



RESEARCH PAPER

OPEN ACCESS

Genome-wide association study based on haplotype model and gene-set enrichment analysis associated with age at first calving in Nelore cattle

H. Mohammadi^{1*}, A. H. Khaltabadi Farahani², M. H. Moradi²

1. Assistant Professor, Department of Animal Sciences, Faculty of Agriculture and Natural Resources, Arak University, Arak, Iran
2. Associate Professor, Department of Animal Sciences, Faculty of Agriculture and Natural Resources, Arak University, Arak, Iran

(Received: 21-06-2021 – Accepted: 06-09-2021)

Introduction: Zebu cattle are highly adapted to tropical regions. However, females reach puberty after taurine heifers, which affects the economic efficiency of beef cattle breeding in tropical regions. A method to identify new loci and confirm existing quantitative trait loci (QTL) is through genome-wide association studies (GWAS). QTL-assisted selection and genomic regions affecting the production and reproduction traits have been considered to increase the efficiency of selection and improve production performance. The GWAS typically focuses on genetic markers with the strongest evidence of association. However, single markers often explain only a small component of the genetic variance and hence offer a limited understanding of the trait under study. A solution to tackle the aforementioned problems, and deepen the understanding of the genetic background of complex traits, is to move up the analysis from the single nucleotide polymorphism (SNP) to the gene and gene-set levels. In a gene-set analysis, a group of related genes that harbor significant SNP previously identified in GWAS is tested for over-representation in a specific pathway. The present study aimed to conduct a GWAS based on a gene-set enrichment analysis for identifying the loci associated with age at first calving trait using the high-density SNPs.

Materials and methods: A total of 2273 Nelore cattle (995 males and 1278 females) genotyped using the Illumina BovineHD BeadChip were used in the current study. The association analysis included females with valid first calving records as well as open heifers. The association analyses were carried out by considering deregressed estimated breeding values (dEBV) for age at first calving as response variables. Before deregression, the estimated breeding values (EBV) were obtained for the dataset by considering both the calved and non-calved heifers. Variance components and EBV were obtained using the DMU software. In the analysis of AFC, a single-trait animal model was run. The gene-set analysis consists of three different steps: the assignment of SNPs to genes, the assignment of genes to functional categories, and finally the association analysis between each functional category and the phenotype of interest. The GWAS was evaluated using the GHap package in the R program. Using the biomaRt2 R package, the SNPs were assigned to genes if they were within the genomic sequence of the gene or within a flanking region of 15 kb up- and downstream of the gene. For the assignment of the genes to functional categories, the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway databases were used. The GO database designates biological descriptors to genes based on attributes of their encoded products and it is further partitioned into three components: biological process, molecular function, and cellular component. The KEGG pathway database contains metabolic and regulatory pathways, representing the actual knowledge of molecular interactions and reaction networks. Finally, a Fisher's exact test was performed to test for overrepresentation of the significant genes for each gene-set.

Results and discussion: The block sizes of five SNPs were chosen to perform association studies. Gene-set enrichment analysis has proven to be a great complement to GWAS. Among available gene-set databases, GO is probably the most popular, whereas KEGG is a relatively new tool that is gaining ground in livestock genomics.

* Corresponding author: H-mohammadi64@araku.ac.ir



We hypothesized that the use of gene set information could improve prediction. However, neither of the gene set SNP classes outperformed the standard whole-genome approach. Gene-sets have been primarily developed using data from model organisms, such as mice and flies; therefore, some of the genes included in these terms may be irrelevant for reproduction. Gene-set enrichment analysis identified candidate genes related to estrogen metabolic process (*HSD17B12*), synapse organization (*PPFIA2* and *PPFIA2*), sensory perception of mechanical stimulus (*MYO3A* and *KCNMA1*), protein tyrosine kinase activity (*IGF1R*), the cell-cell junction (*FRMD4A*), GnRH signaling pathway (*ADCY5*), and focal adhesion (*PPP1R12A*). Some of the genes which were found are consistent with some previous studies and are involved in biological pathways related to fertilization, age at first calving, estrogen biosynthesis, heifer conception rate, early development of the fetus, puberty, and glucose homeostasis in the ovary.

Conclusions: This study supported previous results from GWAS of reproductive traits, and also revealed additional regions in the cattle genome associated with these economically important traits. These findings could potentially be useful for genetic selection in cows.

Keywords: Gene-set enrichment analysis, Reproduction, Cattle, Biological pathways, Haplotype

Ethics statement: Ethical review and approval was not required for the animal study because datasets used in this study were obtained from pre-existing databases based on routine animal recording procedures.

Data availability statement: The data that support the findings of this study were derived from the following resources available in the public domain: Nascimento, A. V. D., Romero, Á. R. D. S., Utsunomiya, Y. T., Utsunomiya, A. T. H., Cardoso, D. F., Neves, H. H. R., Carvalheiro, R., Garcia, J. F., & Grisolia, A. B. (2018). Genome-wide association study using haplotype alleles for the evaluation of reproductive traits in Nelore cattle. *PloS one*, 13(8), e0201876. <https://doi.org/10.1371/journal.pone.0201876>

Conflicts of interest: The authors declare no conflicts of interest.

Funding: The authors received no specific funding for this project.

Acknowledgment: The authors would like to express their gratitude to Dr. Grisolia from UFGD, Brazil, for providing genotypic data and phenotypic records for the present study.

How to cite this article:

Mohammadi H., Khaltabadi Farahani A. H. and Moradi M. H. 2022. Genome-wide association study based on haplotype model and gene-set enrichment analysis associated with age at first calving in Nelore cattle. *Animal Production Research*, 11(2): 69-80. doi: 10.22124/AR.2022.19943.1629



مقاله پژوهشی

مطالعه پویش کل ژنوم بر پایه مدل هاپلوتیپ و تجزیه و تحلیل غنی سازی مجموعه-های ژنی مرتبط با سن اولین زایش در گاو نلور

حسین محمدی^{۱*}، امیر حسین خلت آبادی فراهانی^۲، محمد حسین مرادی^۲^۱- استادیار، گروه علوم دامی، دانشکده کشاورزی و محیط زیست، دانشگاه اراک^۲- دانشیار، گروه علوم دامی، دانشکده کشاورزی و محیط زیست، دانشگاه اراک

(تاریخ دریافت: ۱۴۰۰/۰۳/۳۱ - تاریخ پذیرش: ۱۴۰۰/۰۶/۱۵)

چکیده

در این پژوهش از رکوردهای فنوتیپی صفت سن اولین گوساله‌زایی گاوهای نلور برای مطالعه پویش ژنوم بر پایه تجزیه و تحلیل غنی‌سازی جهت شناسایی ساز و کارهای زیستی استفاده شد. ارزیابی پویش کل ژنوم با بسته نرم افزاری GHap بر پایه مدل هاپلوتیپی در برنامه R انجام شد. تجزیه غنی‌سازی مجموعه‌های ژنی با بسته نرم‌افزاری goseq برنامه R با هدف شناسایی طبقات عملکردی و مسیرهای زیستی ژن‌های نزدیک در مناطق ژنومی کاندیدا انجام شد و در نهایت برای تجزیه و تحلیل بیوانفورماتیکی از پایگاه‌های GO، KEGG، DAVID و PANTHER استفاده شد. با تجزیه و تحلیل غنی‌سازی مجموعه‌های ژنی، مسیرهای هستی‌شناسی (ژن‌های کاندیدا) *(HSD17B12)* Estrogen metabolic process، *(KIRREL3)* Synapse organization و *(PPFIA2)* Protein tyrosine kinase activity، *(MYO3A)* و *(KCNMA1)* Sensory perception of mechanical stimulus، *(ADCY5)* GnRH signaling pathway، *(FRMD4A)* cell-cell junction، *(IGF1R)* و *(PPP1R12A)* Focal adhesion شناسایی شدند. ژن‌های کاندیدا نقش مهمی در باروری، سن اولین گوساله‌زایی، بیوسنتز استروژن، نرخ آبستنی تلیسه‌ها، رشد ابتدایی جنین، سن بلوغ و تنظیم هموستازی گلوکز در تخمدان داشتند. با توجه به تأیید مناطق قبلی پویش ژنومی و شناسایی مناطق ژنومی جدید، استفاده از یافته‌های این پژوهش می‌تواند در انتخاب ژنتیکی گاو مفید باشد.

واژه‌های کلیدی: تجزیه و تحلیل غنی‌سازی، تولیدمثل، گاو، مسیرهای زیستی، هاپلوتیپ

* نویسنده مسئول: H-mohammadi64@araku.ac.ir

مقدمه

بیشتر برنامه‌های اصلاح نژادی گاو بر پایه افزایش صفات تولیدی استوار است، ولی با توجه به همبستگی ژنتیکی نامطلوب بین صفات تولیدی با تولیدمثل، کاهش عملکرد تولیدمثلی در گاوهای شیری مشاهده شده است. این امر مخصوصاً زمانی که انتخاب بر مبنای بیشینه تولید شیر باشد، بیشتر نمود پیدا می‌کند (Boligon *et al.*, 2011).

بازده تولیدمثل یک گاو با معیارهایی از قبیل سن اولین گوساله‌زایی، فاصله گوساله‌زایی، روزهای باز و تعداد تلقیح‌ها به ازای هر آبستنی سنجیده می‌شود. سن اولین گوساله‌زایی شامل دوره‌ای است که یک گاو برای رسیدن به بلوغ و تولیدمثل برای اولین بار نیاز دارد (Hare *et al.*, 2006). سن اولین گوساله‌زایی یکی از عوامل مؤثر بر عملکرد تولیدمثلی است و معمولاً به عنوان یک صفت شاخص برای بلوغ جنسی گاو ماده در برنامه‌های اصلاح نژادی مورد استفاده قرار می‌گیرد. سن اولین گوساله‌زایی و فاصله دو زایش، دو عامل مهم برای نتیجه اقتصاد پرورش گاو شیری هستند (Mota *et al.*, 2017). زایش تلیسه‌ها در سن بیشتر از ۲۴ ماهگی سبب افزایش هزینه‌های تولید در صنعت گاوداری می‌شود. گاو نلور از گونه بوس/یندیکوس یا زبو است و با شرایط آب و هوایی گرمسیری به خوبی سازگار شده است. از خصوصیات بارز نژاد نلور می‌توان به صفات مادری خوب، پایین بودن سقط جنین و بالا بودن طول عمر اقتصادی اشاره کرد. با این حال، مشابه با سایر نژادهای زبو، گاوهای نلور نسبت به نژادهای بوس تاروس دیرتر به بلوغ می‌رسند (Magnabosco *et al.*, 2016). در تلیسه‌های نلور، سن مطلوب برای رسیدن به سن بلوغ حدوداً ۲۳ ماهگی در نظر گرفته می‌شود. وجود تنوع بالای صفت سن اولین گوساله‌زایی در گاوهای نلور، نشان‌دهنده امکان انتخاب ژنتیکی برای این صفت است (Nascimento *et al.*, 2018).

توسعه روش‌های نسل جدید توالی‌یابی و متعاقب آن، تولید تراشه‌های متراکم نشانگری، انجام مطالعات پویش کل ژنومی را به عنوان یک ابزار بسیار مؤثر در بررسی معماری ژنتیکی صفات و همچنین یافتن واریانت‌های ژنتیکی مرتبط با صفات مهم اقتصادی، ممکن ساخته است (Sharma *et al.*, 2015).

اولین بار، داده‌های مورد استفاده در این تحقیق برای صفت سن اولین گوساله‌زایی مورد تجزیه و تحلیل قرار گرفته بود

که در آن تحقیق، با استفاده از مدل هاپلوتیپ تجزیه و تحلیل شده و پنج بلوک هاپلوتیپی معنی‌دار مرتبط با سن اولین زایش روی کروموزوم‌های ۳، ۵، ۶، ۲۱ و ۲۶ شناسایی شد و شامل ۱۹ ژن کاندیدای مرتبط با صفت تولیدمثلی بودند (Nascimento *et al.*, 2018). در یک مطالعه پویش کل ژنومی در تلیسه‌های نلور، مناطق ژنومی روی کروموزوم‌های ۲ و ۱۴ و ژن‌های کاندیدای OPRK1، TMEM68 و ADAMTS18 مؤثر بر صفت اولین سن زایش گزارش شده است (Mota *et al.*, 2020).

یکی از ایرادات تحقیقات مطالعات پویش کل ژنومی، استفاده از آزمون‌های سخت‌گیرانه و محافظه‌کارانه همچون تصحیح بنفرونی برای جلوگیری از بروز خطای نوع اول است. در حالی که، پرهیز از خطای نوع اول سبب افزایش خطای نوع دوم می‌شود. یکی از راه‌کارهای مناسب برای حل این مشکل، انجام مطالعات پویش کل ژنومی بر مبنای مسیر است (Peñagaricano *et al.*, 2013). در واقع در این روش به جای انجام تجزیه برای یک چندشکلی تک نوکلئوتیدی (SNP) یا یک ژن، ارتباط بین صفت مورد مطالعه و واریانت‌های ژنتیکی در یک دسته یا گروه ژنی که به‌طور عملکردی با هم مرتبط هستند بررسی می‌شود. به عبارت دیگر، تجزیه و تحلیل پیوستگی بین یک مجموعه ژنی معنی‌دار زیستی با فنوتیپ مورد آزمون قرار می‌گیرد. در این روش، ژن‌هایی مورد بررسی قرار می‌گیرند که به تنهایی اثر آنها بر صفت مورد نظر معنی‌دار نشده و دارای آثار متوسطی هستند، ولی اثر تجمعی آنها روی صفت دارای اثر معنی‌دار است. برای اینکه بتوان تفسیر درستی از کنار هم قرار دادن این ژن‌ها حاصل شود، از مسیرهای زیستی به عنوان بستری معنی‌دار که عملکرد مجموع ژن‌ها در آنها یک یا چند هدف واحد را پیگیری می‌کند استفاده می‌شود (Marques *et al.*, 2018).

گزارش شده است که تجزیه و تحلیل پویش ژنومی بر مبنای مسیر، دقت شناسایی مناطق ژنومی مؤثر بر صفت نرخ باروری گاوهای نر را بالا می‌برد، زیرا با استفاده از این روش تمامی نشانگرهای معنی‌دار در سطح ۰/۰۵ تجزیه و تحلیل می‌شوند و در نتیجه، میزان خطای نوع اول و بیش برآوردها کاهش پیدا می‌کند (Peñagaricano *et al.*, 2013). اخیراً مطالعه پویش ژنومی بر مبنای مسیر با استفاده از تجزیه غنی‌سازی مجموعه‌های ژنی روی صفات مرتبط با وزن بدن

(Stephens, 2006) و سپس بلوک‌های هاپلوتیپی پنج نشانگری تشکیل شد (Nascimento *et al.*, 2018) و فراوانی آلل‌های هاپلوتیپی کمتر از سه درصد کنار گذاشته شدند. با توجه به متفاوت بودن قابلیت اعتماد ارزش‌های اصلاحی، با استفاده از وزن‌دهی ارزش‌های اصلاحی، dEBVها برای تمام حیوانات برآورد شدند. وزن‌ها بر اساس فرمول پیشنهادی زیر محاسبه شد (Garrick *et al.*, 2009):

$$w = (1 - h^2) / \left[\left(c + \frac{1 - r^2}{r^2} \right) * h^2 \right]$$

در این فرمول، h^2 برابر با وراثت پذیری صفت (۰/۰۹۲) (Nascimento *et al.*, 2018)، r^2 قابلیت اعتماد dEBV و c برابر با عدد ثابت ۰/۵ است. همچنین برای بررسی ارتباط هاپلوتیپ‌ها با صفت سن اولین گوساله‌زایی از مدل آماری زیر استفاده شد:

$$y = Xb + Zu + e$$

در این مدل، y، بردار dEBV، X، ماتریس ضرایب مربوط به آثار ثابت، b، بردار آثار ثابت (گله، سال و فصل تولد، گروه-های هم عصر)، Z، ماتریس ضرایب مربوط به آثار تصادفی، u، بردار آثار تصادفی با واریانس $K\sigma_u^2$ ، که K برابر با ماتریس روابط آلل‌های هاپلوتیپی بود و e، بردار اثر تصادفی باقی-مانده با واریانس-کواریانس ($W\sigma_e^2$). تجزیه و تحلیل پویش کل ژنومی با استفاده از بسته نرم افزاری GHap نسخه ۱/۲/۱ (Utsunomiya *et al.*, 2016) و با استفاده از دستورات gchap.assoc و gchap.imm انجام شد.

تجزیه غنی سازی مجموعه‌های ژنی: اساساً تجزیه غنی-سازی بر پایه تجزیه و تحلیل مجموعه‌های ژنی در سه مرحله انجام می‌شود: (۱) تعیین مکان SNPهای معنی‌دار به ژن، (۲) ارتباط ژن‌ها به طبقات عملکردی و مسیرهای بیوشیمیایی و (۳) پویش کل ژنومی بر پایه تجزیه و تحلیل مسیر.

تعیین مکان SNPها به ژن‌ها: هاپلوتیپ‌هایی که مقدار P-value آن‌ها کمتر و یا مساوی ۰/۰۰۵ بود با استفاده از بسته نرم افزاری *biomaRt2* (Durinck *et al.*, 2009) در محیط نرم افزار R و با استفاده از مرجع ژنومی گاو (نسخه ARS-UCD1.2) به ژن‌هایی که نشانگر SNP مورد نظر در داخل آن ژن و یا ۱۵kb بالادست یا پایین دست آن ژن قرار داشت، ارتباط داده شدند. در این مرحله، ژنی که حداقل

در مرغ انجام شده است. نتایج به‌دست آمده از تجزیه و تحلیل مسیر، منجر به شناسایی ۱۷ طبقه مختلف عملکردی هستی‌شناسی ژن و مسیرهای زیستی KEGG شامل مسیرهای cytoskeletal protein binding، anatomical structure development، Tricarboxylic acid cycle مرتبط با توسعه الیف‌های عضلانی اسکلتی و سوخت و ساز چربی شده بود (Khalatabadi Farahani *et al.*, 2020).

هدف از انجام پژوهش حاضر، شناسایی مناطق ژنومی حاوی مسیرهای زیستی و ژن‌های کاندیدای مرتبط با سن اولین گوساله‌زایی در گاوهای نلور بر اساس تجزیه بر مبنای مسیر و با استفاده از روش غنی‌سازی مجموعه‌های ژنی است.

مواد و روش‌ها

از اطلاعات صفات تولیدمثلی گاوهای نلور که با هدف پویش کل ژنومی براساس مدل خطی هاپلوتیپ بود، استفاده شد (Nascimento *et al.*, 2018). اطلاعات شامل تعداد ۲۲۷۳ رأس گاو شامل ۹۹۵ نر و ۱۲۷۸ ماده بودند و dEBV آنها در دسترس بود. DNA نمونه‌ها با استفاده از آرایه‌های Illumina BovineHD بر اساس پروتکل استاندارد Illumina برای ۷۷۷۹۶۲ جایگاه تعیین ژنوتیپ شده بودند. در مطالعه اولیه از تصحیح بنفرونی برای تعیین آستانه معنی‌داری و جلوگیری از خطای نوع اول استفاده شده بود. برای فیلتراسیون داده‌های تعیین ژنوتیپ شده، ابتدا حیوانات با بیش از ۱۰ درصد ژنوتیپ از دست رفته حذف شدند. در مرحله بعد، SNPهایی که حداقل فراوانی آللی در آنها کمتر از دو درصد و SNPهایی که در کمتر از ۹۵ درصد افراد تعیین ژنوتیپ شده بودند حذف شدند. در نهایت، SNPهای باقیمانده که در تعادل هاردی-واینبرگ (سطح احتمال ۱۰^{-۵}) قرار نداشتند به عنوان معیاری از خطای تعیین ژنوتیپ کنار گذاشته شدند. همچنین SNPهای با موقعیت نامشخص روی کروموزوم، از ادامه تجزیه حذف شدند. پس از کنترل کیفیت، تعداد ۵۱۱۳۷۵ SNP برای تجزیه‌های پویش کل ژنومی بر پایه تجزیه و تحلیل غنی-سازی مجموعه ژنی باقی ماندند.

تجزیه و تحلیل پویش کل ژنومی: در این پژوهش، ابتدا برای فازبندی ژنوتیپ نشانگرها از نرم افزار fastPHASE نسخه ۱/۲ استفاده شد. در این برنامه از روش بیزی بر اساس مدل مخفی مارکوف (HMM) استفاده شده است (Scheet and

نتایج و بحث

پلات‌های Q-Q و منهن مرتب با صفت سن اولین گوساله-زایی به ترتیب در شکل‌های ۱ و ۲ ارائه شده است. چندین روش برای تخمین کنترل تورم ژنومیکی (λ) وجود دارد که در این تحقیق، روش تخمین گر میانه اجرا شد. فاکتور تورم کنترل جمعیتی بزرگتر از ۱ ($\lambda > 1$) وجود لایه‌بندی جمعیتی و یا خطای تعیین ژنوتیپ را نشان می‌دهد. فاکتور لامبدا از راه تجزیه و تحلیل پیوستگی در PLINK برای صفت مورد مطالعه محاسبه شد که برای صفت سن اولین گوساله‌زایی برابر با ۱/۰۳۴ بود که همان طور که مشاهده می‌شود تقریباً برابر با یک بود و نشان‌دهنده عدم وجود لایه‌بندی جمعیتی برای اجرای تجزیه و تحلیل‌های پویش کل ژنومی بود. در مطالعاتی که لامبدا کوچکتر از ۱/۱ یا مساوی ۱ است نیازی به تصحیح آثار ساختار جمعیتی نیست (Hinrichs et al., 2009).

شناسایی ژن‌های کاندیدا در مناطق ژنومی: تعداد ۱۹۷۲۳ ژن از ۲۷۶۰۷ ژن حاشیه نویسی شده در گاو به وسیله نشانگرهای SNP پوشش داده شد که در این میان، تعداد ۱۷۳۵ ژن دارای اثر معنی‌داری بودند، یعنی حداقل یک نشانگر با P -value کمتر از ۰/۰۰۵ در داخل و یا در بالا یا پایین دست این ژن تا فاصله ۱۵ kb قرار گرفت. این ژن‌ها به عنوان ژن‌های معنی‌دار مرتبط با صفات تولیدمثلی برای تجزیه GSEA-SNP انتخاب شدند و باقیمانده ژن‌ها به عنوان ژن‌های پس‌زمینه‌ای مورد استفاده قرار گرفتند.

تعداد مجموعه‌های ژنی حاصل از پایگاه‌های داده‌ای مختلف شامل ۸۶۱ طبقات هستی‌شناسی، ۵۴ مسیر بیوشیمیایی KEGG و ۳۷ مسیر Reactome بود. همان‌طور که در جدول ۱ مشاهده می‌شود تعداد ۱۲ طبقات عملکردی در هستی‌شناسی فرآیندهای زیستی، عملکرد مولکولی، اجزای سلولی و مسیرهای KEGG با صفات تولیدمثلی دارای ارتباط هستند ($P < 0/05$). مسیرهای که بیشتر از سه ژن و کمتر از ۵۰۰ ژن داشتند گزارش شده‌اند.

جزئیات کامل ترم‌های هستی‌شناسی به همراه اسامی ژن-های کاندیدا در جدول ۱ ارائه شده است. نتایج حاصل از این تحلیل نشان‌دهنده این است که ژن‌های HSD17B12 و SULT1E1 با فرآیند Estrogen metabolic process، ژن‌های KIRREL3 و PPFIA2 با فرآیند Synapse

حاوی یک SNP معنی‌دار در سطح ۰/۰۰۵ باشد، ژن معنی-دار محسوب می‌شود.

ارتباط ژن‌ها به طبقات عملکردی و مسیرهای بیوشیمیایی: پایگاه‌های داده‌های هستی‌شناسی ژن GO (<http://www.geneontology.org>) و مسیرهای بیوشیمیایی DAVID (<http://www.david.org>) و PANTHER (<http://www.pantherdb.org>) جهت تعیین طبقات عملکردی و مسیرهای سوخت و ساز و تنظیمی ژن-های معنی‌دار مورد استفاده قرار گرفتند. در این مرحله فرض بر این است که ژن‌هایی که در یک طبقه عملکردی قرار می‌گیرند می‌توانند به عنوان یک گروه از ژن‌هایی که برخی از ویژگی‌های خاص و مشترک دارند، مانند شرکت در سه فرآیند هستی‌شناسی شامل فرآیندهای زیستی، عملکرد مولکولی و اجزای سلولی، در نظر گرفته شوند. پویش کل ژنومی بر پایه تجزیه و تحلیل مسیر: ارتباط‌های معنی‌دار مسیرهای عملکردی با صفت اولین سن زایش با استفاده از توزیع فوق هندسی (Hypergeometric) و آزمون Fisher's exact test مورد آزمون قرار گرفت. P -value مسیرهای عملکردی که تعداد k ژن معنی‌دار در آن قرار دارد با فرمول زیر محاسبه شد (Peñagaricano et al., 2013):

$$P - value = 1 - \sum_{i=1}^{k-1} \frac{\binom{S}{i} \binom{N-S}{m-i}}{\binom{N}{m}}$$

در این فرمول، k برابر با تعداد ژن‌های معنی‌دار در طبقه عملکردی، S برابر با تعداد کل ژن‌های معنی‌دار مرتبط با صفت مورد بررسی، N برابر با کل تعداد ژن‌هایی که در این مطالعه تجزیه و تحلیل شدند و m برابر با تعداد کل ژن‌های موجود در مسیر عملکردی بود. تجزیه غنی‌سازی مجموعه ژنی با استفاده از بسته نرم افزاری goseq (Young et al., 2010) در محیط نرم افزار R انجام شد. بسته goseq از آزمون فوق هندسی استفاده می‌کند. همچنین برای اختلاف طول ژن‌ها نیز تصحیح انجام می‌شود. برای تفسیر بهتر عملکرد ژن‌های به‌دست آمده از پایگاه‌های اطلاعاتی آنالین GeneCards (<http://www.genecards.org>) و g:profiler (<https://biit.cs.ut.ee/gprofiler/gost>) استفاده شد.

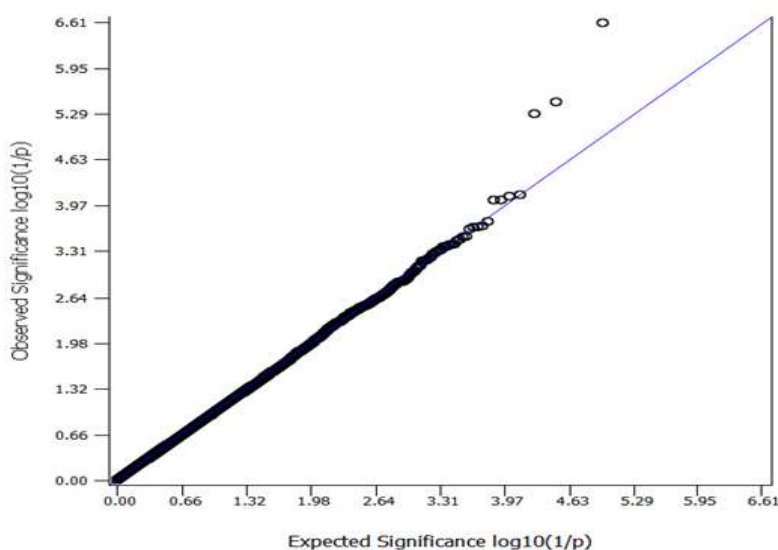


Fig. 1. Q-Q haplotype-based plots for age at first calving traits in Nelore cattle
 شکل ۱- پلات Q-Q بر پایه هاپلوتیپ برای صفت سن اولین گوساله‌زایی در گاو نلور

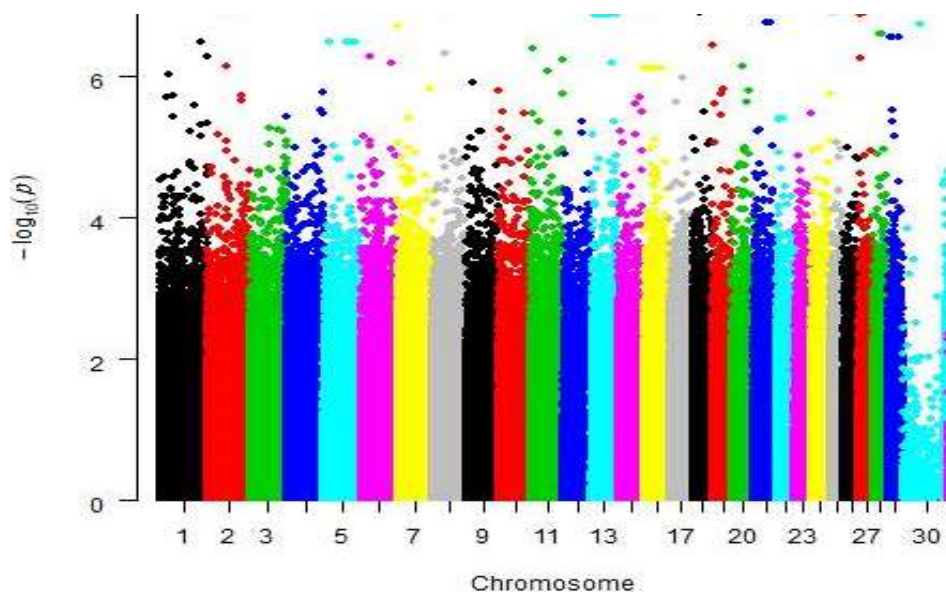


Fig. 2. Manhattan plot for age at first calving trait in Nelore cattle. X axis: Haplotype alleles positions on chromosomes, and Y axis: $-\log_{10} P$ -value

شکل ۲- پلات منهن برای صفت سن اولین گوساله‌زایی در گاو نلور. محور X، مکان آلل‌های هاپلوتیپی روی کروموزوم‌ها و محور Y، منفی لگاریتم بر مبنای ۱۰ ارزش P-value

مختلف بررسی ژن‌های کاندیدای مرتبط با صفات تولیدمثل، به خصوص باروری و سن اولین گوساله‌زایی، در ارتباط هستند.

ژن HSD17B12 نقش کلیدی در بیوسنتز استروژن دارد، به‌طوری که ژن HSD17B12 کد کننده آنزیم تبدیل

organization و ژن FYN در فرایند زیستی Defense response و ژن‌های کاندیدای MYO3A و KCNMA1 با فرایند Sensory perception of mechanical stimulus در ترم فرآیندهای زیستی بوده که مرتبط با صفات تولیدمثل هستند. برخی از ژن‌های این فرایند زیستی در مطالعات

ژن کاندیدای MYO3A برای اولین بار در مطالعه پویش کل ژنومی در گاوهای هلشتاین ایرانی در ارتباط با صفات باروری گاوهای ماده شامل سن اولین تلقیح، سن اولین گوساله‌زایی و فاصله گوساله‌زایی گزارش شده است (Mohammadi *et al.*, 2020).

ژن‌های کاندیدای SH2B2 و FYN در تنظیم تعادل انرژی و وزن بدن نقش کلیدی دارند. ژن کاندیدای FYN در مطالعه پویش کل ژنومی بر پایه تجزیه و تحلیل غنی‌سازی ژنومی در گاوهای شیری هلشتاین در شرایط تنش حرارتی در ارتباط با عملکرد تولیدمثلی شامل باروری و رشد ابتدایی جنین گزارش شده است (Sigdel *et al.*, 2020). همچنین در مطالعه بررسی چندشکلی ژن کاندیدای SH2B2 در گاوهای گوشتی Qinchuan، ارتباط معنی‌داری با صفات اندازه بدن گزارش شده است (Abbas Raza *et al.*, 2020). عملکردهای مولکولی و اجزای سلولی مشاهده شده شامل ژن‌های IGF1R و FGFR4 با فرآیند Protein tyrosine kinase activity و ژن SCN1A با فرآیند Calcium channel activity و ژن FRMD4A با فرآیند سلولی Cell-cell junction بودند.

مسیر cell-cell junction یکی از مهم‌ترین مسیرهای مؤثر بر فرآیندهای تخمدان معرفی شده است. مسیر ذکر شده برای حمایت از سوخت و ساز تخمک در حال رشد، یون‌ها را برای تنظیم pH تخمک تامین کرده و به تامین cGMP مورد نیاز برای نگهداری و حفظ تخمک در حالت میوزی کمک می‌کند. مطالعه پویش کل ژنومی در گاوهای بومی کلمبیا (Romosinuano and Costeño con Cuernos) با عملکرد تولیدمثلی بالا، ارتباط معنی‌داری بین ژن‌های IGF1R و FRMD4A با سن اولین گوساله‌زایی و فاصله گوساله‌زایی مشخص کرد (Fernández *et al.*, 2018). همچنین در مطالعات مختلف پویش کل ژنومی، ارتباط معنی‌داری بین ژن IGF1R با صفات باروری گاو نر و سن بلوغ و همچنین وزن بدن در گوسفند گزارش شده است (Honarvar *et al.*, 2012; Han and Peñagaricano, 2016).

استرون به استرادیول است و همچنین در ساخت اسید آراشیدونیک نقش کلیدی داشته که برای زنده‌مانی جنین لازم است. در یک مطالعه پویش کل ژنوم، ژن کاندیدای HSD17B12 ارتباط معنی‌داری با صفت نرخ آبستنی در گاوهای هلشتاین داشت (Cochran *et al.*, 2013). همچنین در تحقیق دیگر روی تلیسه‌های هلشتاین با هدف شناسایی ژن‌های کاندیدای مرتبط با نرخ آبستنی، ژن HSD17B12 به عنوان شاخصی برای پیش‌بینی شایستگی ژنتیکی عملکرد تولیدمثل گزارش شد (Ortega *et al.*, 2016).

استروژن یکی از دو هورمون جنسی استروئیدی است که به وسیله تخمدان ترشح می‌شود و در رشد جنینی، بروز صفات ثانویه جنسی و چرخه باروری شامل شکل‌گیری فولیکول‌ها و بلوغ تخمک‌ها نقش مهمی ایفا می‌کند. برای تنظیم عملکرد فیزیولوژیکی هورمون‌های استروئیدی وجود گیرنده‌ها در سلول‌های هدف ضروری است. گیرنده‌های استروژن در سلول‌های رحم و واژن و به‌طور اختصاصی در تخمک، سلول‌های گرانولوزا و سلول‌های اپیتلیال تخمدان وجود دارند.

PPFIA2 ژن مهم دیگر مرتبط با تولیدمثل است که تأثیر این ژن بر نرخ باروری در مطالعات اخیر مشاهده شده است. در مطالعه پویش کل ژنومی گاوهای نلور در ارتباط با سن اولین گوساله‌زایی، ژن PPFIA2 روی کروموزوم ۵ گزارش شده است (Mota *et al.*, 2020). ژن PPFIA2 نقش مهمی در بسیاری از فرآیندهای تولیدمثلی و توسعه غدد پستانی بر عهده دارد.

ژن کاندیدای KCNMA1 در بین ژن‌های این مجموعه ژنی قرار داشت و محصول این ژن، پروتئینی به همین نام است که یک کانال پتاسیمی است و ویژگی بارز آن، انتقال مقادیر زیاد پتاسیم از غشای سلولی است. با توجه به نقش گسترده کانال‌های پتاسیمی در بافت‌ها، KCNMA1 می‌تواند با عملکردهای متفاوتی، فرآیند باروری را تحت تأثیر قرار دهد. در مطالعه پویش کل ژنومی گاوهای ایتالیایی، ارتباط معنی‌داری بین ژن KCNMA1 با صفت سن گوساله‌زایی گزارش شده است (Li *et al.*, 2018).

جدول ۱- تجزیه و تحلیل غنی‌سازی مجموعه‌های ژنی معنی‌دار ($P < 0.05$) مرتبط با صفات تولیدمثلی

Table 1. Gene set enrichment analysis which significantly ($P < 0.05$) associated with reproductive traits

GO ID	Term (GO hierarchy level)	No. genes in the GO term	No. significant genes	Related candidate genes	FDR
Biological process					
GO: 0008210	Estrogen metabolic process	38	4	HSD17B12, SULT1E1	1.77E-02
GO: 0050808	Synapse organization	65	15	FXR1, KIRREL3, PPFA2	1.53E-02
GO: 0006952	Defense response	392	12	NOX4, FYN, CD44	4.29E-02
GO: 0006955	Immune response	267	12	SH2B2	1.04E-02
GO: 0050954	Sensory perception of mechanical stimulus	124	22	FAM107B, GRM7, MYO3A, KCNMA1, SOBP, TMEM232	1.69E-02
GO: 0002673	Regulation of acute inflammatory response	159	9	ASH1L, CD59, NLRP3	2.56E-02
GO: 0006940	Regulation of smooth muscle contraction	86	9	SULF2, ADRA1B	1.46E-02
Molecular Function					
GO: 0004713	Protein tyrosine kinase activity	84	18	IGF1R, FGFR4	2.16E-02
GO: 0005262	Calcium channel activity	86	17	SCN1A	4.57E-02
Cellular component					
GO: 0005911	Cell-cell junction	75	17	FRMD4A	1.30E-02
PANTHER Pathways					
P04912	GnRH signaling pathway	37	11	ADCY5, ITPR2, MAPK12, MAP3K1	1.22E-02
P04510	Focal adhesion	92	21	HGF, PDGFD, PPP1R12A,	2.74E-02

ارتباط با تنظیم سیگنالینگ انسولین و گیرنده‌های انسولین دارد. همچنین نقش این ژن در سوخت و ساز و هموستازی انسولین و گلوکز در تخمدان مشخص شده است. در مطالعه پویش کل ژنومی در تلیسه‌های نلور با هدف شناسایی جایگاه‌های ژنی مرتبط با سن اولین گوساله‌زایی، ژن PPP1R12A گزارش شده است (Mota *et al.*, 2020). همچنین با مطالعه پویش کل ژنومی برای صفات شاخص مرتبط با بلوغ جنسی زودتر در گاوهای نلور، ژن کاندیدای PPP1R12A گزارش شده است (Irano *et al.*, 2016).

در یک مطالعه پویش ژنومی صفات مرتبط با سن اولین گوساله‌زایی که بر اساس مدل هاپلوتیپ بر پایه پنجره‌های ۱ Mb انجام شده بود ارتباط سن اولین گوساله‌زایی با یک ناحیه ۳۴۵ kb روی کروموزوم ۲۶ گزارش شده بود که با منطقه شناسایی شده در پژوهش حاضر همخوانی داشت. همچنین بر اساس تجزیه و تحلیل غنی‌سازی مجموعه ژنی، ناحیه معنی‌دار مؤثر روی فرآیند لانه‌گزینی رویان در رحم

مسیرهای زیستی مرتبط با ژن‌های تولیدمثلی با استفاده از پایگاه داده KEGG مورد بررسی قرار گرفت که با نتایج برخی از تحقیقات قبلی مرتبط با صفات تولیدمثلی مطابقت داشت. با بررسی نتایج حاصل مشاهده می‌شود که ژن‌های ADCY5، ITPR2، MAPK12 و MAP3K1 با مسیر زیستی GnRH signaling pathway، و ژن‌های HGF، PDGFD و PPP1R12A با مسیر زیستی Focal adhesion در گاوهای نلور مرتبط هستند.

مطالعه پویش کل ژنومی بر پایه تجزیه و تحلیل ژنی و تفسیر عملکردی بیان ژنی در گاوهای شیری شمال اروپا نشان داد ژن ADCY5 به عنوان ژن کاندید مرتبط با باروری و در مسیر میوزیس اووسیت گزارش شده است (Cai *et al.*, 2019).

مسیر سیگنالی GnRH signaling pathway و ژن‌های مرتبط شناسایی شده از پایگاه برخط KEGG در شکل ۳ ارائه شده است. ژن کاندیدای PPP1R12A نقش کلیدی در

به سوی ژن‌های مؤثر بر صفت سن اولین زایش هدایت کند و بررسی بیشتر نواحی مهم ژنومی شناسایی شده با استفاده از آزمون‌های آزمایشگاهی مختلف می‌تواند در تأیید نتایج به‌دست آمده در این پژوهش مؤثر باشند. به دلیل عدم دسترسی به رکوردهای فنوتیپی و اطلاعات ژنوتیپی مرتبط با صفت سن اولین زایش در گاوهای بومی کشور در این تحقیق، استفاده از نتایج حاصل در گاوهای بومی کشور نیاز به مطالعات بیشتر دارد تا در این جمعیت‌ها نیز تأیید شوند.

تشکر و قدردانی

نویسندگان مراتب قدردانی خود را از خانم دکتر گریسولیا از دانشگاه UFGD کشور برزیل به خاطر در اختیار گذاشتن داده‌های ژنوتیپی و رکوردهای فنوتیپی برای انجام پژوهش حاضر اعلام می‌دارند.

و بارداری مشخص شد که با خانواده ژنی PRDX5 و RGS3 در مطالعه پیشین مشترک بود (Nascimento *et al.*, 2018).

نتیجه‌گیری کلی

بررسی مناطق ژنومی شناسایی شده روی کروموزوم‌های ۵، ۹، ۱۳، ۱۵، ۲۱، ۲۵، ۲۶ و ۲۹ با استفاده از پایگاه داده BioMart، GeneCards و UniProtKB نشان داد که این مناطق با صفت سن اولین زایش مرتبط هستند. با توجه به عملکرد ژن‌های کاندیدای شناسایی شده در مسیرهای زیستی این پژوهش، به نظر می‌رسد می‌توان کارآیی روش تجزیه و تحلیل غنی‌سازی مجموعه ژنی برای پویش ژنومی صفات تولیدی اقتصادی را نیز مورد تأیید قرار داد. مسیرهای شناسایی شده در این تحقیق می‌تواند محققان را

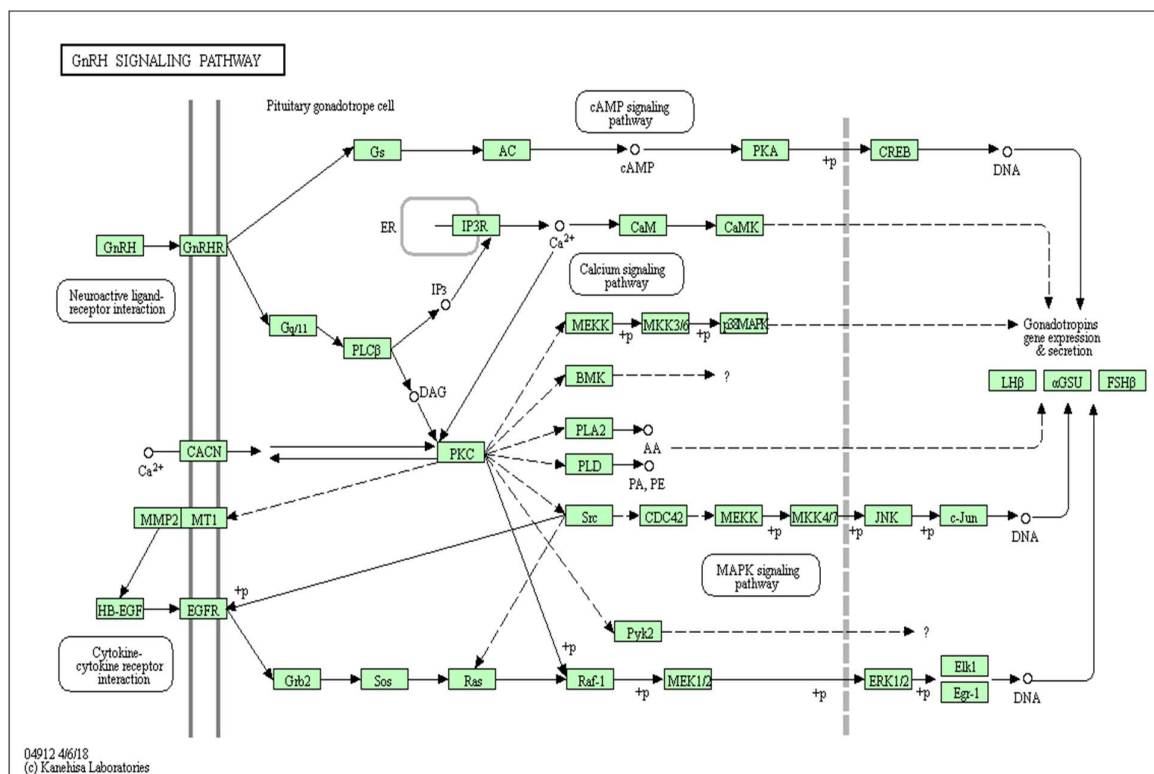


Fig. 3. GnRH signaling pathway and candidate genes related to reproductive traits highlighted (KEGG database)
 شکل ۳- مسیر سیگنالی GnRH و ژن‌های کاندیدای مرتبط با صفت تولیدمثلی که به‌صورت هایپلایت شده مشخص شده‌اند
 (پایگاه داده KEGG)

فهرست منابع

- Boligon A. A. and Albuquerque L. G. 2011. Genetic parameters and relationships of heifer pregnancy and age at first calving with weight gain, yearling and mature weight in Nelore cattle. *Livestock Science*, 141(1): 12-16.
- Cai Z., Guldbrandtsen B., Lund M. S. 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): 255.
- Cochran S. D., Cole J. B., Null D. J. and Hansen P. J. 2013. Discovery of single nucleotide polymorphisms in candidate genes associated with fertility and production traits in Holstein cattle. *BMC Genetics*, 7(14): 49.
- Durinck S., Spellman P. T., Birney E. and Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. *Nature Protocols*, 4: 1184-1191.
- Fernández J. C., Pérez J. E., Herrera N., Martínez R., Bejarano D. and Rocha J. F. 2019. Genomic association study for age at first calving and calving interval in Romosinuano and Costeño con Cuernos cattle. *Genetics and Molecular Research*, 18(2): 1-13.
- Garrick D. J., Taylor J. F. and Fernando R. L. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*, 41(1):1.
- Han Y. and Peñagaricano F. 2016. Unravelling the genomic architecture of bull fertility in Holstein cattle. *BMC Genetics*, 17: 143.
- Hare H. D., Norman J. R. and Wright A. 2006. Trends in calving ages and calving intervals for dairy cattle breed in the United States. *Journal of Dairy Science*, 89: 365-370.
- Honarvar M., Sadeghi M., Moradi-Shahrehabak H., Behzadi S. H., Mohammadi H. and Lavaf A. 2012. Study of Polymorphisms in the 5' flanking region of the Ovine IGF-I gene in Zel sheep. *World Applied Sciences Journal*, 16(5): 726-728.
- Hinrichs A. L., Larkin E. K. and Suarez B. K. 2009. Population stratification and patterns of linkage disequilibrium. *Genetic Epidemiology*, 33: 88-92.
- Khaltabadi Farahani A. H., Mohammadi H., Moradi M. H., Ghasemi H. A. and Hajkhodadadi I. 2020. Gene set enrichment analysis using genome-wide association study to identify genes and pathways associated with litter size in various sheep breeds. *Animal Production Research*, 9(3): 47-57. (In Persian).
- Li J., Liu J., Campanile G., Plastow G., Zhang C., Wang Z., Cassandro M., Gasparrini B., Salzano A., Hua G., Liang A. and Yang L. 2018. Novel insights into the genetic basis of buffalo reproductive performance. *BMC Genomics*, 19(1): 814.
- Marques D. B. D., Bastiaansen J. W. M., Broekhuijse M. L. W. J., Lopes M. S., Knol E. F., Harlizius B., Guimarães S. E. F., Silva F. F. and Lopes P. S. 2018. Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. *Genetics Selection Evolution*, 50(1): 40.
- Magnabosco C. U., Lopes F. B., Magalhaes Rosa G. J. and Sainz R. D. 2016. Bayesian estimates of genetic parameters for reproductive traits in Nelore cows raised on pasture in tropical regions. *Revista Colombiana de Ciencias Pecuarias*, 29(2): 119-129.
- Mohammadi A., Alijani S., Rafat S. A. and Abdollahi-Arpanahi R. 2020. Genome-wide association study and pathway analysis for female fertility traits in Iranian Holstein cattle. *Annals of Animal Science*, 20(3): 825-851.
- Mota R. R., Guimarães S. E. F., Fortes M. R. S., Hayes B., Silva F. F. and Verardo L.L. 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(6): 484-492.
- Mota L. F. M., Lopes F. B., Fernandes G. A., Rosa G. J. M., Magalhães A. F. B., Carneiro R. and Albuquerque L. G. 2020. Genome-wide scan highlights the role of candidate genes on phenotypic plasticity for age at first calving in Nelore heifers. *Scientific Reports*, 10(1): 6481.
- Nascimento A. V. D., Romero A. Ñ. R. D. S., Utsunomiya Y. T., Utsunomiya A. T. H., Cardoso D. F. and Neves H. H. R. 2018. Genome-wide association study using haplotype alleles for the evaluation of reproductive traits in Nelore cattle. *PLoS ONE*, 13(8): e0201876.
- Ortega M. S., Denicol A. C., Cole J. B., Null D. J. and Hansen P. J. 2016. Use of single nucleotide polymorphisms in candidate genes associated with daughter pregnancy rate for prediction of genetic merit for reproduction in Holstein cows. *Animal Genetics*, 47(3): 288-297.
- Peñagaricano F., Weigel K. A., Rosa G. J. and Khatib H. 2013. Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. *Frontiers in Genetics*, 3: 307-314.
- Raza S. H. A., Khan R., Schreurs N. M., Wang X., Mei C., Yang X., Gong C. and Zan L. 2020. Bioinformatics analysis and genetic polymorphisms in genomic region of the bovine SH2B2 gene and their associations with molecular breeding for body size traits in Qinchuan beef cattle. *Bioscience Reports*, 40(3): BSR20192113.

- Sharma A., Lee J. S., Dang C. G., Sudrajat P., Kim H. C., Yeon S. H., Kang H. S. and Lee S. H. 2015. Stories and Challenges of Genome Wide Association Studies in Livestock-A Review. *Asian-Australasian Journal of Animal Science*, 28(10): 1371-1379.
- Scheet P. and Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, 78(4): 629-644.
- Sigdel A., Liu L., Abdollahi-Arpanahi R., Aguilar I. and Peñagaricano F. 2020. Genetic dissection of reproductive performance of dairy cows under heat stress. *Animal Genetics*, 51(4): 511-520.
- Utsunomiya Y. T., Milanesi M., Utsunomiya A. T., Ajmone-Marsan P. and Garcia J. F. 2016. GHap: An R package for Genome-wide Haplotyping. *Bioinformatics*, 32(18): 2861-2862.
- Young M. D., Wakefield M. J., Smyth G. K. and Oshlack A. 2010. Method gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology*, 11: 14-23.