Animal Production Research

Vol. 11, No. 3, 2022 (15-26)
doi: 10.22124/AR.2022.21763.1688
eISSN: 2538-6107 pISSN: 2252-0872

University of Guilan

RESEARCH PAPER **OPEN ACCESS**

Multi-population joint genome-wide association study to detect genomic regions associated with litter size in sheep

M. Gholizadeh[1*], S. M. Esmaeili-Fard[2]

1. Associate Professor, Department of Animal Science and Fisheries, Sari Agricultural Sciences and Natural Resources University, Sari, Iran
2. Former Ph.D. Student, Department of Animal Science and Fisheries, Sari Agricultural Sciences and Natural Resources University, Sari, Iran

(Received: 12-02-2022 – Accepted: 08-05-2022)

Introduction: Reproduction is one of the most important economic traits in sheep with within and between-breeds variation. Reproductive traits normally show low to medium heritability and therefore response to conventional selection methods is not satisfactory for these traits. Considering the genetic information of the genetic variants underlying reproduction variability could efficiently increase the selection efficacy. Genome-wide association studies (GWAS) have been used to identify associations between genotypes and phenotypes as well as candidate genes for economically important traits. Statistical power in GWAS is mostly affected by sample size. The low sample size is hence a main obstacle in GWAS. Combining multiple data sets of different studies for joint (mega) GWAS provides an opportunity to increase the sample size required for GWAS. This study was performed to identify genomic regions affecting litter size in sheep using the mega-analysis of GWAS.

Materials and methods: Multi-population joint GWAS was performed using genotypic and phenotypic data of six sheep breeds retrieved from the database. Quality control was performed using the Plink software. The markers or individuals were removed from the further study based on the following criteria: (1) unknown chromosomal or physical location, call rate <0.95, missing genotype frequency >0.05, minor allele frequency (MAF) < 0.05, and a *P*-value for Hardy–Weinberg equilibrium test less than 10^{-6}. Before analysis, imputation of missing genotypes for combined data set was implemented by LD-kNNi method. Mega-analysis was performed using a mixed linear model in TASSEL software considering kinship and population structure (top five components of principal component analysis (PCA)) as confounding effects. The quantile–quantile (Q–Q) plot was visualized by plotting the distribution of obtained *vs.* expected log10 (*P*-value). The association results along the genome and the significant SNPs were visualized in the Manhattan plot. To account for multiple test problem and identify the genome-wide and chromosome-wide significance level, Bonferroni test was used based on the number of independent SNPs obtained from pairwise linkage disequilibrium analysis. After GWAS analysis, the 300 bp sequence upstream and downstream of the significant SNP was explored to identify the adjacent candidate genes using Ovis aries_v4.0 (UCSC).

Results and discussion: In the present study, we implemented a mega GWAS using six different sheep breed data to identify the genetic mechanisms responsible for litter size in sheep. After quality control, 305 animals and 351,615 SNP markers with a mean MAF of 0.33 were kept for further analysis. The results of the mega-analysis identified one marker on chromosome 21 at the genome-wide level and 10 markers at the chromosome-wide level on chromosomes 1, 2, 3, 14, 17, and 22. The quantile–quantile plot that features the total distribution of the observed *P*-values (−log10 *P*-values) of quality passed SNPs *vs.* the expected values, showed the effective control for confounding effects. Many of the significant SNPs identified in this study were located in or very adjacent to known genes

* Corresponding author: m.gholizadeh@sanru.ac.ir

CC BY

(*OPCML*, *GULP1*, *RBP4*, *MMP2,* and *LPCAT2*) that have been already reported for their contribution to fertility and pregnancy success. It has been reported that *OPCML* is more consistently expressed in cells lining the uterus, oviduct, and rete ovarii. *OPCML* has been reported as a tumor suppressor protein that is frequently inactivated in epithelial ovarian cancer. It has been reported that the *RBP4* gene is expressed during the period of fast elongation of the pig blastocyst which is a crucial period for the survival of the embryos. Also, it has been suggested that *RBP4* has the main contribution in uterine and conceptus physiology during the establishment of pregnancy and therefore can be considered as a candidate gene for litter size. *MMP2* has an essential function during ovulation and pregnancy through extracellular matrix (ECM) components degradation and therefore enabling cell migration and angiogenesis.
Conclusions: Comparison of the results of this study with previous reports showed that the mega-analysis of GWAS, compared to the meta-analysis already reported for GWAS results, had comparable power in identifying genomic regions influencing litter size in sheep but identified fewer genomic regions than individual GWAS for each breed. No previously reported major genes controlling litter size in sheep were identified using our mega GWAS. The results of our research are suggested for further investigations in identifying causal genetic variants or genomic regions underlying the litter size variation in sheep and can be used to understand the genetic mechanism controlling this trait.
Keywords: Prolificacy, Sheep, Genome-wide association study, Mega-analysis, Marker

How to cite this article:

Gholizadeh M. and Esmaeili-Fard S. M. 2022. Multi-population joint genome-wide association study to detect genomic regions associated with litter size in sheep. Animal Production Research, 11(3): 15-26. doi: 10.22124/AR.2022.21763.1688

تحقیقات تولیدات دامی

سال یازدهم/شماره سوم/پاییز ۱۴۰۱ (۲۶-۱۵)

Animal Production Research

مقاله پژوهشی

مطالعه ارتباط ژنومی چند جمعیتی مشترک برای شناسایی مکان‌های ژنومی موثر بر چندقلوزایی در گوسفند

محسن قلی زاده[1*]، سید مهدی اسماعیلی فرد[2]

۱- دانشیار، گروه علوم دامی، دانشگاه کشاورزی و منابع طبیعی ساری

۲- دانش‌آموخته دکتری، گروه علوم دامی، دانشگاه کشاورزی و منابع طبیعی ساری

(تاریخ دریافت: ۱۴۰۰/۱۱/۲۳ - تاریخ پذیرش: ۱۴۰۱/۰۲/۱۸)

چکیده

چندقلوزایی یکی از مهمترین صفات اقتصادی در گوسفند با تنوع داخل و بین نژادی است. این مطالعه به منظور شناسایی مکان‌های ژنومی موثر بر چندقلوزایی در گوسفند با رویکرد مگاآنالیز مطالعه ارتباط ژنومی و با استفاده از داده‌های ژنوتیپی و فنوتیپی شش نژاد گوسفند از پایگاه داده انجام شد. کنترل کیفیت با استفاده از نرم افزار Plink و ایمپیوتیشن با روش LD-kNNi انجام شد. مگاآنالیز با استفاده از مدل خطی مختلط در نرم افزار TASSEL با در نظر گرفتن خویشاوندی و ساختار جمعیت انجام شد. پس از پایان کنترل کیفیت، تعداد ۳۰۵ حیوان و ۳۵۱۶۱۵ نشانگر SNP با متوسط MAF برابر با ۰/۳۳ برای ادامه تجزیه مورد استفاده قرار گرفتند. نتایج مگاآنالیز، یک SNP را روی کروموزوم ۲۱ در سطح ژنوم و ۱۰ SNP در سطح کروموزومی روی کروموزوم های ۱، ۲، ۳، ۱۴، ۱۷ و ۲۲ شناسایی کرد که به‌طور معنی‌داری با چندقلوزایی در گوسفند در ارتباط بودند. ژن‌های *OPCML*، *GULP1*، *RBP4*، *MMP2* و *LPCAT2* شناسایی شده در این تحقیق، نقش موثری در باروری و موفقیت آبستنی دارند. نتایج این تحقیق می‌تواند در درک ساز و کار ژنتیکی کنترل‌کننده چندقلوزایی در گوسفند مورد استفاده قرار گیرد.

واژه‌های کلیدی: چندقلوزایی، گوسفند، مطالعه ارتباط ژنومی، مگاآنالیز، نشانگر

* نویسندهٔ مسئول: m.gholizadeh@sanru.ac.ir

doi: 10.22124/AR.2022.21763.1688

مقدمه

چندقلوزایی یکی از مهمترین صفات اقتصادی در گوسفند است. مطالعه ارتباط گسترده ژنومی (GWAS[1]) و شناسایی معنی‌دارترین نشانگرها و استفاده از این نشانگرها در برآورد ارزش‌های اصلاحی ژنومی یکی از روش‌های استفاده از نشانگرهای مولکولی در اصلاح نژاد دام است (Hayes *et al.,* 2010). تحقیقات متعددی برای شناسایی ساختار ژنوم و مناطق ژنومی موثر بر صفات تولیدمثلی در گوسفند با استفاده از نشانگرهای متراکم انجام شده است (Pasandideh *et al.,* 2017; Abdoli *et al.*, 2019; Esmaeili fard *et al.,* 2019; Taghizade *et al.,* 2020). اندازه جمعیت مرجع که شامل افراد دارای ژنوتیپ و فنوتیپ است یکی از فراسنجه‌های اصلی موثر بر صحت پیش‌بینی ژنومی است (Meuwissen *et al.,* 2001; Daetwyler *et al.,* 2008; VanRaden *et al.,* 2009). در مورد نژادهای با اندازه جمعیت کوچک، دستیابی به یک جمعیت مرجع به اندازه کافی بزرگ ممکن است بسیار دشوار و یا غیرممکن باشد که در این صورت، اضافه کردن افراد از نژادهای دیگر می‌تواند به حل مشکل کمک نماید (Marjanovic and Calus, 2020). در GWAS، اندازه نمونه مهمترین عاملی است که قدرت آماری را تحت تأثیر قرار می دهد و در کنترل محقق است. به‌طور ویژه، اندازه نمونه یک مانع بزرگ برای GWAS در صفاتی است که اندازه‌گیری آن ها دشوار و یا پرهزینه است. یکی از راهکارهای حل محدودیت اندازه نمونه در فهم ساز و کار ژنتیکی صفات، ترکیب کردن نمونه‌های کوچک‌تر برای GWAS مشترک (مگاآنالیز GWAS)، یا ترکیب نتایج GWASهای انفرادی برای متاآنالیز است (Gebreyesus *et al.,* 2019). ترکیب کردن مجموعه داده‌ها برای انجام GWAS مشترک، یک روش کارآمد برای افزایش قدرت GWAS برای مطالعه بیماری‌های انسان و دام (Veerkamp *et al.,* 2012, Bernal Rubio *et al.,* 2015; Bouwman *et al.,* 2018) بوده است. با استفاده از داده‌های شبیه‌سازی شده و واقعی نشان داده شده است که متاآنالیز نتایج GWASهای انفرادی به لحاظ آماری به اندازه مگاآنالیز ترکیب کردن داده‌های انفرادی موثر است (Lin and Zheng, 2010). همچنین مقایسه تجربی نشان داده است که متا و مگاآنالیز در شناسایی آثار متقابل ژن و محیط، عملکرد یکسانی دارند (Sung *et al.,* 2014). با در نظر گرفتن محدودیت‌های کمتر در ادغام نمونه‌های حیوانی، انتخاب روش‌های اجرای مگا-GWAS برای صفات دامی ممکن است فقط وابسته به بازدهی آماری باشد و بنابراین مقایسه مگا و متاآنالیز ممکن است خیلی پراهمیت شود (Gebreyesus *et al.,* 2019).

در یک مطالعه GWAS با استفاده از تراشه‌های SNP گوسفندی 600 k که به‌صورت همزمان ولی مستقل (انفرادی) روی شش نژاد مختلف گوسفند با هدف درک ساز و کار ژنتیکی موثر روی چندقلوزایی در گوسفند انجام شد، ژن‌های متفاوتی در هر نژاد شناسایی شدند. در نژاد وادی، ژن‌های MMP2، FBN1 وBMPR1B، در نژاد هوو، ژن‌های CTNNB1، SMAD1 و GRIA2، در نژاد ایسلندیک، ژن NCOA1، در نژاد فینشیپ، ژن‌های INHBB، NF1، FLT1، PTGS2 و LCB3، در نژاد رومانف، ژن ESR2 و در نژاد تکسل، ژن‌های ESR1، GHR، ETS1، MMP15، FFLI1 و SPP1 گزارش شدند (Xu *et al.,* 2018). شناسایی ژن‌های متفاوت در هر نژاد نشان می‌دهد که ساز و کار ژنتیکی موثر روی چندقلوزایی در گوسفند می‌تواند مختص هر نژاد باشد و لزوم این مطالعه‌ها را برای هر نژاد پررنگ می‌سازد. در متاآنالیز GWAS روی این شش نژاد که با استفاده از ترکیب نتایج GWAS انفرادی هر نژاد انجام شد، هر چند تعداد نشانگرهای معنی‌دار متفاوتی شناسایی شدند، تنها ژن BMPR1B به عنوان ژن موثر بر چندقلوزایی گزارش شد (Gholizadeh and Esmaeili-Fard, 2022). اثر ترکیب کردن مجموعه داده‌های نژادهای مختلف برای انجام GWAS مشترک (مگاآنالیز GWAS) در شناسایی مناطق ژنومی موثر روی چندقلوزایی گزارش نشده است. بنابراین هدف از مطالعه حاضر انجام مگا-GWAS و شناسایی مناطق ژنومی موثر بر چندقلوزایی گوسفند با استفاده از ترکیب نمونه‌های شش نژاد (وادی، هوو، تکسل، فینشیپ، ایسلندیک و رومانف) بود. همچنین عملکرد مگا-GWAS با گزارش‌های قبلی GWAS انفرادی و متا-GWAS روی این نژادها مورد مقایسه قرار گرفت.

1. Genome-wide association study

مواد و روش‌ها

داده‌های ژنوتیپی و فنوتیپی: در این مطالعه از داده‌های ژنوتیپی و فنوتیپی شش نژاد گوسفند با نام‌های وادی (۱۰۰ نمونه)، هوو (۱۱۷ نمونه)، آیسلندیک (۵۴ نمونه)، فینشیپ (۵۴ نمونه)، رومانف (۷۸ نمونه) و تکسل (۵۹ نمونه) استفاده شد. نتایج مطالعه GWAS انفرادی برای این جمعیت‌ها پیش-تر گزارش شده است (Xu *et al.,* 2018). داده‌ها از پایگاه داده بازیابی شدند (AnimalGenome.ORG Data Repository).

کنترل کیفیت داده‌های ژنوتیپی: ابتدا کنترل کیفیت با استفاده از نرم‌افزار Plink (Purcell *et al.,* 2007) در داخل هر جمعیت انجام شد و حیوانات با نرخ تعیین ژنوتیپ کمتر از ۹۵ درصد، SNPها با نرخ تعیین ژنوتیپ کمتر از ۹۹ درصد، فراوانی آلل کمیاب کمتر از پنج درصد و انحراف از تعادل هاردی– واینبرگ با شدت $P< 10^{-6}$ از تجزیه کنار گذاشته شدند. پس از کنترل کیفیت، داده‌های فیلتر شده با استفاده از نرم افزار Plink ادغام شدند و دور بعدی کنترل کیفیت روی داده‌های ادغام شده با فراسنجه‌های مشابه دور قبل انجام شد. در مرحله بعد، قبل از انجام GWAS، ایمپیوتیشن با روش LD-kNNi (Money *et al.,* 2015) انجام شد. ادغام داده‌های جمیعت‌های انفرادی در قالب یک مجموعه داده بزرگ، امکان مگاایمپیوتیشن و مگاآنالیز (Fritsche *et al.*, 2016) و شناسایی واریانت‌های بیشتر نسبت به متاایمپیوتیشن را فراهم می‌نماید (Gorski *et al.*, 2019). ایمپیوتیشن با استنتاج SNP های گمشده یا تعیین ژنوتیپ نشده، ابزاری مفید برای GWAS است که می‌تواند مجموعه SNPهای مشابه برای افراد مورد مطالعه را فراهم نماید و بنابراین منجر به افزایش قدرت GWAS می‌شود (Gorski *et al.*, 2019).

بررسی ساختار جمعیت و مدل کردن داده‌ها: تجزیه GWAS با استفاده از مدل خطی مختلط در نرم افزار TASSEL 5 (Bradbury *et al.,* 2007) انجام شد. دو عامل مختل‌کننده (confounding factors) که می توانند نتایج GWAS را دچار اریبی نمایند شامل خویشاوندی (kinship) و ساختار (لایه-بندی) جمعیت هستند. مدل‌های خطی عمومی یا GLM امکان در نظر گرفتن لایه‌بندی جمعیتی را از راه وارد کردن ابعاد مولفه‌های اصلی در مدل فراهم می نمایند، ولی نمی-توانند خویشاوندی را در نظر گیرند. مدل مختلط این امکان را فراهم می‌سازد که علاوه بر لایه‌بندی جمعیتی، خویشاوندی نیز در قالب ماتریس روابط خویشاوندی در مدل وارد شود. جهت بررسی ساختار جمعیت ادغام شده از تجزیه به مولفه‌های اصلی (PCA) استفاده شد. این تجزیه با استفاده از بسته GenABEL (Aulchenko *et al.,* 2007) در محیط نرم‌افزار R (R Development Core Team, 2011) انجام شد و تعداد ۱۰ مولفه اول ماتریس ژنوتیپی محاسبه شد. جهت بررسی بیشتر از Screen plot استفاده شد. این پلات نسبت واریانس توجیه شده به وسیله هر مولفه اصلی را به واریانس کل نشان می‌دهد و جهت یافتن بیشترین تعداد سطوح مولفه-های اصلی که باید وارد مدل شوند مورد استفاده قرار می‌گیرد. برای رسم این پلات از نرم افزار R استفاده شد. مدل خطی مختلط (MLM) به‌صورت زیر برازش شد:

$$\mathbf{y} = \boldsymbol{X_{SNP}\beta_{SNP}} + PC_{1:5} + \mathbf{Zu} + \boldsymbol{e}$$

در این مدل، **y**، بردار مشاهدات (میانگین تعداد بره)؛ **X**، ماتریس طرح ارتباط دهنده مشاهدات به ژنوتیپ‌ها؛ $\boldsymbol{\beta}$، بردار اثر نشانگرها؛ $PC_{1:5}$، پنج سطح اول تجزیه به مولفه‌های اصلی؛ **u**، بردار آثار ژنتیکی چند ژنی؛ **Z**، ماتریس طرح بردار **u** و **e**، بردار آثار باقی‌مانده است. جهت تصحیح برای آزمون‌های چندگانه و رسیدن به سطح معنی‌داری ژنومی و کروموزومی از آزمون بن‌فرونی بر اساس تعداد SNPهای مستقل استفاده شد. برای استخراج نشانگرهای مستقل از نرم افزار Plink 1.07 و دستور indep-pairwise 50 5 0.05 استفاده شد. برای رسیدن به سطح معنی داری پنج درصد ژنومی/کروموزومی (α) و با تعداد SNPهای مستقل (N)، آستانه معنی‌داری ژنومی، 1.7×10^{-6} برآورد شد (α/N). جهت بررسی کیفیت برازش مدل و نتایج، از پلات QQ استفاده شد. جهت شناسایی ژن‌های مرتبط با نشانگرهای معنی‌دار در بازه ۳۰۰ کیلوباز (Esmaeili-Fard *et al.*, 2021; Gudjonsson *et al.*, 2022) از پایگاه داده UCSC (https://genome.ucsc.edu/) و از اسمبلی Oar_v4.0 (2015) استفاده شد.

نتایج

پس از پایان کنترل کیفیت، تعداد ۳۰۵ حیوان و ۳۵۱۶۱۵ نشانگر SNP با متوسط MAF برابر با ۰/۳۳ در تجزیه GWAS مورد استفاده قرار گرفتند. نتیجه تجزیه به مولفه‌های

اصلی (PCA) جهت بررسی ساختار جمعیت در شکل ۱ نشان داده شده است. همچنین جهت مشاهده واریانس توجیه شده به وسیله هر سطح PC از screen plot استفاده شد که در شکل ۲ نشان داده شده است. همان‌طور که ملاحظه می شود استفاده از پلات، سطح یک مولفه‌های اصلی در برابر سطح دو جمعیت‌ها را از هم تفکیک کرد. همچنین پنج سطح اول بالاترین میزان واریانس ژنتیکی را توجیه کردند و بنابراین برای کنترل ساختار جمعیت در مدل آماری استفاده شدند.

نتایج مگاآنالیز GWAS، یک SNP را روی کروموزوم ۲۱ شناسایی کرد که به‌طور معنی‌داری در سطح ژنوم با میانگین چندقلوزایی در ارتباط بود (شکل ۳، جدول ۱). نشانگر OAR21_33276424 روی کروموزوم ۲۱ درون ژن OPCML قرار دارد. همچنین ۱۰ عدد SNP در سطح کروموزومی (سه SNP روی کروموزوم ۲، دو SNP روی کروموزوم ۱۷ و یک SNP روی کروموزوم ۱، ۳، ۱۴ و۲۲) شناسایی شدند (جدول ۱). آثار افزایشی و غلبه به همراه آزمون معنی‌داری این آثار در جدول ۲ ارائه شده است. همان‌طور که ملاحظه می‌شود این آثار برای بسیاری از SNPهای شناسایی شده معنی‌دار هستند (جدول ۲). پلات QQ نشان می‌دهد مدل خطی مختلط با در نظر گرفتن ساختار جمعیت و روابط خویشاوندی توانسته است خطای نوع اول را کاهش دهد (شکل ۴).

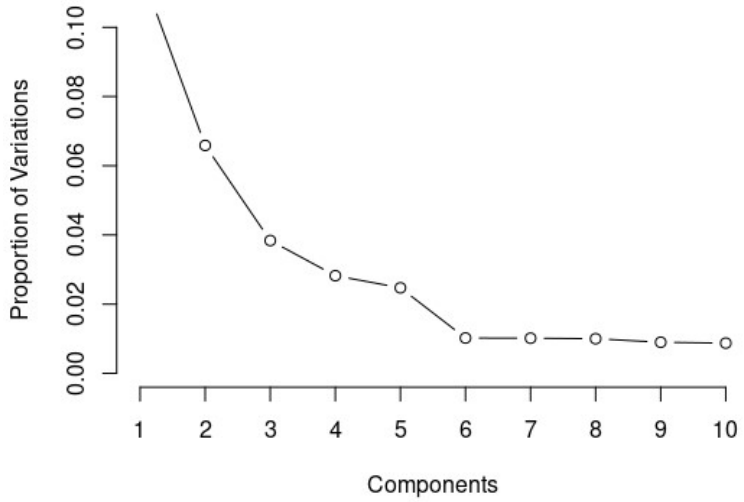

Fig.1. Screen plot of principle component analysis to identify the components explaining most proportion of genetic structure

شکل ۱- اسکرین پلات تجزیه مولفه اصلی برای شناسایی مولفه‌هایی که بیشترین سهم واریانس ژنتیکی را نشان می‌دهند

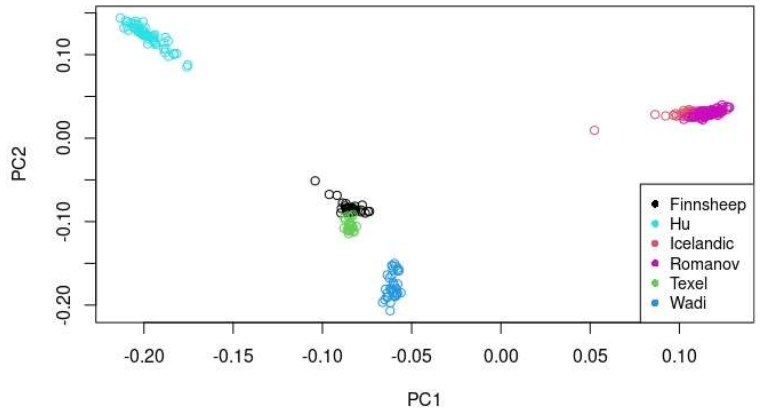

Fig. 2. Principle component analysis to plot population structure of the studied breeds

شکل ۲- تجزیه مولفه اصلی برای ترسیم ساختاری جمعیتی نژادهای مورد مطالعه

جدول ۱- نتایج مگا-GWAS با مدل خطی مختلط برای شناسایی مکان‌های ژنومی موثر بر چندقلوزایی در گوسفند

Table 1. Mega GWAS results obtained from linear mixed model to identify genomic regions associated with prolificacy in sheep

Significance level	SNP	Chromosome	Position (bp)	*P*-value	Gene(s)
Genome	OAR3_OAR21_33276424	21	33212072	1.10E-06	OPCML
Chromosome	OAR3_OAR2_119795639	2	119789298	7.57E-06	GULP1
	OAR3_OAR2_119802214	2	119795873	9.32E-06	GULP1
	OAR3_OAR1_166096660	1	165915826	9.72E-06	
	DU266660_200.1	2	119804089	1.51E-05	GULP1
	OAR3_OAR3_30030279	3	29945859	1.81E-05	
	OAR3_OARX_89164245	27	89052684	2.48E-05	
	OAR17_23141755.1	17	20716892	2.91E-05	
	OAR3_OAR17_18770871	17	18769443	2.93E-05	
	OAR3_OAR14_23181640	14	23115190	4.90E-05	MMP2, CES1, BREH1, LOC508916, SLC6A2, LPCAT2
	OAR22_18241256.1	22	14742914	5.30E-05	MYOF, SLC35G1, FRA10AC1, PDE6C, RBP4, FFAR4, PLCE1

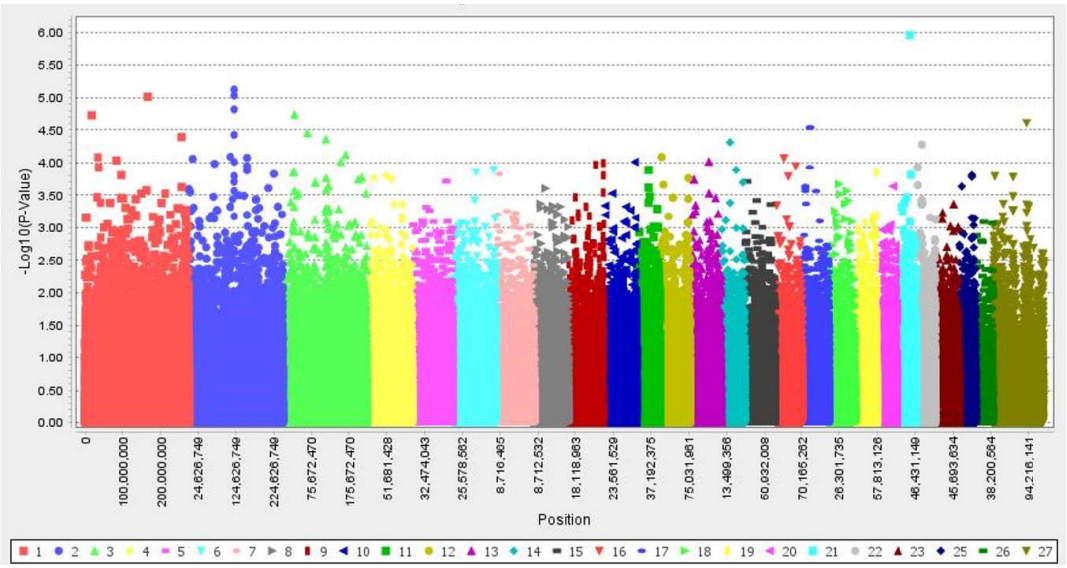

Fig. 3. Manhatan plot of mega GWAS results obtained from linear mixed model to identify genomic regions associated with prolificacy in sheep with genome-wide significance level of 1.7×10^{-6}

شکل ۳- پلات منهتن نتایج مگاآنالیز GWAS با استفاده از مدل خطی مختلط برای شناسایی نواحی ژنومی مرتبط با چندقلوزایی در گوسفند در سطح معنی‌داری ژنومی 1.7×10^{-6}

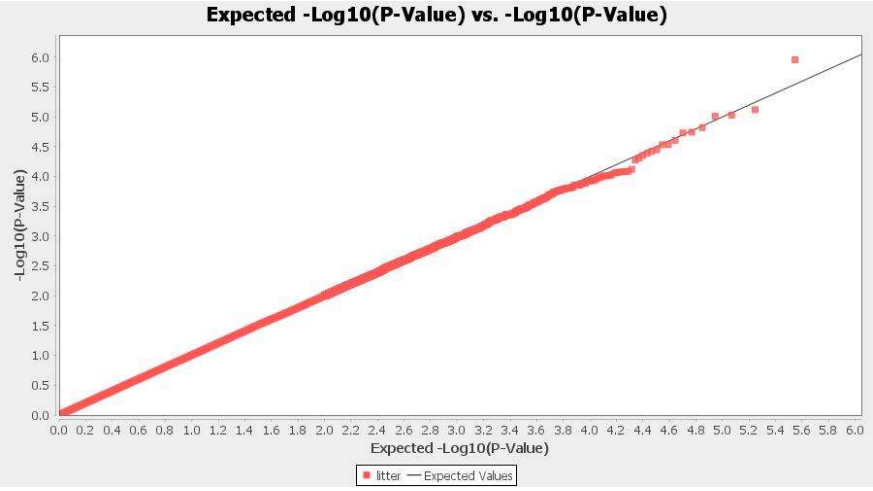

Fig. 4. Q-Q plot of mega GWAS results for assessing quality of results

شکل ۴- پلات Q-Q نتایج مگا-GWAS برای کنترل کیفیت نتایج

جدول ۲- نتایج بررسی آثار افزایشی و غالبیت نشانگرهای معنی‌دار مگا-GWAS با مدل خطی مختلط

Table 2. The results related to the study of additive and dominance effects of significant markers obtained from mega GWAS with linear mixed model

Marker	Chr	Position (bp)	Additive effect	Additive *P*-value	Dominance effect	Dominance *P*-value
OAR3_OAR21_33276424	21	33212072	0.296	0.000	0.109	0.131
OAR3_OAR2_119795639	2	119789298	-0.313	0.000	0.135	0.106
OAR3_OAR2_119802214	2	119795873	-0.298	0.000	0.121	0.131
OAR3_OAR1_166096660	1	165915826	0.401	0.000	0.453	0.000
DU266660_200.1	2	119804089	-0.281	0.000	0.087	0.265
OAR3_OAR3_30030279	3	29945859	-0.050	0.625	0.429	0.001
OAR3_OAR1_16900325	1	16887522	0.046	0.359	0.320	0.000
OAR3_OARX_89164245	27	89052684	-0.143	0.008	0.330	0.000
OAR17_23141755.1	17	20716892	-0.095	0.078	-0.277	0.000
OAR3_OAR17_18770871	17	18769443	-0.250	0.000	-0.017	0.826
OAR3_OAR3_66208809	3	66074756	-0.013	0.886	-0.359	0.001
OAR3_OAR2_119744800	2	119738462	0.364	0.000	0.196	0.043
OAR3_OAR1_256963466	1	256796696	-0.963	0.000	-0.875	0.000
OAR3_OAR3_114596652	3	114432003	0.246	0.000	0.075	0.294
OAR3_OAR14_23181640	14	23115190	0.281	0.000	0.006	0.950
OAR22_18241256.1	22	14742914	0.284	0.000	0.225	0.010

بحث

مگا-GWAS انجام شده در تحقیق حاضر منجر به شناسایی یک نشانگر معنی‌دار در سطح ژنوم و ژن OPCML و ۱۰ نشانگر و ژن‌های مرتبط با این نشانگرها در سطح کروموزومی شد که به‌طور معنی‌داری با چندقلوزایی در ارتباط بودند. بررسی‌های بیشتر در پایگاه‌های داده نشان داد که این ژن‌ها غالباً نقش مهمی در عمکرد تولیدمثلی دارند.

ژن *OPCML* شناسایی شده در این تحقیق در GWAS انفرادی در نژاد رومانف گزارش شده است که در سطح کروموزومی با چندقلوزایی در این نژاد در ارتباط بود (Xu *et al.*, 2018). ژن *OPCML* به عنوان ژن موثر بر تحرک‌پذیری اسپرم در گاو گوشتی آمیخته شناسایی شده است (Sweet *et al.*, 2020). این ژن به عنوان ژن سرکوب‌کننده تومور تخمدانی شناخته می‌شود که در سرطان اپی‌تلیال تخمدان با

جهش آللی و متیلاسون پروموتور غیرفعال می‌شود (Sellar *et al.*, 2003, Tsou *et al.*, 2007). نشان داده شده است که بیشترین واکنش‌پذیری ایمنی این ژن در لوله رحم، پایه رحم و رحم رخ می‌دهد (Fleming *et al.,* 2009).

همچنین نتایج تحقیق حاضر، سه نشانگر را روی کروموزوم ۲ شناسایی کرد که به‌طور معنی‌داری در سطح کروموزومی با چندقلوزایی در ارتباط بودند. تجزیه بیشتر در پایگاه داده نشان داد که این نشانگرها با ژن *GULP1* در ارتباط هستند. نشان داده شده است که این ژن با طول دوره لوتئال در گاو در ارتباط است (Tenghe *et al.,* 2016). همچنین این ژن تنظیم‌کننده کلیدی سیگنال‌دهی فاکتور رشد تغییردهنده بتا (TGF-B) با میانجی‌گری *LRP1* به‌شمار می‌رود و نقش مهمی در تنظیم سیگنال‌های TGF-B در سلول‌های تخمدان دارد (Ma *et al.,* 2012). TGF-B یک مولکول تنظیم‌کننده کلیدی است که آثار چندژنی روی رشد سلول، مهاجرت و تهاجم دارد. در نتیجه، نقص در سیگنال‌دهی مناسب TGF-B منجر به تومورزایی و متاستاز می‌شود (Massague, 1988). همچنین مطالعات ریزآرایه نشان داده است که *GULP* به‌طور معنی‌داری در بسیاری از سرطان‌های تخمدان دچار کاهش بیان می‌شود (Schwartz *et al.,* 2003; Hendrix *et al.,* 2006). بنابراین *GULP* در سلول‌های تخمدانی برای حفظ حساسیت آن‌ها به TGF-B ضروری است که نقش مهم آن‌ها را در پیشرفت سرطان تخمدان ثابت می‌کند (Ma *et al.,* 2012)

ژن *MMP2* روی کروموزوم ۱۴ نقش‌های تنظیمی مهمی در لانه‌گزینی و تشکیل جفت برای اطمینان از باروری موفق ایفا می‌کند (Cohen *et al.,* 2006) و جهش‌های آللی در آن با سقط مکرر خودبخودی در ارتباط است (Behforouz *et al.,* 2020). این ژن در GWAS انفرادی در نژاد وادی به عنوان ژن موثر بر چندقلوزایی گزارش شده است (Xu *et al.,* 2018). این ژن که در فولیکول بالغ بیان می‌شود دیواره فولیکولی پستانداران را طی اوولاسیون تخریب می‌کند و بنابراین فعالیت این ژن به لحاظ ژنتیکی برای اوولاسیون مورد نیاز است (Deady *et al.,* 2015).

ژن *RBP4* روی کروموزوم ۲۲ به عنوان ژن کاندید چندقلوزایی معرفی شده است (Messer *et al.,* 1996). این ژن طی فاز توسعه سریع بلاستوسیت بیان می شود که دوره حیاتی برای زنده‌مانی جنین است. این ژن پروتئینی را کد می‌کند که در رحم و در اوایل آبستنی حضور دارد. این پروتئین‌ها به رتینول متصل می‌شوند، سپس رتینول متصل شده به وسیله سلول‌ها پذیرفته شده و جنین‌زایی را آغاز می‌کند (Yelich, 1997). همچنین گزارش شده است که ژن *LPCAT2* روی کروموزوم ۲۲ با رشد آغازین و زنده‌مانی رویان در ارتباط است و بنابراین برای آبستنی موفقیت‌آمیز در نشخوارکنندگان مورد نیاز است (Abdollahi-Arpanahi *et al.,* 2019)

تجزیه انفرادی انجام شده در این جمعیت‌ها، مناطق ژنومی و ژن‌های بیشتری را شناسایی کرد (Xu *et al.,* 2018). متاآنالیز نتایج GWAS انفرادی در جمعیت‌ها و نژادهای مورد مطالعه در این تحقیق نیز تنها ژن *BMPR1B* را به عنوان ژن موثر در چندقلوزایی معرفی کرد (Gholizadeh and Esmaeili-Fard, 2022). ترکیب کردن مجموعه داده‌های چند جمعیت چالش‌های مخصوص به خود را دارد. یکی از موانع اصلی، عدم تجانس نمونه‌های جمعیت‌های مختلف است. این عدم تجانس‌ها ممکن است به طور مثال ناشی از فاصله ژنتیکی بین جمعیت‌ها، تفاوت‌های اندازه‌گیری صفات، محیط‌های پرورشی متفاوت و تراشه‌های تعیین ژنوتیپ متفاوت باشد (Begum *et al.,* 2012). برخلاف نتایج این تحقیق، نشان داده شده است که GWAS با استفاده از ادغام چندین مجموعه داده، نواحی بیشتری را در مقایسه با تجزیه‌های مختص جمعیتی و همچنین متاآنالیز نتایج GWAS آن جمعیت‌ها دارد. اما شناسایی تعداد نواحی بیشتر با استفاده از متاآنالیز، در مقایسه با تجزیه‌های مختص جمعیتی نشان از سودمندی متاآنالیز در غیاب داده‌های خام مختص جمعیتی برای اجرای GWAS مشترک دارد (Gebreyesus *et al.,* 2019). تاکنون تعداد معدودی GWAS به منظور شناسایی نواحی ژنومی مرتبط با صفت دوقلوزایی در نژادهای مختلف موجود در ایران انجام شده است که منجر به شناسایی و معرفی ژن‌های متفاوتی شده است. در گوسفند نژاد بلوچی نشان داده شده است که واریانت‌های متفاوت ژن‌های *NTRK2*، CTH، *ANKRD13C*، *SRSF11*، *PTGER3*، *LDHB*، *LRRC40* و *KCNMA1* در تنوع موجود در صفت دوقلوزایی در این نژاد نقش دارند (Esmaeili-Fard *et al.*, 2021). مطالعات مولکولی نشان داده

است که ژن NTRK2 در اسمبل شدن و رشد اولیه فولیکول‌ها و زنده ماندن تخمک نقش اساسی دارد. در مطالعه دیگری که روی نژاد لری-بختیاری انجام شده است، ژن *LHCGR* به عنوان ژن کاندید موثر بر دوقلوزایی در نژاد لری-بختیاری معرفی شده است (Abdoli *et al*., 2018). این ژن در فرآیند ساخت هورمون‌های استروئیدی در تخمدان نقش دارد. تفاوت در نتاج مطالعات مختلف می‌تواند به دلیل تفاوت‌های بین نژادی، روش‌های مورد استفاده در تجزیه داده‌ها و اندازه نمونه ژنوتیپ شده و دارای فنوتیپ باشد. این موضوع نشان می‌دهد که ساز و کارهای زیستی متفاوتی در تنوع موجود در صفت دوقلوزایی در بین نژادهای مختلف وجود دارد و می‌توان با GWAS، ژن‌های کاندید موثر بر این صفت را شناسایی کرد و از متا و مگاآنالیز در ترکیب نتایج بهره گرفت.

نتیجه‌گیری کلی

در مجموع، نتایج مگاآنالیز، یک نشانگر را روی کروموزوم ۲۱ در سطح ژنوم و ۱۰ نشانگر در سطح کروموزومی شناسایی کرد که به‌طور معنی‌داری با چندقلوزایی در گوسفند در ارتباط بودند. بررسی منابع نشان داد که ژن‌های *OPCML*، *GULP1*، *RBP4*، *MMP2* و *LPCAT2* شناسایی شده در این تحقیق نقش موثری در باروری و موفقیت آبستنی دارند و می‌توانند در برنامه‌های اصلاح نژادی مورد استفاده قرار گیرند. همچنین مگاآنالیز GWAS قدرت مشابهی در مقایسه با متاآنالیز گزارش شده برای نتایج GWAS داشت، ولی در مقایسه با GWASهای انفرادی، مناطق کمتری را شناسایی کرد.

تشکر و قدردانی

این پژوهش با حمایت مالی دانشگاه علوم کشاورزی و منابع طبیعی ساری تحت قرارداد شماره ۰۵-۱۳۹۹-۰۳ انجام شد که بدین وسیله سپاسگزاری می‌شود.

فهرست منابع

Abdoli R., Mirhoseini S. Z., Ghavi Hossein-Zadeh N., Zamani P. and Gondro C. 2018. Genome-wide association study to identify genomic regions affecting prolificacy in Lori-Bakhtiari sheep. Animal Genetics, 49(5): 488-491.

Abdoli R., Mirhosseini S. Z., Ghavi Hossein-Zadeh N., Zamani P., Ferdosi M. H. and Gondro C. 2019. Genome-wide association study of four composite reproductive traits in Iranian fat-tailed sheep. Reproduction, Fertility and Development, 31(6): 1127-1133.

Abdollahi-Arpanahi R., Carvalho M. R., Ribeiro E. S. and Peñagaricano F. 2019. Association of lipid-related genes implicated in conceptus elongation with female fertility traits in dairy cattle. Journal of Dairy Science, 102(11): 10020-10029.

Begum F., Ghosh D., Tseng G. C. and Feingold E. 2012. Comprehensive literature review and statistical considerations for GWAS meta-analysis. Nucleic Acids Research, 40: 3777-3784.

Behforouz A., Dastgheib S. A., Abbasi H., Karimi-Zarchi M., Javaheri A., Hadadan A. and Neamatzadeh H. 2021. Association of MMP-2, MMP-3, and MMP-9 polymorphisms with susceptibility to recurrent pregnancy loss. Fetal and Pediatric Pathology, 40(5): 378-386.

Bernal Rubio Y. L., Gualdron Duart J. L. E., Bates R. O., Ernst C. W., Nonneman D., ohrer G. A., King D. A., Shackelford S. D., Wheeler T. L., Cantet R. J. and Steibel J. P. 2015. Implementing meta-analysis from genome-wide association studies for pork quality traits. Journal of Animal Science, 93: 5607-5617.

Bouwman A. C., Daetwyler H. D., Chamberlain A. J., Ponce C. Sargolzaei M., Schenkel F. S., Sahana G., Govignon-Gion A., Boitard S., Dolezal M., Pausch H., Brøndum R. F., Bowman P. J., Thomsen B., Guldbrandtsen B., Lund M. S., Servin B., Garrick D. J., Reecy J., Vilkki J., Bagnato A., Wang M., Hoff J. L., Schnabel R. D., Taylor J. F., Vinkhuyzen A. A. E., Panitz F., Bendixen C., Holm L. E., Gredler B., Hozé C., Boussaha M., Sanchez M. P., Rocha D., Capitan A., Tribout T., Barbat A., Croiseau P., Drögemüller C., Jagannathan V., Vander Jagt C., Crowley J. J., Bieber A., Purfield D. C., Berry D. P., Emmerling R., Götz K. U., Frischknecht M., Russ I., Sölkner J., Van Tassell C. P., Fries R., Stothard P., Veerkamp R. F., Boichard D., Goddard M. E. and Hayes B. J. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nature Genetics, 50: 362-367.

Bradbury P. J., Zhang Z., Kroon D. E., Casstevens T. M., Ramdoss Y. and Buckler E. S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19): 2633-2635.

Cohen M., Meisser A. and Bischof P. 2006. Metalloproteinases and human placental invasiveness. Placenta, 27(8): 783-793.

Daetwyler H. D., Villanueva B. and Woolliams J. A. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS One, 3(10): e3395.

Deady L. D., Shen W., Mosure S. A., Spradling A. C. and Sun J. 2015. Matrix metalloproteinase 2 is required for ovulation and corpus luteum formation in Drosophila. PLoS Genetics, 11(2): e1004989.

Esmaeili fard S. M., Hafezian S., Gholizadeh M., Abdolahi Arpanahi R. 2019. Gene set enrichment analysis using genome-wide association study to identify genes and biological pathways associated with twinning in Baluchi sheep. Animal Production Research, 8(2): 63-80. (In Persian).

Esmaeili-Fard S. M., Gholizadeh M., Hafezian S. H. and Abdollahi-Arpanahi R. 2021. Genes and pathways affecting sheep productivity traits: Genetic parameters, genome-wide association mapping, and pathway enrichment analysis. Frontiers in Genetics, 1351.

Fleming J. S., McQuillan H. J., Millier M. J. and Sellar G. C. 2009. Expression of ovarian tumour suppressor OPCML in the female CD-1 mouse reproductive tract. Reproduction, 137(4): 721.

Fritsche L. G., Igl W., Bailey J. N., Grassmann F., Sengupta S., Bragg-Gresham J. L. and Heid I. M. 2016. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. Nature Genetics, 48(2): 134-143.

Gebreyesus G., Buitenhuis A. J., Poulsen N. A., Visker M. H. P. W., Zhang Q., van Valenberg H. J. F. and Bovenhuis H. 2019. Combining multi-population datasets for joint genome-wide association and meta-analyses: The case of bovine milk fat composition traits. Journal of Dairy Science, 102(12): 11124-11141.

Gholizadeh M. and Esmaeili-Fard S. M. 2022. Meta-analysis of genome-wide association studies for litter size in sheep. Theriogenology, 1: 180:103-112.

Gorski M., Günther F., Winkler T. W., Weber B. and Heid I. M. 2019. On the differences between mega- and meta-imputation and analysis exemplified on the genetics of age-related macular degeneration. Genetic Epidemiology, 43(5): 559-576.

Gudjonsson A., Gudmundsdottir V., Axelsson G. T., Gudmundsson E. F., Jonsson B. G., Launer L. J., Lamb J. R., Jennings L. L., Aspelund T., Emilsson V. and Gudnason V. 2022. A genome-wide association study of serum proteins reveals shared loci with common diseases. Nature Communications, 13(1): 1-13.

Hayes B. and Goddard M. 2010. Genome-wide association and genomic selection in animal breeding. Genome, 53(11): 876-883.

Hendrix N. D., Wu R., Kuick R., Schwartz D. R., Fearon E. R. and Cho K. R. 2006. Fibroblast growth factor 9 has oncogenic activity and is a downstream target of Wnt signaling in ovarian endometrioid adenocarcinomas. Cancer Research, 66: 1354-1362.

Lin D. Y. and Zeng D. 2010. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. Genetic Epidemiology, 34: 60‑66.

Ma C. I. J., Martin C., Ma Z., Hafiane A., Dai M., Lebrun J. J. and Kiss R. S. 2012. Engulfment protein GULP is regulator of transforming growth factor-β response in ovarian cells. Journal of Biological Chemistry, 287(24): 20636-20651.

Marjanovic J. and Calus M. P. L. 2020. Factors affecting accuracy of estimated effective number of chromosome segments for numerically small breeds. Journal of Animal Breeding and Genetics,138: 151-160.

Massague J. 1998. TGF-β signal transduction. Annual Review of Biochemistry, 67: 753-791.

Messer L. A., Wang L., Yelich J., Pomp D., Geisert R. D. and Rothschild M. F. 1996. Linkage mapping of the retinol-binding protein 4 (RBP4) gene to porcine chromosome 14. Mammalian Genome, 7: 396-410.

Meuwissen T. H. E., Hayes B. J. and Goddard M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 57(4): 1819‑1829.

Money D., Gardner K., Migicovsky Z., Schwaninger H., Zhong G. Y. and Myles S. 2015. LinkImpute: fast and accurate genotype imputation for nonmodel organisms. G3: Genes, Genomes, Genetics, 5(11): 2383-2390.

Pasandideh M., Rahimi-Mianji G., Gholizadeh M. and Fontanesi L. 2017. Detection of genomic regions affecting reproductive traits in Baluchi sheep using high density markers. Animal Production Research, 6(3): 29-41. (In Persian).

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D. and Sham P. C. 2007. PLINK: a tool set for whole-enome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3): 559-575.

Schwartz D. R., Wu R., Kardia S. L., Levin A. M., Huang C. C., Shedden K. A., Kuick R., Misek D. E., Hanash S.

M., Taylor J. M., Reed H., Hendrix N., Zhai Y., Fearon E. R. and Cho K. R. 2003. Novel candidate targets of β-catenin/T-cell factor signaling identified by gene expression profiling of ovarian endometrioid adenocarcinomas. Cancer Research, 63: 2913-2922.

Sellar G. C., Watt K. P., Rabiasz G. J., Stronach E. A., Li L., Miller E. P., Massie C. E., Miller J., Contreras-Moreira B., Scott D., Brown I., Williams A. R., Bates P. A., Smyth J. F. and Gabra H. 2003. OPCML at 11q25 is epigenetically inactivated and has tumor-suppressor function in epithelial ovarian cancer. Nature Genetics, 34: 337-343.

Sung Y. J., Schwander K., Arnett D. K., Kardia S. L., Rankinen T., Bouchard C., Boerwinkle E., Hunt S. C. and Rao D. C. 2014. An empirical comparison of meta-analysis and mega-analysis of individual participant data for identifying gene-environment interactions. Genetc Epidemiology, 38: 369‑378.

Taghizade K., Gholizadeh M., Moradi M. and Rahimi Mianji G. 2020. Investigation of copy number variation in Baluchi sheep genome using comparative analysis of PennCNV and QuantiSNP algorithms. Animal Production Research, 9(1), 29-44. (In Persian).

Tenghe A. M. M., Bouwman A. C., Berglund B., Strandberg E., de Koning D. J. and Veerkamp R. F. 2016. Genome-wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle. Journal of Dairy Science, 99(7): 5470-5485.

Tsou J. A., Galler J. S., Siegmund K. D., Laird P. W., Turla S., Cozen W., Hagen J. A., Koss M. N. and Laird-Offringa I. A. 2007. Identification of a panel of sensitive and specific DNA methylation markers for lung adenocarcinoma. Molecular Cancer, 6: 70.

VanRaden P. M., van Tassell C. P., Wiggans G. R., Sonstegard T. S., Schnabel R. D., Taylor J. F. and Schenkel F. S. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science, 92(1): 16‑24.

Xu S. S., Gao L., Xie X. L., Ren Y. L., Shen Z. Q., Wang F. and Li M. H. 2018. Genome-wide association analyses highlight the potential for different genetic mechanisms for litter size among sheep breeds. Frontiers in Genetics, 9: 118.

Yelich J. V., Pomp D. and Geisert R. D. 1997. Detection of transcripts for retinoic acid receptors, retinol-binding protein, and transforming growth factors during rapid trophoblastic elongation in the porcine conceptus. Biology of Reproduction, 57: 286-294.