مقاله علمی پژوهشی/نشریه پژوهش‌های علوم دامی/ جلد ۳۴ شماره ۱/ سال۱۴۰۳/ صفحات ۳۱ تا ۴۴
DOI: 10.22034/as.2023.54694.1690

مطالعه پویش کل ژنومی جهت شناسایی مناطق ژنومی مرتبط با ترکیبات پروتئین شیر مبتنی بر روش‌های آماری تحلیل مسیر و غنی سازی ژنی در گاوهای نژاد هلشتاین

حسین محمّدی[۱]*، محمد شمس اللهی[۲] و شینگلی زنگ[۳]

تاریخ دریافت: ۱۴۰۱/۱۰/۹ تاریخ پذیرش: ۱۴۰۲/۱/۲۱

[۱] استادیار گروه علوم دامی، دانشکده کشاورزی و منابع طبیعی، دانشگاه اراک، اراک، ایران
[۲] استادیار گروه علوم دامی، دانشکده کشاورزی، دانشگاه ایلام، ایلام، ایران
[۳] استاد آزمایشگاه مرکزی ژنتیک، اصلاح نژاد دام و تولیدمثل، دانشکده علوم و فناوری دامی، دانشگاه کشاورزی چین، بیژینگ، چین
*مسئول مکاتبه: Email: H-mohammadi64@araku.ac.ir

چکیده

زمینه مطالعاتی: امروزه، در اصلاح نژاد گاوهای شیرده، روش‌های انتخاب مبتنی بر جایگاه‌های کنترل کننده صفات کمّی و مناطق ژنومی مؤثر بر صفات تولیدی برای افزایش بازده انتخاب و بهبود عملکرد تولیدی مورد توجه قرار گرفته است. **هدف**: پژوهش حاضر، با هدف مطالعه پویش ژنومی بر اساس آنالیز مجموعه‌های ژنی برای شناسایی جایگاه‌های ژنی مؤثر بر صفات تولید و ترکیبات پروتئین شیر در یک جمعیت گاو هلشتاین بوده است. **روش کار**: برای هر گاو، هشت صفت شامل مقدار و درصد پروتئین و ترکیبات پروتئین آلفا–اس ۱–کازئین (αs1)، آلفا–اس ۲–کازئین (αs2-CN)، بتا–کازئین (β-CN)، کاپا–کازئین (κ-CN)، آلفا–لاکتوآلبومین (α-LA) و بتا–لاکتوگلوبولین (β-LG) رکورد جمع‌آوری شده بود. در این تحقیق، ابتداً، آنالیز پویش کل ژنومی در برنامه PLINK انجام شد. سپس، با استفاده از بسته نرم افزاری *biomaRt2* تحت برنامه R ژن‌های معنی‌داری که در داخل و یا ۱۵ کیلوباز بالا و پایین دست نشانگرهای معنی‌دار قرار داشتند، شناسایی گردید. در نهایت، آنالیز غنی‌سازی مجموعه‌های ژنی با برنامه KOBAS با هدف شناسایی عملکرد بیولوژیکی ژن‌های نزدیک به مناطق انتخابی و ژن‌های کاندیدا از طریق پایگاه‌های برخط GO، KEGG، Reactome، BioCyc و PANTHER انجام شد. **نتایج**: تجزیه و تحلیل غنی‌سازی مجموعه‌های ژنی تعداد ۲۰ طبقات هستی‌شناسی و مسیرهای بیوشیمیایی با صفات مورد بررسی شناسایی شد (P<۰/۰۵). از این بین طبقاتOxytocin signaling pathway، Glycerolipid metabolism، Response to progesterone، Detection of calcium ion، Complement and coagulation cascades و amino acid binding با داشتن ژن‌های کاندیدای *CDH13*، *P4HTM*، *SPP1*، *CSN1S1*، *CSN2*، *SERPINA1*، *SLC35A3*، *ODC1* و *PAEP* نقش مهمّی در تولید و حجم پروتئین، فسفریله کردن گلیکوپروتئین‌ها، استحکام انعقاد شیر و سنتز لاکتوز داشتند. **نتیجه گیری نهایی**: بعنوان جمع‌بندی، یافته‌های این تحقیق نشان داد پتانسیل انتخاب ژنتیکی برای بهبود کیفیت شیر از جمله ترکیبات پروتئینی شیر امکان پذیر می‌باشد، بطورئیکه بهبود حاصل از انتخاب ژنتیکی به دلیل توارث‌پذیری، تجمعی و دائمی بودن بسیار مفید است.

واژگان کلیدی: آنالیز مسیر، آلفالاکتوآلبومین، بتالاکتوگلوبولین، کازئین، ژن کاندیدا

مقدمه

شیر، بعنوان، یک ماده غذایی ارزشمند، تأمین کننده مواد مغذی مورد نیاز برای سلامت بدن از جمله کلسیم، فسفر،

پروتئین و ریبوفلاوین می‌باشد. بطورئیکه، شیر غنی از پروتئین‌های با کیفیت بالا می‌باشد که دارای ویژگی‌های تغذیه‌ای، عملکردی و بیولوژیکی بی‌نظیری هستند (ددئوسیس و همکاران ۲۰۱۷). پروتئین‌های ویژه شیر شامل: چهار کازئین (آلفا S1، آلفا S2، گاما و کاپا) و دو آب پنیر (آلفا لاکتوآلبومین و بتالاکتوگلوبولین) می‌باشد که از این میان، کازئین یکی از پروتئین‌های اصلی شیر می‌باشد که حدود ۸۰ درصد از کل پروتئین‌های شیر را به خود اختصاص می‌دهد (سولیموا و همکاران ۲۰۰۷). همچنین، امروزه ارزش دام‌های ماده منحصراً، بوسیله‌ی تولید شیر تعیین نمی‌شود، بلکه، شاخص‌های چربی و پروتئین درآمد ناشی از فروش شیر را نیز تحت تأثیر قرار می‌دهند. درصد چربی و پروتئین شیر از نظر ارزش اقتصادی در قیمت‌گذاری شیر مؤثر می‌باشند. علاوه بر این، ترکیبات شیر بر خصوصیات کیفی شیر و فرآوردهای حاصل از آن و روی ارزش تغذیه‌ای شیر نیز اثر دارد. بنابراین، در کنار سایر اهداف اصلاحی، این صفات نیز از اهمیت خاصی برخوردارند. در نتیجه، شناسایی ژنوتیپ پروتئین‌های شیر در دام‌های شیری موقعیت بی‌نظیری فراهم می‌آورد که در آن ژنتیک مولکولی می‌تواند روی صفات کمّی حائز اهمیت اقتصادی اثر مستقیمی داشته باشد (رضوان نژاد و همکاران ۲۰۲۲؛ نجفی و همکاران ۲۰۲۰).

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 است، پیدا کردن جهش‌های علّی یا مسبب می‌باشد که بر فنوتیپ یک صفت اثر می‌گذراند. این اطلاعات می‌تواند برای انتخاب به کمک نشانگر مفید بوده و به درک مکانیسم مولکولی صفات مورد مطالعه کمک نماید (محمّدی و همکاران ۲۰۲۰).

مطابقه یافته‌های تحقیقات مختلف پیشین، در تجزیه و تحلیل پویش ژنومی مبتنی بر مدل‌های خطی و غیر خطی، مشکل بالا بودن نرخ خطای نوع اول و بیش برآورد اثر نشانگرهای SNPها می‌باشد. به‌عبارت دیگر، یکی از ایرادات تحقیقات مطالعات پویش ژنومی در نظر گرفتن آستانه معنی‌داری محافظه‌کارانه مانند تصحیح بونفرونی 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، راهکار جایگزین مناسب برای حل این مشکل، انجام مطالعات پویش کل ژنومی بر مبنای تجزیه و تحلیل مسیر[1] و با استفاده از تجزیه و تحلیل مجموعه‌های ژنی است. در واقع، در این روش، بجای انجام تجزیه برای یک SNP یا یک ژن، ارتباط بین صفت مورد مطالعه و واریانت‌های ژنتیکی را در یک دسته یا گروه ژنی که بطور عملکردی با هم مرتبط هستند بررسی می‌کند. بعبارتی دیگر، تجزیه و تحلیل پیوستگی بین یک مجموعه ژنی معنی‌دار زیستی با فنوتیپ مورد آزمون قرار می‌گیرد (وانگ و همکاران ۲۰۱۱). در حقیقت، در روش پیشنهادی محققان 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 بعنوان، بستری معنی‌دار که عملکرد مجموع ژن‌ها در آنها یک یا چند هدف واحد را پیگیری می‌کند استفاده می‌شود (ددئوسیس و همکاران ۲۰۱۷).

در این راستا، برای اولین بار پناکریگانو 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 تجزیه و تحلیل می‌شوند و در نتیجه میزان خطای نوع اول و بیش برآوردها کاهش پیدا می‌کند. در تحقیق دیگر،

[1] Pathway-based analysis

مطالعه پویش ژنومی بر مبنای مسیر با استفاده از تجزیه و تحلیل غنی‌سازی مجموعه‌های ژنی روی خصوصیات لخته‌شدگی شیر، تولید پنیر و استحکام دلمه انجام شد. براساس این نتایج خروجی ها براساس مسیر، منجر به شناسایی ۲۱ طبقات مختلف عملکردی هستی شناسی ژن و ۱۷ مسیر بیوشیمیایی KEGG معنی‌دار مرتبط با این صفات شد که شامل ژن‌های کاندیدای *CSN1S1*، *LAP3*، *DGAT1*، *RPL8* و *HSF1* بودند (ددئوسیس و همکاران ۲۰۱۷).

در مطالعه پویش کل ژنومی در گاوهای هلشتاین چینی با هدف شناسایی ژن‌های کاندیدای مرتبط با تولید و ترکیبات پروتئین شیر انجام شده است. نتایج این تحقیق نشان داد ژن‌های کاندیدای *ARL6*، *SST*، *EHHADH*، *PCDHB4*، *SLC36A2*، *GALNT14*، *FPGS*، *LARP4B* و *ACADSB* روی کروموزوم‌های شماره های ٦، ١١، ١٣ و ١٤ مرتبط با ترکیبات پروتئین شیر بودند (ژو و همکاران ۲۰۱۹).

هدف از انجام پژوهش حاضر، شناسایی مناطق ژنومی و ژن‌های کاندیدای مرتبط با صفات مقدار، درصد و ترکیبات پروتئین شیر 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 مرتبط با مقدار و درصد پروتئین شیر و ترکیبات پروتئین شیر شامل، آلفا-اس ۱-کازئین[1]، آلفا-اس ۲-کازئین[2]، بتا-کازئین (β-CN)، کاپا-کازئین (κ-CN)، آلفا-لاکتوآلبومین (α-LA) و بتا-لاکتوگلوبولین (β-LG) در گاوهای هلشتاین استفاده گردید. اطلاعات شامل مجموع تعداد ٦١٤ رأس گاو نژاد هلشتاین چینی بود که شجره آنها بطور کامل، در دسترس بود. DNA ژنومی مجموع ۵۹۸ رأس از گاوها با استفاده از آرایه‌های Illumina BovineSNP50 براساس دستورالعمل آزمایشگاهی استاندارد Illumina تعیین ژنوتیپ شده بودند. داده‌های ژنوتیپی جهت افزایش صحت نتایج از تراکم متوسط (54K) به تراکم بالا (777K) با استفاده از نرم افزار BEAGLE (نسخه ۳/۳/۱) امپیوت[3] شده بودند (بروئنینگ و بروئنینگ ۲۰۰۷). داده‌های تعیین ژنوتیپ شده در پایگاه برخط ذخیره ژنومی figshare (https://figshare.com/articles/dataset) با شماره دسترسی (206a2bcbf0cb0e4c2564) در دسترس می-باشد.

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، سپس، SNPهایی که P-Value تعادل هاردی- واینبرگ برای آن SNPها بزرگ‌تر از سطح معنی‌داری بود، کنار گذاشته شدند. سطح معنی‌داری فیلتراسیون تعادل هاردی- واینبرگ، 1×10^{-6} تعیین شده بود. در نهایت، بعد از کنترل کیفیت و انجام ایمپیوتیشن، تعداد ٥٨٦٣٠٤ SNP برای تجزیه و تحلیل مطالعه پویش کل ژنومی بر پایه روش‌های مختلف بیزی باقی ماندند.

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

اساساً ،تجزیه و تحلیل پویش ژنومی بر پایه تجزیه و تحلیل مجموعه‌های ژنی در سه مرحله انجام می‌گردد: ۱) تعیین مکان SNPها 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* (دیورسینک و همکاران ۲۰۰۹) در

[1] αs1-CN
[2] αs2-CN
[3] Imputed

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 ۱۵kb 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) (اشبیونر و همکاران ۲۰۰۰) و مسیرهای بیوشیمیایی KEGG (کانیشیا و گوتو ۲۰۰۰)، Reactome (جسال و همکاران ۲۰۲۰) 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 بعنوان 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 مقدار و ترکیبات پروتئین شیر با استفاده از توزیع فوق هندسی[1] 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 صفات مورد بررسی، N برابر با کل تعداد ژن‌هایی که در این مطالعه تجزیه و تحلیل شدند و m برابر با تعداد کل ژن‌های موجود در مسیر عملکردی. آنالیز غنی‌سازی مجموعه‌های ژنی با استفاده از برنامه KOBAS (بیو و همکاران ۲۰۲۱) 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

آماره‌های توصیفی رکوردهای فنوتیپی مربوط به صفات مقدار، درصد و ترکیبات پروتئینی شیر گاوهای هلشتاین در جدول ۱ ارائه شده است.

Table 1– Descriptive statistics of milk protein composition traits in Holstein population.

Traits	No. cows	Mean	Standard deviation	Max	Min
Protein percentage	614	3.06	0.29	4.28	2.09
Protein yield (kg)	614	0.75	0.32	1.82	0.36
αs1-casein(αs1-CN)	614	35.45	17.46	72.63	1.96
αs2-casein(αs2-CN)	614	16.64	8.62	53.91	1.01
β -casein(β-CN)	614	31.23	10.31	69.59	2.24
k-casein(κ-CN)	614	7.51	1.69	23.48	0.43
α -lactalbumin(α-LA)	614	2.25	0.85	10.10	0.10
β- lactoglobulin (-LG)	614	6.93	3.67	48.64	0.18

The six major milk proteins are expressed as a weight-proportion of the total protein fraction.

[1] Hypergeometric

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 مقدار، درصد پروتئین و ترکیبات پروتئین شیر در گاو نژاد هلشتاین انجام گردید. پلات‌های منهتن مرتبط با این صفات به ترتیب در شکل‌های ۱ تا ۲ ارائه شده است.

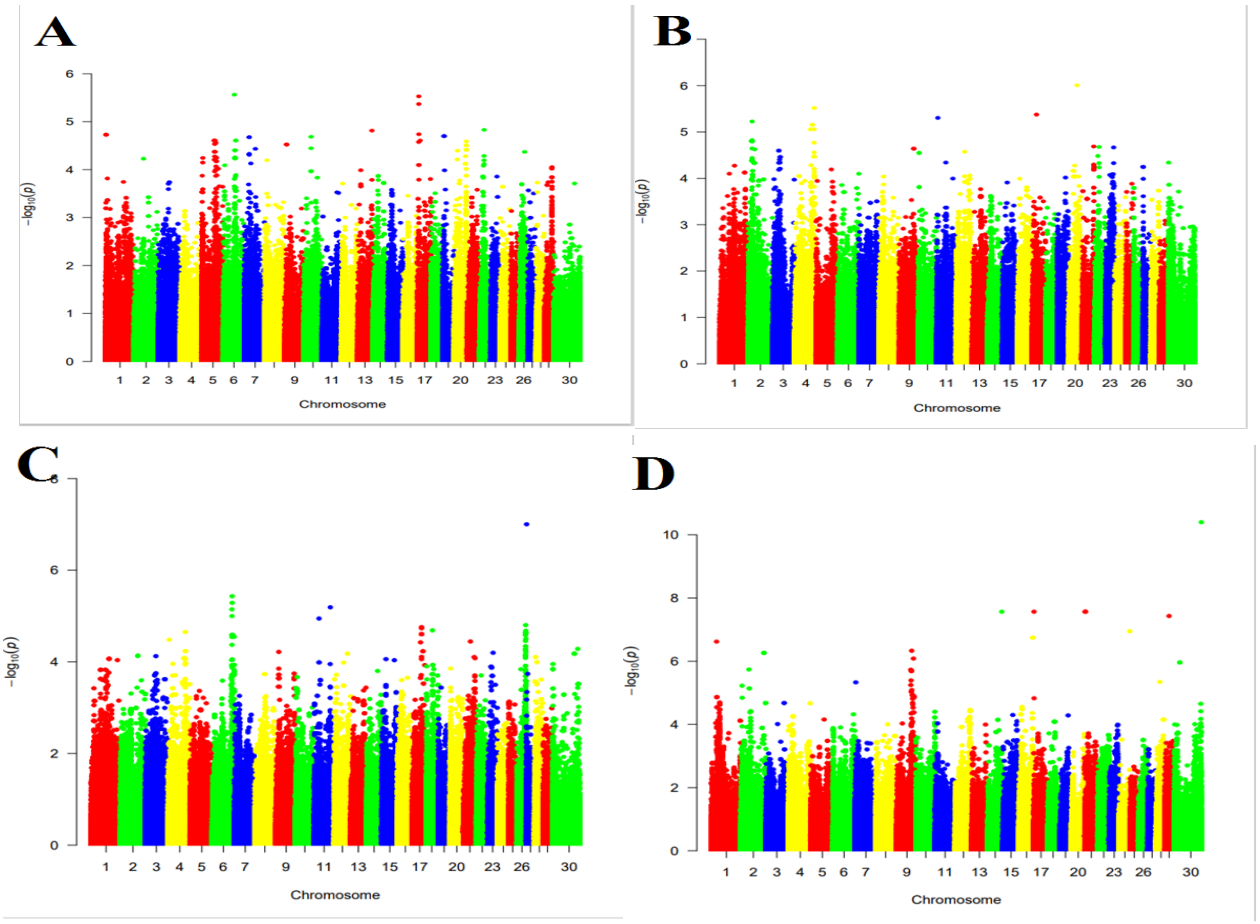

Figure 1- Manhattan plot for A)protein percentage, B)α-lactalbumin, C)β-lactoglobulinin and D) protein yield in Holsteins cattle. X axis, SNPs positions on chromosomes, Y axis, -Log10 P-value

شناسایی ژن‌های کاندیدا در مناطق ژنومی با استفاده از تجزیه GSEA-SNP[6]

در مجموع تعداد ۱۹۸۰۵ عدد ژن از ۲۷۶۰۷ ژن حاشیه نویسی شده در گاو بوسیله، نشانگرهای SNP پوشش داده شد که در این میان تعداد ۱۷۳۵ ژن دارای اثر معنی‌داری بودند، یعنی، حداقل یک نشانگر با P-value کمتر از ۰/۰۵ در داخل و یا در بالا یا پایین دست این ژن تا فاصله ۱۵ kb قرار گرفت (جدول ۲). این ژن‌ها بعنوان، ژن‌های معنی‌دار مرتبط با صفات مقدار، درصد پروتئین و ترکیبات پروتئین شیر برای تجزیه GSEA-SNP انتخاب شدند و باقیمانده ژن‌ها بعنوان ژن‌های پس زمینه‌ای مورد استفاده قرار گرفتند. تعداد مجموعه‌های ژنی حاصل از پایگاه‌های داده‌ای مختلف شامل، ۱۰۶۰ طبقات هستی شناسی، ۱۷۵ مسیر بیوشیمیایی KEGG و ۳٤ مسیر Reactome بود. همانطور که در جدول ۳ مشاهده می‌شود تعداد ۲۰ طبقات عملکردی در هستی شناسی فرآیندهای

[6] Gene Set Enrichment Analysis- SNP

زیستی، عملکرد مولکولی، اجزای سلولی و مسیرهای هستی شناسی KEGG با صفات مقدار، درصد پروتئین و ترکیبات پروتئین شیر دارای ارتباط هستند (P<۰/۰۵).

مسیرهای که بیشتر از ۳ ژن و کمتر از ۵۰۰ ژن داشتند گزارش شده‌اند.

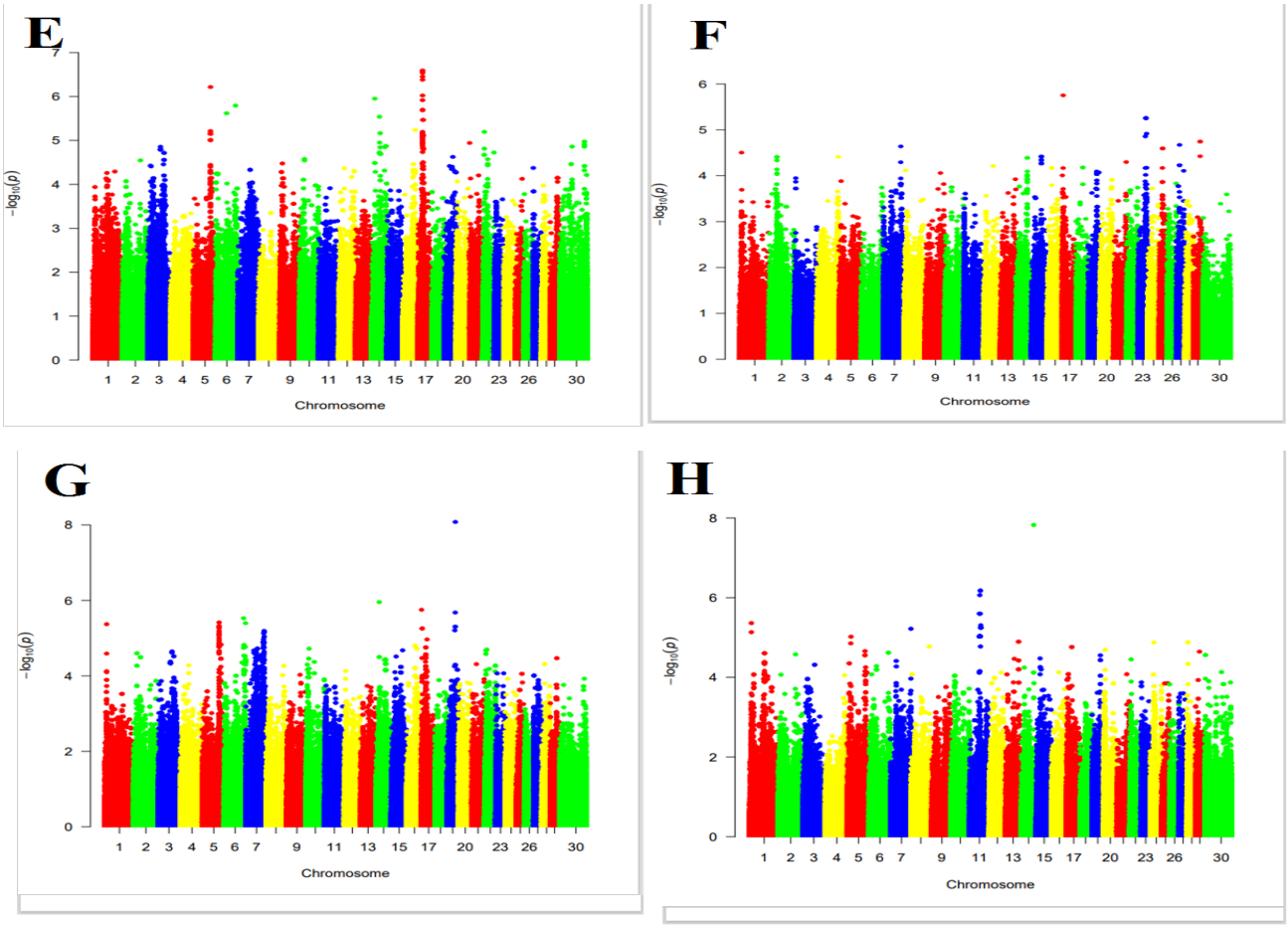

Figure 2- Manhattan plot E)αs1-casein, F)αs2-casein, G)β-casein, H)κ-casein, in Holsteins cattle. X axis, SNPs positions on chromosomes, Y axis, -Log10 P-value

از مسیرهای مهم و معنی‌دار مرتبط با مقدار و درصد پروتئین را می‌توان به مسیر بیولوژیکی Oxytocin signaling pathway اشاره نمود که از بین ۱۰ ژن معنی‌دار، ژن کاندیدای *CDH 13* در مطالعات قبلی پویش ژنومی ارتباط معنی‌داری با مقدار پروتئین تولید شیر در گاوهای نژاد شیرده Vrindavani گزارش شده است (سینق و همکاران ۲۰۲۲).

مسیر Focal adhesion با تعداد ۱۸ ژن معنی‌دار از دیگر مسیرهایی بود که در ارتباط با مقدار و درصد پروتئین شیر شناسایی شد. این مسیر در مسیر KEGG طبقه بندی می‌شود. از بین ژن‌های معنی‌دار در این مسیر، ژن کاندیدای *P4HTM* در مطالعات قبلی پویش ژنومی با مقدار پروتئین شیر گوسفند برپایه توالی‌یابی کل ژنوم گزارش شده است (رضوان نژاد و همکاران ۲۰۲۲). 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 (بلیکووا 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.

Table 2– Number of significant SNP identified from genome-wide association studies (GWAS) and genes mapped by related trait

Traits	No. of significant SNP	No. of significant SNP assigned to genes	No. of significant mapped genes
Protein percentage	1542	643	527
Protein yield	1735	602	484
αs1-CN	996	315	278
αs2-CN	875	450	363
β-CN	622	283	264
κ-CN	902	543	455
α-LA	1240	864	655
β-LG	1045	626	581

از دیگر مسیرهای بیولوژیکی معنی‌دار مرتبط با مقدار و درصد پروتئین 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 ٦ ارتباط معنی‌داری با ترکیبات شیر گزارش شده است (ژانگ و همکاران ۲۰۱۹).

مسیر دیگر معنی‌دار مرتبط با مقدار و درصد پروتئین شیر را می‌توان Glycerolipid metabolism نام برد که از بین ۱۲ ژن، ژن کاندیدای *CMIP* در مطالعات قبلی، ارتباط معنی‌داری با صفات مرتبط با درصد پروتئین شیر در گاوهای شیری گزارش شده است (سینق و همکاران ۲۰۲۲).

کازئین، که مهم‌ترین پروتئین شیر است شامل ٣٦ درصد آلفا کازئین، ۳۷ درصد بتاکازئین، ۱۲ درصد کاپاکازئین و هفت درصد پپتیدها و اسیدهای آمینه می‌باشد (الندری و همکاران ۱۹۹۰). مسیر Response to progesterone که جزء مسیر فرآیند زیستی است از مسیرهای معنی‌دار مهم در این پژوهش بود که با ترکیبات کازئینی شیر به‌دست آمد تعداد ۸ ژن این مسیر معنی‌دار بود. ژن‌های کاندیدای *CSN1S1* و *CSN2* در بین ژن‌های این مجموعه ژنی قرار داشت و محصول این ژن‌ها، پروتئینی به همین نام است که به ترتیب تولید آلفا کازئین و بتاکازئین می‌نمایند. مشخص شده است که چندشکلی تک نوکلئوتیدی در این ژن‌ها با تولید شیر، درصد چربی شیر و همچنین، خصوصیات انعقادی شیر ارتباط دارند (کیشور و همکاران ۲۰۱۳). ژن *CSN1S1* کدکننده پروتئین آلفا S1 کازئین یکی از پروتئین‌های خانواده کازئین می‌باشد که علاوه بر، نقش کلیدی در انتقال کلسیم فسفات دارای نقش مهم فیزیولوژیکی از جمله کاهش فشار خون و مهار آنزیم مبدل آنژیوتنسین می‌باشد (آیمیوتیس ۲۰۰٤). ژن *CSN2* کدکننده پروتئین بتاکازئین، نقش اساسی در انتقال

یون‌های ضروری و مواد معدنی مثل کلسیم و فسفر دارد (Genecards).

از دیگر مسیرهای معنی‌دار مرتبط با بتاکازئین می‌توان به Detection of calcium ion و Bone mineralization اشاره نمود که از بین ژن‌های موجود در مسیرهای نامبرده ارتباط معنی‌داری بین ژن‌های کاندیدای *POU1F1* و *SERPINA1* اشاره کرد. مشخص شده است که بین چندشکلی تک نوکلئوتیدی در ژن *POU1F1* و استحکام انعقاد شیر به پنیر ارتباط معنی‌داری وجود دارد (ددئوسیس و همکاران ۲۰۱۷). در مطالعه پویش ژنومی برپایه تکنیک RNA-Seq در گاوهای هلشتاین، ژن کاندیدای *SERPINA1* مرتبط با مقدار پروتئین شیر گزارش شده است (لی و همکاران ۲۰۱۶).

کاپاکازئین از لحاظ اندازه از بقیه کازئین‌های شیر کوچک‌تر است. این پروتئین علاوه، بر نقش محافظت از میسل کازئین در شیر، در کیفیت پنیر تولید شده، سرعت دلمه بستن و راندمان تبدیل شیر به پنیر اهمیت بسزایی دارد. دو سوم کلسیم شیر یا مستقیماً به کازئین شیر اتصال دارد و یا بخشی از ترکیب کلسیم و فسفات کلوئیدی است که میسل کازئین نامیده می‌شود. کاپاکازئین نسبت به کلسیم غیر حساس بوده و بعنوان تثبیت کننده میسل شناخته می‌شود. کاپاکازئین نقش مهمی در جلوگیری از ته نشین شدن کازئین‌های دیگر دارد. مدت زمان ایجاد لخته و استحکام پنیر از نظر تکنیکی حائز اهمیت است. بین استحکام لخته و اندازه میسل ارتباط منفی وجود دارد. استحکام لخته‌های حاصل از میسل‌های کوچک‌تر بیشتر است و در نتیجه در گاوهایی که شیر آنها کاپاکازئین بالایی دارد، اندازه میسل‌ها کوچکتر و استحکام لخته بیشتر خواهد بود (علی نقی‌زاده و همکاران ۲۰۰۷).

Table 3- Gene set enrichment analysis significantly (P < 0.05) associated with yield and protein percentage and composition traits

Category	Term	No. of genes	genes[1]	P value	FDR
GO_BP	GO:0005513: Detection of calcium ion	5	*CALM2, CALM3,* ***POU1F1****, KCNMB4, STIM1*	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:0032570: Response to progesterone	8	***CSN1S1****, CSN1S2,* ***CSN2****, CSN3, LALBA, TFAP2A, PPP3CA, STAT5B*	2.77E-11	4.96E-08
	GO:0015347: Sodium-independent organic anion transmembrane transporter activity	8	*SLC22A12, SLC22A6, SLCO4A1, SLCO2B1, SLC22A8, SLC22A10, SLC22A9, SLC22A11*	1.04E-04	1.69E-02
	GO:0046983: Protein dimerization activity	7	*HEY1, MYC, ID2, TCF23, STAT3, ANO4, E2F6*	4.1E-2	9.9E-1
	GO:0007595: Lactation	8	*STAT5A, STAT5B, VDR,* ***NEURL1****, ATP2B2, CSN3,CSN2, PRLR*	3.1E-2	1.62E-02
	GO:0030282: Bone mineralization	8	*KLF10, CLEC3B, WNT11, PKDCC, RSPO2, FBXL15, IFITM5,* ***SERPINA1***	2.1E-010	1.63E-06
GO_CC	GO:0030055:cell-substrate junction	16	*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	13	*GRIA1, ANXA3, HOMER1, ANXA4, GRIK3,* ***PAEP****, GRIN2D, CACNG4, GRM4, GRIN3A, DAGLA, HOMER3, GRID1*	1.70E-04	2.14E-02

	GO:0061134:0008066: peptidase regulator activity	11	*C3*, *SERPINA1*, *CAST*, *ANXA2*, *SERPINA5*, ***ODC1***, *MYO10*, *TAGLN*, *CNN1*, *ANG*, *MYH2*	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	10	*CACNA2D2*, *CACNA2D3*, *CACNG1*, *CACNA1D*, *CAMK2A*, *CAMK4*, ***CDH 13***, *GUCY1A2*, *ITPR1*, *MAP2K5*	3.7E-4	2.3E-2
	bta00561:Glycerolipid metabolism	12	*CEL*, *DGKB*, *DGKE*, *DGKG*, *GK2*, *LPIN2*, *MBOAT1*, *PNLIPRP3*, *PNPLA3*, *LOC786474*, ***CMIP***, *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	18	*BRAF*, *FYN*, *ARHGAP5*, *SHC3*, *BIRC3*, *COMP*, *COL4A2*, *COL11A2*, *EGFR*, *FLT1*, *PAK5*, *PARVA*, ***P4HTM***, *PIK3CD*, *PRKCA*, *PRKCB*, ***SPP1***, *TLN2*	3.9E-2	2.7E-1
	bta04960: Aldosterone-regulated sodium reabsorption	3	*INSR*,*NEDD4L*, ***PKHD1***	4.3E-2	2.8E-1
	bta04610: Complement and coagulation cascades	11	*C3*, ***SERPINA1***, ***SLC35A3***, *MET*, *FLNA*, *IGF1R*, *ITGB6*, *COL4A5*, *GARS*, *MARS*, *NARS*	4.7E-2	2.9E-1

مسیر Complement and coagulation cascades که جزء مسیرهای معنی‌دار مرتبط با کاپاکازئین شیر بدست آمد حاوی ۱۱ ژن در این مجموعه ژنی بود که از این بین ژن‌های کاندیدای *SERPINA1* و *SLC35A3* مرتبط با کاپاکازئین می‌توان در نظر گرفت. ژن *SLC35A3* 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. 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). بعلاوه، در مطالعات دیگری ارتباط این خانواده ژنی با صفات تولید و ترکیبات شیر از قبیل چربی و پروتئین گزارش شده است (لیو و همکاران ۲۰۱۶).

ژن بتالاکتوگلوبولین، پروتئین اصلی آب پنیر شیر نشخوارکنندگان بوده که بوسیله سلوهای پوششی غده پستان در طی دوره‌های آبستنی و شیردهی سنتز شده و در کیفیت شیر و فرآیند لخته شدن آن نقش دارد. همچنین نقش این پروتئین در محافظت از رتینول شیر، ایمنی نوزادان و تنظیم سوخت و ساز فسفر در غده پستان شناخته شده است (ژو و دانگ ۲۰۱۳). علاوه بر این گزارش شده است ارتباط معنی‌داری بین چندشکلی در ژن بتالاکتوگلوبولین با مقاومت به بیماری ورم پستان وجود دارد. آلفالاکتوآلبومین، تشکیل دهنده اساسی

آنزیم لاکتوسنتاز بوده که مسئول سنتز لاکتوز در لاکتوز در شیر است. آلفالاکتوآلبومین مستقیم حجم و کیفیت شیر را از طریق سنتز لاکتوز تحت تأثیر قرار می‌دهد و نقش مهمی در تنظیم حجم شیر دارد (ژانگ و همکاران ۲۰۰۷).

از مسیرهای اصلی معنی‌دار مرتبط با پروتئین‌های آب پنیر شیر، peptidase regulator activity و amino acid binding بدست آمد که جزء فرآیندهای عملکرد مولکولی است. که بین ژن‌های کاندیدای موجود در این مسیرها، ژن‌های کاندیدای *ODC1* و *PAEP* بیشترین ارتباط را داشتند. ژن کاندیدای *ODC1* نقش کلیدی در لاکتوژنزیس را داشته و در مطالعه پویش ژنومی با هدف شناسایی ژن‌های کاندیدای مرتبط با تولید شیر، ژن *ODC1* مرتبط با تولید پروتئین شیر گزارش شده است (لی و همکاران ۲۰۲۰). همچنین، ارتباط معنی‌داری بین چندشکلی موجود در ژن کاندیدای *PAEP* با ترکیبات پروتئینی شیر در گاوهای شیری گزارش شده است (کولندا و همکاران ۲۰۲۰).

از مسیرهای هستی شناسی KEGG معنی‌دار مرتبط با پروتئین‌های آب پنیر به مسیر مربوط به Aldosterone-regulated sodium reabsorption اشاره نمود. از بین ۳ ژن معنی‌دار در این مسیر ژن کاندیدای *PKHD1* در مطالعات قبلی در ارتباط با تولید مقدار پروتئین شیر در گاوهای هلشتاین کره‌ای گزارش شده است (کیم و همکاران ۲۰۲۱).

شاید بتوان مسیر Lactation که جزء هستی شناسی اجزای سلولی است را یکی از مهمترین مسیرهای مؤثر بر فرآیند تولید ترکیبات پروتئینی آب پنیر شیر دانست. در این مسیر ژن کاندیدای *NEURL1* در مطالعات قبلی پویش ژنومی با صفت درصد پروتئین شیر گزارش شده است (پدروسا و همکاران ۲۰۲۱).

نتیجه‌گیری کلی

نتایج تجزیه و تحلیل پویش ژنومی برای صفات مرتبط با ترکیبات پروتئین شیر نشان داد تعداد نشانگر SNP قابل توجهی از آستانه پیشنهادی عبور کردند. بررسی این مناطق ژنومی با استفاده از پایگاههای داده BioMart، GeneCards و UniProtKB نشان داد که اکثر این مناطق با صفات تولید و ترکیبات پروتئین شیر مرتبط می‌باشند. با توجه به عملکرد بیولوژیکی مسیرهای شناسایی شده در این پژوهش، به نظر می‌رسد این ژن‌ها در بروز فنوتیپی صفات مرتبط با تولید و ترکیبات پروتئین شیر نقش ایفا می‌کنند، با توجه به بهبود حاصل از انتخاب ژنتیکی به دلیل توارث‌پذیری، تجمعی بودن و دائمی بودن در نتیجه می‌توان کارآیی روش تجزیه و تحلیل غنی‌سازی مجموعه ژنی برای پویش ژنومی صفات تولیدی اقتصادی را نیز مورد تأیید قرار داد.

References

Aimutis WR, 2004. Bioactive properties of milk proteins with particular focus on anticariogenesis. Journal of Nutrition 134(4):89-95.

Aleandari R, Buttazoni G, Schneider JC, Caroli A and Davoli R, 1990. The effect of milk protein polymorphisms on milk components and cheese producing ability. Journal Dairy Science 73:241-255.

Alinaghizadeh R, Mohammad Abadi MR and Moradnasab Badrabadi S, 2007. Kappa-casein gene study in Iranian Sistani cattle breed (Bos indicus) using PCR-RFLP. Pakistan Journal of Biological Sciences 10 (23):4291-4294.

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.

Browning SR and Browning BL, 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics (5):1084-97.

Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu and Guo J, 2021. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Research 49:317–325.

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.

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

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.

Kim S, Lim B, Cho J, Lee S, Dang CG, Jeon JH, Kim JM and Lee J, 2021. Genome-Wide Identification of Candidate Genes for Milk Production Traits in Korean Holstein Cattle. Animals (Basel) 11(5):1392.

Kishore A, Mukesh M, Sobti RC, Mishra BP and Sodhi M, 2013. Variations in the Regulatory Region of Alpha S1-Casein Milk Protein Gene among Tropically Adapted Indian Native (*Bos Indicus*) Cattle. ISRN Biotechnology 14:926025.

Kolenda M, Sitkowska B, Kamola D and Lambert BD, 2021. Composite genotypes of progestogen-associated endometrial protein gene and their association with composition and quality of dairy cattle milk. Animal Bioscience 34(8):1283-1289.

Li C, Cai W, Zhou C, Yin H, Zhang Z, Loor JJ, Sun D, Zhang Q, Liu J and Zhang S, 2016. RNA-Seq reveals 10 novel promising candidate genes affecting milk protein concentration in the Chinese Holstein population. Scientific Reports 6:26813.

Li Q, Liang R, Li Y, Gao Y, Li Q, Sun D and Li J, 2020. Identification of candidate genes for milk production traits by RNA sequencing on bovine liver at different lactation stages. BMC Genetics 21(1):72.

Liu JJ, Liang AX, Campanile G, Plastow G, Zhang C, Wang Z, Salzano A, Gasparrini B, Cassandro M and Yang LG, 2018. Genome-wide association studies to identify quantitative trait loci affecting milk production traits in water buffalo. Journal of Dairy Science 101(1):433–444.

Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M and Haw R, 2020. The reactome pathway knowledgebase. Nucleic Acids Research 48:498–503.

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.

Najafi MH, Mohammadi Y, Najafi A, Shamsolahi M and Mohammadi H, 2020. Lairage time effect on carcass traits, meat quality parameters and sensory properties of Mehraban fat-tailed lambs subjected to short distance transportation. Small Ruminant Research 188:106122.

Pedrosa VB, Schenkel FS, Chen SY, Oliveira HR, Casey TM, Melka MG and Brito LF, 2021. Genome wide Association Analyses of Lactation Persistency and Milk Production Traits in Holstein Cattle Based on Imputed Whole-Genome Sequence Data. Genes (Basel) 12(11):1830.

Peng G, Luo L and Siu H, 2010. Gene and pathway-based second wave analysis of genome-wide association studies. European Journal of Human Genetics 18:111–117.

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 in Genetics 3:307-314.

Playne ML, Bennett L and Smithers G, 2003. Functional dairy foods and ingredients. Australian Journal of Dairy Technology 58(3):242-64.

Rezvannejad E, Asadollahpour Nanaei H, Esmailizadeh A. 2022. Detection of candidate genes affecting milk production traits in sheep using whole-genome sequencing analysis. Veterinary Medical Science 8(3):1197-1204.

Singh A, Kumar A, Gondro C, Pandey AK, Dutt T and Mishra BP, 2022. Genome Wide Scan to Identify Potential Genomic Regions Associated With Milk Protein and Minerals in Vrindavani Cattle. Frontiers in Veterinary Science 9:760364.

Sulimova GE, Abani Azari M, Rostamzadeh J, Mohammad Abani MR and Lazebnyĭ OE, 2007. Allelic polymorphism of kappa-casein gene (CSN3) in Russian cattle breeds and its informative value as a genetic marker. Genetika 43(1): 88-95.

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

Zhang J, Sun D, Womack JE, Zhang Y, Wang Y and Zhang Y, 2007. Polymorphism identification, RH mapping and association of α-lactalbumin gene with milk performance traits in Chinese Holstein. Asian-Australasian Journal of Animal Science 20(9):1327-1333.

Zhang H, Liu A, Li X, Xu W, Shi R, Luo H, Su G, Dong G, Guo G 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.

Zhou JP and Dong CH, 2013. Association between a polymorphism of the α-lactalbumin gene and milk production traits in Chinese Holstein cows. Genetics and Molecular Research 12(3):3375-3382.

Zhou C, Li C, Cai W, Liu S, Yin H, Shi S, Zhang Q and Zhang S, 2019. Genome-Wide Association Study for Milk Protein Composition Traits in a Chinese Holstein Population Using a Single-Step Approach. Frontiers in Genetics 10:72.

Genome wide association study Gene-set enrichment analysis to identify genome region associated with milk protein composition in Holstein cattle breed

H Mohammadi[1], M Shamsollahi[2] and SH Zhang[3]

Received: December 30, 2022 Accepted: April 10, 2023
[1]Assistant Professor, Department of Animal Sciences, Faculty of Agriculture and Natural Resources, Arak University, Arak, Iran.
[2]Assistant Professor, Department of Animal Sciences, Faculty of Agriculture, University of Ilam. Ilam, Iran
[3] Professor, Key Laboratory of Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, China Agricultural University, Beijing, China
*Corresponding author: H-mohammadi64@araku.ac.ir

پژوهش‌های علوم دامی Animal Science Research
Journal of Animal Science/vol.34 No.1/ 2024/pp 31-44
https://animalscience.tabrizu.ac.ir
OPEN ACCESS

DOI: 10.22034/as.2023.54694.1690

Introduction: Genomic selection has provided the dairy industry with a powerful tool to increase genetic gain in economically important traits such as milk production (Taylor et al. 2016). One way to identify new loci and confirm existing QTL is genome-wide association analysis (GWAA). Furthermore, the identification of gene loci with major impacts on economically important traits is one of the most important goals of dairy cattle breeding. It was hypothesized that QTL-assisted selection and genomic regions affecting production traits increase the efficiency of selection and improve production output. Genome-wide association studies typically focus on genetic markers with the strongest evidence of association. However, individual markers often explain only a small component of genetic variance and therefore provide a limited understanding of the trait under study (Dadousis et al., 2017). One solution to address the above issues and deepen the understanding of the genetic background of complex traits is to move the analysis from the SNP to the gene and gene-set level. In a gene set analysis, a group of related genes harboring significant SNPs previously identified in GWAS are tested for over-representation in a particular signaling pathway. Gene set enrichment (GSE) analysis plays an essential role in extracting biological insight from genome-scale experiments. It reduces the complexity of molecular data and improves the interpretability of biological insights (Peñagaricano et al., 2016).
Material and methods: The present study aimed to perform a genome-wide association study (GWAS) based on gene set enrichment analysis to identify the loci associated with milk protein composition traits. For each cow, a total of eight traits including protein yield, protein percentage, αs1-casein, αs2-casein, β-casein, κ-casein, α-lactalbumin and β-lactoglobulin were recorded using plink software and no any correction to adjust the error rate. Gene set analysis essentially consists of three distinct 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. Briefly, nominal P-values < 0.05 from the GWAS analyzes were used for each trait to identify significant SNPs. Using the *biomaRt2* R package, the SNPs were mapped to genes when located within the genomic sequence of the gene or within a 15 kb flanking region upstream and downstream of the gene to include SNPs located in regulatory regions. The Pathway databases Gene Ontology and Kyoto Encyclopedia of Genes and Genomes were used to assign genes to

functional categories. The GO database labels biological descriptors for genes based on attributes of their encoded products and is further divided into 3 components: biological process, molecular function and cellular component. The KEGG Pathway Database contains metabolic and regulatory pathways that represent the current state of knowledge about molecular interactions and reaction networks. Finally, Fisher's exact test was performed to test the over-representation of the significant genes for each gene set. The gene enrichment analysis was performed with the KOBAS platform. to identify over-represented biological processes. In the next step, a bioinformatic analysis was performed to identify the biological pathways performed in the BioMart, DAVID and GeneCards databases

Results and discussion: Gene-set enrichment analysis has proven to be an excellent complement to genome-wide association analysis (Gambra et al., 2013; Abdalla et al., 2016). Among the available geneset databases, GO is probably the most popular, while KEGG is a relatively new tool gaining ground in livestock genomics (Morota et al., 2015, 2016). We hypothesized that using genset information could improve prediction. However, none of the SNP classes of the gene sets outperformed the standard whole genome approach. Gene sets have been developed primarily 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 to milk production. It is likely that a better understanding of the biology underlying milk production in particular, as well as advances in bovine genome annotation, may provide new opportunities for predicting production using gene set information. According to the gene set enrichment analysis, 20 categories from gene ontology and the KEGG pathway were associated with the associated traits (P0.05). These categories include oxytocin signaling, glycerolipid metabolism, response to progesterone, calcium ion detection, complement and coagulation cascades, and amino acid binding, including the significant association of candidate genes C*DH13, P4HTM, SPP1, CSN1S1, CSN2, SERPINA1, SLC35A3, ODC1,* and *PAEP* with protein yield and content, phosphorylation of glycoproteins, coagulation and curd solidification of milk and lactose synthesis. Some of these regulatory regions, such as B. enhancers, are far removed from the genes. Therefore, although the gene could be part of the analysis, the relevant variant would likely not be included in the SNP class of the gene set. Finally, a linkage disequilibrium disrupts the use of biological information in prediction, since irrelevant regions (regions with no biological role) capture some of the information encoded in relevant regions, giving both regions similar predictive abilities. Using very high-density SNP data or even whole genome sequence data could alleviate some of these problems. Finally, it is worth noting that our gene-set enrichment analysis was performed using a panel of SNPs obtained from a single marker regression GWAS based on a simplified theory of the genomic background of traits, e.g. Ignoring the collective effect of SNP. Therefore, other approaches (e.g. GWAS, which studies SNP through SNP interactions) might provide a better basis for analyzing the biological pathway.

Conclusion: Our result showed a potential for genetic selection to improve milk quality in terms of milk protein composition per animal. Because genetic improvements are heritable, cumulative, and permanent, that improvement would be permanent and beneficial.

Keywords: α-lactalbumin, β-lactoglobulin, casein, candidate gene, pathway analysis