

بررسی شبکه تنظیمی ژن‌های مؤثر بر صفات تولید و ترکیبات شیر گوسفند با استفاده از مطالعات پویش کل ژنومی (GWAS)

- رمضانعلی عزیزی^۱، امیرحسین خلت آبادی فراهانی*^۲، محمد حسین مرادی^۳، حسین محمدی^۲، حسینعلی قاسمی^۳
 - ۱- دانش آموخته دانشکده کشاورزی و محیط زیست دانشگاه اراک و عضو هیئت علمی مرکز تحقیقات کشاورزی استان مرکزی
 - ۲- دانشیار دانشکده کشاورزی و محیط زیست دانشگاه اراک، دانشکده کشاورزی و محیط زیست، گروه علوم دامی
 - ۳- استادیار دانشکده کشاورزی و محیط زیست دانشگاه اراک، دانشکده کشاورزی و محیط زیست، گروه علوم دامی

تاریخ دریافت: خرداد ۱۴۰۱ تاریخ پذیرش: دی ۱۴۰۱

شماره تماس نویسنده مسئول: ۰۹۱۲۵۳۵۱۹۵۶

Email: amfarahanikh@gmail.com

شناسه دیجیتال (DOI): 10.22092/ ASJ.2023.358881.2234

چکیده

امروزه به منظور بهره‌گیری از نتایج تحقیقات مرتبط با مطالعات پویش کل ژنومی (GWAS) و استفاده از آن در جهت ترسیم شبکه‌های تنظیم ژنی، مطالعات جدیدی در حال انجام می‌باشد که مبنای آن استفاده از الگوریتم‌های ماتریس AWM است. در مطالعه حاضر، برای ترسیم شبکه‌های ژنی از طریق الگوریتم‌های ماتریس PCIT-AWM و ارتباط SNPها با فنوتیپ‌های تولید و ترکیبات شیر گوسفند استفاده شد. برای این منظور از داده‌های تولید و ترکیبات شیر مربوط به ۴۶۹ رأس میش نژاد Valle del Belice و از اطلاعات ژنوتیپی مربوط به ۳۷۲۲۸ نشانگر SNP استفاده شد. داده‌های فنوتیپی شامل ۵۵۸۶ رکورد روز آزمایش برای ۶ صفت تولید شیر (مقدار تولید، میزان و درصد چربی، میزان و درصد پروتئین و تعداد سلول‌های بدنی شیر) بود و شبکه تنظیم ژنی آن‌ها با استفاده از برنامه cytoscape رسم شد. نتایج نشان داد که با استفاده از الگوریتم‌های ماتریس PCIT-AWM از بین ژن‌های کاندیدای شناسایی شده، که در ارتباط مستقیم و غیرمستقیم با صفات مرتبط با تولید شیر بودند ژن‌های *OSBPL3*، *BAZIA*، *VGLL4*، *ERBB4*، *DDX25*، *CDH23*، *JTSN2*، *FAT3*، *DPY30* و *CAPN10* شناسایی شدند. همچنین نتایج نشان داد که از بین ۱۰ ژن شناسایی شده، ژن *DDX25* ارتباط بیشتری با ژن‌های دیگر دارد. نتایج مطالعه حاضر و سایر مطالعات نشان داد که فرآیند پیچیده تولید و ترکیبات شیر تحت تأثیر شبکه تنظیم ژنی بیش از ۱۰ ژن مهم و ارتباط با تعداد زیادی از ژن‌های دیگر قرار دارد و با توجه به تأیید نتایج حاصل از مطالعه قبلی در زمینه پویش ژنومی صفات تولید و ترکیبات شیر، و شناسایی مناطق ژنومی جدید مرتبط با این صفات، استفاده از یافته‌های این تحقیق، می‌تواند در انتخاب ژنتیکی گوسفند مفید باشد.

واژه‌های کلیدی: مطالعات پویش ژنومی (GWAS) - شبکه تنظیمی ژن - تولید شیر - گوسفند.

Animal Science Journal (Pajouhesh & Sazandegi) No 140 pp: 97-110

Study of gene regulatory network using a genome-wide association study for milk yield and composition traits in sheepBy: R. A. Azizi¹, A. H. Khaltabadi Farahani^{*1}, H. Moradi¹, H. Mohammadi¹, H. A. Ghasemi¹

1: Department of Animal Science, Faculty of Agricultural, University of Arak, Arak, Iran

Received: June 2022**Accepted: January 2023**

Nowadays, new studies are being conducted based on the results of GWAS studies and the mapping of gene regulatory networks based on the use of AWM matrix algorithms. In the present study, to map gene networks using AWM-PCIT matrix algorithms and the association of SNPs with sheep milk production phenotypes and compositions was used. For this purpose, data on milk production traits in Valle del Belice sheep were used. After data quality control, 469 ewes and 37228 SNPs were used for final analysis. Phenotypic data were retained for analysis and contained 5586 Test day records for six milk production traits. Their gene regulation network was plotted using the cytoscape program. The results showed that using AWM-PCIT matrix algorithms, among the candidate genes that are directly and indirectly related to milk production traits, GENES OSBPL3, ERBB4, VGLL4, BAZ1A, DDX25, CDH23, ITSN2, DPY30, FAT3 and CAPN10 were identified. The results also showed that among the 10 genes identified, the *DDX25* gene was more closely related to other genes. The present study and other studies have shown that the complex process of milk production and composition is under the influence of the gene regulation network of more than 10 genes and is associated with many other genes. In total, this study supported previous results from GWAS of milk production and composition traits, also revealed additional regions in the sheep genome associated with these economically important traits. Using these findings could potentially be useful for genetic selection in the breeding programs.

Key words: Gene regulatory network, Milk production, Sheep, GWAS**مقدمه**

بز در سراسر جهان و در مجموع ۱۰۶ نژاد گوسفند وجود دارد که ۲۰/۸ درصد آن‌ها عمدتاً نژادهای شیرده هستند و ۹ نژاد اصلی گوسفند شیرده وجود دارد که عمدتاً در کشورهای مدیترانه‌ای و اروپایی پراکنده شده‌اند و دو سوم جمعیت جهانی گوسفندان شیری را شامل می‌شوند (Li و همکاران، ۲۰۲۲). از طرفی تولید شیر در اثر فرآیندهای فیزیولوژیکی متعددی می‌باشد که تحت-تأثیر تعداد بسیار زیادی از ژن‌ها قرار دارد و وراثت‌پذیری تولید شیر در گوسفند از ۰/۱۶ تا ۰/۵۷ متغیر است (Haile و همکاران، ۲۰۱۹). لذا، بررسی داده‌های ژنومی توالی‌یابی شده توسط دستگاه‌های نسل جدید توالی‌یابی می‌تواند به درک بهتر ساز و کار تولید شیر در گوسفند کمک کند و از طرفی اطلاعات لازم را برای اصلاح نژاد گوسفند در خصوص تولید و ترکیبات شیر در

صنعت پرورش گوسفند شیری، بخش مهمی از صنعت پرورش نشخوارکنندگان کوچک است که منافع اقتصادی بالایی را برای پرورش دهندگان گوسفند به دنبال دارد. فرآورده‌های لبنی یکی از مهم‌ترین مواد غذایی در تغذیه انسان به خصوص کودکان و نوجوانان است. به طوری که استفاده از آن‌ها اثر مستقیم و قابل توجهی بر سلامت انسان دارد. ارزش غذایی شیر گوسفند از شیر بز و شیر گاو بالاتر است و با داشتن چربی، درصد پروتئین و محتوای مواد معدنی بالاتر، فرصت مناسبی برای توسعه صنعت تولید شیر گوسفند فراهم می‌کند (Raynal و همکاران، ۲۰۰۸). گوسفند تقریباً ۱۱۰۰۰ سال پیش در آسیای جنوب غربی اهلی شده است و اکنون در سراسر جهان برای استفاده از گوشت، پشم و شیر آن‌ها پرورش داده می‌شود. حدود ۲۲۰۰ میلیون گوسفند و

مرتبط با تولید شیر (تولید، مقدار و درصد چربی شیر، مقدار و درصد پروتئین شیر و سلول‌های بدنی) در گوسفندانی که با استفاده از آرایه‌های ژنومی با تراکم بالا تعیین ژنوتیپ شده بودند، برای آنالیز استفاده شد.

استخراج داده‌های تعیین ژنوتیپ شده به همراه رکوردهای فنوتیپی از پایگاه‌های داده ذخیره‌ای آنلاین <https://www.animalgenome.org/repository/pub/UPIT2018.0> 80 انجام گرفت.

آنالیز پویش کل ژنومی (GWAS) و کنترل کیفیت داده‌های تعیین ژنوتیپ شده:

برای اجرای آنالیز پویش کل ژنومی، ارزش اصلاحی برآورد شده (EBVs) برای صفات مربوط با تولید و ترکیبات شیر شامل تولید شیر، مقدار و درصد پروتئین، مقدار و درصد چربی و تعداد سلول‌های سوماتیکی شیر به عنوان فنوتیپ حیوانات تعیین ژنوتیپ شده استفاده شد. مطالعه آنالیز پویش کل ژنومی با استفاده از نرم افزار PLINK نسخه ۱/۹ انجام گرفت و برای حصول اطمینان از کیفیت داده‌ها در آنالیزهای بعدی، این داده‌ها توسط نرم افزار PLINK (Purcell و همکاران، ۲۰۰۷) و با اعمال یک سری مراحل کنترل کیفیت، داده‌های دارای اطلاعات کم کیفیت از قبیل ژنوتیپ‌های مونومورف و ژنوتیپ‌های با فراوانی آلی نادر (MAF) کمتر از ۵ درصد از آنالیزهای بعدی حذف شدند. همچنین نشانگرهای SNP که در تعادل هاردی واینبرگ نبودند و یا موقعیت فیزیکی برای آخرین نسخه ژنومی در دسترس نبود از آنالیز نهایی حذف شدند.

تشکیل ماتریس The Association Weight Matrix (AWM)

ماتریس AWM با انتخاب SNP مربوطه از مطالعات پویش کل ژنومی (GWAS) ژن‌ها ایجاد شد. معیارهای انتخاب SNP ها که مشخصاً ژن‌ها را در ماتریس AWM نشان می‌دهد شامل (۱) اهمیت اثر جایگزینی آلل اندازه‌گیری شده برای هر SNP (۲) همبستگی بین فنوتیپ‌ها و (۳) موقعیت ژنومی. معیارهای انتخاب ژن‌های حاوی SNP با ارتباط معنی‌دار در بین فنوتیپ‌های مرتبط بود. این معیارها به صورت متوالی اعمال شدند: ابتدا، اثر جایگزینی

اختیار محققان قرار دهد. بررسی شبکه‌های تنظیم ژنی کاربردهای مختلفی دارد، از جمله با استفاده از این شبکه می‌توان اثرات متقابل بین ژن‌ها و اثرات متقابل احتمالی که بین هر جفت پروتئین حاصل از ژن‌ها وجود دارد را مورد بحث قرار داد و ارتباطات مولکولی که از نظر آماری معنی‌دار باشند را ارائه کرد. علاوه بر این، از شبکه‌های تنظیم ژنی می‌توان جهت یافتن نشانگرهای زیستی برای شناسایی، تشخیص و تخمین اهداف بیولوژیکی نیز بهره برد (Kang و همکاران، ۲۰۱۷). در سال‌های اخیر محققین، استفاده از ماتریس Association Weight Matrix (AWM) را به عنوان یک رویکرد جدید برای ادغام فنوتیپ‌های مرتبط و نظریه شبکه ژنی پیشنهاد کرده‌اند، ماتریس AWM اولین بار توسط Fortes و همکاران (۲۰۱۰) در یک جمعیت گاو گوشتی گزارش شد. استفاده از ماتریس AWM دو جنبه دارد: (۱) همبستگی بین فنوتیپ‌ها را تخمین می‌زند و (۲) تعاملات ژن - ژن را در کل ژنوم بررسی می‌کند. ماتریس AWM برای فنوتیپ‌های پیچیده جهت بررسی اثرات تنظیمی پیشنهاد شده است که با استفاده از رویکرد ماتریس AWM می‌توان نتایج حاصل از مطالعات پویش کل ژنومی (GWAS¹) را توسعه داد (Fortes و همکاران، ۲۰۱۰).

در مطالعه حاضر، شبکه تنظیم ژنی مرتبط با تولید و ترکیبات (پروتئین، چربی و سلول‌های سوماتیکی) شیر گوسفند مورد بررسی قرار خواهند گرفت و هدف از انجام مطالعه حاضر، ایجاد زمینه مورد نیاز جهت به‌کارگیری اطلاعات ژنومی و مطالعات پویش ژنومی (GWAS) جهت شناسایی ژن‌های کاندیدا و شبکه‌های تنظیم ژنی مؤثر بر تولید و ترکیبات شیر با استفاده از الگوریتم‌های ماتریس AWM می‌باشد.

مواد و روش‌ها

استخراج داده‌های تعیین ژنوتیپ شده:

در مطالعه حاضر، از داده‌های مربوط به تولید و ترکیبات شیر گوسفند نژاد Valle del belice برای مطالعات پویش کل ژنومی (GWAS) استفاده شد (Sutera و همکاران، ۲۰۱۸)؛ Sutera و همکاران، ۲۰۱۹). برای این منظور برخی صفات

¹ - genome-wide association studies

$$r(x, y) = \frac{\sum(x - xm)(y - ym)}{\sqrt{\sum(x - xm)^2 \sum(y - ym)^2}} \quad (\text{رابطه 2})$$

در این رابطه، x مقدار اثر محاسبه شده SNP برای یک فنوتیپ و y مقدار اثر SNP محاسبه شده برای فنوتیپ دیگر است.

بهره برداری از ردیف‌های AWM برای ترسیم شبکه اثرات متقابل ژن:

از همبستگی‌های جفتی در ردیف‌های ماتریس AWM برای پیش‌بینی اثرات متقابل ژن-ژن (یا ژن-SNP) برای تشکیل یک شبکه استفاده شد. در این شبکه، هر ژن یا SNP هنگامی که به‌عنوان دسته "خیلی دور" منظور شد یک گره و هر اثر متقابل یک لبه تعیین شد. نهایتاً شبکه ژنی در برنامه نرم افزاری Cytoscape با رویه MCODE و با استفاده از نتایج تجزیه و تحلیل الگوریتم‌های ماتریس AWM و ترکیب آن با الگوریتم‌های PCIT ارتباط ژن و شبکه‌های تنظیمی ژن با استفاده از الگوی شکل ۱ ترسیم شد (Bader و Hogue، ۲۰۰۳).

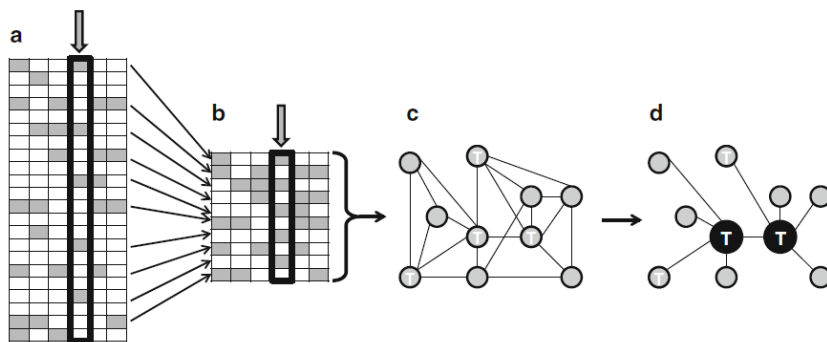
در شکل ۱ نحوه تشکیل ماتریس AWM توضیح داده شده است: (a) مرتب کردن SNP‌ها در ردیف و فنوتیپ‌ها در ستون با توجه به اثرات افزایشی. در این نمودار اثرات مورد نظر با سلول‌های تیره‌تر نشان داده شده و فنوتیپ کلیدی در ستون چهارم برجسته شده است. (b) در مجموع SNP‌های مرتبط با فنوتیپ کلیدی و یا با تعداد زیادی فنوتیپ برای تولید زیرمجموعه‌ای از ماتریس اصلی انتخاب شدند. (c) تجزیه و تحلیل همبستگی برای ایجاد یک شبکه ژنی با تعدادی ژن مهم که به هم مرتبط شده اند انجام شد. (d) با تجزیه و تحلیل نهایی می‌توان شبکه اصلی را با حذف اتصالات غیر ضروری دقیقتر ترسیم کرد.

آلل i امین SNP بر روی j امین فنوتیپ در توزیع Z استاندارد شد تا امکان مقایسه بین فنوتیپ‌ها را به‌شرح رابطه ۱ فراهم کند:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j} \quad (\text{رابطه 1})$$

در این رابطه Z_{ij} اثر استاندارد شده i امین SNP بر روی j امین فنوتیپ است. X_{ij} اثر i امین SNP بر روی j امین فنوتیپ در واحدهای اصلی است. \bar{X}_j میانگین X_{ij} در تمام SNP‌ها است. و S_j انحراف استاندارد X_{ij} در کل SNP‌ها بود.

در مرحله دوم صرف‌نظر از موقعیت ژنومی نشانگرها، SNP با $P < 0.05$ در ۶ فنوتیپ مورد نظر برای تشکیل ماتریس AWM استفاده شد. در مرحله سوم، هر SNP را با توجه به فاصله تعیین شده آن از نزدیک‌ترین ژن شرح داده شد و به‌عنوان نزدیک، دور، خیلی دور یا unmapped طبقه‌بندی شد. SNP در نظر گرفته شده "نزدیک" در $2/5 \text{ Kb} \geq$ از نزدیک‌ترین ژن قرار داشت. SNP در نظر گرفته شده "خیلی دور" $1/5 \text{ Mb} \leq$ از نزدیک‌ترین ژن فاصله داشت. بر این اساس، در فاصله SNP تا ژنی که بین «نزدیک» و «خیلی دور» قرار داشت، به‌عنوان «دور» حاشیه‌نویسی شد. SNP‌های دور یا «unmapped» از تجزیه و تحلیل نهایی کنار گذاشته و فقط گروه‌های «نزدیک» و «خیلی دور» انتخاب شدند. از آنجایی که تولید شیر فنوتیپ اصلی بود، تمام SNP‌های دارای $P < 0.05$ برای صفات دیگر انتخاب شدند. ماتریس AWM با لحاظ کردن همه SNP‌هایی که برای هر ۶ فنوتیپ معنی‌دار بودند، ایجاد شد (Fortes و همکاران، ۲۰۱۰). همبستگی‌های جفتی بین تولید شیر و ۵ فنوتیپ دیگر با استفاده از مقادیر اثر SNP (رابطه ۲) محاسبه شد. در مرحله اول، تمام SNP‌های موجود از ژنوتیپ ۵۰ kb برای این محاسبه استفاده شد، و زیر مجموعه SNP انتخاب شده برای AWM در نمونه دوم استفاده گردید.



شکل ۱- روند ترسیم شبکه تنظیمی ژن با استفاده از ماتریس AWM و الگوریتم‌های PCIT

قرار داشتند، (ب) بسیار نزدیک به ژن‌های شناخته شده قرار داشتند، یا (ج) در فاصله دورتری قرار داشتند، این شرط را می‌توان با گنجاندن SNP‌های unmapped برقرار نمود و برای تعریف این شرط از MAP یا فاصله بیشتر از ۱/۵ Mb استفاده شد. در این مرحله، SNP‌های دور یا unmapped از تجزیه و تحلیل کنار گذاشته شدند و فقط از گروه‌های نزدیک و خیلی دور استفاده شد. اسکرپیت R مربوط به این مرحله به شرح زیر بود:

مرحله ۵: ایجاد رابطه "One SNP-One Gene"

در این مرحله SNP مرتبط با بیشترین تعداد فنوتیپ و دارای کمترین میانگین P-value در همه فنوتیپ‌ها به عنوان نماینده ژن انتخاب شد.

مرحله ۶: تکمیل کردن ماتریس AWM

در این مرحله برای تکمیل ماتریس AWM با استفاده از نتایج GWAS، در هر سلول AWM $\{i, j\}$ ، ارتباط آمین ژن را با ژامین فنوتیپ قرار داده شد. ارتباط آمین ژن را با ژامین فنوتیپ با اثر افزایشی SNP اختصاص داده شده به آن ژن مطابق مراحل قبلی نقشه برداری برای SNP در شرایط (i) و (ii) مرحله ۴ اختصاص داده شد. به منظور امکان مقایسه مستقیم بین فنوتیپ‌ها، اثرات افزایشی SNP با تقسیم هر اثر فردی بر انحراف استاندارد اثر همه SNP در فنوتیپ ستون J ام استاندارد شد. اسکرپیت R مربوط به این مرحله به شرح زیر بود:

ماتریس AWM در شش مرحله به شرح زیر تشکیل گردید (Fortes و Reverter، ۲۰۱۳):

مرحله ۱: انتخاب اولیه SNP

در این مرحله SNP‌هایی انتخاب شد که با فنوتیپ کلیدی مرتبط بودند (تولید شیر) و دارای $P < 0.05$ بودند. اسکرپیت‌های R مربوط به این مرحله به شرح زیر بود:

مرحله ۲: بررسی همبستگی بین فنوتیپ‌ها

برای SNP‌های انتخاب شده در مرحله اول، میانگین تعداد فنوتیپ‌های غیرکلیدی را که به آن‌ها مرتبط هستند، ثبت گردید. در این مرحله برای علامت گذاری، از عبارت AP که این میانگین را نشان دهد استفاده شد. همچنین برای شمارش تعداد فنوتیپ‌هایی که SNP معین به آن‌ها مرتبط است، از همان معنی داری قبلی استفاده شد ($P < 0.05$). A_p تابعی از میزان وابستگی موجود در بین فنوتیپ‌ها می‌باشد اسکرپیت R مربوط به این مرحله به شرح زیر بود:

مرحله ۳: انتخاب SNP ثانویه

در این مرحله SNP‌هایی انتخاب شدند که با حداقل فنوتیپ‌های A_p مرتبط هستند. اسکرپیت R مربوط به این مرحله به شرح زیر بود:

مرحله ۴: بهره برداری از نقشه ژنوم

از SNP‌های گرفته شده در مراحل انتخاب اولیه و ثانویه، آنهایی مورد شناسایی قرار گرفتند که (الف) در مناطق کدگذاری MAP

MCODE برای شناسایی خوشه‌هایی از ژن‌های بسیار بهم پیوسته استفاده شد.

از نرم افزار Cytoscape برای ترسیم شبکه‌ژنی استفاده شد. ژن‌های شناسایی شده در شبکه تنظیم ژنی با استفاده از بانک اطلاعاتی DAVID برای بررسی مسیرهای شناخته شده بیشتر مورد تجزیه و تحلیل قرار گرفتند (Dennis و همکاران، ۲۰۰۳؛ Huang و همکاران، ۲۰۰۹).

نتایج و بحث

کنترل کیفیت داده‌ها و توزیع نشانگرها بر روی کروموزوم‌های مختلف:

در مطالعه حاضر از داده‌های کنترل کیفیت شده مربوط با آرایه-های ژنومی با تراکم نشانگری K50 استفاده شد که در نهایت پس از اجرای مراحل مختلف ویرایش داده‌ها تعداد ۴۲۶ رأس دام و ۴۲۲۸۵ نشانگر SNP جهت پویش ژنومی (GWAS) صفات تولید شیر، درصد و مقدار چربی، درصد و مقدار پروتئین و سلول‌های بدنی شیر استفاده گردید. مراحل مختلف کنترل کیفی در جدول ۱ ارائه شده است.

ترسیم شبکه اثرات متقابل ژن با استفاده از ماتریس AWM:

ساختار جدول مانند ماتریس AWM این امکان را فراهم می‌کند که بتوان از آن به عنوان ورودی در اکثر برنامه‌های نرم افزار تحلیل خوشه‌ای استفاده کرد. برای این منظور، از نرم افزار PermutMatrix که یک برنامه کاربرپسند برای انجام تجزیه و تحلیل گرافیکی می‌باشد استفاده شد (Caraux و Pinloche، ۲۰۰۵).

همبستگی‌های بین ردیف‌ها با استفاده از الگوریتم PCIT شناسایی شد و به عنوان برهم‌کنش‌های ژن-ژن یا ژن SNP در یک شبکه گزارش شد (Reverter و chan، ۲۰۰۸). خروجی PCIT می‌تواند به عنوان ورودی برای برنامه Cytoscape که یک برنامه نرم افزاری قوی برای ساخت و تجزیه و تحلیل شبکه می‌باشد استفاده شد (Shannon و همکاران، ۲۰۰۳). برای این منظور تعدادی پلاگین Cytoscape برای تعیین ویژگی‌های مختلف شبکه‌ها وجود دارد، که در این مطالعه از رویه

جدول ۱- توصیف مراحل مختلف بررسی کنترل کیفیت داده‌های اولیه مورد استفاده در تحقیق حاضر

۴۷۶	تعداد حیوانات
۵۰	حذف نمونه‌هایی با بیش از ۰/۰۵ ژنوتیپ از دست رفته
۴۲۶	تعداد کل نمونه‌های باقی‌مانده
۴۶۸۲۷	تعداد کل نشانگرها
۱۷۱۹	حذف نشانگرهایی با بیش از 0.05 ژنوتیپ از دست رفته
۲۰۶۶	حذف نشانگرها با $MAF < 0.02$
۹۹۹	حذف نشانگرها با انحراف از تعادل هاردی-واینبرگ 0.001
۴۲۲۸۵	تعداد نشانگرهای باقی‌مانده

هم متفاوت بود به طوری که بیشترین تعداد نشانگر روی کروموزوم یک (۶۷۶۹) و کمترین تراکم نشانگر روی کروموزوم ۲۶ (۸۶۸) قرار داشت (شکل ۲). همچنین میانگین فاصله بین نشانگرهای SNP ها نیز روی کروموزوم متفاوت بود به طوری که کمترین فاصله با

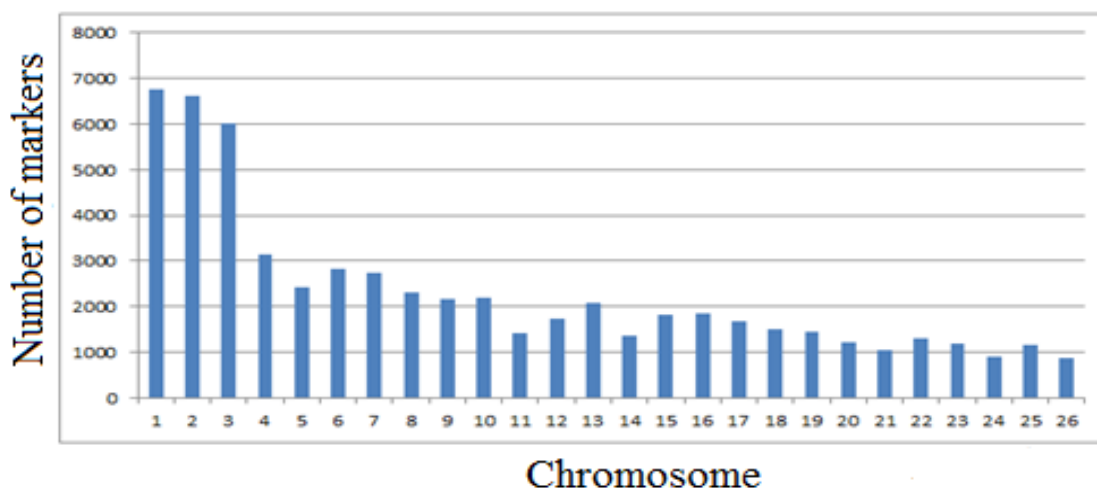
آماره‌های توصیفی رکوردهای فنوتیپی مربوط به صفات تولید و ترکیبات شیر در نژاد del belice Valle در جدول ۲ نشان داده شده است.

در مطالعه حاضر نشان داده شد که توزیع SNP روی کروموزوم‌ها با

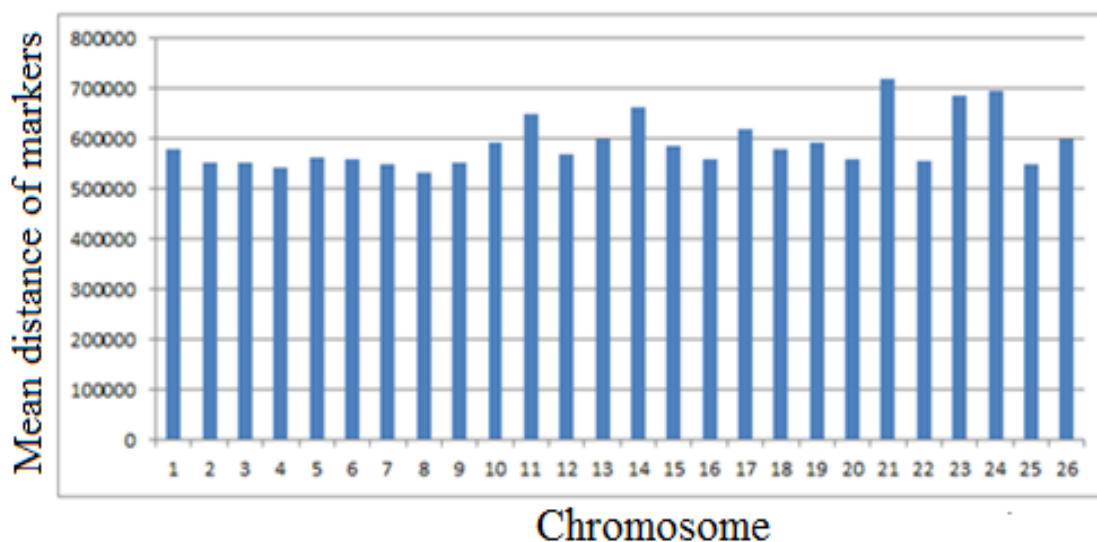
میانگین برابر با (۵۳۴۰۶۸) مربوط به کروموزوم ۸ بیش‌ترین میانگین (۷۲۰۴۸۸) مربوط به کروموزوم ۲۱ بود (شکل ۳).

جدول ۲: آمار توصیفی تولید شیر (MY, g)، تعداد سلول‌های سوماتیکی (SCC×103)، درصد چربی (F) و پروتئین (P) در نژاد گوسفند Valle del Belice.

صفات	واحد	میانگین	انحراف معیار	حداقل	حداکثر
تولید شیر	گرم	۱۱۳۵	۶۰۳	۵۶	۴۳۰۰
چربی	درصد	۷/۲۵	۱/۲۳	۲/۳۳	۱۴/۵۶
پروتئین	درصد	۵/۸	۰/۷۲	۳/۴۶	۱۳/۹
SCC×10 ³	تعداد	۱۵۴۴	۳۵۴۷	۵	۲۹۳۶۸



شکل ۲- تعداد نشانگرهای مورد استفاده در تحقیق حاضر به تفکیک کروموزوم



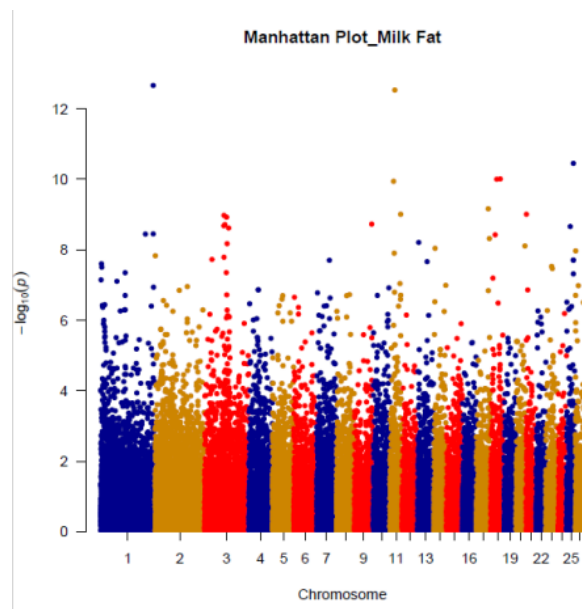
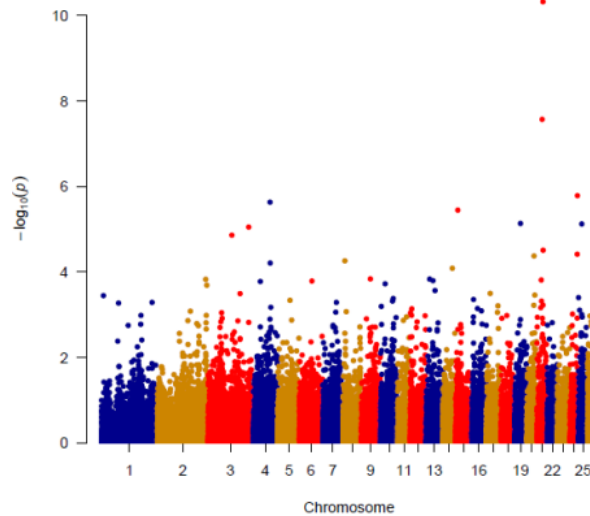
شکل ۳- میانگین فاصله نشانگرهای مورد استفاده در تحقیق حاضر به تفکیک کروموزوم

مطالعه پویش ژنومی (GWAS) برای صفات مختلف

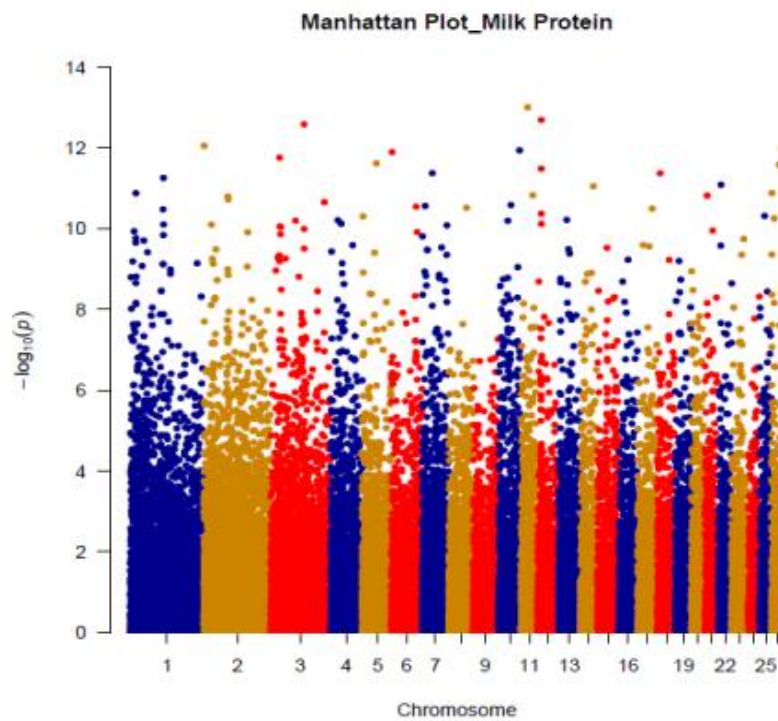
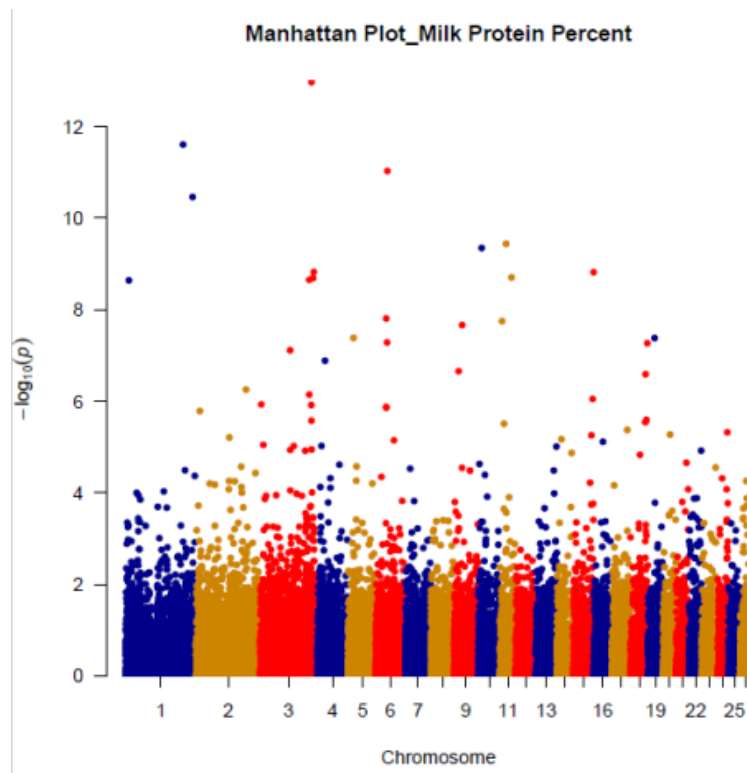
پلات‌های منهن حاصل از مطالعه پیوستگی ژنومی برای صفات تولید شیر (MY)، درصد چربی (F%)، درصد پروتئین (P%) و مقدار تولید پروتئین (PY) رسم شد. در این گراف‌ها بر روی محور X موقعیت SNP‌ها در کروموزوم‌های مختلف و روی

محور Y ارزش‌های منفی لگاریتم P-value نمایش داده شده است. SNP‌هایی که بالاتر از خط قرمز، و بین خط آبی و قرمز هستند، به ترتیب جزو نشانگرهای معنی‌دار و پیشنهادی برای هر صفت محسوب می‌شوند (شکل‌های ۴، ۵ و ۶).

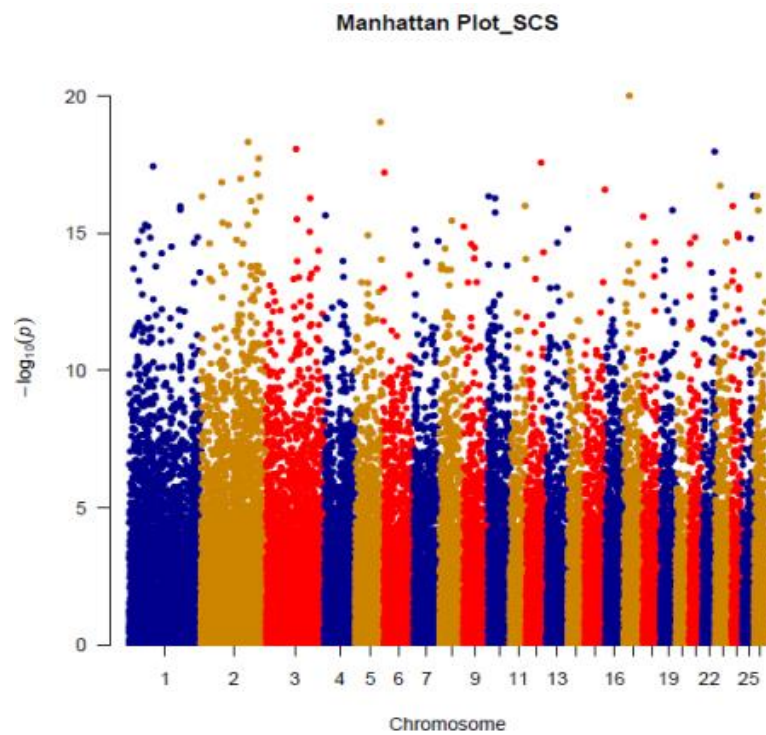
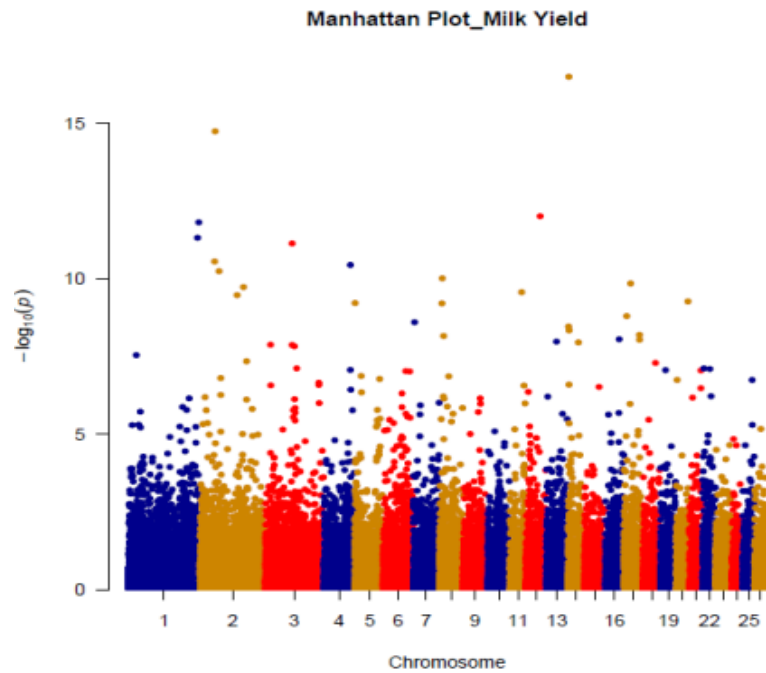
Manhattan Plot_Fat percent



شکل ۴- پلات منهن برای صفات مرتبط با مقدار و درصد چربی.



شکل ۵- پلات منهن برای صفات مرتبط مقدار و درصد پروتئین.



شکل ۶- پلات منهن برای صفت مرتبط عملکرد شیر و سلول های سوماتیک

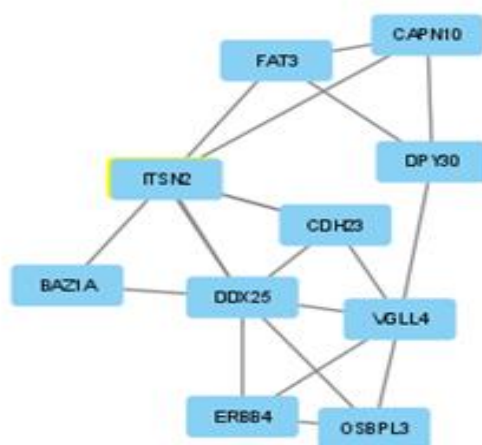
شیر بودند شامل ژن‌های *VGLL4*, *ERBB4*, *OSBPL3*, *FAT3*, *DPY30*, *ITSN2*, *CDH23*, *DDX25*, *BAZIA* و *CAPN10* بودند (شکل ۷ و جدول ۳).

نتایج مطالعه حاضر نشان داد که با استفاده از الگوریتم‌های ماتریس AWM از بین ژن‌های کاندیدای شناسایی شده ژن‌های کاندیدای مهم که در ارتباط مستقیم و غیرمستقیم با صفات مرتبط با تولید

جدول ۳- مشخصات ژن‌های شناسایی شده در شبکه تنظیم ژنی

Gen bank	Gene Abbreviation	The full name of the gene	FPKM ²
BC109867	<i>DDX25</i>	Dead (Asp-Glu-Ala-Asp) Box Polypeptide 25	188.791
BC102970	<i>HSBP1</i> , <i>OSBPL3</i>	heat shock factor binding protein 1	156.88
	<i>ERBB4</i> , <i>TUBB</i> , <i>ERBB2</i>	testis-specific serine kinase 3	10.77
	<i>VGLL4</i>	Dependent anion channel 2	67.005
786886	<i>BAZIA</i>	Chromatin-dependent regulation of transcription.	12.5
AY508164	<i>CDH23</i>	Epithelial cells. Cadherin related 23	
BC110170	<i>ITSN2</i> , <i>CSNK2B</i>	Casein Kinase 2, Beta Polypeptide	110.244
BC102328	<i>SNRNPB2</i>	Ribonucleoprotein Polypeptide B"	109.087
BC109581	<i>Dpy30</i> , <i>DYDC1</i>	Dpy30 Domain Containing 1	154.473
	<i>FAT3</i>	FAT3; FAT atypical cadherin 3, protocadherin Fat 3	
BC109951	<i>CAPN10</i>	Filament) Muscle Z-Line, Alpha 3capping protein (actin	402.076
	<i>CAPZA3</i>	filament) muscle Z-line, alpha 1	
	<i>CAPZA1</i>	capping protein (actin filament) muscle Z-line, beta	
	<i>CAPZB</i>	B-cell receptor-associated protein 29	
BC110030	<i>BCAP29</i>	similar to CAP1 protein; CAP, adenylate cyclase-associated	186.465
	<i>CAP1</i>	protein 1 (yeast)	

ژن *DDX25* موتیف حفاظت شده (Asp-Glu-Ala-Asp (DEAD) می‌باشد که پروتئین هلیکازهای RNA هستند. این ژن در تعدادی از فرآیندهای سلولی شامل تغییر ساختار ثانویه RNA، نقش دارد. اعتقاد بر این است که بر اساس الگوهای توزیع آن‌ها، برخی از اعضای خانواده پروتئین DEAD BOX در تشکیل جنین، تولید اسپرم و رشد و تقسیم سلولی نقش دارند (Suthar و همکاران، ۲۰۱۶). در مطالعه حاضر نشان داده شد که ژن *DDX25* علاوه بر تاثیر بر صفات تولید مثلی بر صفات تولید و ترکیبات شیر نیز موثر است. ژن *OSBPL3* یکی از اعضای خانواده *OSBP* است که در انتقال چربی، متابولیسم لیپید و سیگنالدهی سلولی نقش دارد (Du و همکاران، ۲۰۱۸).



شکل ۷- شبکه تنظیمی ژن ترسیم شده با استفاده از الگوریتم‌های ماتریس AWM و PCIT

(Ramzan و همکاران، ۲۰۲۰).

ژن *BAZIA* فاکتور بازسازی کروماتین *ACF1* را کد می‌کند، *ACF1* یک کمپلکس بازسازی کننده کروماتین وابسته به *ATP*، که فاصله نوکلئوزوم‌ها را با استفاده از *ATP* تنظیم می‌کند تا نوکلئوزوم‌هایی با فاصله یکنواخت در طول کروماتین تولید کند. این ژن همچنین در تسهیل فرآیند تکثیر *DNA* نقش دارد. علاوه بر این، اخیراً نشان داده شده است که ژن *BAZIA* در سه مسیر متمایز (۱) تنظیم ویتامین *D*، (۲) سیگنالدهی *Wnt* و (۳) سیگنالدهی پس‌سیناپسی نقش آفرینی می‌کند (*Zaghlool* و همکاران، ۲۰۱۶).

ژن *FAT3* (*FAT Atypical Cadherin 3*) یک ژن کد کننده پروتئین است. وظیفه اصلی این ژن تنظیم یون کلسیم است. *FAT4* همچنین در مهار تکثیر سلولی نقش دارد (*Okajima* و همکاران، ۲۰۲۰).

ژن *DPY30* مسیرهای پیر سلولی را تنظیم می‌کند. پیری سلولی یک مکانیسم دفاعی ذاتی در برابر تنش‌های سلولی مختلف است. ژن *DPY30* یک تنظیم کننده کلیدی در تکثیر سلول‌های اولیه انسانی است و چنانچه ژن *DPY30* حذف گردد، سلول‌ها نقص شدید در تکثیر نشان می‌دهند و نشانه‌های بروز فنوتیپ پیری را نشان می‌دهند (*Simboeck* و همکاران، ۲۰۱۳). از طرفی وجود ژن *DPY30* برای حفظ ساختار و عملکرد سلول‌های بنیادی خونساز بالغ حیاتی است. ژن *DPY30* در تمایز سلول‌های بنیادی خونساز بسیار موثر است و در صورت نقص در این ژن با کاهش شدید سلول‌های خونساز بالغ همراه خواهد بود (*Yang* و همکاران، ۲۰۱۶).

ژن *CAPN10*، ژن کدکننده سیستئین پروتئاز کالپین-۱۰ است که با دیابت نوع ۲ مرتبط می‌باشد. کالپین‌ها پروتئازهای تنظیم شده با کلسیم هستند که در عملکردهای سلولی شامل پروتئولیز عضلانی هم قبل و هم پس از مرگ دخالت دارند (*Evans* و همکاران، ۲۰۰۱).

ژن *JTSN2* پروتئین *Intersectin-2* را در انسان کدگذاری می‌کند. این ژن یک پروتئین سیتوپلاسمی حاوی دامنه‌های *SH3* را کد می‌کند. این پروتئین عضوی از خانواده پروتئین‌هایی است که در اندوسیتوز با واسطه کلاترین دخیل هستند و در اندوسیتوز اینتگرین‌بتا-۱ (*ITGB1*) و گیرنده ترانسفرین (*TFR*) نقش دارد (*Pucharcos* و همکاران، ۲۰۰۰). در مطالعه حاضر نشان داده شد که این ژن در فرآیند تولید و ترکیبات شیر نیز نقش دارد که احتمالاً در فرآیند انتقال مواد معدنی به داخل شیر از جمله آهن می‌تواند نقش داشته باشد.

ژن *ERBB4* نقش مهمی در افزایش فرآیند لاکتوژنز و تمایز غده پستانی در دوران آبستنی ایفا می‌کند (*Williams* و همکاران، ۲۰۱۷). علاوه بر این، ژن *ERBB2* که عضو دیگر این خانواده می‌باشد و به‌عنوان یک ژن کاندیدا موثر بر غلظت پروتئین شیر در گاوهای شیری گزارش شده است (*Li* و همکاران، ۲۰۱۶). در مطالعه دیگری نشان داده شد که ژن *ERBB2* بر صفت پروتئین شیر در گاوهای شیری موثر است. نتایج مطالعات انجام شده اثرات ژنتیکی قابل توجه ژن *ERBB2* را بر مقدار پروتئین شیر و سایر صفات تولید شیر گاوهای هلشتاین چینی تایید کردند و نشان دادند که واریانس فنوتیپی *SNP* بالای یک درصد ممکن است به‌عنوان نشانگر ژنتیکی در برنامه‌های اصلاح نژاد گاو شیری عمل کند (*Li* و همکاران، ۲۰۱۸).

ژن *VGLL4* به‌همراه ژن‌های *ANKRD23* و *ACOXL* در متابولیسم اسیدهای چرب و انرژی نقش دارند. این پروتئین نقش مهمی در توسعه فیبر عضلانی اسکلتی و هایپرتروفی ایفا می‌کند. *ANKRD23* در قلب موش، ماهیچه اسکلتی و چربی قهوه‌ای بیان می‌شود و با متابولیسم اسیدهای چرب ارتباط دارد (*Dong* و همکاران، ۲۰۰۷).

ژن *CDH23* گلیکوپروتئین چسبنده سلولی وابسته به کلسیم را با چندین حوزه *multiple cadherin-like domains* کد می‌کند. پروتئین کدگذاری شده، *CDH23* به‌ابرخانواده *cadherin* تعلق دارد که خانواده‌ای از پروتئین‌های غشایی را تشکیل می‌دهد که واسطه چسبندگی سلولی وابسته به کلسیم است

نتیجه گیری

ژن‌هایی که سازگار با عوامل بیولوژیکی شناخته شده تولید شیر هستند، بینش‌های جدیدی را برای تحقیقات بیشتر فراهم می‌کنند که ممکن است به درک بهتر فرآیندهای تولید شیر در سایر پستانداران نیز کمک کند. در نهایت ادغام اطلاعات عملکردی و تنظیم ژن با روش‌های آماری GWAS ممکن است دانش مکانیسم‌های ژنومی مؤثر بر فنوتیپ‌های پیچیده را افزایش دهد. در این تحقیق به دلیل عدم دسترسی به رکوردهای فنوتیپی و اطلاعات ژنوتیپی مرتبط با صفت مورد مطالعه تحقیق حاضر در دام‌های بومی کشور، از اطلاعات فنوتیپی و ژنوتیپی کشور چین استفاده شد. لذا استفاده از نتایج این تحقیق در جمعیت‌های گوسفندان کشور نیاز به مطالعات بیشتر دارد تا در این جمعیت‌ها نیز تأیید شوند.

با توجه به چالش‌های موجود در صنعت پرورش گوسفند و تولید شیر در کشور وجود دارد، این تحقیق در راستای شناسایی ژن‌ها و شبکه تنظیم ژنی صفات تولید و ترکیبات شیر گوسفند پایه ریزی شد. آنالیزهای پویس ژنومی منجر به شناسایی مناطق ژنومی کاندیدا برای صفات تولید و ترکیبات شیر گردید. استفاده از روش ماتریس AWM و ترسیم شبکه تنظیم ژنی نهایتاً باعث شناسایی ژن‌های جدید مؤثر بر تولید و ترکیبات شیر گوسفند گردید که به نوعی اثر پلیوتروپی دارند. از نظر کاربردی در توسعه صنعت پرورش گوسفند شیری و تولید شیر و فرآورده‌های شیر گوسفند می‌تواند بسیار مؤثر باشد. همچنین رویکرد مطالعه حاضر باعث تشخیص ژن‌های مؤثر بر فنوتیپ‌های با ساختار ژنتیکی پیچیده شد. نتایج این مطالعه نشان داد که رویکرد ماتریس AWM برای بررسی و ترسیم شبکه‌های تنظیم ژنی در خصوص

منابع

- Bader, G. D., and Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*, 4(1), 1-27.
- Caraux, G., and Pinloche, S. (2005). PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics*, 21(7), 1280-1281.
- Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., and Lempicki, R. A. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome biology*, 4(9), 1-11.
- Du, X., Turner, N., and Yang, H. (2018). The role of oxysterol-binding protein and its related proteins in cancer. In *Seminars in cell and developmental biology* (Vol. 81, pp. 149-153). Academic Press.
- Evans, J. C., Frayling, T. M., Cassell, P. G., Saker, P. J., Hitman, G. A., Walker, M., and Hattersley, A. T. (2001). Studies of association between the gene for calpain-10 and type 2 diabetes mellitus in the United Kingdom. *The American Journal of Human Genetics*, 69(3), 544-552.
- Fortes, M. R., Reverter, A., Zhang, Y., Collis, E., Nagaraj, S. H., Jonsson, N. N., and Hawken, R. J. (2010). Association weight matrix for the genetic dissection of puberty in beef cattle. *Proceedings of the National Academy of Sciences*, 107(31), 13642-13647.
- Haile, A., Hilali, M., Hassen, H., Lobo, R. N. B., and Rischkowsky, B. (2019). Estimates of genetic parameters and genetic trends for growth, reproduction, milk production and milk composition traits of Awassi sheep. *animal*, 13(2), 240-247.
- Kang, Y. Y., Liu, Y., Wang, M. L., Guo, M., Wang, Y., and Cheng, Z. F. (2017). Construction and analyses of the microRNA-target gene differential regulatory network in thyroid carcinoma. *PloS one*, 12(6), e0178331.
- Li, C., Sun, D., Zhang, S., Liu, L., Alim, M. A., and Zhang, Q. (2016). A post-GWAS confirming the SCD gene associated with milk medium-and long-chain unsaturated fatty acids in Chinese Holstein population. *Animal genetics*, 47(4), 483-490.

- Li, C., Wang, M., Cai, W., Liu, S., Zhou, C., Yin, H., and Zhang, S. (2018). Genetic analyses confirm SNPs in HSPA8 and ERBB2 are associated with milk protein concentration in Chinese Holstein cattle. *Genes*, 10(2), 104.
- Li, R., Ma, Y., and Jiang, L. (2022). Research Progress of Dairy Sheep Milk Genes. *Agriculture*, 12(2), 169.
- Okajima, T., Gu, Y., Teruya, R. I., Yano, S., Taketomi, T., Sato, B., and Tsuruta, F. (2020). Atypical cadherin FAT3 is a novel mediator for morphological changes of microglia. *eneuro*, 7(6).
- Pucharcos, C., Estivill, X., and de la Luna, S. (2000). Intersectin 2, a new multimodular protein involved in clathrin-mediated endocytosis. *FEBS letters*, 478(1-2), 43-51.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., and Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3), 559-575.
- Ramzan, K., Al-Numair, N. S., Al-Ageel, S., Elbaik, L., Sakati, N., Al-Hazaa, S. A., and Imtiaz, F. (2020). Identification of novel CDH23 variants causing moderate to profound progressive nonsyndromic hearing loss. *Genes*, 11(12), 1474.
- Raynal-Ljutovac, K., Lagriffoul, G., Paccard, P., Guillet, I., and Chilliard, Y. (2008). Composition of goat and sheep milk products: An update. *Small ruminant research*, 79(1), 57-72.
- Reverter, A., and Chan, E. K. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics*, 24(21), 2491-2497.
- Reverter, A., and Fortes, M. R. (2013). Association weight matrix: a network-based approach towards functional genome-wide association studies. In *Genome-Wide Association Studies and Genomic Prediction* (pp. 437-447). Humana Press, Totowa, NJ.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11), 2498-2504.
- Simboeck, E., Gutierrez, A., Cozzuto, L., Beringer, M., Caizzi, L., M Keyes, W., and Di Croce, L. (2013). DPY30 regulates pathways in cellular senescence through ID protein expression. *The EMBO journal*, 32(16), 2217-2230.
- Sutera, A. M., Portolano, B., Di Gerlando, R., Sardina, M. T., Mastrangelo, S., and Tolone, M. (2018). Determination of milk production losses and variations of fat and protein percentages according to different levels of somatic cell count in Valle del Belice dairy sheep. *Small Ruminant Research*, 162, 39-42.
- Sutera, A. M., Riggio, V., Mastrangelo, S., Di Gerlando, R., Sardina, M. T., Pong-Wong, R., and Portolano, B. (2019). Genome-wide association studies for milk production traits in Valle del Belice sheep using repeated measures. *Animal genetics*, 50(3), 311-314.
- Suthar, M. K., Purva, M., Maherchandani, S., and Kashyap, S. K. (2016). Identification and in silico analysis of cattle DEXH/D box RNA helicases. *Springerplus*, 5(1), 1-13.
- Williams, M. M., Vaught, D. B., Joly, M. M., Hicks, D. J., Sanchez, V., Owens, P., and Cook, R. S. (2017). ErbB3 drives mammary epithelial survival and differentiation during pregnancy and lactation. *Breast Cancer Research*, 19(1), 1-14.
- Yang, Z., Shah, K., Khodadadi-Jamayran, A., and Jiang, H. (2016). Dpy30 is critical for maintaining the identity and function of adult hematopoietic stem cells. *Journal of Experimental Medicine*, 213(11), 2349-2364.
- Zaghlool, A., Halvardson, J., Zhao, J. J., Etemadikhah, M., Kalushkova, A., Konska, K., and Feuk, L. (2016). A role for the chromatin-remodeling factor BAZ1A in neurodevelopment. *Human mutation*, 37(9), 964-975.