

DOI: 10.22034/AS.2022.37887.1545

تجزیه و تحلیل خوشه‌ای صفت عملکرد چربی شیر در گاوهای شیری با استفاده از فراتحلیل مطالعات پویش ژنومی

سمیه بخشعلی‌زاده^۱، سعید زره‌داران^{۲*} و علی جوادمنش^۳

تاریخ دریافت: ۱۳۹۸/۱۰/۳۰

تاریخ پذیرش: ۱۴۰۰/۱۱/۲۰

^۱دانشجوی دکتری اصلاح نژاد دام گروه علوم دامی دانشکده کشاورزی دانشگاه فردوسی مشهد

^۲استاد گروه علوم دامی دانشگاه فردوسی مشهد

^۳استادیار گروه علوم دامی دانشگاه فردوسی مشهد

*مسئول مکاتبه: Email: zerehdaran@um.ac.ir

چکیده

زمینه مطالعاتی: فراتحلیل، یک روش آماری است که نتایج حاصل از مطالعات متعدد علمی را با هم ترکیب می‌کند. هدف: این مطالعه با بررسی فراتحلیل مطالعات پویش ژنومی (GWAS) در تجزیه و تحلیل خوشه‌ای از طریق تجمیع مطالعات انفرادی، سعی در افزایش صحت نتایج حاصل از مطالعات متعدد و درک بهتر از معماری ژنتیکی صفت مورد مطالعه را دارد. **روش کار:** داده‌ها از ۱۹ مطالعه (مقاله‌ها و پایان‌نامه) از سال ۲۰۱۰ تا ۲۰۱۹ از نژادهای مختلف گاو شیری (نژاد هلشتاین و فریژن، جرسی، سوئیسی قهوه‌ای و قرمز سوئدی) جمع‌آوری شد. با استفاده از روش فراتحلیل، تمام ژن‌های قابل دسترس از خلاصه GWAS (۲۲۳ ژن) با تعاملات پروتئین-پروتئین (PPIs) در شبکه‌های مولکولی با هم ترکیب شدند. تجزیه و تحلیل ژن‌ها با استفاده از پلاگین STRING v1.5.0 در نرم‌افزار Cytoscape v3.7.2 اجرا شد. برای خوشه‌بندی ژن‌ها از پلاگین MCODE v1.5.1 در نرم‌افزار Cytoscape استفاده شد. علاوه بر این، برای تعیین اصطلاحات بیولوژیکی (عملکرد مولکولی، فرآیند بیولوژیکی و اجزای سلولی) مرتبط با مناطق ژنومی و برای شناسایی شبکه‌های ژنی از نرم‌افزار DAVID v6.7 استفاده شد. **نتایج:** PPIs منبع مهمی از اطلاعات مربوط به فرآیندهای بیولوژیکی و عملکردهای متابولیکی را در صفت مربوطه فراهم کرد. شبکه برهمکنش پروتئین‌هایی موثر در تنظیم مقدار عملکرد چربی شیر، ترسیم شدند. در این شبکه حدوداً ۷ خوشه شناسایی شدند. پروتئین‌های موجود در خوشه‌ها (به عنوان مثال در خوشه ۱: ARHGAP39، CPSF1، CYHR1، PPP1R16A، GRINA، MROH1 و SMPD5) در PPIs مورد بررسی قرار گرفتند. خوشه‌های ۱ و ۲ از امتیاز بالایی برخوردار بودند. نتیجه‌گیری کلی: فراتحلیل پویش ژنومی می‌تواند به درک وسیعی از تجسم شبکه و تجزیه و تحلیل خوشه‌ای از ژن‌های شناسایی شده در مسیرهای غنی شده نقش مهمی را ایفا کند. ژن‌های شناسایی شده می‌تواند در ارزیابی‌های ژنومی و برنامه‌های اصلاح نژادی گاوهای شیری برای عملکرد چربی شیر مفید باشد.

واژگان کلیدی: فراتحلیل، GWAS، آنالیز خوشه‌ای، گاوهای شیری

مقدمه

شیر گاو برای تغذیه انسان از اهمیت بالایی برخوردار است. در میان اجزای آن، چربی (۳-۵ درصد) اصلی‌ترین منبع انرژی است و بر طعم شیر و لبنیات اثر می‌گذارد (فنلان و جوینه ۱۹۹۹ و مارتینی و همکاران ۲۰۱۶). میزان چربی یکی از مهمترین عوامل اقتصادی مؤثر بر تولید شیر در گاوهای شیری است (اسپلمن و همکاران ۱۹۹۶). با این حال، کبد در حیوانات نشخوارکننده، از جمله گاوهای شیری نقش مهمی را در سوخت و ساز کربوهیدرات‌ها، چربی‌ها، ویتامین‌ها، هورمون‌ها و سایر موارد دیگر دارد. مواد مغذی جذب شده از دستگاه گوارش از طریق کبد عبور می‌کنند، وارد سامانه گردش خون می‌شوند و سرانجام وارد غدد پستانی گاوهای شیری می‌شوند، بنابراین کبد در دوران شیردهی گاو نقش اساسی دارد (گرابر و همکاران ۲۰۱۰ و اسلیگل و همکاران ۲۰۱۲). با این حال، تمام عواملی که کیفیت شیر را تعیین می‌کنند می‌توانند به عنوان صفات کمی در نظر گرفته شوند که توسط بسیاری از ژن‌ها کنترل می‌شوند و تحت تأثیر عوامل محیطی نیز قرار دارند. بنابراین، اگر نشانگرهای ژنتیکی بتوانند بخش قابل توجهی از تغییرات را توضیح دهند می‌توانند به عنوان کاندیدهای ایده‌آل در انتخاب ژنومی در نظر گرفته شوند (شی و همکاران ۲۰۱۹). در گاوهای شیری برخی از ژن‌های عمده با اثر قابل توجه بر چربی شیر در مطالعات قبلی پویش ژنومی^۱ شناسایی شده است (جیانگ و همکاران ۲۰۱۰، استراکن و همکاران ۲۰۱۲، وانگ و همکاران ۲۰۱۲، مردیت و همکاران ۲۰۱۲، راون و همکاران ۲۰۱۴، شاننا و همکاران ۲۰۱۴، پالامبو و همکاران ۲۰۱۵، ایبکا-آومو و همکاران ۲۰۱۶، ایسو-تورو و همکاران ۲۰۱۶، نیری و همکاران ۲۰۱۶، شین و همکاران ۲۰۱۷ و مریت و همکاران ۲۰۱۸). بنابراین، حجم بسیار زیاد مطالعات GWAS در گاوهای شیری شرایطی را فراهم

می‌کند که با استفاده از روش‌هایی مانند فراتحلیل مطالعات انفرادی تجمیع شده و نتایج با صحت بالاتر حاصل شود. در واقع، فراتحلیل تجزیه و تحلیل آماری است که نتایج حاصل از مطالعات متعدد علمی را با هم ترکیب می‌کند (لین و همکاران ۲۰۰۹، مالت و همکاران ۲۰۱۲، هوویجمن و همکاران ۲۰۱۴ و وسترنین و همکاران ۲۰۱۴). این مطالعات به درک فعلی از تنظیم ژنتیکی صفت عملکرد چربی شیر کمک می‌کند. همچنین، شناسایی تغییرات ژنتیکی حاصل از جهش‌های مؤثر بر چربی شیر که قبلاً شناسایی شده‌اند از طریق تجزیه و تحلیل شبکه ژنی امکان‌پذیر شده است (فورتز و همکاران ۲۰۱۰ و ریورتر و فورتز ۲۰۱۳). این روش، درک بهتری از معماری ژنتیکی صفات پیچیده را فراهم می‌کند. با این حال، الگوریتم خوشه‌بندی شبکه و شناسایی خوشه‌ها به عنوان یک ابزار مهم در تجزیه و تحلیل ساختاری شبکه‌ها است. ژن‌ها یا گره‌ها^۲ ممکن است بر اساس توپولوژی شبکه گروه‌بندی شوند. گروهی از گره‌های بهم پیوسته که ممکن است یک خوشه را تشکیل دهند، گروه‌های پروتئینی را که دارای عملکرد مشابهی هستند را مشخص می‌کند. بنابراین، رویکردهای خوشه‌بندی برای PPIs^۳ را می‌توان به صورت گسترده‌ای مبتنی بر فاصله یا مبتنی بر نمودار طبقه‌بندی کرد. خوشه‌بندی مبتنی بر فاصله از روش‌های کلاسیک خوشه‌بندی استفاده می‌کند و بر تعریف فاصله بین پروتئین‌ها تمرکز دارد. در حالی که، روش‌های مبتنی بر نمودار شامل رویکردهایی است که توپولوژی شبکه‌های PPIs را در نظر می‌گیرد (لین و همکاران ۲۰۰۹). در مطالعه حاضر، رویکرد مبتنی بر نمودار اتخاذ گردید. شبکه PPIs یک نمودار بدون وزن است که در آن وزن هر لبه^۴ بین هر یک از دو پروتئین صفر یا یک است که به صورت آشکارا در نمودار ارائه می‌شود. بنابراین، فرآیند خوشه‌بندی یک مجموعه داده را به چنین نموداری تبدیل

¹ Genome-Wide Association Studies (GWAS)

² Nodes

³ Protein-protein interactions (PPIs)

⁴ Edge

ژن‌ها را برای هر مطالعه ارزیابی و داده‌های جمع‌آوری شده به صورت گسترده مورد تنظیم قرار گرفتند که شامل بررسی کلیه ژن‌های مورد بررسی در هر مطالعه بود. با استفاده از روش فراتحلیل نیز تمام ژن‌های قابل دسترس با هم ترکیب، سنتز و ارزیابی شدند. در نهایت، نتایج حاصل از فرا تحلیل GWAS با فعل و انفعالات PPIs در شبکه‌های مولکولی ترکیب گردید که سهم بسزایی در افزایش توان آماری در شناسایی ژن‌های مؤثر بر عملکرد چربی شیر دارد (جیا و ژائو ۲۰۱۴). تعداد ژن‌های به اشتراک گذاشته شده در این مطالعه در حدود ۲۲۳ ژن بود که تجزیه و تحلیل و تجسم این ژن‌ها با استفاده از پلاگین STRING v1.5.0 (دونچوا و همکاران ۲۰۱۹) در نرم افزار Cytoscape v3.7.2 (شانون و همکاران ۲۰۰۳) اجرا شد. از پلاگین MCODE v1.5.1 (بادر و هوگ ۲۰۰۳) و سایتو و همکاران (۲۰۱۲) برای پیش‌بینی مجموعه پروتئین-پروتئین در کل داده‌های شبکه PPIs ادغام شده استفاده شد.

این روش مناطق متراکم و متصل را با وزندهی گره‌ها بر اساس تراکم موقعیت محلی آنها تشخیص می‌دهد (بادر و هوگ ۲۰۰۳). این روش وزندهی یک نوع روش برداری از طریق تراکم موقعیت‌های مجاور است که چندین هسته‌آرا بر اساس وزن زیاد انتخاب می‌کند و مناطق متراکم را طبق پارامترهای داده شده جدا می‌کند. این روش در سه مرحله انجام می‌شود: (۱) وزندهی برداری (۲) پیش‌بینی مجموعه‌های پیچیده (۳) پیش پردازش برای فیلتر یا اضافه کردن پروتئین به مجموعه‌های حاصل با توجه به معیارهای اتصالی خاص. نتیجه شبکه‌های تعاملات پروتئین در قالب یک نمودار نشان داده می‌شوند که در مرکز آنها پروتئین‌ها یا گره‌ها هستند و تعامل‌ها تعاملات پروتئین‌ها را تشکیل می‌دهند. در واقع، نمایش گرافیکی از سیستم‌های بیولوژیکی، روش‌های نظری را در تجزیه و تحلیل برای یافتن خوشه‌ها در نمودار امکان‌پذیر می‌نماید (واتس و استروگاتز ۱۹۹۸). علاوه بر این، از نرم

می‌کند (گولپالی و همکاران ۲۰۱۵). بسیاری از الگوریتم‌های خوشه‌بندی در انواع مختلف برای تجزیه و تحلیل شبکه‌های PPIs استفاده می‌شوند. در این مطالعه از الگوریتم توسعه یافته معروف به MCODE^۱ برای شناسایی مناطق متراکم از یک نمودار PPIs استفاده شده است (بادر و هوگ ۲۰۰۳). هدف کلی از خوشه‌بندی شبکه PPIs، گروه‌بندی ژن‌ها یا پروتئین‌هایی است که بر اساس برخی از مقیاس‌ها مرتبط هستند. شبکه PPIs حاوی پروتئین‌های مختلفی است که در مسیرهای مختلف نقش ایفا می‌کنند، زیرا این ژن‌ها یا پروتئین‌ها بر اساس شباهت متریک خوشه‌بندی می‌شوند و به عنوان فاصله ماتریس شناخته می‌شوند. پیش‌بینی مجموعه‌های مولکولی از داده‌های تعامل پروتئین مهم است زیرا سطح دیگری از حاشیه‌نویسی عملکردی را فراهم می‌کند (گولپالی و همکاران ۲۰۱۵). محمدی و همکاران (۲۰۱۸) با استفاده از تجزیه و تحلیل خوشه‌ای در نرم‌افزار Cytoscape موفق به شناسایی ژن‌های مهم در فرآیندهای بیولوژیکی در بافت عضله جنین گوسفند شدند. هدف از انجام این مطالعه فراتحلیل GWAS در تجزیه و تحلیل خوشه‌ای به منظور شناسایی ژن‌های مؤثر در تولید چربی شیر در گاوهای شیری بود.

مواد و روش‌ها

در این مطالعه، داده‌های مورد استفاده مربوط به گردآوری ژن‌های قابل دسترس در خلاصه‌های آماری GWAS بود (جدول ۱). بدین منظور طیف وسیعی از اطلاعات مربوط به ژن‌ها از ۱۹ مطالعه منتشر شده از سال ۲۰۱۰ تا ۲۰۱۹ در جمعیت‌های مختلفی از نژادهای گاو شیری (نژاد هولشتاین و فریژن، جرسی، سوئیسی قهوه‌ای و قرمز سوئدی) جمع‌آوری گردید. این تحقیق شامل مقالات اصلی و پایان‌نامه (پایان‌نامه‌های معتبر با مقالات چاپ شده) بود. به طور دستی طیف گسترده‌ای از

¹ Molecular Complex Detection (MCODE)

² Seeds

افزار DAVID v6.7 (هوانگ و همکاران ۲۰۰۹) هم برای خوشه‌بندی حاشیه‌نویسی کاربردی مبتنی بر آنالیز تعیین اصطلاحات بیولوژیکی غنی شده مرتبط با مناطق ژنومی و برای شناسایی شبکه‌های ژنی از ابزار مسیریابی‌های غنی‌شده استفاده شد.

Table 1- A summary of the gene list related to the genome-wide association studies (GWAS)

Reference	Gene
Jiang et al. 2010	CYHR1, DGAT1, PPP1R16A
Cole et al. 2011	FUT10, LPP, PGLYRP1, ZBP1, FKBP2, GNAS, HHIP, MGMT, TTC14
Strucken et al. 2012	ERICH3, TTC27
Chamberlain et al. 2012	GHR, GRK5
Guo et al. 2012	ECI1, HS3ST6, MLST8
Meredith et al. 2012	ADAM19, AGO2, ALCAM, ARHGEF16, ASIC2, ATP10B, BIRC6, BRINP2, C10H15orf41, C14H8orf33, C1QTNF3, CACNA1E, CDC42BPA, CLK4, CLNK, CNOT6L, CNTN6, CNTNAP5, COL22A1, CPS1, DISC1, EDIL3, EEF1D, EEPD1, EPPK1, ESRRB, EYA4, FAM102A, FAM114A1, FCRL3, FCRLA, FGGY, FMN2, FTSJ3, FYB1, GNG2, GPC6, GPIHBP1, H2AFY, HOPX, IGFBP7, KCNK9, KCNN2, KHDRBS3, KLHL2, MALT1, MAML3, MED24, MIR2360, MTF, MPP2, MS4A10, MYRIP, NELFCD, NIPBL, NLK, NNT, NPY2R, PDE4D, PIP5KL1, PLXNA4, POLR2B, PRKRIP1, PTX4, PUM1, RBM47, DIEXF, RELL2, APP, RER1, RHPN1N, RPF2, RYR2, SDE2, SEMA6A, SGK1, SLA, SLC6A9, SPOCK1, SRPK2, EIF2C2, SYNM, SYNPR, TERT, TET1, TNRC6A, TRHDE, TRIM9, TSHZ2, UBR4, ULK4, USH2A, ZNF696, AOX4, ATP2B2, KNG1, C18H16orf46, CACHD1, CAPZB, CSMD2, FAT3, GML, HIVEP3, KCNH8, MARK2, MEIS2, MSANTD2, OXCT1, PID1, PRKCE, SATB1, SCARB2, SLC25A12, SLC25A48, SOST, SPHKAP, SSBP2, TENM2, USP44, VDAC1, XKR4
Zielke et al. 2013	RPGRIPL, FTO
Minozzi et al. 2013	EPS8, GPR20, LY6K, MAF1, PTK2, SCRIB, TRAPPC9, ZC3H3, ZNF623, NINJ2, PLBD1, DENND3
Tesfayonas 2014	MROH1
Jiang et al. 2014	FAM135B, GPT, MAPK15
Raven et al. 2014	BTRC, MGST1
Fang et al. 2014	MIR2388, SYN3
Spehar 2015	PAEP, PRKD1, SLC8A1, THRB, ZNF385D, CHD4, IFFO1
Iso-Touru et al. 2016	HSF1, CLEC16A, NEURL1
Nayeri et al. 2016	ZNF16, CPSF1, GRINA, JRK, PSCA, SLURP1, SMPD5, ZNF34, ZNF7, ARHGAP39, EXOSC4, PARP10, RPL8, TONSL
Shin et al. 2017	LDLRAD3, ACADM, ASB17, NABP1, RABGGTB, TMEFF2
Marete et al. 2018	AQP7, FCHSD2, GOLGA7, GOT1, XPO5, KDM4B, LY6G5B, MYO3B, NMI, NOS1AP, STK3, CCDC57, ABCC9, ADARB2, FBX011, GRIN3B, HERC1, LURAP1L, MYOF, NEUROD1, OR6N2, PKNOX2, PREP, PRKDC, RAB1A, STAT5B, USP40, ZFH3
Cecchinato et al. 2019	EPHA5, TECRL
Cai et al. 2019	COX15

نتایج و بحث

تجزیه و تحلیل شبکه و خوشه‌های ژنی

۲۱۳ ژن یا گره و ۲۱۹ تعامل (ارتباط ژنی) بود و شبکه حاصل به صورت یک نمودار نمایش داده شد (شکل ۱) که گره‌ها و تعامل‌ها در شبکه حاصل بهم متصل شده بودند. شبکه PPIs به عنوان مقیاسی است که از قانون توزیع اتصال پیروی می‌کند و مقدار P-value محاسبه شده در شبکه (با استفاده از پلاگین STRING) برای مسیرهای غنی شده در PPIs از لحاظ آماری معنی دار بود (

به طور کلی ۲۲۳ ژن با استفاده از پلاگین STRING در نرم افزار Cytoscape مورد تجزیه و تحلیل قرار گرفت. همچنین، این ژن‌ها حداقل با یک ژن دیگر در ارتباط بوده و دارای همبستگی مستقیم و جزئی بودند. در این شبکه ژنی همبستگی ایجاد شده برای مقدار عملکرد چربی شیر شامل

($P = 2.6e - 09$). مجموعه ژن‌های مهم و مورد علاقه با استفاده از پلاگین MCODE مورد بررسی قرار گرفتند. بر اساس پارامترهای مشخص شده، این الگوریتم ۷ خوشه را پیش‌بینی کرد. خوشه‌هایی که با این روش یافت می‌شوند باید با مشخص کردن یک راس seed، در حالت غیرمستقیم مورد مطالعه قرار گیرند. با این حال، خوشه‌های بزرگ و مترکم توسط MCODE در رتبه بالاتری قرار گرفتند که معمولاً با مجموعه‌های شناخته شده مطابقت دارند. برخی از خوشه‌های MCODE حاوی پروتئین‌های ناشناخته‌ای هستند که بسیاری از آنها به زیر واحدهای پیچیده شناخته شده متصل هستند. احتمال انتخاب تصادفی یک مجموعه‌ی پروتئینی (خوشه) شناخته شده از تعاملات پروتئینی به اندازه خوشه و شبکه بستگی دارد. در شبکه، نحوه انتخاب مدل‌های خوشه‌ای به توپولوژی شبکه بستگی دارد. با افزایش اندازه خوشه، تعداد توپولوژی‌های پیچیده ممکن است به صورت یک شبکه متصل با تراکم مناسب افزایش یابد، به همین ترتیب تعداد زیرگراف‌های احتمالی که می‌توانند نمایانگر یک مجموعه باشند نیز افزایش می‌یابد (قولپالی و همکاران ۲۰۱۵). مطالعات اخیر در تحقیقات مربوط به مدل‌سازی سیستم‌های پیچیده نشان داده‌اند که شبکه‌های PPIs عاری از مقیاس هستند (وانگر و فل ۲۰۰۱ و گالدرج ۱۹۸۴). شبکه‌های بدون مقیاس از ضرایب خوشه‌بندی بزرگ در نمودار برخوردار هستند (قولپالی و همکاران ۲۰۱۵). در این مطالعه، الگوریتم استفاده شده مانند MCODE مناطق خوشه‌دار را پیدا می‌کند. خوشه‌های به دست آمده با اندازه‌های مختلف، پیچیدگی و اهمیت بیولوژیکی مرتبط با آنها را در PPIs نشان می‌دهد. در مجموع، ۷ خوشه بر اساس اندازه و تراکم از کل شبکه به دست آمد و رتبه‌بندی شد (جدول ۲). خوشه‌بندی پروتئین‌ها با استفاده از الگوریتم MCODE در شکل ۲ نمایش داده شده است که به ترتیب هر یک از خوشه‌های موجود به وضوح توضیح

داده می‌شوند. مجموعه ۱ با تراکم ۷ گره متصل به ۱۸ تعامل و با امتیاز ۷ در رتبه ۱ قرار گرفت. مشاهده شد که، خوشه یک برترین خوشه رتبه‌بندی شده در بین همه خوشه‌های موجود بود. این خوشه پروتئین‌هایی را نشان می‌دهد که در فضای داخلی شبکه آندوپلاسمی، اتصال یون‌های فلزی و انتگرال غشایی نقش‌های مهمی را دارند. با این حال، فضای داخلی شبکه آندوپلاسمی و انتگرال غشایی جز اجزای سلولی^۱ هستند. در حالی که، اتصال یون‌های فلزی جز عملکرد مولکولی^۲ است. CPSF^۱ ژن عمده‌ای است که در فضای داخلی شبکه آندوپلاسمی دخیل است. در دوران شیردهی، سلول‌های اپیتلیال ترشحی پستان بسیار متمایز شده و در ساختارهای آلئول توسط سلول‌های میوپیتلیال انقباضی احاطه می‌شوند و در استروما (بافت‌های همبند چربی، رگ‌های خونی و پایانه‌های عصبی) احاطه شده و سازمان‌دهی می‌شوند. سلول‌های اپیتلیال ترشحی پستان مقادیر زیادی شیر تولید می‌کنند، که یک مایع حاوی پروتئین‌ها، گلبول‌های چربی شیر و اجزای محلول مانند لاکتوز و مواد معدنی است (تورنر و همکاران ۱۹۹۲). اگرچه سازوکارهای مختلفی ارائه شده است، اما مدل غالب این است که اجزای لیپید سیتوپلاسمی با تجمع لیپیدهای خنثی (عمدتاً تری‌گلیسرول و استرول) و کلاسترول بین دو جزء غشای آندوپلاسمی تشکیل می‌شوند (والدر و همکاران ۲۰۱۷). بنابراین، لیپید سیتوپلاسمی توسط یک لایه فسفولیپید حاوی کلاسترول که کاملاً متمایز از غشای آندوپلاسمی یا میکرومین‌های غنی از اسفنگولیپید یا کلاسترول است، احاطه شده است. بنابراین، پروتئین دخیل در این فرآیند با اتصال غشاهای داخل سلولی اهمیت زیادی را در ترشح چربی شیر دارد. لی و همکاران (۲۰۰۶) گزارش کردند که اتصال یون‌های فلزی در سلول نقش مهمی را در کنترل هورمونی و تکثیر و تمایز سلولی دارد (لی و همکاران ۲۰۰۶). به احتمال زیاد کنترل سیگنالینگ

¹ Cellular components

² Molecular function

³ Cleavage and polyadenylation specific factor 1

یون‌های فلزی اثرات عمیقی بر فیزیولوژی پستانداران خواهد گذاشت (شان و پکر ۲۰۰۰). چندین ژن در این خوشه شناسایی شده است. به عنوان مثال، ژن CYHR^۱ در اتصال یون‌های فلزی نقش مهمی را دارد. همچنین، پروتئین GRIN^۲ عمده‌ای است که در انتگرال غشایی (پروتئین داخلی یا تراغشایی) دخیل است که نقش مهمی را در فرآیند تشکیل چربی شیر دارد. در واقع، تشکیل چربی‌ها با بسته‌بندی تری‌گلیسرول‌ها به قطرات میکرولیپیدی در شبکه آندوپلاسمی صورت می‌گیرد که بهم می‌پیوندند و از سلول‌های اپیتلیال غدد پستانی جوانه می‌زنند و به شکل قطرات لیپیدی سیتوپلاسمی توسط یک لایه فسفولیپید احاطه می‌شوند. مهاجرت قطرات لیپیدی سیتوپلاسمی به قطب آپیکال سلول اپیتلیال منجر به همجوشی با غشای پلاسمایی و اضافه شدن دو لایه محیطی می‌شود که شامل انواع پروتئین‌های فعال زیستی و چربی‌ها است (پارک ۲۰۰۹). گلبولهای چربی به صورت کاملاً مسطح توسط غشاء پوشانده می‌شوند و در نتیجه چربی شیر در خارج از سلول ترشح می‌شود (پارک ۲۰۰۹ و اسموشینیسکی ۲۰۱۷). خوشه ۲، نشان‌دهنده پروتئین‌های موجود در مسیر پروتئولیز درگیر در فرآیند کاتابولیک پروتئین سلولی (فرآیند بیولوژیکی) و اتصال یونی (عملکرد مولکولی) است. این خوشه شامل ۶ گره و

۱۵ تعامل است. بر اساس نتایج به دست آمده مشاهده شد که پروتئین‌های HERC^۱ و ASB17^۳ نقش مهمی را در پروتئولیز درگیر در فرآیند کاتابولیک پروتئین سلولی دارند. گزارش شده که در طی فرآیند پروتئولیز، پیوندهای پپتیدی که اسیدهای آمینه را پیوند می‌دهند هیدرولیز می‌شوند و اسیدهای آمینه آزاد می‌شوند. این فرآیند توسط یک گروه متنوع از آنزیم‌هایی (HERC1 و ASB17) که پروتئازها نامیده می‌شوند، انجام می‌شود. در طول پروتئولیز، انرژی موجود در سنتز پیوند پپتیدی آزاد می‌شود، وقتی اسیدهای آمینه آزاد می‌شوند می‌توانند برای تولید انرژی در بدن و یا چربی‌ها مورد استفاده قرار بگیرند که این عوامل می‌تواند در فرآیند متابولیسم چربی شیر نقش داشته باشد (گلیکمن و سیچانور ۲۰۰۲). همچنین، در این خوشه پروتئین TRIM9^۴ به عنوان تنظیم‌کننده اتصال یونی است که در فرآیند مختلف مولکولی در عملکرد چربی شیر نقش بسزایی دارد. خوشه ۳، بیانگر اکثر پروتئین‌های درگیر در هموستاز یونی (فرآیند بیولوژیکی)، غشای پلاسمایی (اجزای سلولی) و تنظیم رشد (فرآیند بیولوژیکی) است که شامل ۴ گره و ۶ تعامل است. پروتئین KNG1^۵ از نظر عملکردی نقش مهمی را در هموستاز یونی برای عملکرد چربی شیر دارد.

¹ Cysteine/histidine-rich 1

² Glutamate ionotropic receptor NMDA type subunit associated protein 1

³ Biological process

⁴ HECT and RLD domain containing E3 ubiquitin protein ligase family member 1

⁵ Ankyrin repeat and SOCS box containing 17

⁶ Tripartite motif containing 9

⁷ Kininogen 1

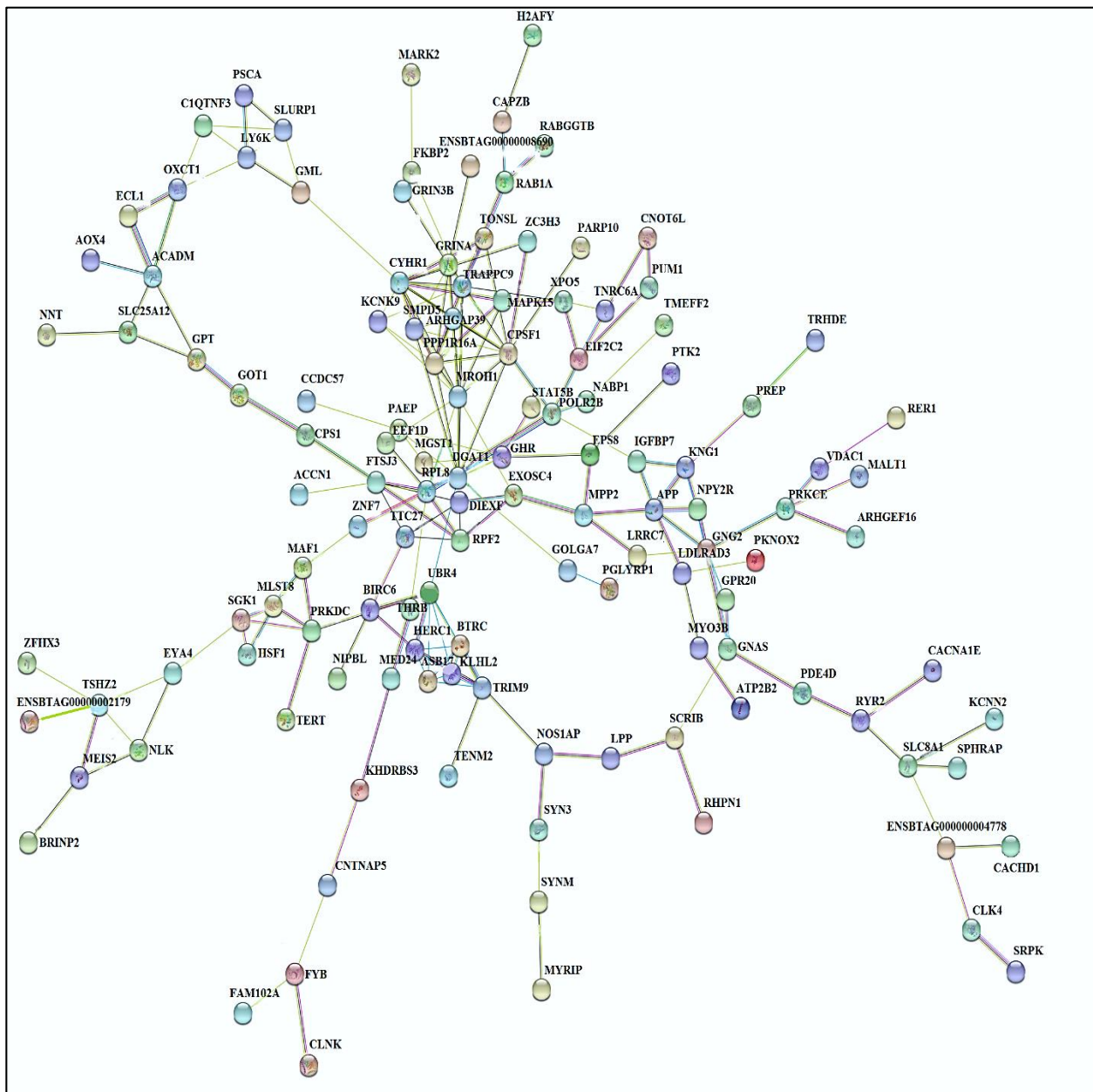


Figure 1- Protein-protein interaction (PPI) network for milk fat yield. The network analysis was performed with a high level of significance, and the transverse lines represent the levels of interaction (wider lines indicate stronger evidence of interactions)

Table 2- Statistics of the seven clusters predicted by the MCODE algorithm in the protein-protein interaction (PPI) dataset on milk fat yield

Cluster Rank	Score	Proteins (nodes)	Interaction (edges)	Interacting protein names	Protein Function
1	7	7	18	ARHGAP39, CPSF1, CYHR1, PPP1R16A, GRINA, MROH1, SMPD5	Internal space of the endoplasmic reticulum, metal ion binding, integral to membrane
2	6	6	15	HERC1, UBR4, ASB17, TRIM9, KLHL2, BTRC	Proteolysis involved in cellular protein catabolic process, ion binding
3	4	4	6	KNG1, NPY2R, GNG2, APP	Ion homeostasis, plasma membrane part, regulation of growth

4	4	4	6	TTC27, DIEXF, FTSJ3, RPF2	Internal space of the endoplasmic reticulum
5	3.3	4	5	MEIS2, NLK, EYA4, TSHZ2	Positive regulation of biosynthetic process, protein kinase activity, ion binding
6	3	3	2	MGST1, PAEP, GHR	Mitochondrial membrane, Extracellular region, oxaloacetate metabolic process
7	3	3	2	TNRC6A, XPO5, EIF2C2	Nucleotide binding

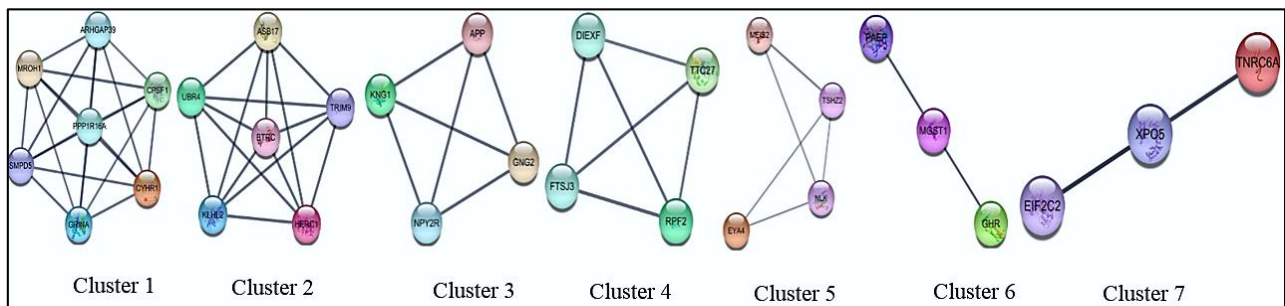


Figure 2- Clusters ranked using the MCODE algorithm for milk fat yield. From left to right represents clusters 1, 2, 3, 4, 5, 6, and 7

تعامل است که نشان‌دهنده فضای داخلی شبکه آندوپلاسمی (اجزای سلولی) است. پروتئین^۴ RPF2 در این خوشه از طریق اثر بر فضای داخلی شبکه آندوپلاسمی نقش مهمی را در عملکرد چربی شیر دارد. خوشه ۵ شامل ۴ گره و ۵ تعامل است، که پروتئین‌های مربوط به این خوشه مرتبط با پردازش تنظیم مثبتی از فرآیند بیوسنتز چربی شیر، فعالیت پروتئین کیناز و اتصالات یونی است. پروتئین درگیر (MEIS^۲) در فرآیند بیوسنتز چربی شیر در فرآیندهای بیولوژیکی نقش دارد. در این مطالعه پروتئین^۵ NLK به عنوان تنظیم‌کننده فعالیت پروتئین کیناز است که در مسیر عملکرد مولکولی نقش دارد. در مطالعات قبلی گزارش شده است که پروتئین کیناز نقش کلیدی را در هموستاز انرژی بدن ایفا می‌کند (کان و همکاران ۲۰۰۵). مطالعات بالینی نشان داده است که پروتئین کیناز از طریق پاسخ به هورمون‌ها و سیگنال‌های مغزی نقش مهمی را در متابولیسم چربی کبد دارد (هاردی ۲۰۰۳ و

پروتئین‌های NPY2R و GNG^۲ در مسیرهای مختلف غشای پلاسمایی نقش بسزایی دارند. همچنین، پروتئین APP به عنوان تنظیم‌کننده رشد در فرآیند بیولوژیکی است که اثر زیادی را روی هورمون‌ها می‌گذارد، مشاهده شده است که تعادل منفی انرژی در اوایل شیردهی، تغییرات عمده‌ای در تنظیم هورمونی در گاوهای شیری پر تولید ایجاد می‌کند (جوکوویچ و همکاران ۲۰۱۴)، که شامل تغییر در غلظت هورمون‌های کلیدی نظیر هورمون رشد و هورمون‌های غده تیروئید است. غلظت هورمون‌های بیان شده در خون نقش مهمی در رشد غدد پستانی دارد (بومن و بروس کوری ۱۹۸۰). به عنوان مثال هورمون رشد هیپوفیز، هورمون‌های غده تیروئید، انسولین، کاکتول آمین‌ها و لپتین‌ها نمونه‌هایی از هورمون‌های تنظیم‌کننده غدد درون‌ریز در متابولیسم چربی‌ها است (بومن و بروس کوری ۱۹۸۰، روچ و همکاران ۲۰۰۹ و جوکوویچ و همکاران ۲۰۱۴). خوشه ۴ شامل ۴ گره و ۶

¹ Neuropeptide Y receptor Y2

² G protein subunit gamma 2

³ Amyloid beta precursor protein

⁴ Ribosome production factor 2 homolog

⁵ Meis homeobox 2

⁶ Nemo like kinase

مطالعه مشاهده شد که خوشه‌های ۱ و ۲ در بین همه‌ی خوشه‌های گزارش شده از امتیاز بالایی برخوردار هستند.

نتیجه‌گیری

این نتایج نشان می‌دهد که با استفاده از داده‌های منابع مختلف می‌توان قابلیت اطمینان و صحت مطالعات GWAS را افزایش داد. ما با استفاده از فراتحلیل داده‌های خلاصه GWAS در تجزیه و تحلیل خوشه‌های توانستیم مهمترین ژن‌ها را در مسیریابی‌های شبکه شناسایی کنیم. مهمترین ژن‌ها با امتیاز بالا در خوشه یک (ARHGAP39، CPSF1، CYHR1، PPP1R16A، GRINA، MROH1 و SMPD5) و دو (HERC1، UBR4، ASB17، TRIM9، KLHL2 و BTRC) شناسایی شدند. بنابراین، ژن‌های مهم شناسایی شده می‌تواند در آینده برای ارزیابی‌های ژنومی و برنامه‌های اصلاح نژادی در مجموعه‌ای از گاوهای شیری برای مقدار چربی شیر مفید باشد.

سپاس‌گزاری: نویسندگان مراتب تقدیر و تشکر خود را از حمایت دانشگاه فردوسی مشهد (طرح پژوهشی: ۳/۴۸۷۰۸) ابراز می‌نمایند.

ویولت و همکاران (۲۰۰۶) که می‌تواند اثرات مستقیمی در متابولیسم چربی شیر داشته باشد. همچنین، پروتئین TSHZ2 در اتصالات یونی در مسیریابی عملکرد مولکولی نقش دارد. خوشه‌ی ۶ دارای ۳ گره و ۲ تعامل است، که شامل ژن‌های MGST1، PAEP و GHR می‌باشد. پروتئین MGST1 نقش مهمی را در غشای میتوکندری در مسیر اجزای سلولی دارد و پروتئین PAEP به عنوان تنظیم‌کننده خارج سلولی در فرآیند اجزای سلولی است. همچنین، پروتئین GHR نقش مهمی را در متابولیسم اگزوالوستات در سنتز پيرووات در چرخه کربس در مسیریابی فرآیند بیولوژیکی دارد که با آزاد کردن ATP نقش مهمی را در سنتز چربی شیر ایفا می‌کند. این یافته‌ها با نتایج قبلی مطابقت دارد (وایت ۲۰۱۵). خوشه ۷ هم شامل ۳ گره و ۲ تعامل است. پروتئین TNRC6A^۴ در اتصال نوکلئوتیدی در مسیر عملکرد مولکولی نقش مهمی دارد. بر اساس یافته‌های قبلی مشاهده شده است که بهترین امتیازدهی خوشه‌ای (Score > 5) برای حاشیه‌نویسی ژن و اصطلاحات آنتولوژی ژن (GO) در فرآیندهای بیولوژیکی قابل توجه هستند (گولپالی و همکاران ۲۰۱۵). در این

منابع مورد استفاده

- Bader GD and Hogue CWV, 2003. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4(1): 2.
- Bauman DE and Bruce Currie W, 1980. Partitioning of Nutrients During Pregnancy and Lactation: A review of mechanisms involving homeostasis and homeorhesis. *Journal of Dairy Science* 63(9): 1514-1529.
- Cai Z, Guldbbrandtsen B, Lund MS, Sahana G, 2019. Dissecting closely linked association signals in combination with the mammalian phenotype database can identify candidate genes in dairy cattle. *BMC Genetics* 20: 1–12.
- Cecchinato A, Macciotta NPP, Mele M, Tagliapietra F, Schiavon S, Bittante G and Pegolo S, 2019. Genetic and genomic analyses of latent variables related to the milk fatty acid profile, milk composition, and udder health in dairy cattle. *Journal of Dairy Science* 102(6): 5254–5265.
- Chamberlain AJ, Hayes BJ, Savin K, Bolormaa S, McPartlan HC, Bowman PJ, Van Der Jagt C, MacEachern S, Goddard ME, Pryce J, Chamberlain AJ, Bowman PJ and Goddard M E, 2012. Validation of single

¹ Teashirt zinc finger homeobox 2

² Microsomal glutathione S-transferase 1

³ Progesterone associated endometrial protein

⁴ Growth hormone receptor

⁵ Trinucleotide repeat containing adaptor 6A

⁶ Gene ontology

- nucleotide polymorphisms associated with milk production traits in dairy cattle. *Journal of Dairy Science* 95(2): 864–875.
- Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, Van Tassell CP, Yang J, Wang S, Matukumalli LK and Da Y, 2011. Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary U.S. Holstein cows. *BMC Genomics* 12: 1–17.
- Djokovic R, Cincovic M, Kurcubic V, Petrovic M, Lalovic M, Jasovic B and Stanimirovic Z, 2014. Endocrine and metabolic status of dairy cows during transition period. *The Thai Journal of Veterinary Medicine* 44(1): 59-66.
- Doncheva NT, Morris J H, Gorodkin J and Jensen LJ. 2019. Cytoscape stringApp: network analysis and visualization of proteomics data. *Journal of Proteome Research* 18 (2): 623–632.
- Esposito G, Masucci F, Napolitano F, Braghieri A, Romano R, Manzo N and Francia, A, 2014. Fatty acid and sensory profiles of Caciocavallo cheese as affected by management system. *Journal of Dairy Science* 97(4): 1918-1928.
- Fang M, Fu W, Jiang D, Zhang Q, Sun D, Ding X and Liu J, 2014. A multiple-SNP approach for genome-wide association study of milk production traits in Chinese Holstein cattle. *Plos One* 9(8): e99544.
- Fenelon MA and Guinee TP, 1999. The effect of milk fat on Cheddar cheese yield and its prediction, using modifications of the Van Slyke cheese yield formula. *Journal of Dairy Science* 82(11): 2287-2299.
- Fortes MRS, Reverter A, Zhang Y, Collis E, Nagaraj SH, Jonsson NN, Prayaga KC, Barris W and Hawken RJ, 2010. Association weight matrix for the genetic dissection of puberty in beef cattle. *Proceedings of the National Academy of Sciences* 107(31): 13642-13647.
- Galderg AV, 1984. Finding maximum density subgraph. Technical report UCB/CSD 84-171.
- Glickman MH and Ciechanover A, 2002. The ubiquitin-proteasome proteolytic pathway: Destruction for the sake of construction. *Physiological Reviews* 82(2): 373-428.
- Gollapalli P, Hanumanthappa M and Pattar S, 2015. Cluster analysis of protein-protein interaction network of mycobacterium tuberculosis during host infection. *Advances in Bioresearch* 6(5):38–46.
- Graber M, Kohler S, Kaufmann T, Doherr MG, Bruckmaier RM and van Dorland HA, 2010. A field study on characteristics and diversity of gene expression in the liver of dairy cows during the transition period. *Journal of Dairy Science* 93(11): 5200-5215.
- Guo J, Jorjani H, Carlborg Ö, 2012. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. *BMC Genetics* 13:82.
- Hardie DG, 2003. Minireview: The AMP-activated protein kinase cascade: the key sensor of cellular energy status. *Endocrinology* 144(12): 5179-5183.
- Hooijmans CR, IntHout J, Ritskes-Hoitinga M and Rovers MM, 2014. Meta-analyses of animal studies: An introduction of a valuable instrument to further improve healthcare. *ILAR Journal* 55(3): 418-426.
- Huang DW, Sherman BT and Lempicki RA, 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.
- Ibeagha-Awemu EM, Peters SO, Akwanji KA, Imumorin IG and Zhao X, 2016. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. *Scientific Reports* 6: 31109.
- Iso-Touru T, Sahana G, Guldbbrandtsen B, Lund MS and Vilkki J, 2016. Genome-wide association analysis of milk yield traits in Nordic Red Cattle using imputed whole genome sequence variants. *BMC Genetics* 17(1): 1–12.
- Jia P and Zhao Z, 2014. Network-assisted analysis to prioritize GWAS results: Principles, methods and perspectives. *Human Genetics* 133(2): 125-138.
- Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y and Zhang Q, 2010. Genome wide association studies for milk production traits in Chinese Holstein population. *Plos One* 5(10): e13661.

- Jiang L, Liu X, Yang J, Wang H, Jiang J, Liu L, He S, Ding X, Liu J and Zhang Q, 2014. Targeted re-sequencing of GWAS loci reveals novel genetic variants for milk production traits. *BMC Genomics* 15: 1–9.
- Kahn BB, Alquier T, Carling D and Hardie DG, 2005. AMP-activated protein kinase: Ancient energy gauge provides clues to modern understanding of metabolism. *Cell Metabolism* 1(1): 15-25.
- Lean IJ, Rabiee AR, Duffield TF and Dohoo IR, 2009. Invited review: Use of meta-analysis in animal health and reproduction: Methods and applications. *Journal of Dairy Science* 92(8): 3545-3565.
- Lee WJ, Monteith GR and Roberts-Thomson SJ, 2006. Calcium transport and signaling in the mammary gland: Targets for breast cancer. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1765(2): 235-255.
- Lin C, Cho YR, Hwang WC, Pei P and Zhang A, 2006. Clustering methods in a protein-protein interaction network. *Knowledge Discovery in Bioinformatics: techniques, methods and application* 1-35.
- Mallett R, Hagen-Zanker J, Slater R and Duvendack M, 2012. The benefits and challenges of using systematic reviews in international development research. *Journal of Development Effectiveness* 4(3): 445-455.
- Marete AG, Gulbrandsen B, Lund MS, Fritz S, Sahana G and Boichard D, 2018. A meta-analysis including pre-selected sequence variants associated with seven traits in three french dairy cattle populations. *Frontiers in Genetics* 9: 522.
- Martini M, Salari F and Altomonte I, 2016. The macrostructure of milk lipids: the fat globules. *Critical Reviews in Food Science and Nutrition* 56(7): 1209-1221.
- Meredith BK, Kearney FJ, Finlay EK, Bradley DG, Fahey AG, Berry DP and Lynn DJ, 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genetics* 13(1): 21.
- Minozzi G, Nicolazzi EL, Stella A, Biffani S, Negrini R, Lazzari B, Ajmone-Marsan P and Williams JL, 2013. Genome wide analysis of fertility and production traits in Italian Holstein cattle. *Plos One* 8 (11): 1–10.
- Mohammadi F, Tahmoorespur M and Javadmanesh A, 2018. Study of differentially expressed genes, related pathways and gene networks in sheep fetal muscle tissue in thin- and fat-tailed breeds. *Animal Science Journal* 123:301-312.
- Nayeri S, Sargolzaei M, Abo-Ismael MK, May N, Miller SP, Schenkel F, Moore SS and Stothard P, 2016. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genetics* 17(1): 75.
- Palombo V, Milanese M, Sgorlon S, Capomaccio S, Mele, M, Nicolazzi, E, Ajmone-Marsan, P, Pilla, F, Stefanon, B and D'Andrea M, 2015. Searching new signals for production traits through gene-based association analysis in three Italian cattle breeds. *Animal Genetics* 46(4): 361-370.
- Park YW, 2009. Overview of bioactive components in milk and dairy products. *Bioactive Components in Milk and Dairy Products* 1–12 p.
- Raven LA, Cocks BG and Hayes BJ, 2014. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genomics* 15(1): 62.
- Reverter A and Fortes MRS, 2013. Breeding and genetics symposium: building single nucleotide polymorphism-derived gene regulatory networks: Towards functional genomewide association studies. *Journal of Animal Science* 91(2): 530-536.
- Roche JR, Friggens NC, Kay JK, Fisher MW, Stafford KJ and Berry DP, 2009. Body condition score and its association with dairy cow productivity, health, and welfare. *Journal of Dairy Science* 92(12): 5769-5801.
- Sahana G, Gulbrandsen B, Thomsen B, Holm LE, Panitz F, Brøndum RF, Bendixen G and Lund MS, 2014. Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle. *Journal of Dairy Science* 97(11): 7258-7275.
- Saito R, Smoot ME, Ono K, Ruschinski J, Wang PL, Lotia S, Pico AR, Bader GD, and Ideker T, 2012. A travel guide to Cytoscape plugins. *Nature Methods* 9 (11): 1069-1076.
- Schlegel G, Ringseis R, Keller J, Schwarz FJ and Eder K, 2012. Changes in the expression of hepatic genes involved in cholesterol homeostasis in dairy cows in the transition period and at different stages of lactation. *Journal of Dairy Science* 95(7): 3826-3836.

- Shannon P, Markiel A, Ozier O, Baliga NS, Jonathan T, Wang DR, 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13 (11): 2498-504.
- Shennan DB and Peaker M, 2000. Transport of milk constituents by the mammary gland. *Physiological Reviews* 80(3): 925-951.
- Shi L, Lv X, Liu L, Yang Y, Ma Z, Han B and Sun D, 2019. A post-GWAS confirming effects of PRKG1 gene on milk fatty acids in a Chinese Holstein dairy population. *BMC Genetics* 20(1): 53.
- Shin D, Lee C, Park KDo, Kim H and Cho KH, 2017. Genome-association analysis of Korean Holstein milk traits using genomic estimated breeding value. *Asian-Australasian Journal of Animal Sciences* 30(3): 309–319.
- Smoczyński M, 2017. Role of phospholipid flux during milk secretion in the mammary gland. *Journal of Mammary Gland Biology and Neoplasia* 22(2): 117-129.
- Spehar M, 2015. Genomic Evaluation and Association Studies of Correlated Traits in Dairy and Dual Purpose Cattle Breeds. Doctoral dissertation, Ljubljana, Univ. of Ljubljana, Biotechnical faculty.
- Spelman RJ, Coppieters W, Karim L, Van Arendonk JAM and Bovenhuis H, 1996. Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* 144(4): 1799-1807.
- Strucken EM, Bortfeldt RH, De Koning DJ and Brockmann GA, 2012. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. *Animal Genetics* 43(4): 375-382.
- Tesfayonas, YG, 2014. Genome wide association study of milk composition traits in Swedish Red cows. 40.
- Turner MD, Rennison ME, Handel SE, Wilde CJ and Burgoyne RD, 1992. Proteins are secreted by both constitutive and regulated secretory pathways in lactating mouse mammary epithelial cells. *The Journal of Cell Biology* 117(2): 269-278.
- Vesterinen HM, Sena ES, Egan KJ, Hirst TC, Churolov L, Currie GL, Antonic A, Howells DW and Macleod MR, 2014. Meta-analysis of data from animal studies: A practical guide. *Journal of Neuroscience Methods* 221: 92-102.
- Viollet B, Foretz M, Guigas B, Horman S, Dentin R, Bertrand L, Hue L and Andreelli F, 2006. Activation of AMP-activated protein kinase in the liver: A new strategy for the management of metabolic hepatic disorders. *The Journal of Physiology* 574(1): 41-53.
- Wagner A and Fell DA, 2001. The small world inside large metabolic networks. *Proceedings of the Royal Society of London. Series B: Biological Sciences*. 268(1478): 1803-1810.
- Wang X, Wurmser C, Pausch H, Jung S, Reinhardt F, Tetens J, Thaller G and Fries R, 2012. Identification and dissection of four major QTL affecting milk fat content in the German Holstein-Friesian population. *Plos One* 7(7): e40711.
- Watts D and Strogatz S, 1998. Collective dynamics of 'small-world' networks. *Nature* 393 (6684): 440–2.
- White HM, 2015. The role of TCA cycle anaplerosis in ketosis and fatty liver in periparturient dairy cows. *Animals* 5(3): 793-802.
- Woalder TC, Chung J and Farese RV, 2017. Lipid droplet biogenesis. *Physiology and Behavior* 6(33): 491-510.
- Zielke LG, Bortfeldt RH, Reissmann M, Tetens J, Thaller G and Brockmann GA, 2013. Impact of Variation at the FTO Locus on Milk Fat Yield in Holstein Dairy Cattle. *Plos One* 8(5): e63406.

Cluster analysis of milk fat yield trait in dairy cows using meta-analysis of the genome-wide association studies

S Bakhshalizadeh¹, S Zerehdaran^{2*} and A Javadmanesh³

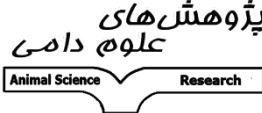

Received: January 20, 2020 Accepted: January 10, 2022

¹PhD Student, Department of Animal Sciences, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

²Professor, Department of Animal Sciences, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

³Assistant Professor, Department of Animal Sciences, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

*Corresponding author: E mail: zerehdaran@um.ac.ir

 <p>پژوهش‌های علوم دامی Animal Science Research</p>	<p>Journal of Animal Science/vol.31 No.4/ 2022/pp 29-42 https://animalscience.tabrizu.ac.ir</p>	 <p>OPEN ACCESS</p>
<p>© 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.2022.37887.1545</p>		

Introduction: Cow's milk is one of the most important animal products for any country and known as an important source for fat and protein originated from animals (Fenelon and Guinee, 1999, Martini et al., 2016). Milk fat has a high impact on taste of milk and other dairy products (Spelman et al., 1996). The liver in ruminant animals including dairy cows plays an important role in the metabolism of carbohydrates, fats, vitamins, hormones, and etc. The absorbed nutrients pass through the liver from the gastrointestinal tract and enter the blood circulation system, and eventually enter the mammary glands of dairy cows. Therefore, the liver plays an essential role in cow lactation (Graber et al., 2010 and Schlegel et al. 2012). All of the components that determine milk quality can be considered as quantitative traits that are controlled by many genes and are influenced by environmental factors. If genetic markers can explain a significant part of the variation, they can be considered as ideal candidates for genomic selection (Shi et al. 2019). Previously, microsatellite markers were frequently used to identify quantitative trait locus (QTL). With the progress of science, the advent of the single-nucleotide polymorphisms (SNP) are used in genome-wide association studies (GWAS) to identify QTL. In dairy cows, some of the major genes with significant effects on milk fat have been identified in previous GWAS studies. Therefore, with the existence of a large number of GWAS in dairy cows, these studies can be combined using meta-analysis to achieve higher power results. These studies contribute to our current understanding of the genetic regulation of milk fat yield traits. This approach provides a better understanding of the genetic architecture of complex traits. The network clustering algorithm and cluster identification are important tools in the structural analysis of networks. Various types of clustering algorithms are used for protein-protein interactions (PPIs) networks analysis. In this study, we used an algorithm known as MCODE to identify dense regions in the PPIs diagram (Bader and Hogue, 2003). The overall purpose of PPIs network clustering is grouping of genes or proteins that according to various scales are related to each other. The network of PPIs contains different proteins important in different pathways. These genes or proteins are clustered based on the similarity of metric and are known as matrix distances. It is also important to predict molecular assemblies of protein interaction data because it provides another level of functional annotation (Gollapalli et al., 2015). The purpose of this study was to conduct a meta-analysis of GWAS in cluster analysis to identify genes that are effective in milk fat yield in dairy cows.

Material and methods: In this study, the data used were GWAS summary data. All data were collected from 19 published studies from 2010 to 2019. This research included main papers and dissertation (valid dissertations with published papers). All available genes were combined, synthesized, and evaluated using a meta-analysis method. The Cytoscape v3.7.2 software was used to analyze and visualize the genes examined by the STRING v1.5.0 plugin and to extract clusters from the MCODE v1.5.1 algorithm. Therefore, the results of the GWAS summary data were combined in molecular networks with PPIs, which have a significant role in increasing the association studies power to identify genes affecting milk fat trait. Also, the DAVID server was used to identify the gene ontology (GO) term enrichment in order to detect enriched biological terms associated with genomic regions and to identify gene networks using functional annotation clustering tools based on enriched pathways analysis.

Results and discussion: In this study, we analyzed 223 genes using the STRING plugin in Cytoscape software. These genes were associated with at least another gene and had a direct and partial correlation with each other. In the gene network, the correlation created for milk fat yield trait included 213 genes or nodes and 219 edges (gene connection). The P-value calculated in the STRING network was statistically significant for enriched pathways in PPIs ($P = 2.6e-09$). The collection of important and popular genes were evaluated using the MCODE plugin. Seven clusters were identified and grouped in this network. For instance, proteins in cluster 1 included: ARHGAP39, CPSF1, CYHR1, PPP1R16A, GRINA, MROH1, and SMPD5 genes. As shown in Table 2, cluster 1 (score=7) was connected with 7 node density to 21 nodes. This cluster showed proteins that play important roles in the internal space of the endoplasmic reticulum (cellular components), metal ion binding (molecular function), and integral to the membrane (cellular components). CPSF1, CYHR1, and GRINA were the major genes involved in the internal space of endoplasmic reticulum, metal ion binding, and membrane integral, respectively. It was found that clusters 1 and 2 have the highest score between all reported clusters.

Conclusion: These results show that using data from different sources can increase the reliability and accuracy of GWAS studies. We were able to identify the most important genes in the network pathways using the GWAS summary data in cluster analysis. This method determines the quality of proteins involved in fat yield while facilitating our understanding of the molecular structures of proteins. The most important genes with high scores were identified in cluster one (*ARHGAP39*, *CPSF1*, *CYHR1*, *PPP1R16A*, *GRINA*, *MROH1*, and *SMPD5*) and two (*HERC1*, *UBR4*, *ASB17*, *TRIM9*, *KLHL2*, and *BTRC*). These clusters based on existing biological knowledge can help data mining and system models understand network interactions and pathways. These protein clusters provide a deep insight into how genes interact with each other in network analysis for fat yield. Moreover, it was observed that meta-analysis of GWAS summary data can play an important role in the wide understanding of network visualization and cluster analysis of identified genes in enriched pathways. Therefore, cluster analysis can improve the identified genes power for economically important traits such as milk fat yield in a population of dairy cows and can be used in future genomic evaluations and breeding programs.

Keywords: Cluster analysis, Dairy cows, GWAS, Meta-analysis