مقاله علمی پژوهشی/نشریه پژوهش‌های علوم دامی/ جلد ۳۳ شماره ۲/ سال۱۴۰۲/ صفحات ۱۰۷ تا ۱۱۹
DOI: 10.22034/AS.2022.51449.1661

مطالعه پویش کل ژنومی بر پایه آنالیز مجموعه‌های ژنی مرتبط با صفت تعداد سرپستانک در گوسفند

حسین محمّدی[۱]*، آرش جوانمرد[۲]، ابوذر نجفی[۳] و محمد شمس اللهی[۴]

تاریخ دریافت: ۱۴۰۱/۲/۶ تاریخ پذیرش: ۱۴۰۱/۷/۹

[۱]استادیار گروه علوم دامی، دانشکده کشاورزی و محیط زیست، دانشگاه اراک
[۲]دانشیار گروه علوم دامی، دانشکده کشاورزی، دانشگاه تبریز، تبریز
[۳]استادیار گروه علوم دام و طیور، پردیس ابوریحان، دانشگاه تهران، پاکدشت
[۴]استادیار گروه علوم دامی، دانشکده کشاورزی، دانشگاه ایلام
*مسئول مکاتبه: Email: H-mohammadi64@araku.ac.ir

چکیده

زمینه مطالعاتی: تعداد سرپستانک‌ها، صفت مهمی در رابطه با قابلیت مادری گوسفندان چند قلوزا است، شناخت عملکرد ژنتیکی و جایگاه‌های ژنی کنترل‌کننده این صفت در ژنوم گونه‌های مختلف حائز اهمیت است. گوسفندان چند قلوزا به صورت بالقوه به ایمنوگلوبولین‌های دریافتی از آغوز به دلیل ساختار جفت در ساعات ابتدایی تولد وابسته هستند. لذا تعداد سرپستانک‌ها، زمانی که در یک زایش تعداد نتاج متولد شده بیشتری نسبت به تعداد پستانک‌ها وجود داشته باشد، نقش مهمی ایفاء می‌کند. **هدف:** پژوهش حاضر، با هدف مطالعه‌ی پویش ژنومی بر اساس آنالیز مجموعه‌های ژنی، برای شناسایی جایگاه‌های ژنی مؤثر بر صفت تعداد سر پستانک، با استفاده از آرایه‌های ژنومی با تراکم بالا است. **روش کار:** از اطلاعات رکوردهای فنوتیپی و ژنوتیپی مرتبط با صفت تعداد سرپستانک ۱۶۰ نمونه از گوسفند هوی استفاده شد. آنالیز پیوستگی در نرم‌افزار GEMMA انجام شد، سپس با استفاده از بسته‌ی نرم افزاری *biomaRt2* ژن‌های معنی‌داری که در داخل و یا ۲۵ کیلوباز بالا و پایین دست نشانگرهای معنی‌دار قرار داشتند، شناسایی گردید. در نهایت، مجموعه‌ی ژنی با بسته نرم افزاری *goseq* برنامه R با هدف شناسایی عملکرد بیولوژیکی ژن‌های نزدیک به مناطق انتخابی و ژن‌های کاندیدا، از طریق پایگاه‌های GO، KEGG، DAVID و PANTHER تفسیر شدند. **نتایج:** در این پژوهش نشانگرهای تک نوکلئوتیدی واقع روی کروموزوم‌های ۳، ۵، ۷، ۸، ۱۲، ۱۴ و ۱۷ شناسایی شدند که با ژن‌های *CDH11*، *NUMB*، *FGF2*، *ESR1*، *LGR5*، *INSR* و *PTGS2* مرتبط بودند. در تفسیر مجموعه ژنی، تعداد ۱۱ مسیر هستی شناسی ژنی و بیوشیمیایی با صفت تعداد سرپستانک شناسایی شد ($P<0.05$). از این بین، مسیرهای Blastocyst development، Mesenchymal cell development، Developmental growth involved in morphogenesis، Muscle cell differentiation، AMPK signaling pathway و Regulation of lipolysis in adipocytes عملکردهای مهمی را در ارتباط با رشد و توسعه‌ی غدد پستانی و فعال‌سازی مسیر سیگنال‌دهی AMPK بر عهده داشتند. **نتیجه گیری نهایی:** ژن‌های شناسایی شده در این مطالعه می‌توانند به عنوان ژن‌های کاندیدا برای مطالعات تکمیلی، درک بهتر مکانیسم ژنتیکی و نهایتاً در انتخاب ژنتیکی تعداد سرپستانک در گوسفند مورد استفاده قرار گیرد.

واژگان کلیدی: آنالیز مسیر، پویش ژنومی، هستی شناسی، سرپستانک اضافی

مقدمه

نوزاد نشخوارکنندگان به صورت کامل به ایمنوگلوبولین‌های دریافتی از آغوز وابسته می‌باشند (به علت عدم عبور ایمونوگلوبولین‌ها مادر به جنین از جفت دارای ٥ تا ٦ لایه‌ای) و با همین استدلال خون نوزاد نشخوارکنندگان تا زمانی که آغوز مصرف نشده است، فاقد ایمنوگلوبولین است و یا مقدار بسیار جزیی ایمنوگلوبولین در خون آن جریان دارد. لذا تعداد سرپستانک یک فاکتور کلیدی است که توانایی مراقبت از بره و سلامتی آن در آینده را تحت تأثیر قرار می‌دهد. از طرف دیگر سرپستانک‌های ضعیف به عنوان محیط مناسب برای رشد باکتری‌های مسبب ورم پستان هستند که منجر به کاهش بهره‌وری در طول دوره شیردهی و در نهایت حذف دام از گله می‌شود (پن و همکاران ۲۰۱۷؛ هنرور و همکاران ۲۰۱۲).

از آنجا که قابلیت مادری در گوسفند تا حد زیادی به شکل و عملکرد غده پستان بستگی دارد، بنابراین طراحی برنامه اصلاحی برای پستانک‌های عملکردی در حیوانات چند قلوزا نقش مهمی در کاهش رقابت فرزندان در رسیدن به منبع آغوز را ایفا می‌کند، همچنین فهم بهتر از کنترل ژنتیکی این صفت، فرصتی برای درک معماری ژنتیکی را نیز فراهم می‌سازد.

تا به امروز تعداد ۱۲۹ QTL در ارتباط با عارضه تعداد سرپستانک در گوسفند گزارش شده است که روی کروموزوم‌های مختلف شامل ۱، ۳، ٤، ٥، ۱۰، ۱۳، ۱۸ و ۲٥ قرار دارند (https://www.animalgenome.org/cgi-bin/QTLdb/OA/qtrait?trait_ID=3318). مطالعات انجام شده روی عارضه تعداد سرپستانک بسیار اندک بوده به‌طوریکه پن و همکاران (۲۰۱۷) با مطالعه پویش ژنومی در گوسفندان نژاد وادی، ژن‌های کاندیدای *BBX* و *CD47* روی کروموزوم ۱ را در ارتباط با تعداد سرپستانک گزارش کردند. ژیو و همکاران (۲۰۲۱) با مطالعه پویش کل ژنومی در گوسفندان هوی، مناطق ژنومی معنی‌دار مرتبط با تعداد سرپستانک روی کروموزوم‌های ۲، ٤، ٥، ۱۰، ۲۳ شناسایی کردند که حاوی ژن‌های کاندیدای *TWIST1*، *LHFP*، *DPYSL2* و *TDP-43* مرتبط با تعداد سرپستانک بودند. همچنین در مطالعه‌ی پویش کل ژنوم براساس تنوع در تعداد کپی (CNV) ژنوم گوسفندان جهان، ژن‌های کاندیدای *ABCB1* و *AKT1* به ترتیب روی کروموزوم‌های ٤ و ۱۸ در ارتباط با تعداد سرپستانک گزارش شده است (صالحیان-دهکردی و همکاران ۲۰۲۱).

هدف از مطالعات پویش ژنومی (GWAS[1]) که با شناسایی ارتباط بین نشانگرهای SNP و صفت مورد مطالعه با استفاده از نشانگرهای با تراکم بالا در سطح ژنوم همراه است، پیدا کردن جهش‌های مؤثر یا مسببی می‌باشد که بر فنوتیپ صفت موردنظر اثر می‌گذارند. این اطلاعات می‌تواند برای انتخاب به کمک نشانگر مفید بوده و به درک مکانیسم مولکولی صفات مورد مطالعه کمک نماید (محمدی و همکاران ۲۰۲۰). یکی از معایب مطالعات GWAS در نظر گرفتن آستانه معنی‌داری محافظه کارانه برای جلوگیری از اشتباه کاذب یا خطای نوع اول است، در حالیکه پرهیز از اشتباه نوع اول باعث افزایش خطای نوع دوم می‌گردد. لذا برای حل این مشکل، یکی از راه حل‌ها استفاده از آنالیزهای غنی‌سازی مجموعه‌های ژنی می‌باشد. در این روش ارتباط بین صفت مورد مطالعه و نشانگرهای ژنتیکی را در یک دسته یا گروه ژنی که به طور عملکردی با هم مرتبط هستند بررسی می‌کند. در حقیقت در این روش به دنبال ژن‌هایی هستیم که به تنهایی اثر آنها بر صفت مورد نظر معنی‌دار نشده، ولی اثر تجمعی آنها روی صفت دارای اثر معنی‌دار است (مارکس و همکاران ۲۰۱۸). علاوه براین، یکی از دلایل اصلی آنالیز غنی سازی مجموعه‌های ژنی تعداد کم بودن SNPهای معنی‌دار می‌باشد که موجب عدم شناسایی مناطق ژنومی مرتبط با صفات مهم اقتصادی می‌گردد. در نتیجه روش پویش ژنومی بر مبنای مسیر کارآیی

[1] Genome Wide Association Study (GWAS)

بهتری برای یافتن مناطق ژنومی، درک بهتر مکانیسم و معماری ژنتیکی را دارا می‌باشد (اسرایکن و همکاران ۲۰۲۰).

برای اولین بار پناگریکانو و همکاران (۲۰۱۳) نشان دادند که تجزیه و تحلیل پویش ژنومی بر مبنای آنالیز مجموعه‌های ژنی دقت شناسایی مناطق ژنومی مؤثر بر صفت نرخ باروری گاوهای نر را بالا برده است، زیرا با استفاده از این روش تمامی نشانگرهای معنی‌دار در سطح ۰/۰۵ و بدون در نظر گرفتن هیچگونه تصحیحات سختگیرانه آنالیز می‌شوند و در نتیجه میزان خطای نوع اول و بیش برآوردها کاهش پیدا می‌کند. خلت‌آبادی فراهانی و همکاران (۲۰۲۰) مطالعه‌ی پویش ژنومی بر مبنای مسیر با استفاده از آنالیز غنی‌سازی مجموعه‌های ژنی روی صفت تولید مثلی تعداد نتاج در نژادهای مختلف گوسفندان جهان انجام دادند. نتایج به‌دست آمده از آنالیز مسیر، منجر به شناسایی طبقات مختلف عملکردی هستی‌شناسی ژن و مسیرهای زیستی KEGG معنی‌دار مرتبط شامل Oocyte differentiation، Ovulation from ovarian follicle، Estrogen signaling pathway و Positive regulation of peptide hormone secretion pathway و ژن‌های کاندیدای *FOXO3*، *ESR1*، *ESR2*، *RBP4*، *AURKA*، *INSR*، *DLG1* و *ZGLP1* شده بود. گزارش شده است که آنالیز پویش ژنومی بر مبنای مسیر دقت شناسایی مناطق ژنومی مؤثر بر صفات تولیدی شامل مجموع وزن بره‌های متولد شده و از شیرگیری شده در گوسفندان نژاد بلوچی را بالا می‌برد و منجر به شناسایی مسیرهای زیستی مرتبط با میزان خوراک مصرفی، بازدهی خوراک و تولید و ترکیبات شیر، همچنین ژن‌های کاندیدای جدید *NR2C1*، *VEZT*، *HSD17B4*، *RSU1*، *CUBN*، *VIM*، *PRLR* و *FTH1* شده است (اسماعیلی‌فر و همکاران- ۲۰۲۱).

داده‌های مورد استفاده در این پژوهش، در مطالعه قبلی برای صفت تعداد سرپستانک در گوسفند نژاد هوی مورد آنالیز قرار گرفته‌اند که در آن تحقیق از مدل خطی مختلط تک نشانگری و از تصحیح نرخ کشف کاذب براساس روش بنجامین-هوشبرگ برای تعیین آستانه‌ی معنی‌داری و جلوگیری از خطای نوع اوّل استفاده شده است (ژیو و همکاران ۲۰۲۲)، که منجر به از دست رفتن مناطق ژنومی مرتبط با تعداد سرپستانک شده بود. بنابراین، هدف از انجام پژوهش حاضر، تجزیه و تحلیل مجموعه‌های ژنی و مسیرهای زیستی مرتبط با صفت تعداد سرپستانک در گوسفند با استفاده از تراشه‌های 600K و براساس پویش کل ژنوم بر مبنای مسیر می‌باشد.

مواد و روش‌ها

در مطالعه حاضر از اطلاعات ۱۶۰ رأس گوسفندان نژاد هوی رکورد گیری شده مرتبط با صفت تعداد سرپستانک استفاده گردید (ژیو و همکاران ۲۰۲۲). تعداد ۷۷ رأس از میش‌ها دارای دو سرپستانک نرمال (شاهد) و تعداد ۸۳ رأس دارای دو سرپستانک نرمال و یک یا دو سرپستانک اضافی (موردی) بودند. نمونه‌ها با استفاده از آرایه‌های شرکت افی‌متریکس با تراشه‌های شامل ۵۶۶۱۲۹ نشانگر تعیین ژنوتیپ شدند.

برای فیلتراسیون داده‌های تعیین ژنوتیپ شده، ابتدا نمونه‌هایی که فراوانی نرخ تعیین ژنوتیپ آنها کمتر از ۹۰٪ بود، شناسایی و حذف شد. در مرحله بعد نشانگرهایی که حداقل فراونی آللی در آنها کمتر از ۱٪ بود حذف شدند. سپس نشانگرهایی که نرخ تعیین ژنوتیپ آنها در نمونه‌ها کمتر از ۹۰٪ بود شناسایی و حذف شدند. در نهایت برای SNPهای باقی‌مانده، آن‌هایی که در تعادل هاردی-واینبرگ قرار نداشتند به عنوان معیاری از خطای تعیین ژنوتیپ کنار گذاشته شدند. مراحل مختلف فیلتراسیون با استفاده از نرم افزار PLINK (v1.90; http://pngu.mgh.harvard.edu/purcell/plink) انجام شد (پورسل و همکاران ۲۰۰۷). در نهایت بعد از کنترل کیفیت تعداد ۴۸۹۹۵۸ SNP برای آنالیزهای پویش کل ژنومی بر پایه تجزیه و تحلیل غنی‌سازی مجموعه ژنی باقی ماندند. همچنین در این تحقیق برای بررسی وجود

یا عدم وجود لایه‌بندی جمعیتی با استفاده از روش تجزیه و تحلیل مؤلفه‌های اصلی (PCA[1]) با تابع prcomp در نرم افزار R بررسی شد.

جهت انجام تجزیه و تحلیل پویش کل ژنومی با استفاده از نرم افزار GEMMA (ژو و استیفنز ۲۰۱۲) نسخه ۰/۹۸ استفاده شد. این نرم افزار توانایی در نظر گرفتن هر کدام مشاهدات فنوتیپی داده‌های باینری را به صورت متغییر جداگانه دارا می‌باشد (زوانگ و همکاران ۲۰۲۰). مدل مورد استفاده بر پایه مدل خطی مختلط تک صفته به شکل زیر بود:

$$\boldsymbol{y} = \boldsymbol{Wa} + \boldsymbol{x\beta} + \boldsymbol{u} + \boldsymbol{e}$$

که **y**: بردار مشاهدات فنوتیپی، **W** ماتریس ضرایب مربوط به دو مؤلفه اصلی با بیشترین واریانس توجیه شده، $\boldsymbol{x}$: بردار ژنوتیپ‌های نشانگر تحت آزمون، **u** بردار اثرات پلی‌ژنیک، **e** بردار اثر باقی‌مانده‌های تصادفی، **α** و **β** برادر ضرایب ارتباط دهنده.

آنالیز پویش کل ژنومی براساس مجموعه‌های ژنی (GSEA-SNP)

اساساً آنالیز پویش ژنومی بر پایه تجزیه و تحلیل مجموعه‌های ژنی در سه مرحله انجام می‌گردد: ۱) تعیین مکان SNPهای معنی‌دار با ژن ۲) ارتباط ژن‌ها به طبقات عملکردی و مسیرهای زیستی ۳) پویش کل ژنومی بر پایه آنالیز مسیر

۱-**تعیین مکان SNPها با ژن‌ها**: SNPهایی که مقدار *P*-value آنها کمتر از ۰/۰۵ بود با استفاده از بسته نرم افزاری *biomaRt2* در محیط R (نسخه۳/۶/۱) با دستور getBM() و با استفاده از رفرانس ژنومی گوسفند به ژن‌هایی که نشانگر SNP مورد نظر در داخل آن ژن و یا kb ۲۵ بالادست یا پایین دست آن ژن قرار داشت، ارتباط داده شدند.

۲- **ارتباط ژن‌ها به طبقات عملکردی و مسیرهای بیوشیمیایی**: جهت تعیین طبقات عملکردی ژنی و مسیرهای متابولیکی و تنظیمی ژن‌های معنی‌دار از ۵ پایگاه‌های اطلاعاتی شامل هستی شناسی ژن (GO, /http://www.geneontology.org)، مسیرهای بیوشیمیایی (KEGG, /http://www.genome.jp/kegg)، Panther (http://www.pantherdb.org)، Metacyc (http://www.metacyc.org) و Reactome (http://www.reactome.org) جهت تعیین طبقات عملکردی و مسیرهای بیوشیمیایی استفاده گردید.

۳- **پویش کل ژنومی بر پایه آنالیز مسیر**: ارتباط‌های معنی‌دار مسیرهای عملکردی با صفت تعداد سرپستانک با استفاده از توزیع فوق هندسی و آماره 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:

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

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 آنالیز شدند و m برابر با تعداد کل ژن‌های موجود در مسیر عملکردی. تجزیه غنی‌سازی مجموعه ژنی با استفاده از بسته نرم افزاری *goseq* در محیط نرم افزار R (نسخه۳/۶/۱) انجام گردید. برای تفسیر بهتر عملکرد ژن‌های به دست آمده از پایگاه‌های اطلاعاتی آنلاین GeneCards (http://www.genecards.org) و UniProtKB (http://www.uniprot.org) استفاده شد.

نتایج و بحث

هدف از تجزیه به مؤلفه‌های اصلی آن است که واریانس موجود در داده‌های چند متغیره را به مؤلفه‌هایی تجزیه کند که اولین مؤلفه تا آنجا که ممکن است علت بیشترین واریانس موجود در داده‌ها باشد. دومین مؤلفه علت

[1] Principal Component Analysis

بیشترین واریانس ممکن بعد از مؤلفه اول و الی آخر باشد. همچنین لایه‌بندی جمعیتی که ناشی از تفاوت در فراوانی آللی زیر جمعیت‌ها به دلیل تفاوت ژنتیکی جد مشترک و همچنین میزان خویشاوندی می‌تواند نتایج یک مطالعه ارتباط ژنومی را با مشکل مواجه نماید چرا که در مطالعات پویش ژنومی فرض می‌شود که جامعه مورد مطالعه همگن می‌باشد. علاوه بر این، در نظر گرفتن اثرات خویشاوندی می‌تواند منجر به کاهش نتایج مثبت کاذب (خطای نوع اول) شود (لیو و همکاران ۲۰۲۲).

نتایج آنالیز PCA براساس PCA1 و PCA2 نشان داد که برای این جمعیت، دو مؤلفه اول به ترتیب ۹/۷۸ و ۸/۰۲ درصد از کل واریانس را توجیه می‌کنند. در نتیجه برای کنترل اختلاط و ارتباط ژنتیکی بین این حیوانات، دو مؤلفه اصلی که بیشترین میزان واریانس توجیه شده را داشتند در مدل آنالیزی قرار داده شدند (ژیو و همکاران، ۲۰۲۲). پلات منهتن مرتبط با صفت تعداد سر پستانک در شکل ۱ ارائه شده است.

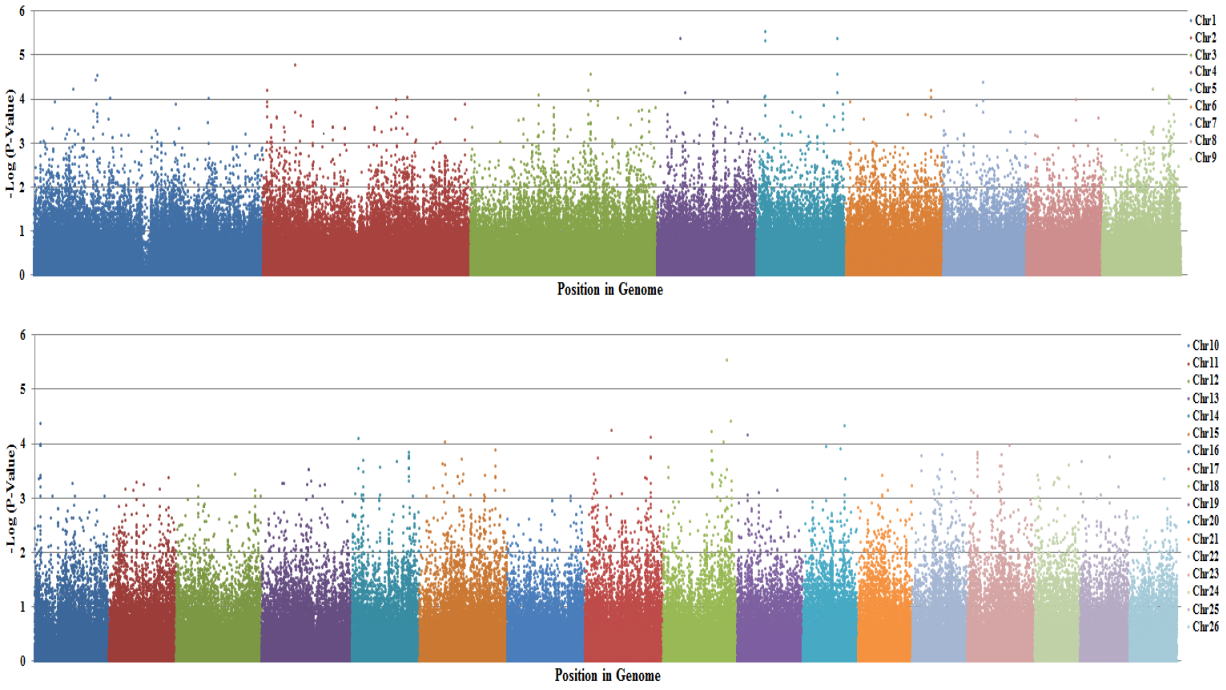

Figure 1- Manhattan plot for number of teats in Hu sheep. X axis, SNP positions on chromosomes, Y axis, -Log10 P-value

شناسایی ژن‌های کاندیدا در مناطق ژنومی کاندیدا

تعداد ۲۵۷۰۰ ژن از ۲۷۰۵٤ ژن‌های کدکننده حاشیه نویسی شده در ژنوم گوسفند به وسیله نشانگرهای SNP پوشش داده شد که در این میان مجموع ۱۱۸۹ ژن دارای اثر معنی‌دار روی صفت تعداد سرپستانک بودند، یعنی حداقل یک نشانگر با P-value کمتر از ۰/۰۵ در داخل و یا در بالا یا پایین دست این ژن تا فاصله ۲۵ kb قرار گرفت.

تعداد مجموعه‌های ژنی حاصل از پایگاه‌های داده‌ای مختلف شامل ۱٦۱ طبقات هستی شناسی، ۲٤ مسیر بیوشیمیایی KEGG و ۷ مسیر Reactome بود. مسیرهای که بیشتر از ۳ ژن و کمتر از ۵۰۰ ژن داشتند گزارش شده‌اند. جزئیات کامل طبقات هستی شناسی به همراه اسامی ژن‌های کاندیدا در جدول ۱ ارائه شده است. همانطور که در جدول ۱ مشاهده می‌شود تعداد ۱۱ طبقات عملکردی در هستی شناسی فرآیندهای زیستی، عملکرد

مولکولی، اجزای سلولی و مسیرهای KEGG با صفت تعداد سرپستانک دارای ارتباط هستند (P<0.05).

از مسیرهای زیستی مهم و معنی‌دار مرتبط با صفت تعداد سرپستانک می‌توان به مسیرهای Blastocyst development و Mesenchymal cell development اشاره نمود که از بین ژن‌های معنی‌دار در این مسیرها، ژن‌های کاندیدای *CDH11* و *NUMB* بیشترین ارتباط را با صفت مورد مطالعه داشتند.

ارتباط معنی‌داری بین ژن *CDH11* با رشد و توسعه غدد پستانی در دام‌های شیری و همچنین بیماری ورم پستان گزارش شده است (یانگ و همکاران ۲۰۱۸؛ بختیاری‌زاده و همکاران ۲۰۲۰). مطالعه‌ی پویش کل ژنومی با هدف شناسایی ژن‌های کاندیدای مرتبط با صفات ساختاری سیستم پستانی در گاوهای هلشتاین، ژن کاندیدای *CDH11* جزء ژن‌های کاندیدای مؤثر بر طول سرپستانک‌های جلویی گزارش شده است (نظر و همکاران ۲۰۲۱). همچنین ژن کاندیدای *NUMB* دارای نقش اساسی در تنظیم رشد سیستم پستانی دارد (چارترجی و همکاران ۲۰۱۹). در مطالعه‌ی پویش کامل ژنوم خوک‌های نژاد Duroc با هدف شناسایی مناطق ژنومی مرتبط با تعداد سرپستانک، ژن کاندیدای *NUMB* گزارش شده است (لی و همکاران ۲۰۲۱).

Table 1. Gene set enrichment analysis significantly (P < 0.05) associated with number of teats

GO ID	Term (GO hierarchy level)	No. genes in the GO term	No. significant genes	Related candidate genes	*P* value
Biological process					
GO: 0001824	Blastocyst development	14	4	*PLPP4, PPP4R4, MED21,* ***CDH11***	0.0015
GO: 0014031	Mesenchymal cell development	8	3	*CDH2, NRP1,* ***NUMB,***	0.0071
GO: 0009653	Anatomical structure morphogenesis	55	7	*ANGPTL3, EGR2,* ***FGF2,*** *FBN1, FGFR2, GDF3, SEMA3A*	0.0148
GO: 0060560	Developmental growth involved in morphogenesis	17	3	***ESR1****, FSTL4, NRP1*	0.0192
GO: 0042692	Muscle cell differentiation	28	7	*ACTN3,* ***CAMK2A****, CAPN2, GDF3, HOMER1, PPARD, SYNE1*	0.0174
GO:0048858	Cell projection morphogenesis	38	5	*MYCBP2, DRD2,* ***LGR5****, EGR2, FLOT1*	0.0215
GO:0060249	Anatomical structure homeostasis	31	3	*ERCC6, PPP1R10, HOMER1*	0.0348
GO:0044255	Cellular lipid metabolic process	36	4	*PLPPR2, ADIPOR1, ANGPTL3, SORL1*	0.0347
GO:0060612	Adipose tissue development	5	2	*ATF2, PPARD,*	0.0436
PANTHER Pathways					
P04152	AMPK signaling pathway	12	2	***INSR,*** *ADIPOR1*	0.0103
P04923	Regulation of lipolysis in adipocytes	7	4	*NPY,* ***PTGS2****, ADRB3, ADCY8*	0.0296

دیگر مسیرهای مهم و معنی‌دار مرتبط با تعداد سرپستانک را می‌توان به مسیر بیولوژیکی Anatomical structure morphogenesis Cycle اشاره نمود که ژن کاندیدای *FGF2* در این مسیر قرار داشت. با مطالعه پویش ژنومی براساس رویکرد سیستم بیولوژی، در سه نژاد گاوهای شیری فرانسوی مرتبط با صفات ساختاری پستان، ژن کاندیدای *FGF2* مرتبط با صفت مورفولوژی پستان گزارش شده است (مرتی و همکاران ۲۰۱۸). وانگ

و همکاران در سال ۲۰۰۸ بیان ژن *FGF2* را در مراحل مختلف رشد غدد پستان گاو شیری مورد آزمایش قرار داده و گزارش کردند که ژن *FGF2* در طول توسعه غده پستان بیان متغیری داشته و سبب تحریک رشد سلول در غده پستان می‌شود.

مسیر Developmental growth involved in morphogenesis که جزء هستی‌شناسی فرآیند زیستی است را می‌توان یکی از مهمترین مسیرهای مؤثر بر فرآیند رشد و توسعه غدد پستانی در ارتباط با صفت تعداد سرپستانک دانست. از بین ژن‌های معنی‌دار در این مسیر، می‌توان به ژن کاندیدای *ESR1* اشاره نمود. در مطالعه پویش کل ژنومی با هدف شناسایی ژن‌های کاندیدای مرتبط با تعداد بره متولد شده در نژادهای مختلف گوسفند، ژن کاندیدای *ESR1* در نژاد تکسل گزارش شده است (زو و همکاران ۲۰۱۸).

استروژن یکی از دو هورمون جنسی استروئیدی است که به‌وسیله تخمدان ترشح می‌شود و در رشد جنینی و بروز صفات ثانویه جنسی نقش مهمی ایفاء می‌کند. برای تنظیم عملکرد فیزیولوژیکی هورمون‌های استروئیدی وجود گیرنده‌ها در سلول‌های هدف ضروری می‌باشد. بیان ژن گیرنده استروژن در آندومتر و غده پستانی بسیار زیاد است و یک فاکتور رونویسی می‌باشد و که واسط عملکرد هورمون استروژن در بسیاری از فرآیندهای فیزیولوژیکی است.

از دیگر مسیرهای زیستی مهم و معنی‌دار مرتبط با تعداد سرپستانک می‌توان به مسیرهای Cell projection morphogenesis و Muscle cell differentiation اشاره نمود که جزء مسیرهای فرآیندهای زیستی مرتبط با سیستم پستانی می‌باشند که ژن‌های کاندیدای *LGR5* و *CAMK2A* دارای نقش بیولوژیکی مستقیمی با صفات مرتبط با رشد و توسعه غدد پستانی داشت که بیشتر بحث خواهد شد. ژن کاندیدای *LGR5* به عنوان ژن کاندیدای مرتبط با رشد و تفرق سلولی و توسعه غدد پستانی گزارش شده است (نیوسی و کلیورس ۲۰۱۷). در مطالعه ژنومی با هدف شناسایی ژن‌های کاندیدای مرتبط با چند قلوزایی در گوسفندان هوی، میزان بیان ژن کاندیدای *LGR5* با نرخ باروری بالا ارتباط مستقیمی گزارش شده است (گایو و همکاران ۲۰۲۱).

همچنین در مطالعه پویش ژنومی برپایه فرا-تحلیل در تلیسه‌های گاوهای بوس ایندکیوس مرتبط با مسیرهای تولیدمثلی، ژن کاندیدای *CAMK2A* گزارش شده است (تهیر و همکاران ۲۰۲۱). گزارش شده است که ژن *CAMK2A*نقش اساسی در مسیر سیگنال‌دهی GnRH داشته و میزان ترشح هورمون‌های تولیدمثلی LH و FSH را تنظیم می‌کند (بلیس و همکاران ۲۰۱۰).

مسیرهای بیوشیمیایی مرتبط با ژن‌های دخیل در رشد و توسعه غدد پستانی با استفاده از پایگاه داده KEGG نیز مورد بررسی قرار گرفته شد که با نتایج برخی از تحقیقات قبلی مرتبط با رشد و توسعه سیستم پستانی مطابقت داشت (دسین و همکاران ۲۰۲۱). جزئیات کامل این مسیرهای زیستی به همراه اسامی ژن‌های کاندیدا در جدول ۱ ارائه شده است. با بررسی نتایج حاصل شده مشاهده می‌شود که ژن‌های *ADIPOR1*، *INSR*، *MAPK12* و *MAP3K1* با مسیر زیستی AMPK signaling pathway، ژن‌های *NPY*، *PTGS2*، *ADRB3* و *ADCY8* با مسیر زیستی Regulation of lipolysis in adipocytes در گوسفندان هوی مشاهده شدند.

در شکل ۲ مسیر سیگنالی AMPK signaling pathway و ژن‌های مربوطه شناسایی شده از پایگاه برخط KEGG ارائه شده است. ژن کاندیدای *INSR* نقش کلیدی در ارتباط با تنظیم کننده سیگنالینگ انسولین دارد. نقش این ژن در ارتباط با تنظیم هموستازی گلوکز می‌باشد (کیم و همکاران ۲۰۱٤). مطالعه‌ی پویش کل ژنومی در گوسفندان بومی اتیوپی، با هدف شناسایی ژن‌های مرتبط با ذخیره چربی، ژن کاندیدای *INSR* گزارش شده است (اهبرا و همکاران ۲۰۱۹).

همچنین در شکل ۳ مسیر Regulation of lipolysis in adipocytes و ژن‌های مربوطه ارائه شده است. ژن

کاندیدای *PTGS2* نقش کلیدی در سیگنال‌دهی LH و فولیکوژنوسیز دارد (دینگ و همکاران، ۲۰۰۶). با مطالعه پویش کل ژنومی در سه نژاد گوسفند لهستانی مرتبط با چند قلوزایی، ژن کاندیدای *PTGS2* گزارش شده است (اسملوچه و همکاران ۲۰۲۱).

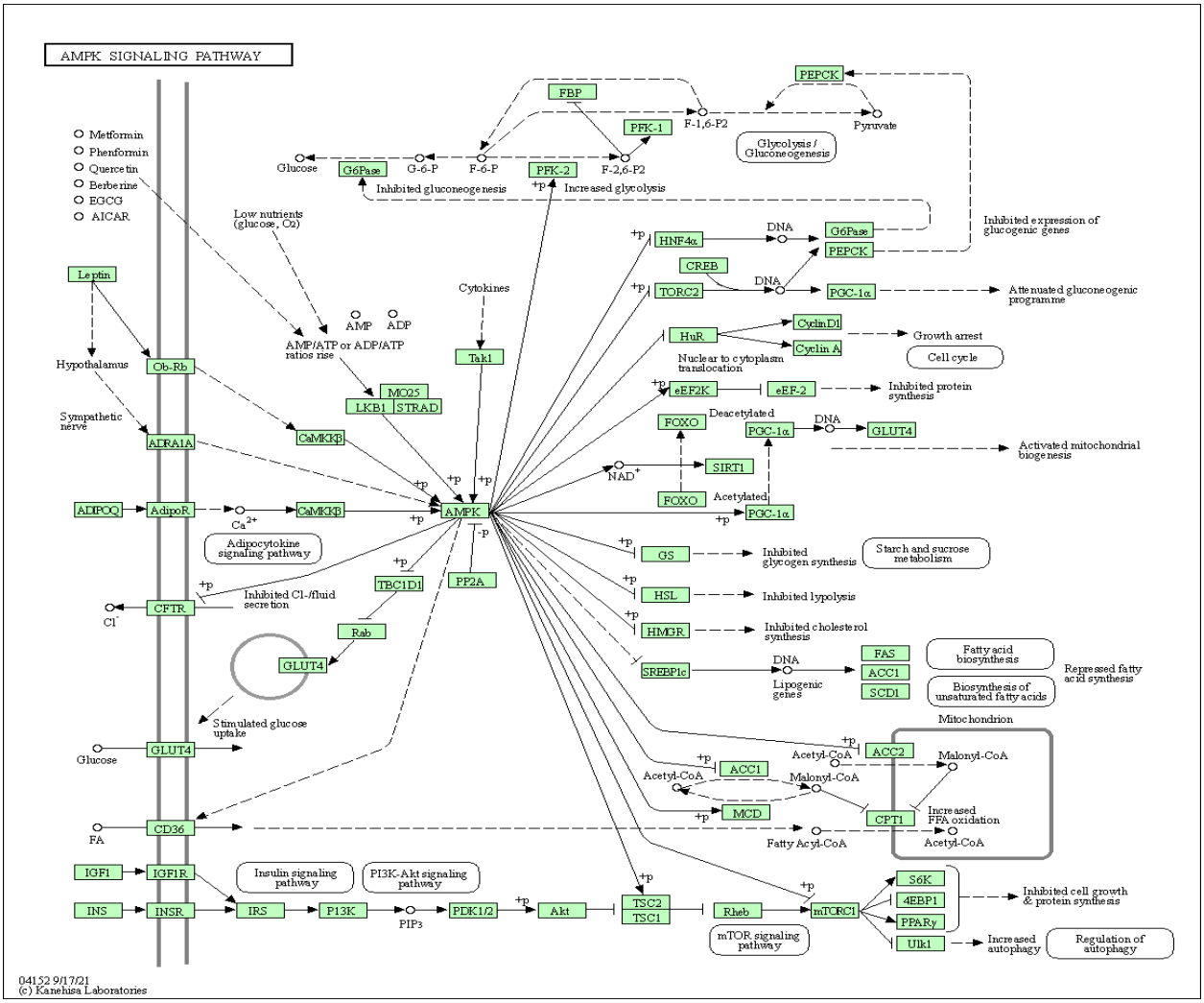

Figure 2- AMPK signaling pathway and candidate genes related to number of teats (KEGG database)

نتیجه‌گیری

با بررسی منابع انجام شده، پژوهش حاضر اولین مطالعه پویش کل ژنومی برپایه آنالیز مسیر مرتبط با صفت تعداد سرپستانک در گوسفند بوده است، لذا در تجزیه و تحلیل‌ها تلاش گردید از آزمون‌های پر کاربرد و سختگیرانه برای شناسایی مسیرهای زیستی معنی‌دار استفاده گردد. بررسی مناطق ژنومی شناسایی شده روی کروموزوم‌های ۳، ۵، ۷، ۸، ۱۲، ۱۴ و ۱۷ با استفاده از پایگاه داده نشان داد که این مناطق با صفت تعداد سرپستانک مرتبط می‌باشند. البته باید توجه کرد که یکی از محدودیت‌های اجرای مطالعات پویش ژنومی، کافی نبودن تعداد رکوردهای فنوتیپی و جمعیت غیر خویشاوند برای آنالیز پیوستگی می‌باشد که منجر به افزایش خطای نوع اول یا به عبارتی شناسایی نشانگرهای معنی‌دار مثبت کاذب می‌شود. با توجه به تعداد کم نمونه مورد استفاده در پژوهش حاضر نیز باید در استفاده از ژن‌های کاندیدای شناسایی شده در برنامه‌های اصلاحی با احتیاط عمل کرد. البته می‌توان با ادغام داده‌های تحقیقات مشابه جدید و استفاده از

آنالیزهای آماری جامع‌تر برای تأیید نتایج پژوهش حاضر استفاده کرد. از طرفی در این تحقیق به دلیل عدم دسترسی به رکوردهای فنوتیپی و اطلاعات ژنوتیپی مرتبط با صفت مورد مطالعه تحقیق حاضر در دام‌های بومی کشور، از اطلاعات فنوتیپی و ژنوتیپی کشور چین استفاده شد. لذا استفاده از نتایج این تحقیق در جمعیت‌های گوسفندان کشور نیاز به مطالعات بیشتر دارد تا در این جمعیت‌ها نیز تأیید شوند.

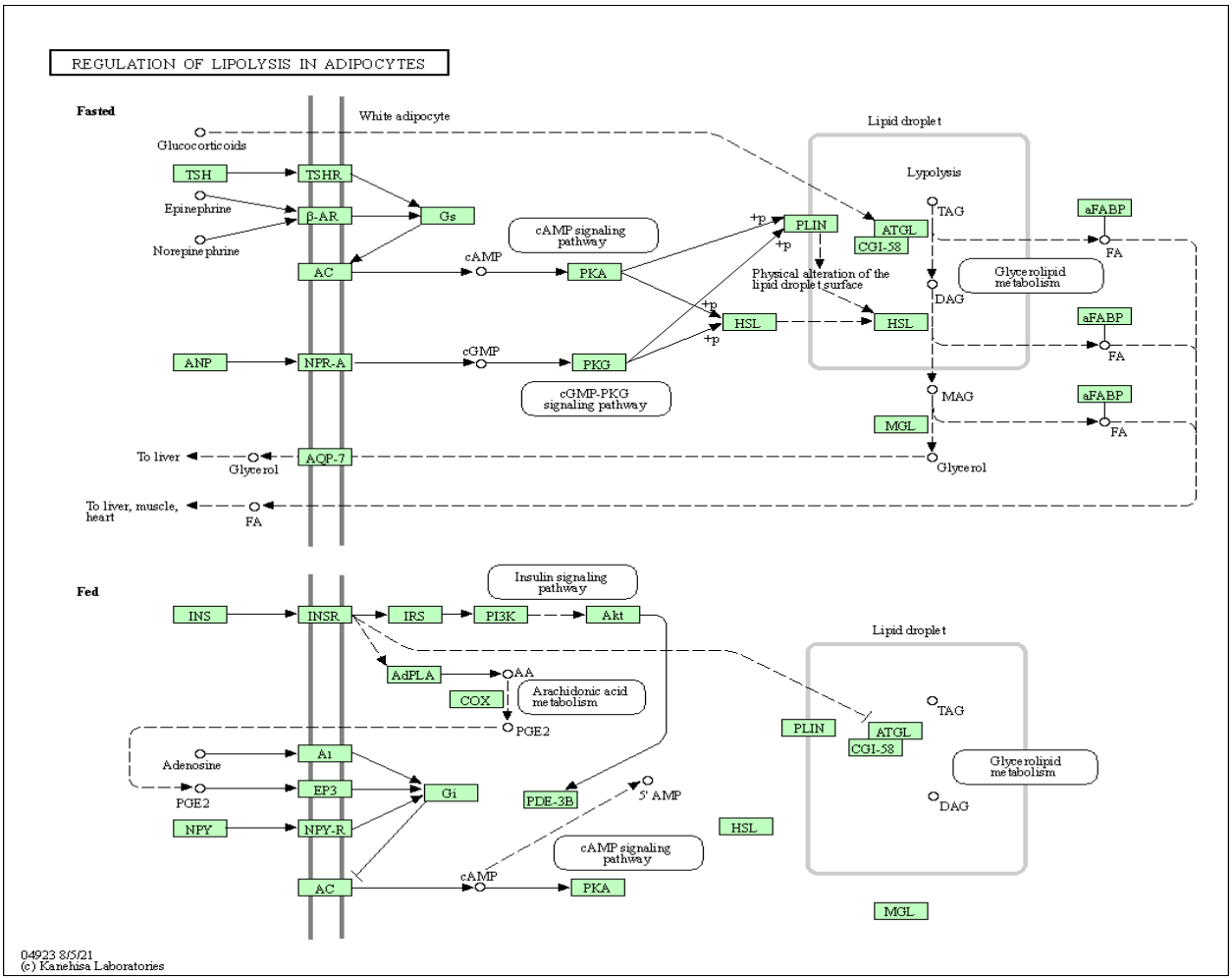

Figure 3- Regulation of lipolysis in adipocytes and candidate genes related to number of teats (KEGG database)

منابع مورد استفاده

Ahbara A, Bahbahani H, Almathen F, Al Abri M, Agoub MO, Abeba A, Kebede A, Musa HH, Mastrangelo S, Pilla F, Ciani E, Hanotte O and Mwacharo JM, 2019. Genome-Wide Variation, Candidate Regions and Genes Associated With Fat Deposition and Tail Morphology in Ethiopian Indigenous Sheep. Frontiers in Genetics 9:699.

Bakhtiarizadeh MR, Mirzaei S, Norouzi M, Sheybani N and Sadi MSV, 2020. Identification of Gene Modules and Hub Genes Involved in Mastitis Development Using a Systems Biology Approach. Frontiers in Genetics 11:722.

Bliss SP, Navratil AM, Xie J and Roberson MS, 2010. GnRH signaling, the gonadotrope and endocrine control of fertility. Frontiers in Neuroendocrinology 31: 322–340.

Chatterjee SJ, Halaoui R, Deagle RC, Rejon C and McCaffrey L, 2019. Numb regulates cell tension required for mammary duct elongation. Biology Open 8:42341.

Dysin AP, Barkova OY and Pozovnikova MV, 2021. The Role of microRNAs in the Mammary Gland Development, Health, and Function of Cattle, Goats, and Sheep. Noncoding RNA 4:78.

Esmaeili-Fard SM, Gholizadeh M, Hafezian SH and Abdollahi-Arpanahi R, .2021. Genome-wide association study and pathway analysis identify NTRK2 as a novel candidate gene for litter size in sheep. PLoS ONE 16(1): e0244408.

Gao X, Yao X, Li X, Liang Y, Liu Z, Wang Z, Li K, Li Y, Zhang G and Wang F, 2021. Roles of WNT6 in Sheep Endometrial Epithelial Cell Cycle Progression and Uterine Glands Organogenesis. Veterinary Sciences 12:316.

Honarvar M, Sadeghi M, Moradi-Shahrebabak H, Behzadi SH, 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.

Huang S, He Y and Ye S, 2018. Genome-wide association study on chicken carcass traits using sequence data imputed from SNP array. Journal of Applied Genetics 59: 335-344.

Khaltabadi Farahani AH, Mohammadi H and Moradi MH, 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 22(3): 325-335.

Kim SC, Jang HC, Lee SD, Jung HJ, Park JC, Lee SH, Kim TH and Choi BH, 2014. Changes in expression of insulin signaling pathway genes by dietary fat source in growing-finishing pigs. Journal of Animal Science and Technology 56:12.

Li Y, Pu L, Shi L, Gao H, Zhang P, Wang L and Zhao F, 2021. Revealing New Candidate Genes for Teat Number Relevant Traits in Duroc Pigs Using Genome-Wide Association Studies. Animals (Basel) 3:806.

Liu Z, Li H, Zhong Z, Jiang S, 2022. A Whole Genome Sequencing-Based Genome-Wide Association Study Reveals the Potential Associations of Teat Number in Qingping Pigs. Animals (Basel). 12(9): 1057.

Marete A, Lund MS, Boichard D and Ramayo-Caldas Y, 2018. A system-based analysis of the genetic determinism of udder conformation and health phenotypes across three French dairy cattle breeds. PLoS One 7:e0199931.

Marques DBD, Bastiaansen JWM, Broekhuijse MLWJ, Lopes MS, Knol EF, Harlizius B, Guimarães SEF, Silva FF and Lopes PS, 2018. Weighted single-step GWAS and gene network analysis reveal new candidate genes for semen traits in pigs. Genetics Selection Evolution 50(1): 40.

Mohammadi H, Rafat SA, Moradi Shahrbabak H, Shodja J and Moradi MH, 2020. Genome-wide association study and gene ontology for growth and wool characteristics in Zandi sheep. Journal of Livestock Science and Technologies 8(2): 45-55.

Nazar M, Lu X, Abdalla IM, Ullah N, Fan Y, Chen Z, Arbab AAI, Mao Y and Yang Z, 2021. Genome-Wide Association Study Candidate Genes on Mammary System-Related Teat-Shape Conformation Traits in Chinese Holstein Cattle. Genes (Basel). 12:2020.

Nusse R and Clevers H, 2017. Wnt/b-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. Cell 16: 985-999.

Peng WF, Xu SS, Ren X, Lv FH, Xie XL, Zhao YX, Zhang M, Shen ZQ, Ren YL, Gao L, Shen M, Kantanen J and Li MH, 2017. A genome-wide association study reveals candidate genes for the supernumerary nipple phenotype in sheep (Ovis aries). Animal Genetics 5: 570-579.

Peñagaricano F, Weigel KA, Rosa GJ, Khatib H, 2013. Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. Frontiers in Genetics 3:307-314.

Pértille F, Moreira GC and Zanella R, 2017. Genome-wide association study for performance traits in chickens using genotype by sequencing approach. Scientific Reports 7:41748.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR and Bender D, 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. The American Journal of Human Genetics 81: 559-575.

Salehian-Dehkordi H, Xu YX, Xu SS, Li X, Luo LY, Liu YJ, Wang DF, Cao YH, Shen M, Gao L, Chen ZH, Glessner JT, Lenstra JA, Esmailizadeh A, Li MH, Lv FH, 2021. Genome-Wide Detection of Copy Number

Variations and Their Association with Distinct Phenotypes in the World's Sheep. Frontiers in genetics 12: 670582.

Smołucha G, Gurgul A, Jasielczuk I, Kawęcka A and Miksza-Cybulska A, 2021. A genome-wide association study for prolificacy in three Polish sheep breeds. Journal of Applied Genetics 2: 323-326.

Srikanth K, Lee SH, Chung KY, Park JE, Jang GW, Park MR, Kim NY, Kim TH, Chai HH, Park WC and Lim D. 2020. A Gene-Set Enrichment and Protein-Protein Interaction Network-Based GWAS with Regulatory SNPs Identifies Candidate Genes and Pathways Associated with Carcass Traits in Hanwoo Cattle. Genes (Basel)11(3):316.

Tahir MS, Porto-Neto LR, Gondro C, Shittu OB, Wockner K, Tan AWL, Smith HR, Gouveia GC, Kour J and Fortes MRS, 2021. Meta-Analysis of Heifer Traits Identified Reproductive Pathways in Bos indicus Cattle. Genes (Basel) 5:768.

Wang XC, Maltecca R, Tal-Stein EL and Khatib H, 2008. Association of bovine fibroblast growth factor2 (FGF2) gene with milk fat and productive life: an example of the ability of the candidate pathway strategy to identify quantitative trait genes. Journal of Dairy Science 91: 2475-2480.

Xu SS, Gao L, Xie XL, Ren YL, Shen ZQ, Wang F, Shen M, Eyþórsdóttir E, Hallsson JH, Kiseleva T, Kantanen J and Li MH, 2018. Genome-Wide Association Analyses Highlight the Potential for Different Genetic Mechanisms for Litter Size Among Sheep Breeds. Frontiers in Genetics 9:118.

Yang B, Jiao B, Ge W, Zhang X, Wang S, Zhao H and Wang X, 2018. Transcriptome sequencing to detect the potential role of long non-coding RNAs in bovine mammary gland during the dry and lactation period. BMC Genomics 19: 1–14.

Zhao Y, Pu Y, Liang B, Bai T, Liu Y, Jiang L and Ma Y, 2022. A study using single-locus and multi-locus genome-wide association study to identify genes associated with teat number in Hu sheep. Animal Genetics 2: 203-211.

Zhou X and Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44:821.

Zhuang Z, Ding R, Peng L, Wu J, Ye Y, Zhou S, Wang X, Quan J, Zheng E, Cai G, Huang W, Yang J, Wu Z, 2020. Genome-wide association analyses identify known and novel loci for teat number in Duroc pigs using single-locus and multi-locus models. BMC Genomics 21(1): 344.

Genome-wide association study based on pathway analysis related to teat numbers in sheep

H Mohammadi[1*], A Javanmard[2], A Najafi[3] and M Shamsollahi[4]

Received: April 26, 2022 Accepted: October 1, 2022

[1]Assistant Professor, Department of Animal Sciences, Faculty of Agriculture and Environmental Science, Arak University.
[2]Assistant Professor, Department of Animal Science, Faculty of Agricultural Sciences, University of Tabriz., Tabriz, Iran.
[3]Assistant Professor, Department of Animal and Poultry Science, College of Aburaihan, University of Tehran, Tehran, Iran.
[4]Assistant Professor, Department of Animal Sciences, Faculty of Agriculture, University of Ilam. Ilam, Iran.
[*]Corresponding author: E-mail: H-mohammadi64@araku.ac.ir

پژوهش‌های علوم دامی Animal Science Research

Journal of Animal Science/vol.33 No.2/ 2023/pp 107-119
https://animalscience.tabrizu.ac.ir

OPEN ACCESS

© 2009 Copyright by Faculty of Agriculture, University of Tabriz, Tabriz, Iran
This is an open access article under the CC BY NC license (https://creativecommons.org/licenses/by-nc/2.0/)
DOI: 10.22034/AS.2022.51449.1661

Introduction: The number of teats is an important trait in relation to the maternal ability of high litter size sheeps, it is important to know the genetic function and controlling positions of this trait in the genomes of different species. High litter size domestic animals are potentially dependent on colostrum-derived immunoglobulins due to placental structure in the early hours of life. Therefore, the number of teat plays a critical role when there are more offspring born at birth than the number of teats. In some cases, particularly the number of the extra teat and even their position is very important. A method to identify new loci and confirm existing QTL is through genome-wide association studies (GWAS). QTL assisted selection and genomic regions affecting the production traits have been considered to increase the efficiency of selection and improve production performance. Genome wide association studies typically focus 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 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 genome wide association studies (GWAS) based on Gene-set enrichment analysis for identifying the loci associated with teat number trait using the high-density SNPs.
Materials and methods: Hu sheep, a descendant of Mongolian sheep, is a famous lambing breed in China. In ewes of the Hu sheep, TT (individuals with two normal teats) and MT (individuals with two normal teats and one or two supernumerary teats) account for 76–62% and 38–24% respectively. For MT ewes, the supernumerary teats are smaller than the normal teats, but some can produce milk. In the current study, we analyzed the complex genetic mechanism of differences in teat number using single-locus GWAS in a total of 160 Hu sheep. The gene set analysis basically 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. Genome wide association study was performed with birth weight and biometric traits using GEMMA

software. Using the biomaRt2 R package, the SNP were assigned to genes if they were within the genomic sequence of the gene or within a flanking region of 25 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 3 components: biological process, molecular function, and cellular component. The KEGG pathway database contains metabolic and regulatory pathways, representing the actual knowledge on 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. The gene enrichment analysis was performed with the *goseq* R package. In the next step, a bioinformatics analysis was implemented to identify the biological pathways performed in BioMart, Panther, DAVID and GeneCards databases.

Results and discussion: Gene set enrichment analysis has proven to be a great complement of genome-wide association analysis (Gambra et al., 2013; Abdalla et al., 2016). 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 (Morota et al., 2015, 2016). We had 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, so it is possible that some of the genes included in these terms are irrelevant for meat production. A better understanding of the biology underlying meat production specifically, and an advance in the annotation of the ovine genome, can provide new opportunities for predicting production using gene set information.

In this research, 7 SNP markers on chromosomes 3, 5, 7, 8, 12, 14 and 17 located in *CDH11*, *NUMB*, *FGF2*, *ESR1*, *LGR5*, *INSR* and *PTGS2* genes were identified. Some of the found genes, are consistent with some of the previous studies related to teat number traits. According to pathway analysis, 11 pathways from gene ontology and biological pathways were associated with teat number ($P<0.05$). Among these pathways, Blastocyst development, Mesenchymal cell development, Developmental growth involved in morphogenesis, Muscle cell differentiation and AMPK signaling pathway have important functions in development of mammary gland and activation of the AMPK signaling pathway. Finally, it is worth noting that our gene-set enrichment analysis was conducted using a panel of SNP obtained from a single marker regression GWAS, which relies on a simplified theory of the genomic background of traits, without considering the joint effect of SNP. Hence, other approaches (e.g., GWAS exploring SNP by SNP interactions) might provide a better basis for biological pathway analysis.

Conclusions: In total, this study supported previous results from GWAS of teat number, also revealed additional regions in the sheep genome associated with these economically important traits. Using these findings could potentially be useful for genetic selection in the breeding programs.

Keywords: Extra teat, Genome scan, Gene ontology, Pathway-based analysis.