

## Application of a random forest algorithm to estimate marker effects and identify candidate genes for reproductive traits in Iranian Holstein dairy cattle

J. Jabbari Turchi<sup>1</sup>, S. Alijani<sup>2\*</sup>, S. A. Rafat<sup>2</sup>, M. A. Abbasi<sup>3</sup>

1. Ph.D. Student in Animal Breeding and Genetics, Department of Animal Science, Faculty of Agricultural Sciences, University of Tabriz, Tabriz, Iran
2. Professor, Department of Animal Sciences, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
3. Associate Professor, the National Animal Breeding Center and Promotion of Animal Products, Karaj, Iran

(Received: 13-09-2023 – Revised: 29-11-2023 – Accepted: 02-12-2023)

**Introduction:** The genome-wide association study (GWAS) is a powerful approach to identify genomic regions associated with fertility traits that explain a significant portion of the genetic variance associated with these traits and identify the relevant causal mutations. Evaluating the correlation between each genotyped marker and trait is an essential strategy for GWAS studies that examine the effects of all markers by considering their possible interactions, environmental factors, and even mutual effects between markers. Recently, machine learning methods have been introduced to genomic topics, and the basis of these methods is different from the common methods of genomic evaluation. The machine learning method is used to estimate the genomic breeding values of the candidate animals by considering the training data (genotypic and phenotypic information of the reference population). One of the key advantages of this method is the ability to analyze large data. Machine learning is a branch of artificial intelligence whose goal is to achieve machines that can extract knowledge (learning) from the environment. A variety of machine learning methods (random forest, boosting, and deep learning) are used to model genetic variance and environmental factors, study gene networks, GWAS, study epistasis effects, and genomic evaluation. Random forest is one of the machine learning methods that has been successfully used in various fields of science. This research was conducted to identify markers and genes related to reproductive traits such as calving interval (CI), days open (DO), daughter pregnancy rate (DPR), and age at first calving (AFC) in Iranian Holstein dairy cattle. These traits have already been investigated with the ssGBLUP method and using a smaller sample size. However, in the present research, by using more genotyped animals, a random forest algorithm was used to identify markers and genes related to reproductive traits.

**Materials and methods:** The records used in this research were provided by the National Animal Breeding Center and Promotion of Animal Products of Iran and included AFC, DO, CI, and DPR related to the genotyped bulls' daughters. In this research, the pedigree information of 2774183 animals was used. The genotypic information of the markers related to 2419 Holstein bulls was used. Genomic data quality control was performed using factors such as the number of genotyped SNPs per animal (ACR), the number of genotyped animals per SNP (CR), Hardy-Weinberg equilibrium (HWE), and minor allele Frequency (MAF). When filtering genomic data, the markers whose MAF was less than 5% were removed, and then the samples whose genotyped frequency was less than 90% were identified and removed. Then, the markers whose genotyping rate was less than 95% in the samples were identified and removed. Finally, the SNPs that deviated from the HWE test ( $P < 10^{-6}$ ) were excluded from the analysis as a measure of genotyping error. To control the quality of genomic data, PLINK 1.9 software was used. Then Ranfog software was used in the Linux environment to perform analysis through random forest algorithm.

**Results and discussion:** By using the random forest algorithm, a total of 21 important SNPs were observed, then important fertility trait candidate genes were identified by the gene ontology method, and 62 genes were within

\* Corresponding author: sad-ali@tabrizu.ac.ir



250 Kb of these SNPs. The most significant SNP was observed for AFC. The main SNP for AFC is in ARS-BFGL-NGS-22647 BTA3, for CI is in ARS-BFGL-NGS-114194 (BTA11), for DO is in BTA-74076 -no-rs (BTA5), and for DPR is in ARS-BFGL-NGS-32553 (BTA26). The researchers, who studied fertility traits in Nellore cattle using machine learning methods, identified *MPZLL1* and *CD247* genes on chromosome number 3 and this gene was associated with age at first calving. Many pathways of cell biology affect the performance of reproductive traits. Research has reported the relationship between the *CD247* gene and pathways of biology, including cell development and function. Research has shown that the *IFFO2* gene plays an important role in the molecular structure of cells, as well as in the mechanism of blastocyst formation, embryos, and the length of gestation in cattle. In a study conducted on the mouse population on the structure of the flagellum and the sperm maturation process, the role of the *ALDH4AI* gene in the sperm maturation process was reported. The association of the *RPS6KCI* gene with pregnancy rate and antral follicle number in Nellore heifers has been reported. The *KAT2B* gene is a transcriptional activator that plays an essential role in regulating the correction of histone acetylation and plays an important role in improving carcass quality, muscle and fat development, and metabolism in native Chinese cattle. In addition, they play a key role in regulating biological processes and are related to cell growth, metabolism and immune system function.

**Conclusions:** According to the objectives of this research, new information on markers and candidate genes related to reproductive traits in Iranian Holstein dairy cattle was reported. The markers and candidate genes identified in the present research can be used in genomic selection to improve the reproductive traits of Holstein dairy cattle.

**Keywords:** Random forest algorithm, Genotype, Dairy cow, Marker, Machine learning

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

**Funding:** The authors received no specific funding for this work.

**Acknowledgments:** The authors would like to thank the National Animal Breeding Center and Promotion of Animal Products of Iran for providing genomic data and information.

#### How to cite this article:

Jabbari Turchi, J., Alijani, S., Rafat, S. A., & Abbasi, M. A. (2024). Application of a random forest algorithm to estimate marker effects and identify candidate genes for reproductive traits in Iranian Holstein dairy cattle. *Animal Production Research*, 13(1), 95-109. doi: 10.22124/AR.2024.25557.1792



## کاربرد الگوریتم جنگل تصادفی در برآورد آثار نشانگرها و تعیین ژن‌های کاندیدا برای صفات تولیدمثلی در گاو شیری هلشتاین ایران

جیران جباری تورچی<sup>۱</sup>، صادق علیجانی<sup>۲\*</sup>، سید عباس رأفت<sup>۲</sup>، مختارعلی عباسی<sup>۳</sup>

۱- دانشجوی دکتری ژنتیک و اصلاح نژاد دام، گروه علوم دامی، دانشکده کشاورزی، دانشگاه تبریز

۲- استاد، گروه علوم دامی، دانشکده کشاورزی، دانشگاه تبریز

۳- دانشیار، مؤسسه تحقیقات علوم دامی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی

(تاریخ دریافت: ۱۴۰۲/۰۶/۲۲ - تاریخ بازنگری: ۱۴۰۲/۰۹/۰۸ - تاریخ پذیرش: ۱۴۰۲/۰۹/۱۱)

### چکیده

روش یادگیری ماشین، رویکرد قدرتمندی برای مطالعات ژنومی است. هدف تحقیق حاضر، استفاده از روش یادگیری ماشین (جنگل تصادفی) برای پیش‌بینی ژنومی پیشنهادی صفات تولیدمثلی شامل سن در زمان اولین زایش (AFC)، روزهای باز (DO)، فاصله گوساله‌زایی (CI) و نرخ آبستنی دختران (DPR) در گاوهای هلشتاین ایران بود. اطلاعات لازم از مرکز اصلاح نژاد و بهبود تولیدات دامی کشور اخذ شد. اطلاعات ژنوتیپی شامل نشانگرهای چند شکلی تک نوکلئوتیدی (SNP) مربوط به ۲۴۱۹ رأس گاو هلشتاین نر بود. فایل داده مشتمل بر رکوردهای ثبت شده سال‌های ۱۳۶۰ تا ۱۳۹۸ شامل ۲۷۷۴۱۸۳ رأس دام بود. با توجه به تفاوت تراکم در اطلاعات ژنومی گاوهای نر، تعداد نشانگرهای آن‌ها نیز با یکدیگر متفاوت بود. برای یکسان‌سازی نشانگرها از نرم افزار FImpute برای جانهی ژنوتیپ استفاده شد. در این تحقیق با استفاده از الگوریتم جنگل تصادفی که نمونه‌ای از الگوریتم‌های با نظارت و از نوع رگرسیونی است، در مجموع، ۲۱ نشانگر با میزان اهمیت بالا برای صفات مختلف تولیدمثلی مشخص شد. سپس، با استفاده از روش هستی‌شناسی ژن، ژن‌های پیشنهادی مهمی برای این صفات شناسایی شدند. ژن‌های MPZL1 و CD247 شناسایی شده روی کروموزوم ۳ در ارتباط با صفت AFC و ژن‌های RPS6KC1 و FAM170A در ارتباط با صفت DPR برای بهبود عملکرد تولیدمثلی گاوهای شیری، مهم بوده و می‌توانند مورد استفاده قرار گیرند. نشانگرها و ژن‌های شناسایی شده در این تحقیق می‌توانند اطلاعات جدیدی را در مورد معماری ژنتیکی صفات تولیدمثلی برای بهبود ژنومی آن‌ها ارائه دهد و در طراحی تراشه‌ها برای ارزیابی صفات تولیدمثلی مورد استفاده قرار گیرد.

واژه‌های کلیدی: الگوریتم جنگل تصادفی، ژنوتیپ، گاو شیری، نشانگر، یادگیری ماشین

\* نویسنده مسئول: sad-ali@tabrizu.ac.ir

## مقدمه

در سال‌های اخیر، روش‌های اصلاح نژاد دام تا حد زیادی تغییر یافته است، به طوری که پیشرفت قابل توجهی در زمینه فناوری زیستی حاصل شده و ابزار قدرتمندی را برای مطالعات ژنتیکی موجودات زنده و به تبع آن، گونه‌های دامی فراهم آورده است (Pryce *et al.*, 2010). با این نشانگرهای انبوه تک نوکلئوتیدی در سراسر ژنوم، کاربردهای بسیار زیادی در مطالعات مختلف ژنومی از جمله انتخاب به کمک نشانگرها (MAS) برای شناسایی جایگاه‌های ژنی مؤثر بر صفات کمی (QTL) به وجود آمده است. انتخاب ژنومی شکلی از MAS است که در آن از تمام نشانگرهای ژنتیکی که کل ژنوم را پوشش می‌دهند به طور همزمان استفاده می‌شود (Meuwissen *et al.*, 2001). حیوانات اهلی، منابع ارزشمندی برای مطالعات معماری مولکولی صفات پیچیده هستند. اگرچه نقشه‌یابی QTLها در دهه‌های اخیر نتایج خوبی به همراه داشته است، ولی تمامی تغییرات ژنتیکی صفات کمی به دلیل استفاده از تراشه‌های SNP با تراکم پایین به دست نیامده است. در نتیجه، از نشانگرهایی با تراکم بالا در مطالعات پیوستگی در سطح ژنوم (GWAS) استفاده شده است. در حال حاضر از مطالعات GWAS برای برآورد ارزش اصلاحی در حیوانات اهلی و شناسایی نشانگرهای معنی‌دار مؤثر بر صفات کمی استفاده می‌شود که باعث افزایش روند پیشرفت ژنتیکی در این زمینه شده است (Zhang *et al.*, 2012). مطالعات GWAS یک رویکرد قدرتمند برای شناسایی مناطق ژنومی است که بخش قابل توجهی از واریانس ژنتیکی صفات را تبیین می‌کند و جهش‌های علی‌مربوطه را شناسایی می‌کند. مزیت مطالعات GWAS در پرورش گاوهای شیری، شناسایی SNPها و مناطق ژنومی مرتبط با صفات مورد نظر است. ارزیابی همبستگی بین هر نشانگر تعیین ژنوتیپ شده و صفت مورد مطالعه، یک راهبرد مهم برای مطالعات GWAS است که آثار همه نشانگرها را با در نظر گرفتن تعامل احتمالی آنها، عوامل محیطی و حتی آثار متقابل بین نشانگرها بررسی می‌کند (Bolormaa *et al.*, 2010; Jiang *et al.*, 2019).

اخیراً روش‌های یادگیری ماشین به مباحث ژنومی وارد شده است و اساس این روش‌ها با روش‌های رایج ارزیابی ژنومی متفاوت است. روش یادگیری ماشین، رویکرد قدرتمندی برای پیش‌بینی و انتخاب ژنومی است که به طور همزمان

کلیه نشانگرها را به عنوان آثار تصادفی در مدل در نظر می‌گیرد. این روش با در نظر گرفتن داده‌های آموزشی (اطلاعات ژنوتیپی، فنوتیپی و شجره در جمعیت مرجع) برای برآورد ارزش‌های اصلاحی ژنومی حیوانات کاندید انتخاب استفاده می‌شود. یکی از مزیت‌های کلیدی این روش، توانایی تجزیه و تحلیل داده‌های بزرگ است. یادگیری ماشین، شاخه‌ای از هوش مصنوعی است که هدف آن دستیابی به ماشین‌هایی است که قادر به استخراج دانش (یادگیری) از محیط هستند (Bureau *et al.*, 2005). الگوریتم‌های یادگیری ماشین می‌توانند مسایل پیچیده‌ای را که انجام دستی آنها غیرعملی و حتی غیرممکن است را حل کنند و توزیع‌ها، الگوها و همبستگی‌ها را بیاموزند تا اطلاعات و الگوهای درون داده‌ها را آشکار کنند. انواع روش‌های یادگیری ماشین (جنگل تصادفی (Random Forest)، یادگیری عمیق (Deep learning) و Boosting) جهت مدل‌سازی واریانس ژنتیکی و عوامل محیطی، مطالعه شبکه‌های ژنی، مطالعات پیوستگی در سطح ژنوم، مطالعه آثار اپیستازی و ارزیابی ژنومی مورد استفاده قرار گرفته است (Yang *et al.*, 2010). جنگل تصادفی، یکی از الگوریتم‌های یادگیری ماشین است که در عرصه‌های مختلف علوم به طور موفقیت‌آمیز مورد استفاده قرار گرفته است. بر اساس نوع نظارتی که در داده‌ها اتفاق می‌افتد، سیستم‌های یادگیری می‌توانند در چهار طبقه مختلف قرار بگیرند از جمله یادگیری با نظارت، یادگیری بدون نظارت، یادگیری نیمه نظارتی و یادگیری تقویتی. الگوریتم جنگل تصادفی، نمونه‌ای از الگوریتم‌های با نظارت هستند که تعدادی داده به آنها داده می‌شود. هر کدام از این الگوریتم‌ها بر اساس روش خاص خود به کمک داده‌ها، الگوی لازم را تشخیص می‌دهند. سپس، داده‌های جدید به آنها داده می‌شود و می‌توانند پیش‌بینی کنند. در این الگوریتم، مجموعه یا جنگلی از درختان مورد استفاده قرار می‌گیرد که هر درخت از ریشه، گره‌ها و برگ‌ها تشکیل شده است. این الگوریتم شامل صدها یا هزاران درخت است که از نمونه‌های تصادفی از اطلاعات ساخته می‌شود. اگرچه یک درخت ممکن است به تنهایی عملکرد خوبی نداشته باشد، تجمعی از درختان دارای عملکرد بهتری نسبت به یک درخت انفرادی است (Bureau *et al.*, 2003). الگوریتم در جمعیت مرجع آموزش می‌بیند و بر جمعیت تأیید (حیوانات کاندیدای انتخاب) اعمال می‌شود. برای ایجاد یک

اولین زایش شناسایی شد (Aloisio *et al.*, 2014). با بررسی پویای ژنومی روی جمعیت گاو نلور در مجموع، ژن‌های پیشنهادی مهم در ارتباط با باروری، سن در زمان اولین زایش، بیوسنتز استروژن، نرخ آبستنی تلیسه‌ها، رشد ابتدایی جنین و سن بلوغ گزارش شد (Mohammadi *et al.*, 2022). در تحقیقی که روی صفت رشد گاوهای گوشتی با استفاده از الگوریتم جنگل تصادفی انجام شد، نشان داده شد که این الگوریتم روشی کارآمد برای نمونه‌گیری از زیر مجموعه‌های بهینه از SNPها است که می‌تواند نشانگرهای مهم در رابطه با صفت مورد نظر را مشخص کند (Li *et al.*, 2018). در مطالعه‌ای که روی بروز ورم پستان در گاوهای شیری انجام گرفت، نقش عملکرد بالقوه الگوریتم جنگل تصادفی در استخراج فراسنجه‌های شیردوشی از جمله لاکتوز و حجم شیر که در بروز این بیماری نقش اساسی داشت را با دقت بالا پیش‌بینی کرد (Ebrahimi *et al.*, 2018). همچنین، نقش اساسی روش یادگیری ماشین در بررسی تولید متان (Zheng *et al.*, 2016) و تولید شیر (Wallén *et al.*, 2018) در گاوهای شیری نشان داده شده است.

این مطالعه با هدف شناسایی نشانگرها و ژن‌های مرتبط با صفات تولیدمثلی از قبیل فاصله گوساله‌زایی، روزهای باز، نرخ آبستنی دختران و سن در زمان اولین زایش در گاوهای هلستاین ایران انجام شد. این صفات قبلاً با روش GBLUP تک مرحله‌ای و با استفاده از اندازه نمونه کمتر مورد بررسی قرار گرفته است (Mohammadi *et al.*, 2020). با این حال در تحقیق حاضر، با استفاده از تعداد حیوانات تعیین ژنوتیپ شده بیشتر، برای شناسایی نشانگرها و ژن‌های مرتبط با صفات مذکور از الگوریتم جنگل تصادفی استفاده شد. نشانگرها و ژن‌های کاندید شناسایی شده در این مطالعه می‌توانند کاندیدای خوبی برای ورود به آزمون‌های ژنومی برای بهبود عملکرد تولیدمثلی گاوهای ماده و نر باشند و اطلاعات جدیدی را در مورد معماری ژنتیکی صفات تولیدمثلی برای بهبود ژنومی آن‌ها ارائه دهد.

### مواد و روش‌ها

**داده‌های فنوتیپی:** رکوردهای مورد بررسی در این مطالعه از مرکز اصلاح نژاد و بهبود تولیدات دامی کشور در سه فایل مجزا مربوط به اطلاعات تاریخ زایش، تاریخ تولد و تاریخ تلقیح تهیه شد. این رکوردها مربوط به صفات تولیدمثلی

جنگل تصادفی با تعداد T درخت از یک سری اطلاعات ژنوتیپی و فنوتیپی، ابتدا نمونه‌هایی از کل اطلاعات با جایگزینی گرفته می‌شود که شامل مقادیر فنوتیپی به‌همراه اطلاعات ژنوتیپی حیوانات برای SNPها در جمعیت مرجع است. با توجه به اینکه نمونه‌گیری به‌صورت جایگزینی است، ممکن است برخی اطلاعات ژنوتیپی و فنوتیپی هرگز نمونه‌گیری نشوند و یا اینکه در چند مرحله نمونه‌گیری، انتخاب شوند. در مرحله بعد، درخت روی نمونه گرفته شده شروع به رشد می‌کند و در هر گره، انتخاب ویژگی اتفاق می‌افتد که در آن از SNPهای از پیش تعیین شده یک SNP انتخاب می‌شود و گره بر اساس ژنوتیپ آن SNP منشعب می‌شود. در مراحل بعد و در گره‌های دیگر از SNPها، یک SNP دیگر انتخاب شده و انشعاب گره‌ها ادامه می‌یابد تا درخت کامل شود. یک ویژگی قابل توجه الگوریتم جنگل تصادفی، برآورد نشانگرها بر اساس میزان اهمیت آن‌هاست. در این الگوریتم، انتخاب نشانگرهای مهم بر اساس رتبه‌بندی میزان اهمیت آن‌هاست. یکی از روش‌های انتخاب نشانگرها از راه شاخص میزان اهمیت نسبی است که جهت برآورد میزان اهمیت استفاده می‌شود (Briec *et al.*, 2018). الگوریتم جنگل تصادفی، یک روش ناپارامتری هست و یکی از مباحث مهم در انتخاب ژنومی برای برآورد آثار نشانگرها است و در مقایسه با روش‌های رایج در آینده نزدیک به چالش کشیده خواهد شد. نتایج مطالعه‌ای که روی گاوهای جرسی صورت گرفت نشان داد که شبکه‌های عصبی ممکن است برای پیش‌بینی صفات پیچیده با استفاده از اطلاعات ژنومی با ابعاد بالا، زمانی که تعداد مجهول‌ها از حجم نمونه بیشتر است، مفید باشند (Gianola *et al.*, 2011). بررسی صفات تولیدمثلی در گاوهای ماده از راه روش یادگیری ماشین همراه با بررسی شبکه‌های ژنی، با دقت بالایی بیان نه ژن در اپیتلیوم اندومتر گاوها که باعث تمایز باروری و زنده‌مانی جنین می‌شود را نشان داد (Diniz *et al.*, 2022). نتایج تحقیقی که با هدف ارزیابی توانایی پیش‌بینی روش‌های مختلف یادگیری ماشین و مدل‌های پارامتریک برای پیش‌بینی ژنومی صفات تولیدمثلی در گاو نلور انجام شد نشان داد که رگرسیون شبکه‌های عصبی، روشی مناسب برای پیش‌بینی ژنومی صفات تولیدمثلی است (Alves *et al.*, 2021). در تحقیقی که روی گاوهای نلور با هدف انجام GWAS با استفاده از الگوریتم جنگل تصادفی انجام شد، در مجموع ۱۱۷ SNP برای صفت سن در زمان

زایش،  $\alpha_i$ ، اثر ژنتیکی افزایشی مربوط به ژامین حیوان و  $e_{ij}$  آثار تصادفی باقیمانده مربوط به  $y_{ij}$  هستند.

با استفاده از مدل تکرارپذیری زیر، ارزش‌های اصلاحی برای صفات (روزهای باز، فاصله گوساله‌زایی و نرخ آبستنی دختران) مربوط به گاوهای ماده برآورد شد:

$$y_{ijkl} = \mu + HYS_i + b(\text{Age}_{ijk} - \overline{\text{Age}}) + P_j + Pe_k + \alpha_k + e_{ijkl}$$

که در این مدل،  $y_{ijkl}$ ، رکورد صفت مورد نظر،  $\mu$  میانگین کل، آثار ثابت شامل:  $HYS_i$ ، اثر ژامین گله-سال-فصل زایش،  $P_j$ ، ژامین شکم زایش،  $Age_{ijk}$ ، سن در زمان زایش (عامل متغیر همبسته)،  $Pe_k$ ، اثر تصادفی محیط دائمی،  $\alpha_k$ ، اثر ژنتیکی افزایشی مربوط به  $k$  ژامین حیوان و  $e_{ijkl}$ ، آثار تصادفی باقیمانده مربوط به  $y_{ijkl}$  هستند. اطلاعات آماری (تعداد، کمینه (Min.)، بیشینه (Max.)، میانگین (Mean)، ضریب تغییرات (C.V.) و انحراف استاندارد (S.D.)) داده-های مورد تجزیه برای صفات تولیدمثلی در جدول ۲ ارائه شده است.

محاسبه ارزش اصلاحی غیر رگرسیونی گاوهای نر: ارزش اصلاحی غیر رگرسیونی (deregressed proofs) به‌عنوان ورودی مطالعات GWAS در نظر گرفته می‌شود. DRP با استفاده از بسته نرم افزاری "DRP" و کد دستوری مربوطه در محیط R محاسبه شد (Garrick et al., 2009).

داده‌های ژنومی: در تحقیق حاضر، اطلاعات ژنوتیپی نشانگرهای SNP مربوط به ۲۴۱۹ رأس گاوهای هلشتاین نر از مرکز اصلاح نژاد و بهبود تولیدات دامی کشور اخذ شد.

دختران گاوهای نر دارای اطلاعات ژنوتیپی بود. برای این تحقیق از رکوردهای سال ۱۳۸۰ تا ۱۳۹۸ (بر اساس سال زایش) استفاده شد. دلیل این انتخاب این بود که داده‌های سال‌های اخیر از صحت و دقت بیشتری برخوردار بودند. رکوردها مربوط به تلیسه‌ها و گاوهای شکم اول تا سوم، در دامنه سنی ۴۴-۲۱ ماه (شکم اول)، ۵۶-۳۲ ماه (شکم دوم) و ۷۰-۴۵ ماه (شکم سوم) بودند. صفات مورد بررسی در این مطالعه شامل سن در زمان اولین زایش (AFC)، روزهای باز (DO)، فاصله گوساله‌زایی (CI) و نرخ آبستنی دختران (DPR) بود. اطلاعات فایل شجره شامل اطلاعات سال‌های ۱۳۶۰ تا ۱۳۹۸ و کل حیوانات موجود در شجره نیز ۲۷۷۴۱۸۳ رأس بود. ساختار شجره مورد استفاده در این مطالعه در جدول ۱ نشان داده شده است.

مدل‌های آماری: برای آماده‌سازی و حذف داده‌های پرت از نرم افزارهای SAS (9.4)، R (4.0.2)، RENUMF90 و جهت آماده کردن فایل شجره از نرم افزار CFC (Sargolzaei et al., 2006) استفاده شد. مؤلفه‌های واریانس با روش بیزی و با استفاده از نرم‌افزار Gibbs2f90 از خانواده برنامه BLUPF90 برآورد شدند. ارزش اصلاحی برای صفت سن در زمان اولین زایش برای تلیسه‌ها با استفاده از مدل آماری زیر تخمین زده شد:

$$y_{ij} = \mu + HYS_i + \alpha_j + e_{ij}$$

که در این مدل،  $y_{ij}$ ، رکورد صفت مورد نظر،  $\mu$  میانگین کل، آثار ثابت شامل:  $HYS_i$ ، اثر ژامین گله-سال-فصل

جدول ۱- اطلاعات مربوط به فایل شجره

Table 1. Information about the pedigree

Parameter	Number of animals
Individuals in total	2774183
Inbred animals in total	2020760
Sires in total	28498
Dams in total	1332556
Individuals with progeny	1361054
Individuals without progeny	1413129
Founders	278081

جدول ۲- اطلاعات آماری داده‌های مورد تجزیه برای صفات مختلف تولیدمثلی

Table 2. Statistical information of analyzed data for different reproductive traits

Trait	Number of Records	Mean	Min.	Max.	S.D.	C.V.(%)
CI (day)	943480	412.45	300.00	700.00	83.371	20.21
AFC (day)	1355824	773.88	630.00	1200.0	97.340	12.57
DO (day)	1283493	132.43	40.000	350.00	71.750	54.17
DPR (%)	902031	25.163	-27.750	48.000	17.792	70.54

CI: Claving interval; AFC: Age at first calving; DO: Days open; DPR: Daughter pregnancy rate

جانهی انجام شد. سپس، مجدداً معیارهای کنترل کیفیت اعمال شد و در نهایت، تعداد نشانگرهای تمامی ۲۴۱۹ گاو نر، یکسان و برابر با ۴۱۰۹۸ عدد بود.

رتبه‌بندی SNPها بر اساس میزان اهمیت با الگوریتم *Random Forest* مهم‌ترین ویژگی الگوریتم جنگل تصادفی توانایی آن در رتبه بندی نشانگرها بر اساس میزان اهمیت آن‌ها در ارتباط با فنوتیپ است. از این ویژگی می‌توان جهت غربالگری نشانگرها بر اساس میزان اهمیت بهره برد. میزان اهمیت بدین صورت محاسبه می‌شود که برخی از اطلاعات هرگز نمونه‌گیری نمی‌شوند، این داده‌ها به‌عنوان یک اعتبارسنج داخلی برای هر درخت عمل می‌کنند که این اعتبارسنجی از راه برآورد خطای OOB (خارج از کیسه) انجام می‌شود. اختلاف بین دو خطای پیش‌بینی در دو حالت حضور و عدم حضور SNP مورد نظر، میزان اهمیت یا تأثیر SNP مربوطه را نشان می‌دهد. هدف از این کار مشخص کردن نشانگرهایی است که ارتباط معنی‌داری با صفت مورد نظر دارند و همچنین فیلتر کردن و ایجاد مجموعه‌ای کوچک از نشانگرها است که با صحت بالایی فنوتیپ را پیش‌بینی می‌نمایند. این عمل در اصطلاح، انتخاب متغیر، انتخاب ویژگی و انتخاب زیرمجموعه نامیده می‌شود (Gregorutti et al., 2017). در نتیجه، اهمیت هر نشانگر را می‌توان به‌صورت میزان تأثیری که روی صحت پیش‌بینی دارد، مشخص نمود. در رابطه با نحوه تجزیه با استفاده از الگوریتم جنگل تصادفی از نرم افزار *Ranfog* (Breiman, 2001) در محیط لینوکس استفاده شد. فراسنجه‌های بسیار مهم در روش جنگل تصادفی با توجه به بررسی منابع انجام شده و مقادیر استفاده شده در مقالات، مقادیر دست بالای مناسبی برای تعداد متغیر انتخاب شده در هر گره درخت (mtry)، تعداد درخت (ntree) و کمترین اندازه یا کمترین مشاهدات در گره پایانی یا برگ‌ها (nodesize) انتخاب شد.

### نتایج

با توجه به اطلاعات جدول ۲ مشاهده می‌شود که مقادیر ضریب تغییرات در خصوص این صفات بالا است که دلیل این نتیجه این است که بیشتر صفات، قابل رکوردبرداری نیستند و به‌طور غیرمستقیم و از راه انجام محاسبات و از روی یک سری از شاخص‌های دیگر محاسبه می‌شوند.

اطلاعات ژنومی گاوهای نر با یکدیگر متفاوت بود. در این مطالعه، تعداد ۱۱۳ گاو نر تعیین ژنوتیپ شده با تراکم پایین (<20K)، تعداد ۱۳۵۶ گاو نر تعیین ژنوتیپ شده با تراکم متوسط (<60K و >20K) و ۹۵۰ گاو نر تعیین ژنوتیپ شده با تراکم بالا (>60K) وجود داشت.

جانهی ژنوتیپ‌های گم شده: با توجه به تفاوت تراکم در اطلاعات ژنومی گاوهای نر، تعداد نشانگرهای شناسایی شده آن‌ها نیز با یکدیگر متفاوت بود. برای اینکه بتوان تعداد نشانگرهای شناسایی شده در گاوهای نر را یکسان نمود از نرم افزار *FImpute* (Sargolzaei, 2014) استفاده شد. پس از انجام عمل جانهی مجدداً کنترل کیفیت روی داده‌ها انجام شد و در نهایت، ۴۱۰۹۸ عدد نشانگر و ۲۴۱۹ گاو نر باقی ماند.

کنترل کیفیت داده‌های ژنومی: برای انجام کنترل کیفیت، یک سری از گاوهای نر شامل گروه اول دارای ۵۰۹۴۳ نشانگر بودند و گروه دوم شامل گاوهای نری بود که تعدادی از آن‌ها کمتر از ۵۰۰۰۰ و تعدادی نیز بیشتر از ۵۰۰۰۰ نشانگر دارند. برای کنترل کیفیت اولیه داده‌ها از نرم افزار *PLINK* (Purcell et al., 2007) و معیارهایی از قبیل تعداد SNPهای تعیین ژنوتیپ شده برای هر حیوان، تعداد حیوانات تعیین ژنوتیپ شده به‌ازای هر SNP، تعادل هاردی-واینبرگ (HWE) و فراوانی آلل نادر (MAF) مورد استفاده قرار گرفتند. در فیلتراسیون داده‌های ژنومی مربوط به گروه اول، نشانگرهایی که فراوانی آلل نادر آن‌ها کمتر از پنج درصد بود حذف شدند (Wiggans et al., 2010). نشانگرهایی که فراوانی تعیین ژنوتیپ شده آن‌ها کمتر از ۹۰ درصد بود شناسایی و حذف شدند. نشانگرهایی که نرخ تعیین ژنوتیپ شده آن‌ها در نمونه‌ها کمتر از ۹۵ درصد بود شناسایی و حذف شدند (Purfield et al., 2016). در نهایت، نشانگرهایی که سطح معنی‌داری آزمون HWE آن‌ها از  $10^{-6}$  کمتر بود (Panetto et al., 2017)، به‌عنوان معیاری از خطای تعیین ژنوتیپ از تجزیه‌ها کنار گذاشته شدند. بدین ترتیب با توجه به معیارهای مزبور، تعداد ۹۵۵۵ نشانگر حذف و از تعداد ۴۱۳۸۸ نشانگر استفاده شد. سپس، برای فیلتراسیون داده‌های ژنومی مربوط به گروه دوم، نشانگرهایی در نظر گرفته شد که با نشانگرهای گروه اول مشترک بودند و نشانگرهایی که مشترک نبودند از گروه دوم حذف شدند. برای اینکه بتوان تعداد نشانگرهای شناسایی شده در گروه دوم را با گروه اول یکسان نمود عمل

همچنین، عوامل محیطی و مدیریتی در پراکندگی این صفات، اثر بیشتری دارند.

یادگیری ماشین-جنگل تصادفی: نمودارهای منهن برای نشان دادن ارتباط SNP‌های مرتبط با صفات تولیدمثلی گاوهای هلشتاین ماده و تلیسه‌ها در شکل ۱ نشان داده شده است. با توجه به بررسی منابع انجام شده و در نظر گرفتن دامنه شاخص میزان اهمیت در تجزیه‌های انجام شده در این تحقیق که بیشترین مقدار آن برابر با  $2/80$  بود، مقدار شاخص اهمیت به میزان  $1/5$  به عنوان پایه مقایسات انتخاب شد تا تعداد نشانگرهای با اهمیت بالا در حد معقولی قرار گیرند. لازم به ذکر است با توجه به اینکه در تجزیه‌های الگوریتم جنگل تصادفی، معنی‌داری آثار نشانگرها مورد بررسی قرار نمی‌گیرد بلکه میزان اهمیت نشانگرها گزارش می‌شود، بنابراین فقط میزان اهمیت نشانگرها مورد بررسی قرار گرفته است و نشانگرهایی که میزان اهمیت آن‌ها بزرگتر از  $1/5$  بودند به عنوان نشانگرهای مهم مشخص شدند.

نشانگرهای مهم مرتبط با صفات مورد بررسی با استفاده از الگوریتم جنگل تصادفی در جدول ۳ ارائه شده‌اند. در مجموع، ۲۱ نشانگر با میزان اهمیت بالا مشاهده شد. تعداد بیشتری از این SNP‌های مهم در رابطه با صفت AFC مشاهده شدند. از میان SNP‌های با میزان اهمیت بالا برای صفات مختلف تولیدمثلی، نشانگر واقع در کروموزوم ۳ برای صفت AFC (ARS-BFGL-NGS-22647)، نشانگر واقع در کروموزوم ۱۱ برای صفت CI (ARS-BFGL-NGS-114194)، نشانگر واقع در کروموزوم ۵ برای صفت DO (BTA-74076-no-rs) و نشانگر واقع در کروموزوم ۲۶ برای صفت DPR (ARS-BFGL-NGS-32553) مشاهده شد. نتایج حاصل از روش هستی‌شناسی ژنی در مجموع ۶۲ ژن در داخل و یا در فاصله 250-Kb این نشانگرها را شناسایی کرد.

#### بحث

بر اساس یافته‌های قبلی، در تحقیقی که روی گاوهای هلشتاین ایران با روش ssGBLUP انجام شد (Mohammadi et al., 2020)، تعداد نشانگرهای شناسایی شده در رابطه با صفات باروری برابر با ۳۳ SNP معنی‌دار

گزارش شد. لازم به ذکر است جمعیتی که از راه این روش مورد بررسی قرار گرفته بود دارای حیوانات تعیین ژنوتیپ شده به نسبت کمتر از تحقیق حاضر بود. نتایج حاصل از این تحقیق دارای نشانگرهای مشابه (-Hapmap54493، rs29018712 و Hapmap47041-BTA-54700) و ژن‌های مشابه (IFFO2، ALDH4A1، TAS1R2، PAX7، KLHDC7A و CRIP1) و DPP6 بود، که با یافته‌های حاضر مطابقت داشت. تفاوت در تعداد نشانگرهای مشاهده شده را می‌توان به اندازه نمونه متفاوت، ماهیت غیرپارامتری روش جنگل تصادفی در مطالعه حاضر در مقایسه با روش پارامتری ssGBLUP و تعیین حد آستانه متفاوت در این دو روش اشاره کرد. بر اساس یافته‌های تحقیق حاضر، ژن‌های مرتبط با صفات تولیدمثلی به شرح زیر است.

صفت سن در زمان اولین زایش: نتایج حاصل از الگوریتم جنگل تصادفی در ارتباط با صفت سن در زمان اولین زایش نشان داد که نشانگر ARS-BFGL-NGS-22647 روی کروموزوم ۳ با میزان اهمیت برابر با  $2/80$ ، نقش اساسی و مهمی روی این صفت ایفا می‌کند. یکی از ژن‌هایی که در فاصله 250-Kb نشانگر ARS-BFGL-NGS-22647 وجود داشت ژن MPZL1 است که در گاوهای گوستی سیمینتال برای صفت AFC به وسیله محققین دیگر (An et al., 2020) گزارش شده است. در کروموزوم ۲، ژن‌های مختلفی با صفت AFC ارتباط داشتند که از بین آن‌ها، ژن PAX7 نقش اساسی در رشد و توسعه جنین داشت. وجود این ژن برای حفظ باروری در اسپرماتوژنز طبیعی در موش‌ها، ضروری گزارش شده است (Aloisio et al., 2014). ژن IFFO2 از لحاظ زیستی، نقش اساسی در ساختار مولکولی سلول‌ها دارد. این ژن در گاوهای ماده در ساز و کار مولکولی به‌ویژه تخمک‌ها، نقش مهمی در ارتباط با تشکیل بلاستوسیت، جنین و طول مدت آبستنی ایفا می‌کند (Kordowitzki et al., 2021). در بررسی صفت AFC گاوهای نلور با استفاده از الگوریتم یادگیری ماشین، ژن‌های MPZL1 و CD247 در موقعیت فیزیکی 1207469kb نشانگر ARS-BFGL-NGS-22647 روی کروموزوم ۳ شناسایی شد که با این صفت در ارتباط بودند. این ژن‌ها در گاوهای ماده و نر باعث افزایش باروری و بهبود عملکرد رشد می‌شوند (Alves et al., 2022).



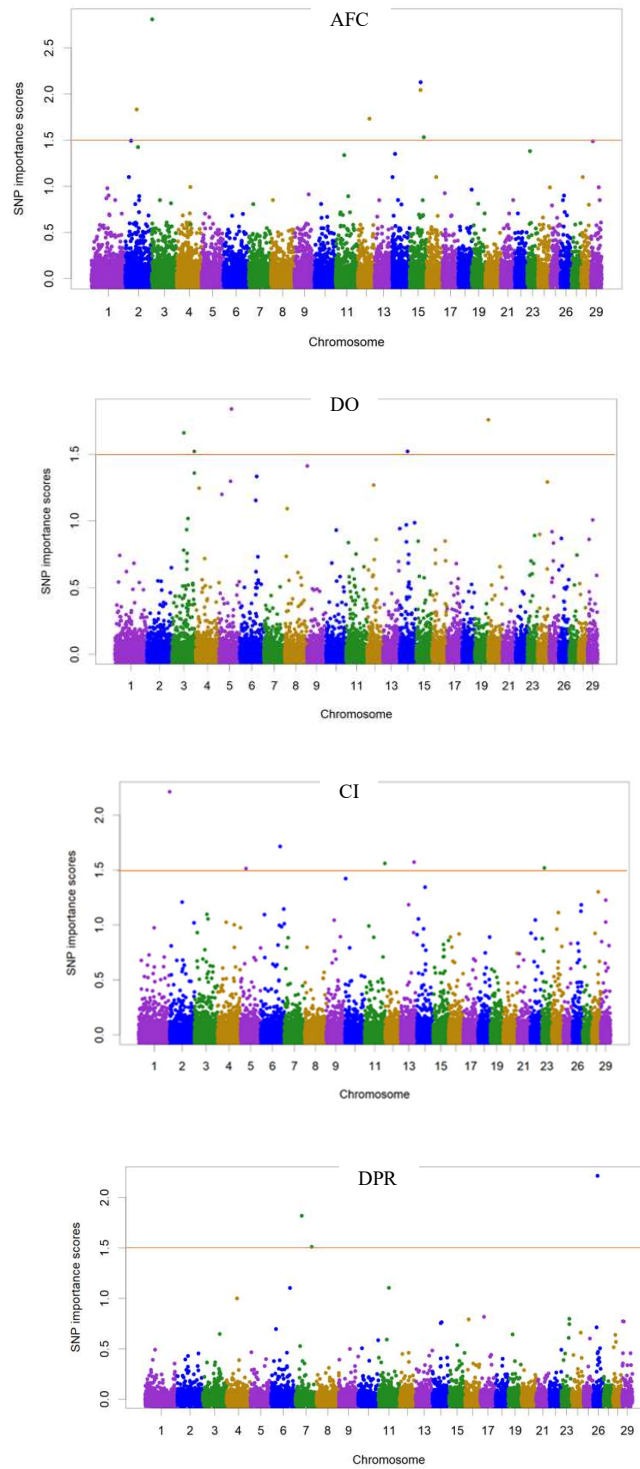


Fig. 1. Manhattan plots (GWAS) for age at first calving (AFC), days open (DO), calving interval (CI), and daughter pregnancy rate (DPR). The colored dots indicate the markers and the red line is the threshold of the markers with high importance score.

شکل ۱- نمودارهای منهن (GWAS) برای صفات، سن در زمان اولین زایش (AFC)، روزهای باز (DO)، فاصله گوساله‌زایی (CI) و نرخ آبستنی دختران (DPR). نقاط رنگی نشان‌دهنده نشانگرها و خط قرمز، حد آستانه نشانگرهای با میزان اهمیت بالا است.

جدول ۳- میزان اهمیت SNP‌های شناسایی شده برای صفات تولیدمثلی گاوهای ماده و تلیسه‌ها  
Table 3. The importance score of identified SNPs for reproductive traits of cows and heifers

Trait	BTA	SNP	Position	Importance score	Symbols of nearby genes
	3	ARS-BFGL-NGS-22647	1207469	2.80	MPZL1, RCSD1, CREG1, CD247, POU2F1
	2	ARS-BFGL-NGS-37797	133986945	2.12	IFFO2, ALDH4A1, TAS1R2, PAX7, KLHDC7A
	4	ARS-BFGL-NGS-102938	116467552	2.04	DPP6, Vault
	16	BTB-01240789	70565112	1.83	RPS6KC1, ANGEL2, VASH2, FLVCR1
AFC	12	Hapmap54493-rs29018712	10648444	1.73	OLFM4
	15	Hapmap47041-BTA-54700	49958907	1.53	OR52AB4, OR52J1, OR52J3D, OR51P1, OR51L1C, OR52S27
	1	Hapmap52737-rs29021682	157579130	2.21	KAT2B, SGO1
	6	BTA-03189-rs29010955	102495127	1.71	HSD17B11, NUDT9, SPARCL1, DSPP, DMP1
	13	ARS-BFGL-NGS-14434	71987095	1.57	SRSF6, L3MBTL1, SGK2, IFT52
CI	11	BFGL-NGS-114194	106245832	1.56	LCN12, PTGDS, CLIC3, PAXX, ABCA2
	23	Hapmap30534-BTA-137081	14580054	1.51	LRFN2
	27	Hapmap24404-BTA-139489	38672897	1.51	PSD3
DO	5	BTA-74076-no-rs	76019053	1.84	ELFN2, MFNG, CADD10, U6, USP15
	20	ARS-BFGL-NGS-46098	2951322	1.75	-
	14	Hapmap51382-BTA-107906	38044172	1.52	PI15
	15	BTB-00266941	70325855	1.52	LRRC4C, 5S_rRNA
	23	ARS-BFGL-NGS-84065	46963476	1.66	LYPD6, LYPD6B, KIF5C, SLC35B3
DPR	26	ARS-BFGL-NGS-32553	22977848	2.21	ARMH3, HPS6, LDB1, PPRC1
	7	BTA-95616-no-rs	34011987	1.81	5S_rRNA, FAM170A
	7	Hapmap33885-BES9 Contig513_1348	85176647	1.51	U2, 5S_rRNA, CRIP1

AFC: Age at first calving; DO: Days open; CI: Calving interval; DPR: Daughter pregnancy rate

CREG1 که روی کروموزوم ۳ معنی‌دار بود، در ارتباط با بیان *miR-31* که به‌طور غیرمستقیم باعث تقویت تکثیر سلولی و مهار تمایز سلولی می‌شود، موثر معرفی شد (Wang *et al.*, 2012). در مطالعات GWAS روی گاو نژاد نلور، نقش ژن‌های CREG1 و POU2F1 با تأثیر بر عوامل رونویسی فعالیت چندین ژن و فعال کننده مسیرهای متابولیکی دخیل در رشد، تکثیر و فرآیندهای تمایز سلولی گزارش شده است (Carreño *et al.*, 2019). با بررسی که روی تنظیم بیان پروتئین *Octl* که جزء پروتئین‌های اصلی شیر هست و همچنین باعث توسعه غدد پستانی در حیوان می‌شود، نقش ژن POU2F1 واقع بر کروموزوم ۶ از دیگر ژن‌های معنی‌دار در ارتباط با صفت AFC روی این پروتئین گزارش شده است (Abdel-Shafy *et al.*, 2020). مطالعه‌ای

در تحقیقی که روی بسیاری از مسیرهای زیستی سلولی از جمله توسعه سلولی، عملکرد و حفظ سلولی انجام شد، نقش ژن CD247 به‌عنوان ژن کلیدی مؤثر بر مسیرهای زیستی معرفی شد (Liao *et al.*, 2015). در مطالعه‌ای که روی سلول‌های سوماتیک شیر در گوسفندان مقاوم و حساس به ورم پستان انجام شد، نقش ژن CD247 به‌عنوان یکی از ژن‌های کاندید برای صفت AFC روی کروموزوم ۳ در پاسخ ایمنی و مسیرهای سیگنال‌دهی سلول‌های T گزارش شد (Bonfont *et al.*, 2010). در ارتباط با درصد چربی شیر در گاوهای هلشتاین چینی، نقش ژن RCSD1 بر این صفت، مؤثر نشان داده شد (Lu *et al.*, 2022). همچنین، در ارتباط با الگوهای بیان *MicroRNA* در غده‌های پستانی گاو که تحت تأثیر مراحل شیردهی قرار می‌گیرند، نقش ژن

ژن SPARCL1 به‌عنوان ژن پیشنهادی صفت CI واقع در کروموزوم ۶ در تنظیم مسیرهای سیگنال‌دهی سلولی به‌وسیله فاکتور تبدیل رشد بتا (*TGFB*) و پروتئین مورفوژنتیک استخوان ۷ (*BMP7*) که باعث تمایز سلولی می‌شود، گزارش شد (Wang *et al.*, 2019). در بررسی که روی گاوهای هلشتاین چینی انجام شد، نقش ژن‌های DSPP و DMP1 به‌عنوان ژن پیشنهادی مؤثر در تنظیم رشد و نمو استخوان گزارش شده است (Lu *et al.*, 2021). تحقیقی که روی جمعیت باکتری اشریکلای صورت گرفت، نقش ژن لیپوکالین ۱۲ (*Lcn12*)، که یک پروتئین اختصاصی اپیدیدیم است و همچنین در فرآیند فیزیولوژی باروری گونه‌های نر نقش مهمی ایفا می‌کند، گزارش شده است (Peng *et al.*, 2010). با انجام مطالعات GWAS روی گاوهای هلشتاین آمریکایی، تأثیر ژن LRFN2 در ناحیه BT23 کروموزومی در ارتباط با قابلیت پیش‌بینی توان انتقال برای صفت مرده‌زایی دختران گزارش شده است (Cole *et al.*, 2011). در تحقیقی روی صفت مقاومت کنه در گاو، نقش ژن LRFN2 گزارش شده است (Santos *et al.*, 2022).

صفت روزهای باز: یکی از مهمترین ژن‌هایی که از مسیر بررسی هستی‌شناسی ژنی برای صفت روزهای باز به‌دست آمد ژن USP15 است که در فاصله 250kb نشانگر -BTA-74076-no-rs با میزان اهمیت برابر با ۱/۸۴ روی کروموزوم ۵ شناسایی شد. با مطالعه روی جمعیت گاوها، نقش ژن USP15 به‌عنوان نشانگرهای زیستی نوین مؤثر بر سلول‌های بنیادی/پیش‌ساز پستان گاو گزارش شده است (Choudhary *et al.*, 2021). همچنین، با بررسی مسیر سیگنالینگ TGF- $\beta$ ، نقش کلیدی این ژن مشخص شده است (Eichhorn *et al.*, 2012).

صفت نرخ آبستنی دختران: در اصلاح نژاد گاو شیری، انتخاب صحیح گاو نر از مهمترین عوامل تأثیرگذار بر بهبود ژنتیکی گله‌های گاو برای صفات تولیدی و تولیدمثلی است و معیاری که بتواند با وزن‌دهی مناسب به صفات مختلف در رسیدن به هدف اصلاح نژاد و بیشترین سودآوری کمک کند، شاخص انتخاب است. در حال حاضر، صفت نرخ آبستنی دختران در شاخص انتخاب گنجانده شده است. در تحقیق حاضر، ژن‌های ARMH3، PPRC1، LDB1 و HPS6 در فاصله 250Kb نشانگر -ARS-BFGL-NGS-32553 روی کروموزوم ۲۶ با میزان اهمیت برابر با ۲/۲۱

که روی جمعیت موش‌ها در رابطه با ساختار تاژک و فرآیند بلوغ اسپرم صورت گرفت، نقش ژن ALDH4A1، که عضوی از خانواده آلدئید دهیدروژناز است، به‌عنوان آنزیم کلیدی در فرآیند بلوغ اسپرم گزارش شده است (Xiao *et al.*, 2022). با بررسی صفت رشد در بسیاری از پستانداران از جمله گاو، نقش ژن TAS1R2 گزارش شده است (Zhang *et al.*, 2012). با مطالعه‌ای که روی تلیسه‌های نلور انجام شد، ارتباط ژن RPS6KC1 با نرخ آبستنی و تعداد فولیکول‌های آنترال در این تلیسه‌ها مشخص شده است (Junior *et al.*, 2017). در بررسی مسیرهای زیستی مؤثر بر مراحل مختلف عملکرد تخمدانی گاو، نقش کلیدی ژن VASH2 مشخص شد (Berisha *et al.*, 2017). ژن اولفاکتومدین ۴ (OLFM4) یک گلیکوپروتئین ماتریکس خارج سلولی است که به خانواده اولفاکتومدین تعلق دارد که ارتباط آن با صفت AFC، معنی‌دار بود. این ژن در تکثیر سلولی، تمایز، آپوپتوز، چسبندگی و پاسخ التهابی به پاتوژن‌ها در گاو نقش اساسی ایفا می‌کند (Liu *et al.*, 2012). همچنین، ژن OLFM4 با سرعت شیردوشی در گاوهای قهوه‌ای سوئیس (Kramer *et al.*, 2014) و با رنگ پوشش بدنی در گاوهای هندی (Choudhary *et al.*, 2021) مرتبط است.

صفت فاصله گوساله‌زایی: ژن KAT2B که ارتباط آن با صفت فاصله گوساله‌زایی روی کروموزوم ۱ با میزان اهمیت برابر با ۲/۲۱ مشاهده شد، به‌عنوان فعال‌کننده عوامل رونویسی، در تنظیم و تصحیح استیل‌اسیون هیستون‌ها، نقش اساسی دارد و در بهبود کیفیت لاشه در گاوهای بومی نژاد گوشتی چینی، نقش مهمی ایفا می‌کند (Xiaoding *et al.*, 2022). همچنین، این ژن دارای نقش مهمی در تنظیم فرآیندهای زیستی است و با رشد و بازسازی سلولی، سوخت و ساز و عملکرد سیستم ایمنی مرتبط است (Lin *et al.*, 2022). در مطالعه‌ای که روی گاوهای نژاد گوشتی انجام شد، نقش ژن HSD17B11 روی کروموزوم ۶ بر پروتئین هیدروکسی استروئید ۱۷، مؤثر نشان داده شد. این آنزیم در مسیرهای زیستی مختلف تولیدمثل در نرها و ماده‌ها بسیار مهم است و غلظت داخل سلولی استروئیدهای غیرفعال را تنظیم می‌کند (Zhang *et al.*, 2022). همچنین، در تحقیقی که روی صفت رشد پشم و مو گوسفند مرینو انجام شد نقش ژن HSD17B11، مؤثر گزارش شده است (Sun *et al.*, 2022). در بررسی که روی ترمیم آسیب عضلانی در موش‌ها صورت گرفته بود، نقش

به‌طور کلی، مطالعه حاضر، ژن‌های پیشنهادی صفات تولیدمثلی مربوط به تلیسه‌ها و گاوهای ماده هلشتاین را شناسایی و اطلاعات جدیدی را در مورد معماری ژنتیکی این صفات برای بهبود ژنومی آن‌ها ارائه می‌دهد. ژن‌های شناسایی شده در این مطالعه برای صفت سن در زمان اولین زایش در گاوهای شیری مهم بوده و می‌توانند مورد استفاده قرار گیرند. همچنین، نشانگرهای تأیید شده در مطالعه حاضر، کاندیداهای خوبی برای ورود به آزمون‌های ژنومی برای صفات مختلف باروری در گاوهای شیری هستند و نشانگرها می‌توانند در طراحی تراشه‌ها برای ارزیابی صفات تولیدمثلی مورد نظر مورد استفاده قرار گیرند. با استفاده از نتایج حاصله و شناسایی جایگاه‌های ژنی مؤثر بر این صفات می‌توان به پیشرفت ژنتیکی صفات باروری در گاوهای هلشتاین دست یافت.

#### تشکر و قدردانی

بدین وسیله از مرکز اصلاح نژاد و بهبود تولیدات دامی کشور به‌واسطه فراهم نمودن داده‌ها و اطلاعات ژنومی تشکر و قدردانی می‌شود.

برای این صفت شناسایی شدند. در مطالعه‌ای که روی گاوهای شیری نیوزلند صورت گرفت، تأثیر ژن ARMH3 روی صفات تولیدمثلی و همچنین عملکرد زیستی این ژن مرتبط با تنظیم عملکرد دستگاه گلژی در سازمان‌دهی انتقال پروتئین و لیپیدها به‌طور معنی‌دار گزارش شده است (Jayawardana *et al.*, 2023). در تحقیقی، ژن FAM170A با باروری در نرها مرتبط گزارش شده است. این ژن دارای نقش اساسی در مورفولوژی و اسپرم‌سازی است (Devlin *et al.*, 2020). همچنین، در تحقیقی که روی زرده تخم مرغ صورت گرفت، ژن ARMH3 در تولید لیپید فسفاتیدیل‌اینوزیتول ایفای نقش می‌نماید (Zhang *et al.*, 2023).

#### نتیجه‌گیری کلی

در مجموع، با استفاده از الگوریتم جنگل تصادفی، ۲۱ نشانگر با میزان اهمیت بالا برای صفات مختلف تولیدمثلی مشاهده شد. سپس، با استفاده از روش هستی‌شناسی ژن، ژن‌های کاندید مهمی برای این صفات شناسایی شدند. این ژن‌ها در فرآیندها و مسیرهای زیستی مختلف برای صفات مورد بررسی و سایر خصوصیات در پستانداران نقش دارند.

#### فهرست منابع

- Abdel-Shafy, H., Awad, M. A., El-Regalaty, H., Ismael, A., El-Assal, S. E. D., & Abou-Bakr, S. (2020). A single-step genomic evaluation for milk production in Egyptian buffalo. *Livestock Science*, 234, 103977. doi: 10.1016/j.livsci.2020.103977
- Aloisio, G. M., Nakada, Y., Saatcioglu, H. D., Peña, C. G., Baker, M. D., Tarnawa, E. D., & Castrillon, D. H. (2014). PAX7 expression defines germline stem cells in the adult testis. *The Journal of Clinical Investigation*, 124(9), 3929-3944. doi: 10.1172/JCI75943
- Alves, A. A. C., da Costa, R. M., Fonseca, L. F. S., Carvalheiro, R., Ventura, R. V., Rosa, G. J. D. M., & Albuquerque, L. G. (2022). A random forest-based genome-wide scan reveals fertility-related candidate genes and potential inter-chromosomal epistatic regions associated with age at first calving in Nelore cattle. *Frontiers in Genetics*, 13, 834724. doi: 10.3389/fgene.2022.834724
- An, B., Xu, L., Xia, J., Wang, X., Miao, J., Chang, T., & Gao, H. (2020). Multiple association analysis of loci and candidate genes that regulate body size at three growth stages in Simmental beef cattle. *BMC Genetics*, 21(1), 1-11. doi: 10.1186/s12863-020-0837-6
- Berisha, B., Schams, D., Rodler, D., Sinowatz, F., & Pfaffl, M. W. (2017). Expression pattern of HIF 1alpha and vasohibins during follicle maturation and corpus luteum function in the bovine ovary. *Reproduction in Domestic Animals*, 52(1), 130-139. doi: 10.3389/fgene.2022.834724
- Bolormaa, S., Pryce, J. E., Hayes, B. J., & Goddard, M. E. (2010). Multivariate analysis of a genome-wide association study in dairy cattle. *Journal of Dairy Science*, 93(8), 3818-3833. doi: 10.3168/jds.2009-2980
- Bonnefont, C., Toufeer, M., Caubet, C., Foulon, E., Tasca, C., Aurel, M. R., & Rupp, R. (2011). Transcriptomic analysis of milk somatic cells in mastitis resistant and susceptible sheep upon challenge with *Staphylococcus epidermidis* and *Staphylococcus aureus*. *BMC Genomics*, 12(1), 1-16. doi: 10.1186/1471-2164-12-208
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. doi: 10.1023/A:1010933404324.
- Brieuc, M. S., Waters, C. D., Drinan, D. P., & Naish, K. A. (2018). A practical introduction to Random Forest for genetic association studies in ecology and evolution. *Molecular Ecology Resources*, 18(4), 755-766. doi: 10.1111/1755-0998.12773

- Bureau, A., Dupuis, J., Falls, K., Lunetta, K. L., Hayward, B., Keith, T. P., & Van Eerdewegh, P. (2005). Identifying SNPs predictive of phenotype using random forests. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, 28(2), 171-182. doi: 10.1002/gepi.20041
- Bureau, A., Dupuis, J., Hayward, B., Falls, K., & Van Eerdewegh, P. (2003). Mapping complex traits using Random Forests. *BioMed Central*, 4, 1-5. doi: 1471-2156/4/s1/S64
- Carreño, L. O. D., da Conceição Pessoa, M., Espigolan, R., Takada, L., Bresolin, T., Cavani, L., & Da Fonseca, R. (2019). Genome association study for visual scores in Nelore cattle measured at weaning. *BMC Genomics*, 20, 1-9. doi: 10.1186/s12864-019-5520-9
- Choudhary, R. K., & Capuco, A. V. (2021). Expression of NR5A2, NUP153, HNF4A, USP15 and FNDC3B is consistent with their use as novel biomarkers for bovine mammary stem/progenitor cells. *Journal of Molecular Histology*, 52(2), 289-300. doi: 10.1007/s10735-020-09948-8
- Cole, J. B., Wiggans, G. R., Ma, L., Sonstegard, T. S., Lawlor, T. J., Crooker, B. A., & Da, Y. (2011). Genome-wide association analysis of thirty-one production, health, reproduction and body conformation traits in contemporary US Holstein cows. *BMC Genomics*, 12(1), 1-17. doi: 1471-2164/12/408
- Devlin, D. J., Nozawa, K., Ikawa, M., & Matzuk, M. M. (2020). Knockout of family with sequence similarity 170 member A (Fam170a) causes male subfertility, while Fam170b is dispensable in mice. *Biology of Reproduction*, 103(2), 205-222. doi: 10.1093/biolre/iaaa082
- Diniz, W. J., Banerjee, P., Rodning, S. P., & Dyce, P. W. (2022). Machine learning-based Co-expression network analysis unravels potential fertility-related genes in beef cows. *Animals*, 12(19), 2715. doi: 10.3390/ani12192715
- Ebrahimie, E., Ebrahimi, F., Ebrahimi, M., Tomlinson, S., & Petrovski, K. R. (2018). Hierarchical pattern recognition in milking parameters predicts mastitis prevalence. *Computers and Electronics in Agriculture*, 147, 6-11. doi: 10.1016/j.compag.2018.02.003
- Eichhorn, P. J., Rodón, L., González-Juncà, A., Dirac, A., Gili, M., Martínez-Sáez, E., & Seoane, J. (2012). USP15 stabilizes TGF- $\beta$  receptor I and promotes oncogenesis through the activation of TGF- $\beta$  signaling in glioblastoma. *Nature Medicine*, 18(3), 429-435. doi: 10.1038/nm.2619
- Garrick, D. J., Taylor, J. F., & Fernando, R. L. (2009). Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*, 41, 1-8. doi: 10.1186/1297-9686-41-55
- Gianola, D., Okut, H., Weigel, K. A., & Rosa, G. J. (2011). Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genetics*, 12, 1-14. doi: 1471-2156/12/87
- Jayawardana, J. M. D. R., Lopez-Villalobos, N., McNaughton, L. R., & Hickson, R. E. (2023). Genomic regions associated with milk composition and fertility traits in spring-calved dairy cows in New Zealand. *Genes*, 14(4), 8-60. doi: 10.3390/genes14040860
- Jiang, J., Ma, L., Prakapenka, D., VanRaden, P. M., Cole, J. B., & Da, Y. (2019). A large-scale genome-wide association study in US Holstein cattle. *Frontiers in Genetics*, 412. doi: 10.3389/fgene.2019.00412
- Júnior, G. O., Perez, B. C., Cole, J. B., Santana, M. H. D. A., Silveira, J., Mazzoni, G., & Ferraz, J. B. S. (2017). Genomic study and medical subject headings enrichment analysis of early pregnancy rate and antral follicle numbers in Nelore heifers. *Journal of Animal Science*, 95(11), 4796-4812. doi: 10.2527/jas2017.1752
- Kordowitzki, P., Haghani, A., Zoller, J. A., Li, C. Z., Raj, K., Spangler, M. L., & Horvath, S. (2021). Epigenetic clock and methylation study of oocytes from a bovine model of reproductive aging. *Aging Cell*, 20(5), 33-49. doi: 10.1111/accel.13349.
- Kramer, M., Erbe, M., Seefried, F. R., Gredler, B., Bapst, B., Bieber, A., & Simianer, H. (2014). Accuracy of direct genomic values for functional traits in Brown Swiss cattle. *Journal of Dairy Science*, 97(3), 1774-1781. doi: 10.3168/jds.2013-7054
- Li, B., Zhang, N., Wang, Y. G., George, A. W., Reverter, A., & Li, Y. (2018). Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Frontiers in Genetics*, 9, 2-37. doi: 10.3389/fgene.2018.00237
- Liao, S. F., Boling, J. A., & Matthews, J. C. (2015). Gene expression profiling indicates an increased capacity for proline, serine, and ATP synthesis and mitochondrial mass by the liver of steers grazing high vs. low endophyte-infected tall fescue. *Journal of Animal Science*, 93(12), 5659-5671. doi: 10.2527/jas.2015-9193
- Lin, X., Li, B., Chen, Y., Chen, H., & Liu, M. (2022). KAT2B Gene Polymorphisms Are Associated with Body Measure Traits in Four Chinese Cattle Breeds. *Animals*, 12(15), 19-54. doi: 10.3390/ani12151954
- Liu, R. H., Yang, M. H., Xiang, H., Bao, L. M., Yang, H. A., Yue, L. W., & Huang, Y. (2012). Depletion of OLFM4 gene inhibits cell growth and increases sensitization to hydrogen peroxide and tumor necrosis factor-alpha induced-apoptosis in gastric cancer cells. *Journal of Biomedical Science*, 19(1), 1-11. doi: 10.1186/1423-0127-19-38
- Lu, X., Abdalla, I. M., Nazar, M., Fan, Y., Zhang, Z., Wu, X., & Yang, Z. (2021). Genome-wide association study on reproduction-related body-shape traits of chinese Holstein cows. *Animals*, 11(7), 19-27. doi: 10.3390/ani11071927

- Lu, X., Arbab, A. A. I., Abdalla, I. M., Liu, D., Zhang, Z., Xu, T., & Yang, Z. (2022). Genetic parameter estimation and genome-wide association study-based loci identification of milk-related traits in Chinese Holstein. *Frontiers in Genetics, 12*, 799-664. doi: 10.3389/fgene.2021.799664
- Meuwissen, T. H., Hayes, B. J., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics, 157*(4), 1819-1829. doi: 10.1093/genetics/157.4.1819
- Mohammadi, A., Alijani, S., Rafat, S. A., & 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. doi: 10.2478/aoas-2020-0031
- Mohammadi, H., Khaltabadi Farahani, A. H., & 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 [In Persian]
- Panetto, J. C. D. C., Machado, M. A., da Silva, M. V. G., Barbosa, R. S., dos Santos, G. G., de MH Leite, R., & Peixoto, M. G. C. (2017). Parentage assignment using SNP markers, inbreeding and population size for the Brazilian Red Sindhi cattle. *Livestock Science, 204*, 33-38. doi: 10.1016/j.livsci.2017.08.008
- Peng, Y., Liu, J., Liu, Q., Yao, Y., Guo, C., Zhang, Y., & Lin, D. (2010). Conformational and biochemical characterization of a rat epididymis-specific lipocalin 12 expressed in Escherichia coli. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1804*(11), 2102-2110. doi: 10.1016/j.bbapap.2010.07.020
- Pryce, J. E., Bolormaa, S., Chamberlain, A. J., Bowman, P. J., Savin, K., Goddard, M. E., & Hayes, B. J. (2010). A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *Journal of Dairy Science, 93*(7), 3331-3345. doi: 10.3168/jds.2009-2893
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics, 81*(3), 559-575. doi: 0.1086/519795
- Purfield, D. C., McClure, M., & Berry, D. P. (2016). Justification for setting the individual animal genotype call rate threshold at eighty-five percent. *Journal of Animal Science, 94*(11), 4558-4569. doi: 10.2527/jas.2016-0802
- Santos, E. M. S., Almeida, A. C., Santos, H. O., Cangussu, A. R., Costa, K. S., Alves, J. N., & Aguiar, R. W. S. (2019). Mechanism of Brassica oleracea performance in bovine infectious mastitis by bioinformatic analysis. *Microbial Pathogenesis, 129*, 19-29. doi: 10.1016/j.micpath.2019.01.029
- Sargolzaei, M., Chesnais, J. P., & Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics, 15*(1), 1-12. doi: 1471-2164/15/478
- Sargolzaei, M., Iwaisaki, H., & Colleau, J. J. (2006). CFC: A tool for monitoring genetic diversity. In Proceedings of the 8th world congress on genetics applied to livestock production. Pp. 13-18. doi: 10.1017/S1751731112001723
- Sun, X., Jiang, J., Wang, G., Zhou, P., Li, J., Chen, C., & Ren, H. (2023). Genome-wide association analysis of nine reproduction and morphological traits in three goat breeds from Southern China. *Animal Bioscience, 36*(2), 191. doi: 10.5713/ab.21.0577
- Wallén, S. E., Prestløkken, E., Meuwissen, T. H. E., McParland, S., & Berry, D. P. (2018). Milk mid-infrared spectral data as a tool to predict feed intake in lactating Norwegian Red dairy cows. *Journal of Dairy Science, 101*(7), 6232-6243. doi: 10.3168/jds.2017-13874
- Wang, M., Moisés, S., Khan, M. J., Wang, J., Bu, D., & Looor, J. J. (2012). MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation. *Journal of Dairy Science, 95*(11), 6529-6535. doi: 10.2012-5748.3168
- Wang, Y., Liu, S., Yan, Y., Li, S., & Tong, H. (2019). SPARCL1 promotes C2C12 cell differentiation via BMP7-mediated BMP/TGF- $\beta$  cell signaling pathway. *Cell Death & Disease, 10*(11), 852. doi: 10.1038/s41419-019-2049-4
- Wiggans, G. R., VanRaden, P. M., Bacheller, L. R., Tooker, M. E., Hutchison, J. L., Cooper, T., & Sonstegard, T. S. (2010). Selection and management of DNA markers for use in genomic evaluation. *Journal of Dairy Science, 93*(5), 2287-2292. doi: 10.3168/jds.2009-2773
- Xiao, Y., Wen, Z. Z., Wu, B., Zhu, H. X., Zhang, A. Z., Li, J. Y., & Gao, J. G. (2022). Deletion of Aldh4a1 leads to impaired sperm maturation in mice. *Molecular Biology, 56*(4), 543-550. doi: 10.1134/S002689332204015X
- Xuan, R., Wang, J., Zhao, X., Li, Q., Wang, Y., Du, S., & Chao, T. (2022). Transcriptome analysis of goat mammary gland tissue reveals the adaptive strategies and molecular mechanisms of lactation and involution. *International Journal of Molecular Sciences, 23*(22), 14424. doi: 10.3390/ijms232214424
- Yang, P., Hwa Yang, Y., B Zhou, B., & Y Zomaya, A. (2010). A review of ensemble methods in bioinformatics. *Current Bioinformatics, 5*(4), 296-308. doi: 10.2174/157489310794072508
- Zhang, H., Wang, Z., Wang, S., & Li, H. (2012). Progress of genome wide association study in domestic animals. *Journal of Animal Science and Biotechnology, 3*(1), 1-10. doi: 10.1186/2049-1891-3-26

- Zhang, R., Li, X., Ma, Y., Liu, Y., Zhang, Y., Cheng, X., & Ning, Z. (2023). Identification of candidate genomic regions for thermogelled egg yolk traits based on a genome-wide association study. *Poultry Science*, *102*(3), 102-402. doi: 10.1016/j.psj.2022.102402
- Zhang, T., Wang, T., Niu, Q., Xu, L., Chen, Y., Gao, X., & Xu, L. (2022). Transcriptional atlas analysis from multiple tissues reveals the expression specificity patterns in beef cattle. *BMC Biology*, *20*(1), 79. doi: 10.1186/s12915-022-01269-4