć P, Brockmann GA, König S and May K** (2021) Genome-wide association study using whole-genome sequence data for fertility, health indicator, and endoparasite infection traits in German black pied cattle. *Genes* **12**, 1163.
- Yamamoto I, Kaiya H, Tsutsui C, Sakai T, Tsukada A, Miyazato M and Tanaka M** (2008) Primary structure, tissue distribution, and biological activity of chicken motilin receptor. *General and Comparative Endocrinology* **156**, 509–514.
- Ye W, Xu L, Li Y, Liu L, Ma Z, Sun D and Han B** (2022) Single nucleotide polymorphisms of ALDH18A1 and MAT2A genes and their genetic associations with milk production traits of Chinese Holstein cows. *Genes* **13**, 1437.
- Zoheir KM, Darwish AM, Liguó Y and Ashour AE** (2021) Transcriptome comparisons detect new genes associated with apoptosis of cattle and buffaloes preantral follicles. *Journal of Genetic Engineering and Biotechnology* **19**, 1–9.