مقاله علمی پژوهشی/نشریه پژوهش‌های علوم دامی/ جلد ۳۰ شماره ۱/ سال۱۳۹۹/ صفحات ۷۹ تا ۹۲

DOI: 10.22034/AS.2020.11009

مطالعه پویش کل ژنومی صفات مرتبط با تولید شیر بر پایه تجزیه و تحلیل غنی‌سازی مجموعه‌های ژنی در گاو هلشتاین

نوشین عزیز پور[۱]، امیر حسین خلت آبادی فراهانی[۲]*، محمد حسین مرادی[۲] و حسین محمّدی[۳]

تاریخ دریافت: ۹۸/۰۷/۱۶ تاریخ پذیرش: ۹۸/۰۸/۱۲

[۱]دانش آموخته کارشناسی ارشد گروه علوم دامی، دانشکده کشاورزی و منابع طبیعی، دانشگاه اراک

[۲]استادیار گروه علوم دامی، دانشکده کشاورزی و منابع طبیعی، دانشگاه اراک

[۳]دانش آموخته دکتری گروه علوم دامی، دانشکده کشاورزی، دانشگاه تبریز

*مسئول مکاتبه: Email: amfarahanikh@gmail.com

چکیده

زمینه مطالعاتی: اخیراً انتخاب به کمک QTLها و مناطق ژنومی مؤثر بر صفات تولیدی برای افزایش بازده انتخاب مورد توجه قرار گرفته است. **هدف:** پژوهش حاضر به منظور مطالعه پویش ژنومی بر مبنای تجزیه و تحلیل غنی‌سازی مجموعه ژنی جهت شناسایی ژن‌های مؤثر بر صفات مرتبط با تولید شیر گاو هلشتاین با استفاده از تکنیک توالی‌یابی GGRS بوده است. **روش کار:** بدین منظور، مطالعه پویش ژنومی از ۱۰۹۲ رأس گاو هلشتاین و رکورد مرتبط با تولید شیر، درصد چربی و امتیاز سلول‌های بدنی در برنامه Plink ارزیابی شد. آنالیز غنی‌سازی ژنی با استفاده از بسته نرم افزاری goseq برنامه R و شناسایی عملکرد بیولوژیکی ژن‌های نزدیک در مناطق انتخابی کاندیدا از طریق پایگاه‌های برخط GO، KEGG و Panther انجام گردید. **نتایج:** تجزیه و تحلیل بیوانفورماتیکی نشان داد که مناطق ژنومی شناسایی شده به طور مستقیم و غیر مستقیم با ژن‌های مؤثر بر امتیاز سلول‌های بدنی، تولید و چربی شیر همپوشانی دارند. در این پژوهش تعداد ۱۱ نشانگر تک نوکلئوتیدی واقع روی کروموزوم‌های ۵، ۶، ۷، ۸، ۱۴، ۱۹، ۲۲، ۲۴، ۲۵، ۲۷ و ۲۸ شناسایی شدند که با ژن‌های *ASIC2*، *ANXA3*، *CCL2*، *CCL11*، *CCL24*، *IL33*، *TLR3*، *WWOX*، *EGFR*، *PRKCA*، *CAMK2A*، *KCNMA1*، *FABP2*، *SPP1*، *THBS4*، *HSP90B1* و *ITPR1* مرتبط بودند. برخی از این ژن‌ها در مناطق معنی‌دار با مطالعه قبلی هم‌خوانی داشت. در آنالیز غنی سازی مجموعه ژنی تعداد ۲۵ مسیر هستی شناسی ژنی و مسیر KEGG با صفات تولید و امتیاز سلول‌های بدنی مرتبط بودند ($P<0.01$). از این بین، مسیرهای positive regulation of inflammatory response و defense response نقش مهمی در سیستم ایمنی داشتند. همچنین در ارتباط با تولید و چربی شیر مسیرهای PPAR signaling pathway، Oxytocin signaling pathway و Focal adhesion ارتباط معنی‌داری داشتند. **نتیجه گیری نهایی:** با توجه به تأیید مناطق قبلی پویش ژنومی صفات تولیدی، همچنین شناسایی مناطق ژنومی جدید استفاده از یافته‌های این تحقیق می‌تواند باعث تسریع در پیشرفت ژنتیکی برنامه‌های اصلاح نژادی گاو شود.

واژگان کلیدی: پویش ژنوم، تولید شیر، امتیاز سلول‌های بدنی، آنالیز غنی سازی، گاو

مقدمه

در طول چند سال گذشته اولین آزمایشات موفقیت آمیز از مطالعات پویش کل ژنومی (GWAS) در حیوانات اهلی از جمله گاو (دایت‌وایلر و همکاران ۲۰۰۸) گزارش شده‌اند. هدف نهایی از مطالعات پویش ژنومی که به منظور شناسایی وابستگی بین یک نشانگر SNP و یک صفت با استفاده از نشانگرهای با تراکم بالا در سطح ژنوم می‌باشد، پیدا کردن جهش‌های علّی یا مسبب بوده که بر فنوتیپ یک صفت اثر می‌گذراند (گودارد و هیز ۲۰۰۹ و محمدی و همکاران ۲۰۱۸). این اطلاعات می‌تواند برای انتخاب به کمک نشانگر مفید بوده و در افزایش بازده انتخاب صفات مورد مطالعه کمک نماید (بهلولی و همکاران ۲۰۱۳).

داده‌های مورد استفاده در این تحقیق اولین بار توسط چن و همکاران (۲۰۱۸) برای صفات تولید شیر، درصد چربی شیر و امتیاز سلول‌های بدنی مورد آنالیز قرار گرفته بودند که در آن تحقیق با استفاده مدل هاپلوتیپ و روش بیزی تجزیه شده و ۱۷ بلوک هاپلوتیپی معنی‌دار مرتبط با تولید شیر روی کروموزوم‌های ۱، ۲، ۵، ۱۰، ۱۴، ۱۷، ۱۹، ۲۴ و ۲۶ شناسایی شده بودند.

مطابقه یافته‌های مطالعه مدنظر و مقالات مختلف در آنالیزهای پویش ژنومی مبتنی بر مدل‌های خطی و غیر خطی، مشکل بالا بودن نرخ خطای نوع اول و بیش برآورد اثر نشانگرهای SNPها می‌باشد. به عبارت دیگر یکی از ایرادات تحقیقات مطالعات پویش ژنومی در نظر گرفتن آستانه معنی‌داری محافظه‌کارانه مانند تصحیح بنفرونی برای جلوگیری از بروز خطای نوع اول است. در حالیکه پرهیز از خطای نوع اول سبب افزایش خطای نوع دوم یعنی در نظر نگرفتن SNPهای دارای اثر معنی‌دار پایین‌تر از آستانه در نظر گرفته شده می‌شود (پنگ و همکاران ۲۰۱۰). در نتیجه یک جایگزین مناسب برای حل این مشکل، انجام مطالعات پویش کل ژنومی بر مبنای مسیر[1] و با استفاده از تجزیه و تحلیل مجموعه‌های ژنی است. در واقع در این روش به جای انجام تجزیه برای یک SNP یا یک ژن، ارتباط بین صفت مورد مطالعه و واریانت‌های ژنتیکی را در یک دسته یا گروه ژنی که به طور عملکردی با هم مرتبط هستند بررسی می‌کند. به عبارتی دیگر آنالیز پیوستگی بین یک مجموعه ژنی معنی‌دار زیستی با فنوتیپ مورد آزمون قرار می‌گیرد (وانگ و همکاران ۲۰۱۱). 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. برای اینکه بتوان تفسیر درستی از کنار هم قرار دادن این ژن‌ها حاصل شود، از مسیرهای زیستی به عنوان بستری معنی‌دار که عملکرد مجموع ژن‌ها در آنها یک یا چند هدف واحد را پیگیری می‌کند استفاده می‌شود (مونی و ویلموت ۲۰۱۵).

برای اولین بار پناگریکانو و همکاران (۲۰۱۳) نشان دادند که تجزیه و تحلیل پویش ژنومی بر مبنای مسیر 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 آنالیز می‌شوند و در نتیجه میزان خطای نوع اول و بیش برآوردها کاهش پیدا می‌کند. در تحقیق دیگر مطالعه پویش ژنومی بر مبنای مسیر با استفاده از آنالیز غنی‌سازی مجموعه‌های ژنی روی خصوصیات لخته‌شدگی شیر، تولید پنیر و استحکام دَلمه انجام شد. براساس این نتایج آنالیز براساس مسیر، منجر به شناسایی ۲۱ طبقات مختلف عملکردی هستی شناسی ژن و ۱۷ مسیر بیوشیمیایی KEGG معنی‌دار مرتبط با این صفات شد (دِدئوسیس و همکاران ۲۰۱۷).

همچنین در مطالعه‌ای که با استفاده از داده‌های حساسیت به ویروس لکومی از یک جمعیت گاو شیری هلشتاین انجام شده بود روش پویش ژنومی بر مبنای مسیر کارآیی بهتری برای یافتن مناطق ژنومی، درک بهتر مکانیسم بیماری و معماری ژنتیکی بیماری نسبت به آنالیز بهترین پیش بینی نااریب خطی ژنومی داشت (ابدلا و همکاران ۲۰۱۶).

[1] Pathway-based analysis

هدف از انجام پژوهش حاضر، شناسایی مناطق ژنومی و ژن‌های کاندیدای مرتبط با صفات تولید شیر، درصد چربی شیر و امتیاز سلولهای بدنی در یک جمعیت گاو نژاد هلشتاین براساس تجزیه بر مبنای مسیر و با استفاده از روش غنی‌سازی مجموعه ژنی می‌باشد. شناسایی این جایگاه‌ها از دیدگاه علمی و اقتصادی می‌تواند دارای اهمیت زیادی باشد.

مواد و روش‌ها

در این پژوهش از اطلاعات صفات تولیدی مرتبط با شیر و امتیاز سلول‌های بدنی گاوهای هلشتاین با هدف پویش کل ژنومی که براساس مدل خطی هاپلوتیپ بود، استفاده گردید (چن و همکاران ۲۰۱۸). اطلاعات شامل تعداد ۱۰۹۲ رأس گاو که در طول سالیان ۲۰۰۱ تا ۲۰۱۲ برای تولید شیر انتخاب شده بودند، شامل رکوردهای فنوتیپی صفات تولید شیر، درصد چربی شیر و امتیاز سلول‌های بدنی در شکم اول تا سوم زایش بودند. DNA نمونه‌ها با استفاده از تکنولوژی GGRS[1] توالی یابی شده بودند.-SNPهایی که از تمام مراحل کنترل کیفیت (نشانگرهای با حداقل فراوانی آللی بالاتر از ۰/۰۵ و میزان فراخوانی آللی بالاتر از ۰/۹۰) عبور کردند، برای آنالیزهای بعدی نگه داشته شدند البته این داده‌ها دارای ژنوتیپ‌های از دست رفته بودند، در نتیجه ژنوتیپ‌های از دست رفته با استفاده از نرم افزار iBLUP ایمپیوت شدند (یانگ و همکاران ۲۰۱۴). در نهایت بعد از کنترل کیفیت و ایمپیوتیشن تعداد ۱۰۹۲ رأس گاو و تعداد ۱۶۴۳۱۲ SNP برای آنالیزهای مطالعه پویش کل ژنومی بر پایه آنالیز غنی‌سازی مجموعه ژنی باقی ماندند. جهت ارتباط فنوتیپ‌ها با ژنوتیپ‌ها از نرم افزار Plink (پورسل و همکاران ۲۰۰۷) استفاده شد.

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 به ژن ۲) ارتباط ژن‌ها به طبقات عملکردی و مسیرهای بیوشیمیایی ۳) پویش کل ژنومی بر پایه آنالیز مسیر

۱-**تعیین مکان SNPها به ژن‌ها**: SNPهایی که مقدار *P*-value آنها کمتر و یا مساوی ۰/۰۵ بود با استفاده از بسته نرم افزاری *biomaRt2* (دیورینسک و همکاران، ۲۰۰۹) در محیط نرم افزار R و با استفاده از رفرانس ژنومی گاو نسخه (*ARS-UCD1.2*) به ژن‌هایی که نشانگر SNP مورد نظر در داخل آن ژن و یا ۲۰kb بالادست یا پایین دست آن ژن قرار داشت، ارتباط داده شدند. در این مرحله ژنی که حداقل حاوی یک SNP معنی‌دار در سطح ۰/۰۵ باشد، ژن معنی‌دار به شمار می‌آید.

۲- **ارتباط ژن‌ها به طبقات عملکردی و مسیرهای بیوشیمیایی**: پایگاه‌های داده‌های هستی شناسی ژن (GO) (اشبورنر و همکاران ۲۰۰۰) و مسیرهای بیوشیمیایی KEGG (کانیسیو و گوتو ۲۰۰۰) جهت تعیین طبقات عملکردی و مسیرهای متابولیکی و تنظیمی ژن‌های معنی‌دار مورد استفاده قرار گرفتند. در این مرحله فرض بر اینست که ژن‌هایی که در یک طبقه عملکردی قرار می‌گیرند می‌توانند به عنوان یک گروه از ژن‌هایی که برخی از ویژگی‌های خاص و مشترک دارند مانند شرکت در ۳ فرآیند هستی شناسی شامل فرآیندهای زیستی، عملکرد مولکولی و اجزای سلولی در نظر گرفته شوند.

۳- **پویش کل ژنومی بر پایه آنالیز مسیر**: ارتباط‌های معنی‌دار مسیرهای عملکردی با صفات مرتبط با تولید شیر، درصد چربی شیر و امتیاز سلول‌های بدنی با استفاده از توزیع فوق هندسی[2] و آزمون Fisher's exact test مورد آزمون قرار گرفت. P-value مسیرهای عملکردی که تعداد k ژن معنی‌دار در آن قرار دارد با فرمول زیر محاسبه شد:

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

در این فرمول، k برابر با تعداد ژن‌های معنی‌دار در طبقه عملکردی، S برابر با تعداد کل ژن‌های معنی‌دار مرتبط با

[1] Genotyping by genome reducing and sequencing

[2] Hypergeometric

صفات 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 (دِدئوسیس و همکاران، ۲۰۱۷). تجزیه غنی‌سازی مجموعه ژنی با استفاده از بسته نرم افزاری *goseq* (یانگ و همکاران ۲۰۱۰) در محیط نرم افزار R انجام گردید. بسته *goseq* از آزمون فوق هندسی استفاده می‌کند. همچنین برای اختلاف طول ژن‌ها نیز تصحیح انجام می‌شود.

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. در نهایت آنالیزهای عملکردی ژن‌های کاندیدا از طریق آنالیز هستی شناسی و مسیرهای زیستی ژن‌ها (KEGG) با استفاده از نرم افزار آنلاین Panther (http://pantherdb.org/) انجام شد.

نتایج و بحث

در این پژوهش مطالعه پویش کل ژنومی با تجزیه و تحلیل غنی سازی و مجموعه ژنی جهت شناسایی طبقات عملکردی و ساز و کارهای مولکولی مرتبط با تولید و درصد چربی شیر و امتیاز سلول‌های بدنی در گاو نژاد هلشتاین انجام گردید. پلات‌های منهتن مرتبط با تولید و درصد چربی شیر و امتیاز سلول‌های بدنی به ترتیب در شکل‌های ۱ تا ۳ ارائه شده است.

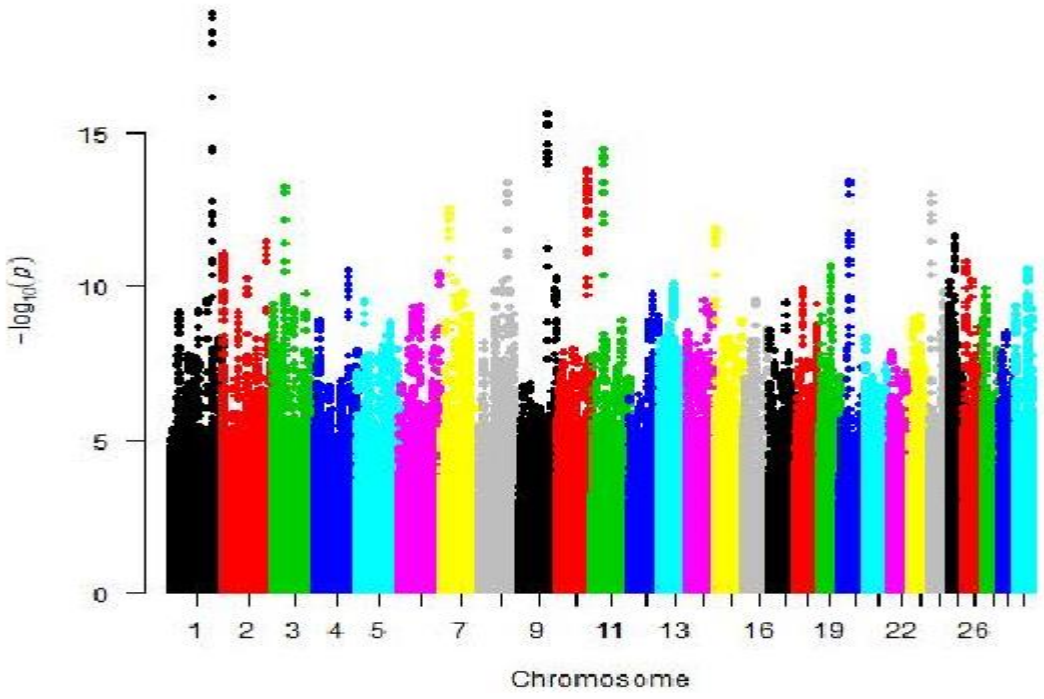

Figure 1- Manhattan plot for milk yield in Holsteins cattle. X axis is the SNPs positions on chromosomes, Y axis is the -Log10 P-value

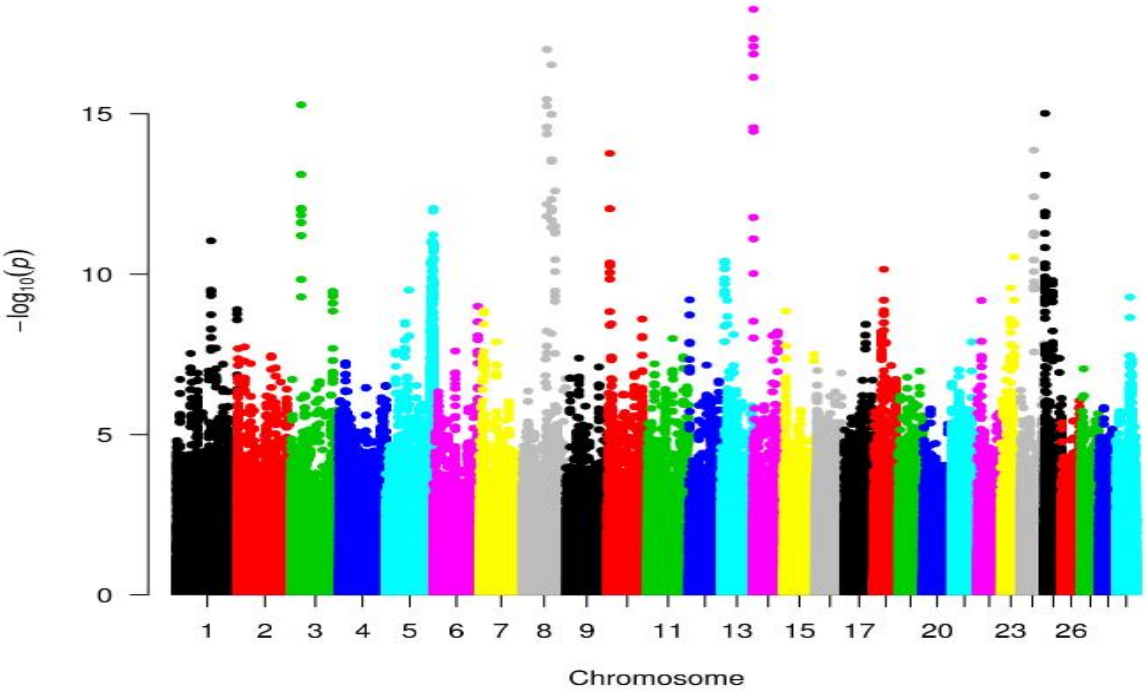

Figure 2- Manhattan plot for fat percentage in Holsteins cattle. X axis is the SNPs positions on chromosomes, Y axis is the -Log10 P-value

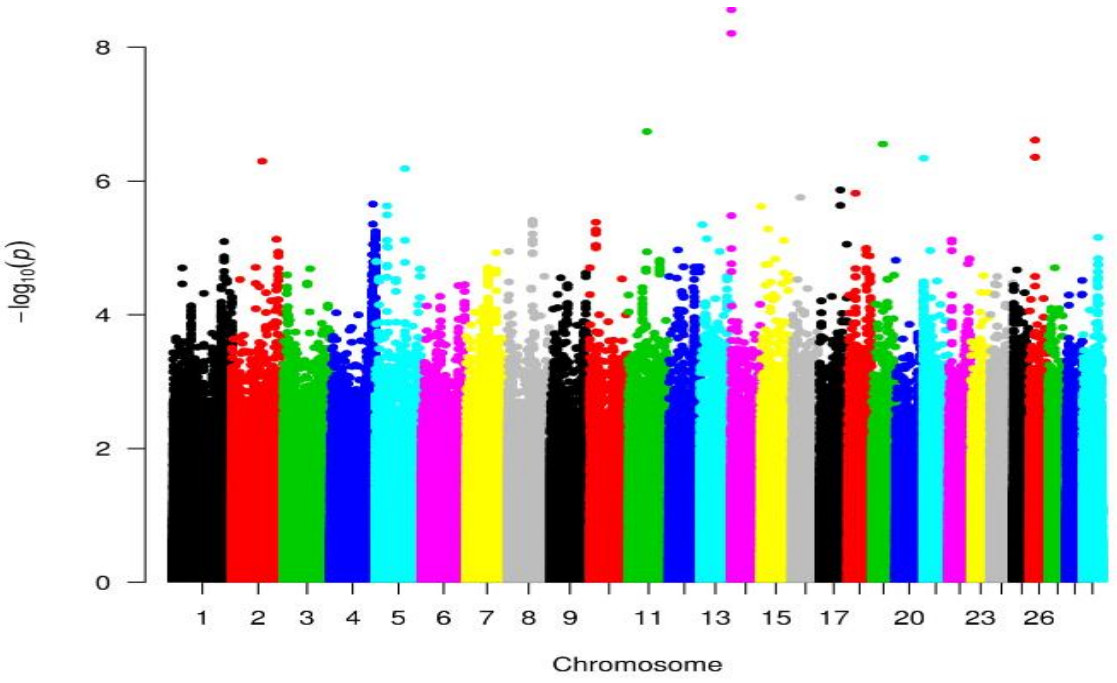

Figure 3- Manhattan plot for somatic cell score in Holsteins cattle. X axis is the SNPs positions on chromosomes, Y axisis the -Log10 P-value

شناسایی ژن‌های کاندیدا در مناطق ژنومی با استفاده از تجزیه GSEA-SNP[1]

تعداد ۱۹۷۲۳ ژن از ۲۷۶۰۷ ژن حاشیه نویسی شده در گاو به وسیله نشانگرهای SNP پوشش داده شد که در این میان تعداد ۱۶۴۵ ژن دارای اثر معنی‌داری 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 kb ۲۰ قرار گرفت (جدول ۱). این ژن‌ها به عنوان ژن‌های معنی‌دار مرتبط با صفات تولید و چربی شیر و امتیاز سلول‌های بدنی برای تجزیه GSEA-SNP انتخاب شدند و باقیمانده ژن‌ها به عنوان ژن‌های پس زمینه‌ای مورد استفاده قرار گرفتند.

Table 1- Number of significant SNP identified from genome-wide association studies (GWAS) and genes mapped by related trait

Trait	No. of significant SNP	No. of significant SNP assigned to genes	No. of significant mapped genes
Milk yield	1645	567	401
Fat percentage	1589	511	436
Somati cell score	1242	487	369

تعداد مجموعه‌های ژنی حاصل از پایگاه‌های داده‌ای مختلف شامل ۲۷۱۴ طبقات هستی شناسی، ۱۶۴ مسیر بیوشیمیایی KEGG و ۱۱۴ مسیر Reactome بود. همانطور که در جدول ۲ مشاهده می‌شود تعداد ۱۹ طبقات عملکردی در هستی شناسی فرآیندهای زیستی، عملکرد مولکولی، اجزای سلولی و مسیرهای KEGG با صفات تولید و چربی شیر و امتیاز سلول‌های بدنی دارای ارتباط هستند (P<0.01). مسیرهای که بیشتر از ۳ ژن و کمتر از ۵۰۰ ژن داشتند گزارش شده‌اند.

با بهبود ژنتیکی گاوهای شیری تولید شیر افزایش یافته و افزایش تولید موجب حساسیت دام به برخی بیماری‌ها نظیر ورم پستان شده است. ورم پستان یک بیماری تورمی غدد پستانی است که به وسیله عفونت داخل پستان ایجاد می‌شود. این بیماری به وسیله طیف وسیعی از پاتوژن‌های باکتریایی و قارچی به وجود می‌آید که سبب فعال شدن سیستم ایمنی و دفاعی در پستان و افزایش سلول‌های سوماتیک در شیر می‌شود. کاهش تولید و کیفیت شیر از اثرات سوء این بیماری است. افزایش مقاومت ژنتیکی به این بیماری که پر هزینه‌ترین بیماری در پرورش گاو شیری است، می‌تواند در برنامه های اصلاح نژادی مد نظر قرار گیرد (سوردیلو و استرایکر ۲۰۰۸). از مسیرهای اصلی معنی‌دار مرتبط با امتیاز سلول‌های بدنی positive regulation of inflammatory response بدست آمد که جزء فرآیندهای زیستی است. اینترلوکین‌ها جزوی از خانواده سایتوکین-ها[2]می‌باشند. سایتوکین‌ها خانواده بزرگی از مولکول‌های پروتئینی قابل حل در آب هستند که نقش تنظیمی در پاسخ به التهاب دارند. به طوریکه هم در ایمنی ذاتی و هم در ایمنی اکتسابی نقش دارند. ژن IL33 هم جزو گروه سایتوکین‌ها (IL1a, IL1ß, IL4R, IL5, IL7, IL15, IL23A, IFN-c) نقش کلیدی در پاسخ‌های ایمنی دارند (Genecards).

[1] Gene set enrichment analysis- SNP

[2] Cytokines

Table 2- Gene set enrichment analysis significantly (P < 0.01) associated with milk yield, fat percentage, and somatic cell score

Category	Term	No. of genes	genes[1]	P value	FDR
GO_BP	GO:0007606:sensory perception of chemical stimulus	3	*OR52N1*, ***ASIC2***, *ASIC1*	7.28E-10	6.51E-07
	GO:0007600:sensory perception	7	***KCNIP4***, *COL11A2*, *OR52N1*, *ASIC2*, *COL11A2*, *ASIC1*, *LTBP3*	1.72E-09	1.02E-06
	GO:0050907:detection of chemical stimulus involved in sensory perception	5	*OR52N1*, *ASIC1*, *LTBP3*, ***SLC22A16, SLC25A31***	2.77E-11	4.96E-08
	GO:0098609:cell-cell adhesion	18	***ANXA3***, *PKP3*, *CEL*, *TENM2*, *CDH12*, *ANXA6*, *CDH1*, *CDH2*, *NPTN*, *PKP4*, *SDK1*, *PALLD*, *CD274*, *LRRC7*, *TNFRSF21*, *CNTN6*, *SDK2*	1.04E-04	1.69E-02
	GO:0050729 :positive regulation of inflammatory response	7	***CCL2, CCL11, CCL24,*** *ETS1*, ***IL33***, *PDE2A*, ***TLR3***	4.1E-2	9.9E-1
	GO:0030593:neutrophil chemotaxis	8	***CCL2, CCL11, CCL24***, *FCER1G*, *CSF3R*, *PDE4D*, *SYK*, *VAV3*	3.1E-2	1.62E-02
	GO:0006952:defense response	7	*FCER1G*, *WASF1*, *CSF3R*, ***IL33,*** ***IL6,*** *PDE4D*, *SYK*,	2.1E-010	1.63E-06
GO_CC	GO:0030055:cell-substrate junction	15	*ADAM17*, *TNS4*, *SYNPO2*, *LIMD1*, *FERMT2*, *WASF1*, *TNS4*, ***WWOX***, *CLASP1*, *GSN*, *SYNE2*, *TLN2*, *EPB41L5*, *AFAP1*, *LMLN*, *NUP214*	1.78E-04	1.22E-02
GO_MF	GO:0016597:amino acid binding	12	*GRIA1*, *ANXA3*, ***HOMER1***, *ANXA4*, *GRIK3*, *GRIN2D*, *CACNG4*, *GRM4*, *GRIN3A*, *DAGLA*, *HOMER3*, *GRID1*	1.70E-04	2.14E-02
	GO:0008066:glutamate receptor activity	12	*GRIA1*, *ANXA3*, *HOMER1*, *ANXA4*, *GRIK3*, *GRIN2D*, *CACNG4*, *GRM4*, *GRIN3A*, *DAGLA*, *HOMER3*, *GRID1*	2.96E-06	1.50E-03
	GO:0016595:glutamate binding	12	*GRIA1*, *ANXA3*, *HOMER1*, *ANXA4*, *GRIK3*, *GRIN2D*, *CACNG4*, *GRM4*, *GRIN3A*, *DAGLA*, *HOMER3*, *GRID1*	2.96E-06	7.48E-04
KEGG pathway	bta04921:oxytocin signaling pathway	20	*ADCY3*, *ADCY5*, *ADCY8*, *ADCY9*, *CACNA2D2*, *CACNA2D3*, *CACNG1*, *CACNA1D*, *CAMK2A*, *CAMK4*, ***EGFR***, *GUCY1A2*, *ITPR1*, *MAP2K5*, *PLA2G4D*, *PRKAA1*, ***PRKCA***, *PRKCB*, *RYR2*, *RYR3*	3.7E-4	2.3E-2
	bta00561:glycerolipid metabolism	11	*CEL*, *DGKB*, *DGKE*, *DGKG*, *GK2*, ***LPIN2***, *MBOAT1*, *PNLIPRP3*, *PNPLA3*, *LOC786474*, *PLPP1*	2.2E-3	8.8E-2
	bta04911:insulin secretion	12	*ATF6B*, *ADCY3*, *ADCY5*, *ADCY8*, *ADCY9*, *CACNA1D*, ***CAMK2A***, *GLP1R*, ***KCNMA1***, *PRKCA*, *PRKCB*, *RYR2*	5.2E-3	1.0E-1
	bta00564:glycerophospholipid metabolism	13	*CDS2*, *CHAT*, *CHKB*, *CHPT1*, *DGKB*, *DGKE*, *DGKG*, *GPD1*, *LOC613966*, ***LPIN2***, *MBOAT1*, *PLA2G4D*, *PLPP1*	6.9E-3	1.2E-1
	bta04918:thyroid hormone synthesis	9	*ATF6B*, *ADCY3*, *ADCY5*, *ADCY8*, *ADCY9*, ***HSP90B1***, *ITPR1*, *PRKCA*, *PRKCB*	3.5E-2	2.6E-1
	bta03320:PPAR signaling pathway	9	*ACSL5*, *APOA2*, *CPT1B*, ***FABP2***, *GK2*, *RXRA*, *RXRB*, *RXRG*, *SLC27A2*	3.8E-2	2.7E-1
	bta04510:focal adhesion	19	*BRAF*, *FYN*, *ARHGAP5*, *SHC3*, *BIRC3*, *COMP*, *COL4A2*,	3.9E-2	2.7E-1

			*COL11A2, EGFR, FLT1, PAK5, PARVA, **ITPR2**, PIK3CD, PRKCA, PRKCB, **SPP1**, TLN2, **THBS4**, VAV3*		
bta04915:estrogen pathway	signaling	11	*SHC3, ATF6B, ADCY3, ADCY5, ADCY8, ADCY9, EGFR, **HSP90B1**, ITPR1, OPRM1, PIK3CD*	4.3E-2	2.8E-1
bta04912:GnRH pathway	signaling	10	*ADCY3, ADCY5, ADCY8, ADCY9, CACNA1D, CAMK2A, EGFR, **ITPR1**, MMP14, PLA2G4D*	4.7E-2	2.9E-1

[1]The most important gene(s) related to traits is labeled bold.

از مسیرهای KEGG معنی‌دار مرتبط با شمارش سلول-های بدنی می‌توان به مسیر GnRH signaling pathway اشاره نمود. از بین ۱۰ ژن معنی‌دار در این مسیر ژن کاندیدای ITPR1 در مطالعات قبلی در ارتباط با مقاومت و یا حساسیت به بیماری‌های مختلف ارتباط معنی‌داری گزارش شده است بطوریکه در مطالعه کیسر و همکاران (۲۰۱۶) ارتباط معنی‌داری بین ژن ITPR1 با ایمنی ذاتی با بیماری مایوباکتریوم آویوم[1]زیر گروه پاراتوبرکلوزیس در نژاد گاو شیری جرسی را گزارش نمودند.

کموکاین‌ها خانواده‌ای از سایتوکین‌های کموتراکتنت ۱ را تشکیل می‌دهند که همراه با سایتوکین‌ها و پروتئازها برای مهاجرت مستقیم لوکوسیت‌ها در طول پروسه التهابی ضروری هستند (Genecards). کموکاین‌ها نقش مهمی در فراخوانی مونوسیت‌ها، نوتروفیل‌ها و لنفوسیت‌ها دارند (UniProtKB).

از ژن‌های کاندیدای در زمینه مقاومت می‌توان ژن‌های TLR (گیرنده های پروتئین شبه T) را نام برد که خانواده-ای از گیرنده‌های سلولی را کد می‌کند. تا کنون ۱۳ نوع ژن TLR در پستانداران شناسایی شده است. این گیرنده-ها توالی مولکولی خاصی از عوامل بیماریزا (PAMP) را تشخیص می‌دهند و در آغاز پاسخ‌های ایمنی بدن به پاتوژن‌های میکروبی نقش محوری دارند (Genecards).

شاید بتوان مسیر cell-substrate junction که جزء هستی شناسی اجزای سلولی است را یکی از مهمترین مسیرهای مؤثر بر فرآیند سیستم ایمنی در ارتباط با امتیاز سلول های بدنی دانست. در این مسیر پروتئین‌های واسطه جهت ارتباط دادن اسکلت سلولی یک سلول به اسکلت سلولی سلول مجاور یا یک پروتئین در ماتریکس خارج سلولی از سطح غشای پلاسمایی گسترش می‌یابند. این مسیر دارای چندین chiled Term است که همه دارای عملکردهای مهمی در فرآیندهای ارتباط و اتصال سلول به هم و یا به بافت همبند پایه هستند. مسیر Gap junction یکی از این مسیرهاست. از بین ژن‌های معنی‌دار در این مسیر مرتبط با سیستم ایمنی می‌توان به ژن کاندیدای WWOX را نام برد. بیان ژن WWOX در انسان سرکوب کننده انواع مختلفی از سرطان شامل سینه، پروستات، شش، معده و پانکراس می‌باشد (UniProtKB).

نشان داده شده است که یکی از عوامل مؤثر در بروز مقاومت به بیماری‌ها ژن گیرنده فاکتور رشد اپیدرمال (EGFR) روی کروموزوم ۲۲ گاو است. این خانواده ژنی جزء گروهی از گیرنده‌های داخل سلولی به نام گیرنده-های فعال کننده پراکسیزومی می‌باشد و در تنظیم متابولیسم گلوکز، چربی و همچنین در تمایز سلولی نقش دارند. ارتباط معنی‌داری بین ژن کاندیدای EGFR با مقاومت به انگل‌های داخلی از طریق ایمنی اکتسابی و ذاتی[2]در گاوهای نژاد سیاه آلمانی گزارش شده است (می و همکاران ۲۰۱۹).

زمانیکه موجود با تنش گرمایی مواجه می‌شود، سلول-های آسیب دیده پروتئین‌هایی تولید می‌کنند که از آنها در مقابل آسیب محافظت می‌کند. یکی از پاسخ‌های حمایتی تولید پروتئین‌های شوک حرارتی (HSP) می‌باشد این خانواده ژنی از ۶ زیر گروه به نام‌های HSP40، HSP60، HSP70، HSP90، HSP100 و HSPB تشکیل شده‌اند. یک ژن از همین خانواده HSP90B1 به طور مستقیم از آپوپتوسیز (مرگ و میر سلولی برنامه ریزی شده) جلوگیری می‌کند.

[1] Mycobacterium avium

[2] Host defenses

ژن پروتئین شبه T فرم ۴ (TLR4) و گیرنده‌های پروتئین شبه T فرم ۳ (TLR3) در سطح سلول می‌باشند. جزو ژن‌های کاندیدای مؤثر بر صفات مرتبط با ایمنی سلولی و ذاتی در پاسخ‌های دفاعی در پستانداران است (UniProtKB).

TLRها، خانواده‌ای از گیرنده‌های شناسایی الگو هستند که در سیر تکامل، بدون تغییر مانده و حفظ شده‌اند و در بسیاری از انواع سلول‌ها بیان می‌شوند و نقش اساسی را در پاسخ‌های ایمنی ذاتی بر ضد میکروب‌ها ایفا می‌کنند.

همانطور که مشاهده می شود مسیر defense response جزء معنی‌دار ترین مسیر زیستی مرتبط با امتیاز سلول‌های بدنی در این تحقیق است که نشان دهنده نقش سیستم ایمنی در این فرآیند است. عمده ژن‌های موجود در این مسیر مربوط به انواع اینترفرون (آلفا، بتا، کاپا و امگا) پروتئین‌های دیفنسین بتا و اینترلوکین‌ها است. ژن‌های اینترلوکین معنی‌دار موجود در این مسیر شامل IL-6، IL33 و IL-5 بودند. نقش اینترلوکین‌ها در امتیاز سلول‌های بدنی بارها مورد تأکید قرار گرفته است (اُگروسی و همکاران ۲۰۰۹)

از مسیرهای معنی‌دار مرتبط با بیماری مختلف و امتیاز سلول‌های بدنی می‌توان به sensory perception of chemical stimulus و Cell-cell adhesion اشاره نمود که ارتباط معنی‌داری بین ژن کاندیدای ANXA3 با امتیاز سلول‌های بدنی گزارش شده است (چن و همکاران ۲۰۱۵).

از مسیرهای مهم و معنی‌دار مرتبط با تولید شیر را می‌توان به مسیر بیولوژیکی Oxytocin signaling pathway اشاره نمود که از بین ۲۰ ژن معنی‌دار، ژن کاندیدای PRKCA در مطالعات قبلی ارتباط معنی‌داری با تولید شیر در گاوهای نژاد شیری هلشتاین گزارش شده است (رینکان و همکاران ۲۰۰۹).

مسیر Focal adhesion با تعداد ۱۹ژن معنی‌دار از دیگر مسیرهایی بود که در ارتباط با تولید شیر شناسایی شد. این مسیر در مسیر KEGG طبقه بندی می‌شود. در مطالعه‌ی قبلی پویش ژنومی صفات مرتبط با تولید شیر که براساس مدل هاپلوتیپ روش بیزی انجام شده بود در ارتباط با تولید شیر روی کروموزوم ۱۰ یک ناحیه ۹ kb گزارش شده بود که با منطقه شناسایی شده در پژوهش حاضر همخوانی داشت. همچنین براساس آنالیز غنی‌سازی مجموعه ژنی ناحیه معنی‌دار مؤثر روی تولید شیر در ناحیه ۷۵ مگابازی بدست آمد که ژن کاندیدای ITPR2 با مطالعه قبلی مشترک بود (چن و همکاران ۲۰۱۸).

از ژن‌های معنی‌دار دیگر در مسیر Focal adhesion ژن کاندیدای SPP1 بود. ژن SPP1 که نام دیگر آن استئوپتین (OPN) می‌باشد و در فسفریله کردن گلیکوپروتئین‌ها و تشکیل استخوان در انسان و موش از طریق تفرق سلولهای استئوبلاست نقش کلیدی دارد (Gencards). ژن SPP1 ارتباط معنی‌داری با صفات مرتبط با تولید شیر در گاوهای فلخویه جمهوری چک گزارش شده است (بولکووآ و همکاران ۲۰۱۲). همچنین در مطالعه کیولاج و همکاران (۲۰۱۹) ارتباط معنی‌داری بین چند شکلی در ژن SPP1 با صفات مرتبط با تولید شیر در گاوهای هلشتاین فریزین گزارش نمودند.

از دیگر مسیرهای بیولوژیکی معنی‌دار مرتبط با تولید شیر می توان به مسیر sensory perception اشاره نمود که جزء مسیرهای فرآیندهای زیستی مرتبط با تولید می‌باشد که از بین ۷ ژن معنی‌دار، از این بین، ژن کاندیدای KCNIP4 دارای نقش بیولوژیکی مستقیمی با صفات مرتبط با تولید داشت که بیشتر بحث خواهد شد. ژن KCNIP4 جزو خانواده ژنی پروتئین‌های کانال‌های دریچه‌دار پتاسیمی وابسته به ولتاژ و پروتئین‌های اتصالی کلسیمی است (Genecards). کانال‌های دریچه‌دار پتاسیمی دارای نقش گسترده عملکردی شامل تنظیم مثبت نوروترانسمیتر، انقباض عضلات صاف، تصحیح ضربان قلب و ترشح انسولین می‌باشند (UniProtKB). بنابراین فرض می‌شود ژن KCNIP4 می‌تواند بر صفات تولید شیر تأثیر داشته باشد. اخیراً در گاو شیری حاوی ژن KCNIP4 روی کروموزوم ۶ ارتباط معنی‌داری با ترکیبات شیر گزارش شده است (ژانگ و همکاران ۲۰۱۹).

مسیر دیگر معنی‌دار مرتبط با تولید شیر را می‌توان Glycerolipid metabolism نام برد که از بین ۱۱ ژن، ژن کاندیدای LPIN2 در مطالعات قبلی ارتباط معنی‌داری با

صفات مرتبط با تولید شیر در نژادهای مختلف گوسفندان شیری گزارش شده است (ژرلندو و همکاران ۲۰۱۹).
مسیر Insulin secretion که جزء مسیر KEGG است از دیگر مسیرهای معنی‌دار در این پژوهش بود که تعداد ۱۲ ژن از ۱۱۳ ژن این مسیر معنی‌دار بود. این پروتئین‌ها در یکپارچگی ساختاری سلول‌ها، اتصال سلول به سلول، تمایز و تکثیر سلولی نقش دارند (Gencards). ژن کاندیدای KCNMA1 در بین ژن‌های این مجموعه ژنی قرار داشت و محصول این ژن، پروتئینی به همین نام است که یک کانال پتاسیمی است و ویژگی بارز آن انتقال مقادیر زیاد پتاسیم از غشای سلولی است. این کانال با تغییر پتانسیل الکتریکی غشاء و یا با افزایش غلظت درون سلولی کلسیم باز شده و پتاسیم به درون سلول وارد می‌شود و سبب برقراری تعادل الکتریکی در دو طرف غشاء می‌شود (GeneCards). این کانال در بافت‌های زیادی بیان شده است و بیشترین فراوانی آن در سلول-های مغز و اندام‌های دارای ماهیچه صاف است. همچنین در بافت‌های دیگری شامل پستان، پانکراس و غدد آدرنال نیز وجود دارد. با توجه به نقش گسترده کانال‌های پتاسیمی در بافت‌ها KCNMA1 می‌تواند با عملکردهای متفاوتی فرآیند تولید شیر و ترکیبات مرتبط با آن (چربی و پروتئین) را تحت تأثیر قرار دهد. در مطالعه مَرتی و همکاران (۲۰۱۸) در مطالعه پویش کل ژنومی مرتبط با مورفولوژی پستان ارتباط معنی‌داری بین ژن کاندیدای KCNMA1 در سه نژاد گوسفند شیری فرانسوی گزارش کردند. همچنین از بین ژن‌های کاندیدا در این مسیر مرتبط با حساسیت به بیماری تب شیر به ژن کاندیدای CAMK2A روی کروموزوم شماره ۷ گاو اشاره نمود که در مطالعه قبلی گزارش شده است (پُچیکو و همکاران ۲۰۱۸).
مسیر PPAR signaling pathway که جزء مسیرهای معنی دار مرتبط با صفات ترکیب شیر بدست آمد حاوی ۹ ژن در این مجموعه ژنی بود که از این بین ژن کاندیدای FABP2 مرتبط با چربی شیر می‌توان در نظر گرفت. ژن FABP2 جزو خانواده ژنی پروتئین‌های اتصالی اسیدهای چرب است که نقش کلیدی در دریافت، حمل و متابولیسم اسیدهای چرب دارند. FABP2 در بافت‌هایی که نیاز بالایی به اسیدهای چرب دارند مانند ماهیچه قلب، ماهیچه‌های اسکلتی و غدد پستانی، کبد و یا بافت چربی یافت می‌شود (UniProtKB).
ژن‌های SLC22A16 و SLC25A31 جزئی از خانواده ژنی SLC با ۲۳ عضو می‌باشد که دارای نقش انتقالی هستند. خانواده ژنی SLC که بیشتر در جمعیت‌های انسانی مورد تمایز و انتخاب قرار گرفته‌اند، نقش‌های متنوعی در مکانیسم‌های زیستی دارند. فرآیندهای از قبیل بالا بردن (SLC27A4) و اکسیداسیون (SLC25A20) اسیدهای چرب نشان داده شده است. این خانواده ژنی نقش‌های بیولوژیکی بسیار گسترده‌ای از قبیل سیگنال‌دهی پرولاکتین، ترشح انسولین، جذب طیف وسیعی از مواد مغذی، سنتز هورمون تیروئید و مسیرهای متابولیکی مختلف دیگر دارند (UniProtKB). به علاوه در مطالعات دیگری ارتباط این خانواده ژنی با صفات تولید و ترکیبات شیر از قبیل چربی و پروتئین گزارش شده است. از این میان می‌توان به ارتباط ژن‌های SLC39A8، SLC34A2، SLC1A4، SLC30A4، SLC7A5، SLC6A2، SLCA3 و SLC35B1 با صفات مربوط به تولید و کیفیت شیر اشاره کرد. ولی با توجه به طیف فعالیتی وسیعی ژن‌های انتخابی این خانواده ژنی، اختصاص دقیق ژن‌های انتخابی این خانواده ژنی به مسیر بیولوژیکی خاصی امکان پذیر نیست.

نتیجه‌گیری نهایی

در انتها بیان می‌گردد بررسی این مناطق ژنومی با استفاده از پایگاه داده BioMart، GeneCards و UniProtKB نشان داد که اکثر این مناطق با صفات امتیاز سلول‌های بدنی، سیستم ایمنی، تولید و چربی شیر مرتبط می‌باشند و حاوی ژن‌های *ASIC2*، *ANXA3*، *CCL2*، *CCL11*، *CCL24*، *IL33*، *TLR3*، *WWOX*، *EGFR*، *PRKCA*، *HSP90B1*، *THBS4*، *SPP1*، *FABP2*، *KCNMA1*، *CAMK2A* و *ITPR1* هستند. با توجه به عملکرد بیولوژیکی مسیرهای شناسایی شده در این پژوهش، به نظر می‌رسد این ژن‌ها در بروز فنوتیپی صفات مرتبط با تولید شیر نقش ایفا می‌کنند، در نتیجه می‌توان کارآیی روش تجزیه و تحلیل

غنی‌سازی مجموعه ژنی برای پویش ژنومی صفات تولیدی اقتصادی را نیز مورد تأیید قرار داد.

سپاسگزاری

نویسندگان مقاله از دکتر چن و همکاران در گروه علوم دامی دانشگاه شنگهای چین به خاطر فراهم نمودن اطلاعات مورد نیاز این تحقیق صمیمانه تقدیر و تشکر می‌نمایند.

منابع مورد استفاده

Abdalla EA, Peñagaricano F, Byrem TM, Weigel KA and Rosa GJ, 2016. Genome-wide association mapping and pathway analysis of leukosis incidence in a US Holstein cattle population. Animal Genetics 47: 395–407.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G, 2000. Gene ontology: Tool for the unification of biology. Nature Genetics 25: 25–29.

Boleckova J, Matejickova J, Stipkova M, Kyselova J, Bartonand L and Vyzkumny J, 2012. The association of five polymorphisms with milk production traits in Czech Fleckvieh cattle. Czech Journal of Animal Science (2): 45–53.

Bohlouli M, Mohammadi H and Alijani S, 2013. Genetic evaluation and genetic trend of growth traits of Zandi sheep in semi-arid Iran using random regression models. Small Ruminant Research 114: 195–201.

Chen Z, Yao Y, Ma P, Wang Q and Pan Y, 2018. Haplotype-based genome-wide association study identifies loci and candidate genes for milk Yield in Holsteins. PLoS ONE 13(2): e0192695.

Chen X, Cheng Z, Zhang S, Werling D and Wathes DC, 2015. Combining Genome Wide Association Studies and Differential Gene Expression Data Analyses Identifies Candidate Genes Affecting Mastitis Caused by Two Different Pathogens in the Dairy Cow. Open Journal of Animal Sciences 5: 358-393.

Daetwyler HD, Schenkel FS, Sargolzaei M and Robinson JA, 2008. A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. Journal of Dairy Science 91: 3225–3236.

Dadousis C, Pegolo S, Rosa GJM, Gianola D, Bittante G and Cecchinato A, 2017. Pathway-based genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield, and curd nutrient recovery in dairy cattle. Journal of Dairy Science 100: 1223-1231.

Durinck S, Spellman PT, Birney E and Huber W, 2009. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. Nature Protocols 4: 1184–1191.

Gerlando R, Sutera AM, Mastrangelo S, Tolone M, Portolano B, Sottile G, Bagnato A, Strillacci MG and Sardina MT, 2019. Genome-wide association study between CNVs and milk production traits in Valle del Belice sheep. PLoS ONE 14: e0215204.

Goddard ME and Hayes BJ, 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics 10: 381-388.

Kanehisa M and Goto S, 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28: 27–30.

Kiser JN, Neupane M, White SN and Neibergs HL, 2018. Identification of loci associated with susceptibility to Mycobacterium Avium subspecies paratuberculosis (Map) tissue infection in cattle. Mammalian Genome 298: 539-549.

Kułaj D, Pokorska J, Ochrem A, Dusza M and Makulska J, 2019. Effects of the c.8514C > T polymorphism in the osteopontin gene (OPN) on milk production, milk composition and disease susceptibility in Holstein Friesian cattle. Italian Journal of Animal Science 18: 546-553.

May K, Scheper C, Brügemann K, Yin T, Strube C, Korkuć P, Brockmann GA and König S, 2019. Genome-wide associations and functional gene analyses for endoparasite resistance in an endangered population of native German Black Pied cattle. BMC Genomics 20: 277-236.

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 13: e0199931.

Mooney MA and Wilmot B, 2015. Gene Set Analysis: A Step-By-Step Guide. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 168: 517-527.

Mohammadi H, Rafat SA, Moradi shahrebabak H, Shodja J and Moradi MH, 2018. An assessment of population stratification and haplotype based Genome-wide association for wool quality traits in Zandi sheep breed. Journal of Animal Science Researches (Agricultural science) 28(2): 193-204.

Ogorevc J, Kunej T, Razpet A and Dovc P, 2009. Database of cattle candidate genes and genetic markers for milk production and mastitis. Animal Genetics 40: 832–851.

Pacheco HA, da Silva S, Sigdel A, Mak CK, Galvão KN, Texeira RA, Dias LT and Peñagaricano F, 2018. Gene Mapping and Gene-Set Analysis for Milk Fever Incidence in Holstein Dairy Cattle. Frontiers Genetics 9: 465-478.

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

Peng G, Luo L, Siu H, Zhu Y, Hu P, Hong S, Zhao J, Zhou X, Reveille JD, Jin L, Amos CI and Xiong M, 2010. Gene and pathway-based second wave analysis of genome-wide association studies. European Journal of Human Genetics 18: 111–117.

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.

Rincón G, Islas-Trejo A, Casellas J, Ronin Y, Soller M, Lipkin E and Medrano JF, 2009. Fine mapping and association analysis of a quantitative trait locus for milk production traits on Bos taurus autosome 4. Journal of Dairy Science 92: 758–764.

Sordillo LM and Streicher KL, 2002. Mammary gland immunity and mastitis susceptibility. Journal Mammary Gland Biology 7: 135-146.

Wang L, Jia P, Wolfinger RD, Chen X and Zhao Z, 2011. Gene set analysis of genome-wide association studies: Methodological issues and perspectives. Genomics 98: 1–8.

Yang Y, Wang Q, Chen Q, Liao R, Zhang X, Yang H, Zheng Y, Zhang Z and Pan Y, 2014. New genotype imputation method with tolerance to high missing rate and rare variants. PLoS ONE 9: e101025.

Young MD, Wakefield MJ, Smyth GK and Oshlack A, 2010. Method gene ontology analysis for RNA-seq: Accounting for selection bias. Genome Biology 11: 14-23.

Zhang H, Liu A, Li X, Xu W, Shi R, Luo H, Su G, Dong G, Guo G and Wang Y, 2019. Genetic analysis of skinfold thickness and its association with body condition score and milk production traits in Chinese Holstein population. Journal of Dairy Science 102: 2347-2352.

Genome-wide association study based on gene-set enrichment analysis associated with milk yield in Holstein cattle

N Azizpour[1], A H Khaltabadi Farahani[2*], M H Moradi[2] and H Mohammadi[3]

Received: October 8, 2019 Accepted: November 3, 2019
[1]MSc Graduated, Department of Animal Sciences, Faculty of Agriculture and Natural Resources, Arak University.
[2]Assistant Professor, Department of Animal Sciences, Faculty of Agriculture and Natural Resources, Arak University.
[3]Ph.D Graduated, Department of Animal Science, Faculty of Agricultural Sciences, University of Tabriz.
*Corresponding author: E-mail: amfarahanikh@gmail.com

پژوهش‌های علوم دامی Animal Science Researches | Journal of Animal Science/vol.30 No.1/ 2020/pp 79-92 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.2020.11009

Introduction: Genomic selection has provided the dairy industry with a powerful tool to increase genetic gains on economically important traits such as milk production (Taylor et al. 2016). One way to identify new loci and confirm existing QTL is through genome-wide association analysis (GWAA). In addition, identifying of genes' loci with large effects on economically important traits has been one of the important goals to dairy cattle breeding. Quantitative Trait Loci 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 aim of the present study was genome wide association studies (GWAS) based on gene set enrichment analysis for identifying the loci associated with milk yield and somatic cell score traits in Holstein cattle breed using the high-confidence SNPs that enable us to study 164312 SNP markers simultaneously.
Material and methods: In this study, the 1092 Holstein cattle and 164312 markers were analyzed with milk yield, fat percentage, and somatic cell score using plink software with no corrections to adjust the error rate. The gene set analysis consists basically in three different steps: (1) the assignment of SNPs to genes, (2) the assignment of genes to functional categories, and finally (3) the association analysis between each functional category and the phenotype of interest.
In brief, for each trait, nominal P-values < 0.05 from the GWAS analyses were used to identify significant SNP. Using the *biomaRt* R package the SNP were assigned to genes if they were within the genomic sequence of the gene or within a flanking region of 20 kb up- and downstream of the gene, to include SNP located in regulatory regions. For the assignment of the genes to functional categories, the gene ontology and Kyoto encyclopedia of genes and genome pathway databases were used. The GO database designates biological descriptors to genes based on attributes of their encoded products and it is further partitioned into three components: biological process, molecular function,

and cellular component. The KEGG pathway database contains metabolic and regulatory pathways, representing the actual knowledge 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). It was 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 milk production. It is likely that a better understanding of the biology underlying milk production specifically, plus an advance in the annotation of the bovine genome, can provide new opportunities for predicting production using gene set information. Eleven SNP markers were identified on chromosomes 5, 6, 7, 8, 14, 19, 22, 24, 25, 27, and 28 located in ASIC2, *ANXA3*, *CCL2*, *CCL11*, *CCL24*, *IL33*, *TLR3*, *WWOX*, *EGFR*, *PRKCA*, *CAMK2A*, *KCNMA1*, *FABP2*, *SPP1*, *THBS4*, *HSP90B1*, and *ITPR1* genes. Some of the found genes are consistent with some previous studies and are involved in biological pathways related to milk yield and immune systems. According to pathway analysis, 25 pathways from gene ontology and KEGG pathway were associated with the milk yield and somatic cell score traits ($P<0.01$). Among those pathways, the sensory perception of chemical stimulus, positive regulation of inflammatory response, and defense response biological pathway have an important role in the immune system and somatic cell score. Also, the GnRH signaling pathway, PPAR signaling pathway, oxytocin signaling pathway, and focal adhesion had a significant association with milk yield and fat percentage traits. Some of these regulatory regions, such as enhancers, are located far from the genes. Therefore, the gene might be part of the analysis, but the relevant variant would probably not be included in the gene set SNP class. Finally, linkage disequilibrium interferes with the use of biological information in prediction, because irrelevant regions (regions without any biological role) capture part of the information encoded in relevant regions, causing both regions to exhibit similar predictive abilities. The use of very high density SNP data or even whole genome sequence data could alleviate some of these issues. Finally, it worth's 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 for instance 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.

Conclusion: This study supported previous results from GWAS of milk yield and somatic cell score, also revealed additional regions in the cattle genome associated with these economically important traits. So, using these findings can accelerate the genetic progress in the breeding programs.

Keywords: Cattle, Gene-set enrichment analysis, Genome scan, Milk yield, Somatic cell score