

استخراج رابطه علی- معلولی ترانسکرپتومی در بافت غدد پستانی گاو شیری با استفاده از شبکه بیزی

• امین مرتضوی

دانشجوی دکتری، ژنتیک و اصلاح دام، گروه علوم دامی، دانشگاه کردستان، سنندج، ایران

• امیر رشیدی (نویسنده مسئول)

استاد ژنتیک و اصلاح دام، گروه علوم دامی، دانشگاه کردستان، سنندج، ایران.

• مصطفی قادری

استادیار زیست سامانه‌های محاسباتی، گروه علوم دامی دانشگاه یاسوج، یاسوج، ایران

• پرهام مرادی

دانشیار رایانه، گروه مهندسی کامپیوتر، دانشگاه کردستان، سنندج، کردستان، ایران

• محمد رزم کبیر

استادیار ژنتیک و اصلاح دام، گروه علوم دامی، دانشگاه کردستان، سنندج، ایران

تاریخ دریافت: اسفند ۱۳۹۶ تاریخ پذیرش: شهریور ۱۳۹۷

شماره تماس نویسنده مسئول: ۰۹۱۸۸۷۱۰۳۴۲

Email: arashidi@uok.ac.ir

شناسه دیجیتال (DOI): 10.22092/asj.2018.121024.1656

چکیده

هدف از این مطالعه، شناسایی ژن‌های تنظیمی مؤثر در بروز بیماری ورم پستان در گاو شیری با استفاده از داده‌های ریزآرایه DNA بود. بدین منظور فایل بیان ژنی با بیشترین تعداد آرایه و متعلق به پلت فرم افی-متریکس GPL1221 با شماره دسترسی GSE24560 از پایگاه GEO استخراج شد. برای آزمون کنترل کیفیت داده از بسته‌ی ArrayQualityMetrics و برای پیش‌پردازش داده‌ها از تابع Threestep بسته‌ی AffyPLM در محیط R استفاده شد. پس از شناسایی ژن‌های متفاوت بیان‌شده، الگوریتم جستجوی تابو در بسته bnlearn برای شناسایی ژن‌های تنظیمی مورد استفاده قرار گرفت. نتایج حاصل از این پژوهش بیانگر نقش علی و تنظیمی ژن‌های BCL2A، CCL2، S100A12، AOX1 و MGP بر بیان سایر ژن‌ها در بافت پستانی گاو شیری بود. بررسی هستی‌شناسی ژن‌ها، تفاوت معنی‌داری را در ۷ گروه عملکردی مولکولی، ۴۸ گروه فرآیندهای زیستی و ۱۱ گروه اجزای سلولی در شرایط بروز بیماری ورم پستان در بافت غدد پستانی نشان داد. همچنین نتایج غنی‌سازی مجموعه داده‌های بیان ژنی نشان داد بیشتر ژن‌های متفاوت بیان‌شده در این پژوهش به طور معنی‌داری ($P < 0.05$) در مسیرهای متابولیسمی (GO:0009605، GO:0002376 و GO:0006954) فعال بوده و در فرآیندهای پاسخ به بروز عوامل بیماری‌زا، پاسخ ایمنی و پاسخ به التهاب در بافت پستانی گاو شیری نقش دارند.

Animal Science Journal (Pajouhesh & Sazandegi) No 123 pp: 197-216

Extraction of cause-and-effect transcriptomic relationship in mammary gland tissue of dairy cattle using Bayesian network.By: Amin Mortazavi¹, Amir Rashidi^{2*}, Mostafa Ghaderi-Zefrehei³, Parham Moradi⁴, Mohammad Razm Kabir⁵¹Ph.D. Student, Department of Animal Science, University of Kurdistan, Sanandaj, Iran,²Professor, Department of Animal Science, University of Kurdistan, Sanandaj, Iran,³Assistant Professor, Department of Animal Science, Yasouj university, Yasouj, Iran,⁴Associate professor, Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran,⁵ Assistant Professor, Department of Animal Science, University of Kurdistan, Sanandaj, Iran,**Received: March 2018****Accepted: September 2018**

The aim of this study was to identify regulatory genes affecting mastitis in dairy cattle using DNA microarray data. To reach this goal, the gene expression data with the largest number of arrays pertained to GPL1221 Platform with accession number GSE24560 was extracted from the GEO database. For quality control of data, ArrayQualityMetrics package and for preprocessing of data, three step function in AffyPLM, an add-in package in R environment were used. After identifying differentially expressed genes, a Tabu search algorithm was used to determine regulatory genes using bnlearn package in R environment. The results of this study revealed the causative and regulatory role for BCL2A, CCL2, S100A12, AOX1 and MGP genes on expression of other genes in mastitis of dairy cattle. Gene anthology analysis revealed significant differences in 7 groups of molecular function, 48 groups of biological process and 11 groups of cellular components. Also, the results of the enrichment of the gene expression data set showed that most of the differentially genes expressed in this study that were significantly ($P < 0.05$) active in metabolic pathways (GO: 0009605, GO: 0002376 and GO: 0006954) involved in response to pathogens, immune response and response to inflammation in the mammary tissue of dairy cattle.

Key words: DNA Microarray, Mastitis, Regulatory Genes, Anthology.**مقدمه**

کاهش تولید شیر ناشی از ورم پستان به‌ازای هر گاو در هر دوره‌ی شیردهی را ۳۸۶ کیلوگرم گزارش کرده‌اند (Bar و همکاران، ۲۰۰۸). برای بهبود ژنتیکی و کاهش بروز بیماری ورم پستان، شناسایی گروه‌های ژنی و ژن‌های تنظیمی مؤثر در دام‌های مقاوم به بیماری موردنیاز است.

از روش‌های ارائه‌شده برای مدل‌سازی شبکه‌های تنظیمی ژن می‌توان به مدل‌های خطی، شبکه عصبی، معادلات دیفرانسیل، شبکه‌های بولین و شبکه‌های بی‌زی اشاره نمود. از نظر ریاضی، شبکه‌های بی‌زی به صورت $G = (V = x_1, x_2, \dots, x_n; E \subseteq V \times V)$ نشان داده می‌شوند که در آن V برابر تعداد متغیرها (مقادیر بیان ژن‌ها)

پیشرفت‌های فناوری اخیر در حوزه زیست‌شناسی مولکولی امکان اندازه‌گیری سطوح بیان هزاران ژن به صورت هم‌زمان را فراهم ساخته است. فناوری ریزآرایه DNA، قادر به آشکارسازی سطوح بیانی ژن‌ها در شرایط و زمان‌های متفاوت و هم‌چنین قادر به شناسایی ارتباط بین سطوح بیان ژن و فنوتیپ می‌باشد (Adomas و همکاران، ۲۰۰۸). واکنش التهابی غدد پستان در پاسخ به بیمارگرها را ورم پستان می‌گویند. از جمله باکتری‌های ایجاد ورم پستان می‌توان به باکتری *E. coli* اشاره کرد که قادر به ایجاد پاسخ ایمنی عمومی و پاسخ ایمنی موضعی در دام‌ها می‌باشد (Jorgensen و همکاران، ۲۰۱۲). برخی از پژوهشگران

قابل ملاحظه باشد. با توجه به انطباق کامل سطوح تصادفی داده‌های ریزآرایه DNA با ماهیت احتمالاتی شبکه‌های بیزی از یک سو و از سوی دیگر کمبود تحقیق جامع در مورد شناسایی روابط علی- معلولی و شناسایی ژن‌های تنظیمی در داده‌های بیان ژن در بافت پستانی در گونه گاو، اهداف مهم این پژوهش آشنایی بیشتر با ابزارهای رایج در مدل‌سازی شبکه ژنی به روش بیزی، شناسایی گروه‌های ژنی و ژن‌های تنظیمی مؤثر بر بروز بیماری ورم پستان در بافت غدد پستانی در یک پژوهش بیان ژنی در گونه گاو با شماره دسترسی GSE24560 و کاربرد ضمنی و احتمالی نتایج حاصله در اصلاح نژاد گاو شیری می‌باشد.

مواد و روش‌ها

در این پژوهش ابتدا فایل‌های CEL آزمایش‌های ریزآرایه DNA از پایگاه GEO که مربوط به گونه گاو بودند، استخراج و برحسب نوع بافت مورد بررسی، آزمایش‌ها گروه‌بندی شدند. جدول ۱ پیوست فراوانی آزمایش‌های انجام‌شده مربوط به هر بافت با شماره دسترسی آزمایش‌ها را به همراه تعداد آرایه‌های به‌کاررفته در آن تا زمان انجام این پژوهش نشان می‌دهد. سعی گردید که آزمایش‌ها متعلق به یک پلات فرم یکسان (GPL1221 شرکت افی‌متریکس) و تعداد آرایه‌های آن زیاد باشند. معمولاً شبکه‌های بیزی، تعداد زیادی فراسنجه دارند. لذا به هراندازه تعداد مشاهده به ازای هر ژن بالاتر باشد، امکان اریب بودن نتایج کاهش می‌یابد. به همین منظور برای انجام این پژوهش، از آزمایش با شماره دسترسی GSE24560 مربوط به بیماری ورم پستان ناشی از باکتری *E. coli* و استافیلوکوکوس اورئوس در بافت غدد پستانی استفاده شد. برای پیاده‌سازی GSE24560 از بسته نرم‌افزاری GEOquery که متکی به محیط R است، استفاده شد (Davis and Meltzer, 2007). به منظور پیش- پردازش داده‌ها (آزمون کنترل کیفیت) از بستهی ArrayQualityMetrics استفاده شد (Kauffmann و همکاران، 2009). برای بررسی کیفیت داده‌ها گراف مرجع Heatmap مورد استفاده قرار گرفت. با بررسی این گراف نمونه‌هایی که دارای کیفیت نامطلوب بودند از مراحل بعدی آنالیزها

و E بیانگر وجود وابستگی شرطی و مستقیم بین ژن‌ها (اتصالات) می‌باشند. بنابراین، ساختار شبکه بیزی توسط یک گراف بدون دور جهت‌دار معرفی می‌شود (Scutari, 2014). در مدل‌سازی شبکه‌های تنظیمی ژنی به روش بیزی می‌توان به پیچیدگی زیاد محاسباتی اشاره نمود که ناشی از تعداد زیاد ژن‌ها در برابر تعداد کم نمونه‌ها می‌باشد. به منظور غلبه بر این محدودیت، لازم است با استفاده از الگوریتم‌های مدل‌سازی شبکه، خواه به روش بهینه- سازی (الگوریتم‌های نمره‌گرا هم‌چون جستجوی تابو) یا روش محدودسازی (الگوریتم‌های محدودیت‌گرا) به شناسایی تعدادی ژن تنظیمی مهم با اتصالات بالا پرداخت (Scutari, 2014). الگوریتم جستجوی تابو با استفاده از بهینه‌سازی توابع امتیاز و تکنیک‌های تکرارشونده، سعی بر دستیابی به بهترین مدل با استفاده از دست‌کاری در ساختار شبکه دارد. مدل‌سازی شبکه تنظیمی ژنی می‌تواند در شناسایی ژن‌های مرتبط با بیماری‌ها و صفات پیچیده، نشان دادن رابطه بین ژن‌ها و بررسی کلی روابط و فرآیندهایی که ژن‌ها در آن دخالت دارند، به اصلاح‌گران کمک شایانی بنماید (Scutari, 2014). اخیراً پژوهش‌های زیادی برای استخراج ژن‌های تأثیرگذار و شبکه تنظیمی ژنی با استفاده از داده‌های ریزآرایه DNA بر بافت‌های مهم در گونه گاو انجام‌شده است. به عنوان نمونه می‌توان به پژوهش‌های (Strunz, 2014)؛ Ramayo-Caldas و همکاران، (2014)؛ Moran و همکاران، (2015)؛ Weber و همکاران، (2015)؛ Alexandre و همکاران، (2015) و Fortes و همکاران، (2016) اشاره نمود. در این پژوهش‌های شبکه‌محور معمولاً سعی بر شناسایی ماژول‌ها از طریق استفاده از الگوریتم‌های نظریه اطلاعات دو طرفه و شبکه‌های هم‌بیان ژنی وزن‌دار بوده است. تمامی این الگوریتم‌های یادشده غیر جهت‌دار بوده، بنابراین هیچ‌گونه استنباطی در مورد استخراج روابط علی معلولی و شناسایی ژن‌های تنظیمی حاصل نشده است. حتی در سایر داده‌های تولیدشده سطوح مختلف اومیکس در ژنوم گاو، ارتباط علی- معلولی بین ژن‌ها بدست نیامده است. درحالی‌که استخراج این ارتباط می‌تواند در شناخت صفات و اهمیت آن‌ها در اصلاح نژاد

Genome Browser و از نسخه Bos_taurus_UMD_3.1.1/bosTau8 Assembly استفاده شد (Kent و همکاران، 2002). برای جمع‌آوری اطلاعات موردنیاز در مورد عملکرد زیستی ژن‌های شناسایی‌شده پایگاه ژنوم <http://www.genecards.org/> مورد استفاده قرار گرفت (Rebhan و همکاران، 1997).

نتایج و بحث

نتایج ارائه‌شده در جدول ۱ پیوست نشان می‌دهد بیشترین میزان پژوهش‌های بیان ژن با استفاده از ریزآرایه DNA در گاو شیری مربوط به بافت پستان (۳۵٪) است. بافت‌های رحمی (۲۶٪) و تخمدان (۱۷٪) از لحاظ تعداد آزمایش‌های ریزآرایه DNA در مراحل بعدی قرار دارند. این موضوع در وهله اول اهمیت ژن‌هایی که در این بافت‌ها بیان می‌شوند را برای پژوهشگران نشان می‌دهد. از بین تمامی سری آزمایش‌های ریزآرایه DNA موجود در پایگاه بیان ژن NCBI و متعلق به بافت پستانی، سری آزمایشی ریزآرایه با شماره دسترسی GSE24560 دارای بیشترین میزان آرایه (۸۸) بود، به همین دلیل در این پژوهش اطلاعات این سری آزمایش مورد استفاده قرار گرفت. در جدول ۲ پیوست فهرستی از نرم‌افزارهای برخط برای استنباط شبکه بیزی و نیز اطلاعات بیشتر در مورد آدرس تارنمای اینترنتی، روش و مدل‌های مورد استفاده ارائه شده است. در جدول ۳ پیوست بسته‌های متعدد نرم‌افزار R برای استخراج شبکه بیزی ارائه شده‌اند. از آن جمله می‌توان به deal، Pcalg، GRbase، catnet، bnstruct و bnlearn اشاره کرد. در طول دو دهه اخیر، توجه وافر به گسترش مدل الگوریتم‌های استخراج شبکه‌های بیزی شده است (Heckerman, 1997). از مزایای استخراج شبکه تنظیمی ژنی به روش بیزی می‌توان به موارد زیر اشاره نمود: این شبکه‌ها به دلیل ارائه در قالب مدل گرافیکی قابلیت درک بیشتری دارند و قابلیت دست‌کاری بر روی ساختار شبکه حاصله مثل اضافه کردن، حذف و تغییر جهت تنظیمی یال‌ها با استفاده از اطلاعات پیشین برای شبکه علی ژنی وجود دارد. از این قابلیت می‌توان در حل مسائل

حذف شدند. به‌منظور پیش‌پردازش داده‌ها از تابع Threestep بسته‌ی AffyPLM در محیط R استفاده شد (Bolstad و همکاران، 2005). برای نرمال‌سازی از روش Quantile، برای تصحیح پس‌زمینه روش IdealMM و برای خلاصه‌سازی از روش Tukey bi-weight استفاده شد. یک درصد ژن‌هایی که بیشترین واریانس را ایجاد کرده بودند به‌عنوان ژن‌های متفاوت بیان‌شده انتخاب شدند (Langfelder and Horvath, 2008). در مرحله بعد از الگوریتم جستجوی تابو موجود در بسته bnlearn برای برآزش شبکه بیزی بر روی داده‌های موجود استفاده شد (Scutari, 2014). از حالت استاندارد الگوریتم جستجوی تابو (اجرای الگوریتم به‌تنهایی) برای شناسایی ژن‌های تنظیمی استفاده شد. ژن‌های با بیشترین نسبت فراسنجه درجه خارج به درجه داخل به‌عنوان ژن‌های تنظیمی شناسایی شدند. توپولوژی شبکه تنظیمی با استفاده از نرم‌افزار Cytoscape تعیین گردید (Shannon و همکاران، 2003). با در نظر گرفتن تعداد ژن‌های متفاوت بیان‌شده به‌عنوان مجموعه ژن، نمودار و نتایج غنی‌سازی مجموعه ژنی در داده‌های ریزآرایه بافت غدد پستانی با استفاده از نرم‌افزار برخط AgriGO ترسیم شد (Tian و همکاران، 2017). در این نرم‌افزار از روش SEA برای شناسایی گروه‌های عملکردی با ژن‌های متفاوت بیان‌شده استفاده شد. هم‌چنین برای برآورد میزان معنی‌داری غنی‌سازی ژنی و p-value تصحیح‌شده از آماره Fisher exact test استفاده شد. تمام ژن‌های مورد بررسی بر اساس سه گروه اصلی فرآیندهای زیستی، اجزای سلولی و عملکرد مولکولی گروه‌بندی و تجزیه و تحلیل شدند. اطلاعات مربوط به SNP‌های مرتبط با ژن‌های تنظیمی شناسایی شده هم-چون مکان SNP‌ها و نوع جهش‌ها از پایگاه اطلاعاتی NCBI استخراج شد. اطلاعات حاشیه‌نویسی عملکردی ژن‌های متفاوت بیان‌شده و ژن‌های تنظیمی شناسایی‌شده با استفاده از نرم‌افزار DAVID صورت پذیرفت (Huang و همکاران، 2009). جهت استخراج اطلاعات حاشیه‌نویسی ژن‌های شناسایی‌شده در ژنوم گونه گاو (نام RefSeq، PubMed، Entrez ID، تعداد جفت باز و محل ژن بر روی کروموزوم) از سایت UCSC

شده است که در آن مقاومت گروه‌های تیمار با عفونت باکتریایی *E. coli* و استافیلوکوکوس اورئوس به ورم پستان ارزیابی می‌شوند، بنابراین منطقی می‌باشد که در مواجهه با عفونت ورم پستانی با باکتری‌های مدنظر بیان ژن‌های درگیر در مسیرهای التهابی و ایمنی افزایش یابد. اطلاعات کامل‌تر در این مورد در پژوهش شریفی و همکاران، (1396) موجود است. نمودار غنی‌سازی مجموعه ژن‌های متفاوت بیان‌شده در مسیرهای زیستی در بافت پستانی در شکل ۲ پیوست ارائه شده است. در کل ۶۲ مسیر زیستی معنی‌دار بر بروز بیماری ورم پستان در بافت پستان غنی‌سازی شده است که از مهم‌ترین مسیرهای غنی‌شده می‌توان به مسیرهای دخیل در فرآیند تکامل و رشد و نمو، پاسخ‌های ایمنی، پاسخ به محرک-ها نام برد. در جدول ۱ نیز هستی‌شناسی ژن‌های متفاوت بیان‌شده به همراه واژه‌های GO آن به صورت خلاصه ارائه شده است. نتایج غنی‌سازی مجموعه داده بیان ژنی نیز نشان داد GO:0009605، GO:0002376 و GO:0006954 در پاسخ به محرک‌های خارجی، پاسخ‌های ایمنی و پاسخ به التهاب در بافت پستانی در گاو شیری معنی‌دار بودند ($P < 0.05$).

استنباطی مختلف مثل تجزیه و تحلیل‌های ارتباطی و نمایش روابط مستقل / غیرمستقل میان ژن‌ها استفاده کرد (Gómez و همکاران، 2011). با توجه به جدول ۳ پیوست، قابلیت‌های متعدد بسته نرم-افزاری bnlearn در ارائه انواع خدمات برای مدل‌سازی شبکه بیزی پس از آزمون و خطاهای متعدد مشاهده می‌شود. توضیحات بیشتر قابلیت بسته نرم‌افزاری bnlearn در مدل‌سازی شبکه بیزی توسط (Scutari, 2014) ارائه شده است. نمایش گرافیکی حاشیه‌نویسی ژن‌های متفاوت بیان‌شده در مسیرهای کارکرد مولکولی و مؤلفه سلولی در شکل ۱ پیوست قابل مشاهده است. تعداد ژن‌های هستی‌شناسی شده ۸۸ ژن بودند. دو فرآیند نهایی برای ژن‌های متفاوت بیان‌شده در مسیرهای فوق به ترتیب شامل فرآیند تنظیم مرگ برنامه‌ریزی شده سلولی (۸ ژن) و فرآیند بلوغ پروتئین‌ها (۵ ژن) بودند. هم‌چنین در بخش فرآیندهای زیستی ۴۸ دسته ژنی غنی‌شده بودند، از جمله مهم‌ترین این گروه‌ها می‌توان به پاسخ به محرک‌های خارجی ($2.01e-23$)، فرآیندهای سیستم ایمنی ($1.23e-22$)، پاسخ به التهاب ($1.67e-21$)، فرآیندهای زیستی تنظیمی مثبت ($1.44e-20$) اشاره نمود. در این تحقیق از داده ریزآرایه DNA با شماره دسترسی GSE24560 استفاده

جدول ۱- حاشیه‌نویسی عملکردی ژن‌های متفاوت بیان‌شده در بافت پستانی

FDR	p-value	Description	GO Term
2e-23	5.8e-26	response to external stimulus	GO:0009605
1.2e-22	7e-25	immune system process	GO:0002376
1.7e-21	1.4e-23	inflammatory response	GO:0006954
1.4e-20	1.6e-22	positive regulation of biological process	GO:0048518
7.7e-19	1.1e-20	response to wounding	GO:0009611
5.5e-16	9.4e-18	cell differentiation	GO:0030154
5.6e-16	1.1e-17	positive regulation of cellular process	GO:0048522
7.9e-16	1.8e-17	response to stimulus	GO:0050896
6.4e-15	1.7e-16	anatomical structure development	GO:0048856
8.7e-15	2.5e-16	acute inflammatory response	GO:0002526

2.3e-14	7.1e-16	defense response	GO:0006952
4.8e-14	1.6e-15	system development	GO:0048731
4.2e-13	1.6e-14	cellular developmental process	GO:0048869
5.7e-13	2.3e-14	organ development	GO:0048513
5.7e-13	2.4e-14	multicellular organismal process	GO:0032501
6.2e-12	2.8e-13	antigen processing and presentation	GO:0019882
1.1e-11	5.5e-13	response to chemical stimulus	GO:0042221
5.9e-11	3.1e-12	immune response	GO:0006955
1.4e-10	7.5e-12	negative regulation of biological process	GO:0048519
1.6e-10	9.4e-12	multicellular organismal development	GO:0007275
1.8e-10	1.1e-11	response to stress	GO:0006950
2.1e-10	1.4e-11	regulation of multicellular organismal process	GO:0051239
2.1e-10	1.4e-11	locomotion	GO:0040011
7.6e-10	5.7e-11	taxis	GO:0042330
7.6e-10	5.7e-11	locomotory behavior	GO:0007626
7.6e-10	5.7e-11	chemotaxis	GO:0006935
1.3e-09	9.9e-11	behavior	GO:0007610
4.6e-09	3.7e-10	developmental process	GO:0032502
1.8e-08	1.5e-09	humoral immune response	GO:0006959
2.8e-08	2.4e-09	cell-cell signaling	GO:0007267
4.1e-08	3.9e-09	regulation of apoptosis	GO:0042981
4.1e-08	3.9e-09	regulation of programmed cell death	GO:0043067
4.1e-08	3.9e-09	regulation of cell death	GO:0010941
4.6e-08	4.6e-09	programmed cell death	GO:0012501
4.6e-08	4.6e-09	apoptosis	GO:0006915
4.9e-08	5.1e-09	negative regulation of cellular process	GO:0048523
6.1e-08	6.6e-09	death	GO:0016265
6.1e-08	6.6e-09	cell death	GO:0008219
1.3e-07	1.5e-08	anatomical structure morphogenesis	GO:0009653
1.5e-07	1.8e-08	response to biotic stimulus	GO:0009607

1.5e-07	1.8e-08	protein maturation by peptide bond cleavage	GO:0051605
4e-07	5e-08	protein processing	GO:0016485
4e-07	5e-08	protein maturation	GO:0051604
2.9e-06	3.7e-07	innate immune response	GO:0045087
4.3e-06	5.6e-07	growth	GO:0040007
2.4e-05	3.2e-06	cell adhesion	GO:0007155
2.4e-05	3.2e-06	biological adhesion	GO:0022610
0.00065	8.9e-05	regulation of biological quality	GO:0065008
3e-07	4.2e-09	extracellular matrix structural constituent	GO:0005201
1.8e-06	5.1e-08	protein binding	GO:0005515
7.8e-06	3.3e-07	receptor binding	GO:0005102
0.00076	4.2e-05	cytokine receptor binding	GO:0005126
0.0011	7.4e-05	calcium ion binding	GO:0005509
0.032	0.0027	endopeptidase inhibitor activity	GO:0004866
0.044	0.0043	peptidase inhibitor activity	GO:0030414
5.9e-36	5.6e-38	extracellular region part	GO:0044421
4.7e-27	8.9e-29	extracellular region	GO:0005576
4.2e-07	1.2e-08	plasma membrane	GO:0005886
5e-06	2.3e-07	cytoplasm	GO:0005737
5e-06	2.4e-07	proteinaceous extracellular matrix	GO:0005578
8.1e-06	4.7e-07	extracellular matrix	GO:0031012
1.2e-05	7.8e-07	extracellular matrix part	GO:0044420
7.4e-05	5.6e-06	plasma membrane part	GO:0044459
0.0097	0.00093	intrinsic to plasma membrane	GO:0031226
0.0097	0.00093	integral to plasma membrane	GO:0005887
0.033	0.0034	protein complex	GO:0043234

سلولی (۶ ژن) و اتصال‌دهنده‌های پروتئینی (۳۱ ژن) می‌باشد. نتایج هستی‌شناسی ژن‌های متفاوت بیان‌شده در بافت غدد پستانی، نشان داد که در بخش کارکردهای مولکولی ۷ دسته ژنی غنی شده

فرآیند نهایی برای ژن‌های افزایش بیان یافته در مسیر کارکرد مولکولی شامل فرآیند اتصال یون کلسیم (۸ ژن) بود. فرآیندهای معنی‌دار در این مسیر شامل ترکیبات ساختاری ماتریکس خارج

شیری انجام شده است، فرآیندهای نهایی و معنی دار در هر سه بخش فرآیندهای زیستی، کارکرد مولکولی و مؤلفه های سلولی همگی به نوعی بیانگر مسیرها و ژن های مؤثر بر بروز این بیماری در بافت غدد پستانی می باشند. پس با شناسایی این گروه های ژنی می توان شناخت دقیق تری از سازوکار ژن های موجود در بافت غدد پستانی و مؤثر بر بیماری ورم پستان داشت. نتایج دسته بندی گروه های عملکردی سلولی، مولکولی و زیستی در این پژوهش با نتایج هستی شناسی ژن در تحقیق شریفی و همکاران، (1396) مطابقت دارد که آن ها در یک پژوهش فراتحلیل بر روی بیان ژن داده های ورم پستان با باکتری *E. coli* توانستند بیشترین تعداد مسیرهای زیستی را به سامانه ایمنی، التهاب و پاسخ مرحله حاد ارتباط دهند.

بودند؛ از جمله مهم ترین گروه ها می توان به ترکیبات ساختاری ماتریکس خارج سلولی ($3.03e-7$)، اتصال پروتئین ($1.83e-06$) و اتصال به گیرنده ($7.82e-06$) اشاره نمود. از سوی دیگر، فرآیندهای نهایی برای ژن های متفاوت بیان شده در مسیر مؤلفه سلولی شامل ۵ ژن بود که جزء اصلی غشای پلاسمایی را تشکیل می دهند. فرآیند اصلی و معنی دار در بخش مؤلفه های سلولی شامل نواحی خارج سلولی (۳۰ ژن) بود. نتایج هستی شناسی ژن های متفاوت بیان شده در بافت غدد پستانی، نشان داد که در بخش مؤلفه های سلولی ۱۱ دسته ژنی غنی شده بودند؛ از جمله مهم ترین گروه ها می توان به قسمتی از ناحیه خارج سلولی ($5.93e-36$)، خارج سلولی ($4.7e-27$) و غشای پلاسمایی ($4.22e-07$) اشاره نمود. از آنجا که این پژوهش بر بروز بیماری ورم پستان در گاوهای

جدول ۲- خوشه بندی ژن های متفاوت بیان شده در بافت غدد پستانی بر اساس حاشیه نویسی عملکردی

FDR	BJ	BF	FE	P	C	Term
Enrichment Score: 4.91						خوشه ۱
1.5E-7	1.3E-8	1.3E-8	33.16	1.4E-10	9	bta05332:Graft-versus-host disease
2.0E-5	4.5E-7	1.8E-6	9.77	1.8E-8	12	bta05168:Herpes simplex infection
Enrichment Score: 4.21						خوشه ۲
7.6E-6	2.2E-7	6.7E-7	10.73	6.8E-9	12	bta05164:Influenza A
2.0E-5	4.5E-7	1.8E-6	9.77	1.8E-8	12	bta05168:Herpes simplex infection
Enrichment Score: 3.99						خوشه ۳
3.3E-4	5.0E-6	3.0E-5	12.89	3.0E-7	9	bta04668:TNF signaling pathway
0.74	0.002	0.06	5.20	6.7E-4	8	bta04060:Cytokine-cytokine receptor interaction
Enrichment Score: 2.79						خوشه ۴
0.001	1.9E-5	1.5E-4	18.36	1.6E-6	7	bta05150:Staphylococcus aureus infection
11.77	0.03	0.67	8.36	0.01	4	bta04610:Complement and coagulation cascades
Enrichment Score: 2.19						خوشه ۵
2.06	0.007	0.16	9.211	0.001	5	bta04974:Protein digestion and absorption

*C:count, P:p value, FE: Fold Enrichment, BF: Bonferroni, BJ: Benjamini

به باکتری‌ها و ویروس‌ها نقش داشتند. حاشیه‌نویسی ساختاری ژن-های هاب مؤثر بر بروز بیماری ورم پستان در بافت غدد پستانی در جدول ۳ قابل مشاهده هست. با استفاده از داده‌های ریزآرایه DNA، در کل شبکه برازش داده‌شده برای بافت غدد پستانی با استفاده از معیار درجه، ۵ ژن هاب شناسایی شد (MGP، AOX1، S100A12، CCL2 و BCL2A). هم‌چنین نام‌های دیگر ژن (Entrez ID، RefSeq ID، PubMed ID) و مکان آن‌ها روی کروموزوم نیز در جدول ۳ قابل مشاهده می‌باشند.

در جدول ۲ نتایج خوشه‌بندی ژن‌های متفاوت بیان‌شده در بافت پستانی ارائه شده است. ژن‌ها تنها بر اساس مسیر KEGG خوشه‌بندی شده و مورد مقایسه قرار گرفتند. خوشه اول با دارا بودن ژن‌های مؤثر بر مقاومت به بیماری (۹ ژن) و عفونت ویروسی (۱۲ ژن) بیشترین مقدار ضریب غنی‌سازی (۴/۹۱) و مقدار P (1.4e-10) را به خود اختصاص دادند. ژن‌ها در سایر خوشه‌ها به‌جز خوشه ۳ (مسیرهای پیام‌رسانی و تعاملات گیرنده سیتوکین) و خوشه ۵ (جذب و هضم پروتئین‌ها) نیز به‌طور مستقیم در مقاومت

جدول ۳- حاشیه‌نویسی ساختاری ژن‌های تنظیمی شناسایی شده در بافت پستانی گاو شیری در سری آزمایش GSE24560

PubMed Product ID	مکان	جفت باز	RefSeq ID	Entrez ID	PubMed ID
Matrix Gla Protein, mRNA	کروموزوم ۵	۳۵۰۹	۱۷۴۷۰۷/۳	۲۸۲۶۶۰	MGP
Aldehyde Oxidase 1, mRNA	کروموزوم ۲	۷۱۵۲۴	۱۷۶۶۶۸/۳	۳۳۸۰۷۴	AOX1
S100 Calcium Binding Protein A12, mRNA.	کروموزوم ۳	۱۱۲۵	۱۷۴۶۵۱/۳	۲۸۲۴۶۷	S100A12
Chemokine (C-C Motif) Ligand 2, mRNA	کروموزوم ۱۹	۱۸۶۷	۱۷۴۰۰۶/۲	۲۸۱۰۴۳	CCL2
BCL2 Related Protein A1, mRNA.	کروموزوم ۲۱	۹۶۵۶	۰۰۱۰۳۷۱۰۰/۳	۲۸۲۱۵۱	BCL2A1

علاوه بر فراخوانی سلول‌های ایمنی به محل التهاب، سبب ایجاد مکانیسم اولیه‌ی دفاع بدنی در برابر عفونت و التهاب می‌شوند. این پروتئین در سلول‌های اپیتلیال پستانی در طول دوره‌ی شیردهی و تحریک یا عفونت باکتری تولید می‌شود. سایر محققان نیز نقش ژن S100A12 را در بروز بیماری ورم پستان مورد تأیید قرار داده‌اند (Regennhard و همکاران، 2012؛ Yang و همکاران، 2011).

ژن CCL2 یک کدکننده پروتئین و حاوی ساختار موتیف CC شیموکین لیگاند ۲ می‌باشد. این سیتوکین‌ها یک ابرخانواده از سیتوکین‌های ترشح‌شده مؤثر در فرآیندهای تنظیمی ایمنی و التهابی می‌باشند (Rebhan و همکاران، 1997). در پژوهش Gronlund و همکاران، (2003) ژن CCL2 به‌عنوان شاخص

نتایج نشان داد ژن MGP یک ژن کدکننده پروتئین با نام پروتئین ژلاماتریکس می‌باشد، که در فرآیند تشکیل ماتریکس ارگانیک استخوان‌ها و غضروف دخالت دارد (Rebhan و همکاران، 1997). ژن AOX1 (آلدهید اکسیداز یک) یک ژن کدکننده پروتئین می‌باشد که در حالات عادی سبب تولید هیدروژن پراکسید و در شرایط خاص سبب کاتالیز تشکیل سوپراکسیدازها می‌شود (Rebhan و همکاران، 1997). ژن S100A12 نیز یک ژن کدکننده پروتئین می‌باشد که حاوی موتیف‌های اتصال‌ی کلسیم EF-2¹ می‌باشد. این ژن یک پروتئین متصل‌شونده به عناصر کلسیم، روی و مس می‌باشد که نقشی اساسی در تنظیم فرآیندهای التهابی و پاسخ‌های ایمنی دارد (Rebhan و همکاران، 1997). SAA در هنگام التهاب شدید ترشح شده و

های نامبرده به عنوان زیست‌نشانگر و شاخص در درمان بیماری ورم پستان معرفی شده‌اند. با این حال سه ژن تنظیمی شناسایی شده دیگر در این تحقیق برای اولین بار به عنوان ژن‌های تنظیمی مؤثر بر بیماری ورم پستان شناخته شد (MGP, BCL2A, AOX1). از آنجایی که علاوه بر باکتری گرم منفی *E. coli* سایر باکتری‌های دیگری نیز در بروز بیماری ورم پستان نقش دارند، می‌بایست در تعمیم نتایج این پژوهش با احتیاط برخورد نمود. در جدول ۴ اطلاعات حاشیه‌نویسی عملکردی ژن‌های تنظیمی قابل مشاهده است. نتایج نشان دادند که ژن‌های هاب شناسایی شده در مسیرهای ترشحي، سيتوپلاسم و پاسخ به التهاب‌ها نقش عملکردی ایفا می‌کنند. در کل نتایج نشان داد اکثر ژن‌های تنظیمی شناسایی شده در بروز بیماری ورم پستان مربوط به ژن‌های سامانه ایمنی نقش دارند. هم‌چنان که Schwerin و همکاران، (2003) گزارش کردند در شرایط بروز بیماری ورم پستان اغلب ژن‌های متفاوت بیان شده در مسیرهای ایمنی نقش داشته و معمولاً بیان این ژن‌ها طی بروز این بیماری تغییر می‌کند.

ژنتیکی در بروز بیماری ورم پستان گزارش شد. اعضای خانواده ژنی CCL2 از گروه‌های شیموکینی هستند که به وسیله‌ی مونوسیت‌ها و ماکروفازها ترشح شده و در هنگام بروز ورم پستان با فراخوانی سلول‌های بنیادی خون‌ساز به بازسازی بافت پستان می‌پردازند. ژن BCL2A نیز ژن کدکننده پروتئین و تنظیم‌کننده مرگ برنامه‌ریزی شده سلولی می‌باشد. این ژن یک نوع از پروتئین غشایی خارجی میتوکندریایی را رمزگذاری می‌کند که از مرگ برنامه‌ریزی شده برخی سلول‌ها از جمله لیمفوسیت و سلول‌های عصبی ممانعت می‌کند (Rebhan و همکاران، 1997). در پژوهش Abad و همکاران (2015) با استفاده از داده‌های ریزآرایه DNA به شناسایی ژن‌های هاب مؤثر بر بروز بیماری ورم پستان ناشی از باکتری *E. coli* پرداختند و به‌غیر از ژن خانواده CCLX ژن‌های متفاوتی را به عنوان ژن‌های هاب معرفی کردند (IL6, ICAM1, JFIH1, PARP14, IL1B, JSG15, IRF9, NOS2, GRO1IFIT3, CCL5). ژن‌های تنظیمی S100A2 و CCL2 در پژوهش-

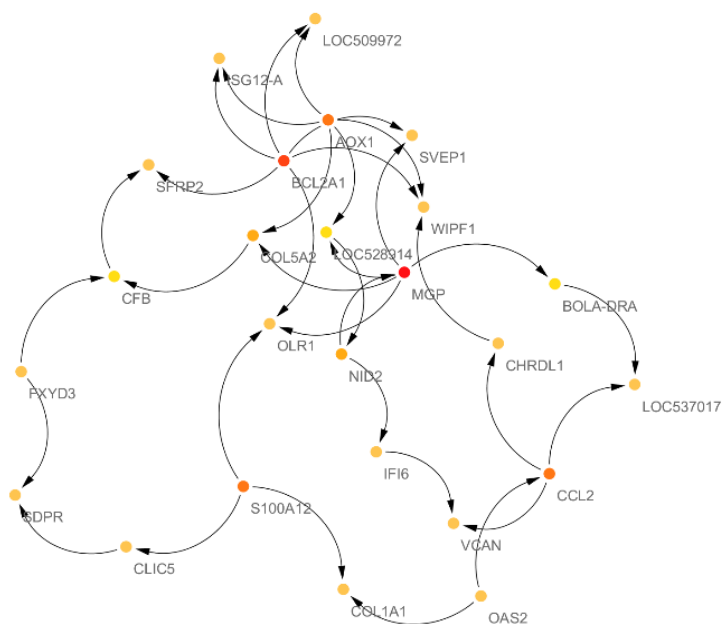
جدول ۴- حاشیه‌نویسی عملکردی ژن‌های تنظیمی شناسایی شده در بافت غدد پستانی

FDR	BJ	BF	FE	P	%	C	Term	Genes List
10.89	0.34	0.34	12.30	0.01	60.0	3	Secreted	CCL2, MGP, S100A12
35.30	0.54	0.79	6.18	0.04	60.0	3	Cytoplasm	AOX1, BCL2A1, S100A12
46.16	0.97	0.97	25.71	0.05	40.0	2	inflammatory response	CCL2, S100A12

*C: count, P: p value, FE: Fold Enrichment, BF: Bonferroni, BJ: Benjamini

ژن‌ها نقش مستقیمی در بروز بیماری ورم پستان ندارد می‌بایست نقش و عملکرد آن‌ها از طریق مطالعات آزمایشگاهی بیشتر مورد بررسی قرار گیرد. هم‌چنین می‌بایست به بررسی تأثیر الگوریتم‌های متفاوت شبکه‌بیزی و معیارهای توپولوژیکی متفاوت برای شناسایی نوع ژن‌های هاب نیز توجه خاصی مبذول داشت. بین سایر ژن‌های متفاوت بیان شده در این شبکه هیچ‌گونه روابطی شناسایی نشد. ژن‌های بارنگ زرد معرف سایر ژن‌ها می‌باشند. بنابراین از نمایش گرافیکی سایر ژن‌ها اجتناب گردید.

در شکل ۱ نمایش گرافیکی ژن‌های تنظیمی شناسایی شده در بروز بیماری ورم پستان در بافت غدد پستانی نشان داده شده است. در این شکل تنها به نمایش قسمتی از شبکه ژنی اکتفا شده است. ژن‌های با رنگ قرمز و نارنجی معرف ژن‌های هاب مؤثر در بروز بیماری ورم پستان می‌باشند (MGP, AOX1, CCL2, BCL2A1 و S100A12). ژن‌های MGP و BCL2A1 بیشترین مقدار اتصالات را به خود اختصاص داده و به نسبت سه ژن هاب دیگر نقش تنظیمی بالاتری را بر روی سایر ژن‌های متفاوت بیان شده دارند. از آنجاکه این



شکل ۱- نمایش روابط علی- معلولی و شناسایی ژن‌های تنظیمی (رنگ تیره) مؤثر بر ورم پستان در بافت پستانی

3'UTR قرار گرفته‌اند. نتایج حاصل از این پژوهش با نتایج گزارش شده توسط پژوهشگران مختلف در رابطه با کاربرد SNPها در سطح ژنومیکس، که اثرات معنی‌دار SNPهای شناسایی شده در این ناحیه برای صفات مختلف را شناسایی کرده‌اند (Bartel و همکاران، ۲۰۰۹؛ Li و همکاران، ۲۰۱۵؛ Kgwatalala و همکاران، ۲۰۰۹) مطابقت دارد.

در جدول ۵ اطلاعات SNP منطبق بر ژن‌های هاب شناسایی شده در بافت پستانی قابل مشاهده می‌باشد. مکان قرارگیری جهش‌ها و نوع جهش‌های منتهی به ایجاد SNP به خوبی در این جدول قابل مشاهده می‌باشد. ژن AOX1 دارای بیشترین تعداد SNP (۱۹۱۷۵) می‌باشد که از این تعداد بیشتر آنها در اینترون ژن قرار گرفته‌اند. تعداد خیلی کمی از این SNPها در ناحیه

جدول ۵- فهرست انواع SNPهای منطبق بر ژن‌های هاب شناسایی شده در بافت غدد پستانی*

FSM	SNM	MSM	NSM	Intron	5NG	5'UTR	3NG	3'UTR	تعداد کل SNP**	Gene ID
۳	۱۰۸	۱۵۵	۳	۲۸	۵۶۰	۱۰۶	۸۵	۲۸	۱۰۰۹	BCL2A
۴	۳۳	۵۳	۱	۲۶۹	۴۱۶	۳۲	۱۲۳	۸۴	۱۰۰۸	CCL2
۶	۴۰	۷۳	۲	۴۰۷	۴۶۳	۱۹	۱۲۹	۳۶	۱۱۶۰	S100A12
۴۰	۳۷۴	۹۸۸	۴۷	۱۷۱۴۰	۴۴۴	۵۸	۸۳	۱۱۵	۱۹۱۷۵	AOX1
۶	۴۴	۱۰۸	۱۰	۶۹۹	۴۳۴	۵۹	۶۴	۲۶۹	۱۶۷۳	MGP

NG: 3' Near gene, 5NG: 5' Near gene, NSM: Nonsense, MSM: Missense, SNM: Synonymous, FSM: Frameshift Mutation

**به دلیل وجود چندین جهش برای یک نوکلئوتید جمع کل SNPها لزوماً با جمع تک تک اجزا یکسان نمی‌باشد.

با این حال نتایج برخی پژوهش‌ها نشان دادند که SNP موجود در ناحیه 5'UTR تأثیر زیادی بر روی صفات مورد مطالعه ندارند. Hajibemani و همکاران (2014) با مطالعه SNP‌های موجود در ناحیه 5'UTR ژن لاکتوفرین و ارتباط آن‌ها با بروز بیماری ورم پستان، نتوانستند هیچ‌گونه رابطه‌ای بین وجود جهش در این ناحیه و بروز بیماری ورم پستان تشخیص دهند. هم‌چنین نتایج این پژوهش نشان داد بیشترین نوع جهش‌های منتج به ایجاد SNP از نوع جهش‌های دگر معنی هستند. نتایج تحقیقات سایر پژوهشگران نیز بر این موضوع دلالت دارند (Grisart و همکاران، 2001). با وجود اهمیت جهش‌های بی‌معنی در این پژوهش، این نوع جهش‌ها تعداد خیلی کمتری از SNP‌ها را به خود اختصاص دادند. با در جدول ۵ مشاهده می‌گردد که در ناحیه 5'UTR تعداد بسیار زیادی از SNP‌ها شناسایی شده است. با توجه به اینکه miRNAها معمولاً در حیوانات عالی به این نواحی متصل شده و سبب تصحیحات تنظیمی ژنی می‌شوند، می‌توان گفت که در ژن‌های مورد نظر، فراژنتیک نیز می‌تواند بر عملکرد این ژن‌ها اثرگذار باشد. نتایج تجزیه و تحلیل تعداد SNP‌های مرتبط با ژن‌های تنظیمی نشان داد که بیشترین تعداد SNP‌ها در ناحیه اینترون و ناحیه نزدیک به 5'UTR وجود دارند. با توجه به نتایج پژوهش‌های پیشین توصیه می‌شود که با توجه به تعداد کم و اهمیت بالای SNP‌های موجود در ناحیه 3'UTR، توجه خاصی به SNP‌های مرتبط با ژن‌های تنظیمی موجود در این ناحیه داشت. هم‌چنین با توجه به وجود تعداد زیاد جهش‌های بی‌معنی برای ژن‌های تنظیمی شناسایی شده، می‌بایست به بررسی بروز این نوع جهش‌ها در ناحیه خود ژن و یا پروموتور ژن‌های تنظیمی توجه خاصی مبذول داشت. تاکنون برای استخراج روابط علی بیشتر از روش‌های تنظیمی استفاده شده است. با این حال استفاده از روش مسائل معکوس برای یافتن پاسخ به سئوالات زیستی که هم‌چون داده‌های بیان ژنی ریزآرایه DNA که ماهیت احتمالاتی و تصادفی داشته کمتر مورد استفاده قرار گرفته است. روش‌های تنظیمی روش‌های مورد استفاده اغلب پژوهشگران در مواجهه با چنین مسائلی می‌باشد. وجود روابط تنظیمی زیستی خود باعث

ایجاد ابهامات زیادی در چگونگی انتخاب نوع روش برای حل مشکلات است. روش‌های تنظیمی همواره یکنواختی را بر مدل مورد نظر اعمال می‌کنند (Gholami and Siahkoochi, 2009). بنابراین روش‌های تنظیمی می‌توانند سبب نادیده انگاشتن ارتباط‌های ژنی موقت در داده‌های ترانسکریپتومی شوند. اگر ژنی باعث تغییر بیان ژن دیگری شود، آیا با در دست داشتن اطلاعات ژن تحت تأثیر واقع شده (معلول)، می‌توان فهمید کدام ژن باعث ایجاد آن تغییر شده است (علت)؟ یا چه بیماری می‌تواند باعث القای این بیان در ژن تحت تأثیر واقع شده (معلول) باشد؟ شبکه‌های تنظیمی تا حدودی توانسته‌اند به این نوع سئوالات زیستی پاسخ دهند. هر چند کاربرد مسائل معکوس در داده‌های ژنی اخیراً مرسوم گشته است (Li و همکاران، 2018). در بیشتر پژوهش‌های بیان ژنی، هدف استخراج زیرمجموعه‌هایی از ژن‌هاست که می‌توانند در حل یک مسئله به‌عنوان نشانگرهای زیستی پیش‌بینی‌کننده عمل کنند (Mungloo- Dilmohamud و همکاران، 2017). در عمل این روش‌ها تنها سبب شناسایی دسته‌های ژنی و یا ژن‌هایی با بیان مشابه می‌شوند و هیچ‌کدام از آن‌ها اشاره‌ای در رابطه با کشف رابطه علی و معلولی بین ژن‌ها و شناسایی ژن‌های تنظیمی نداشته‌اند. از یک سو از نگرش زیست‌شناسی سامانه‌ای، شبکه‌ی پیچیده‌ای از برهمکنش خطی و غیرخطی تعداد بسیاری از ژن‌ها و محصولات آن‌ها در تعامل پیچیده با محیط بر روی بروز فنوتیپی صفات در اصلاح نژاد تأثیر دارند. از سوی دیگر، اساس علم اصلاح نژاد بر اساس ارزیابی و انتخاب ژنتیکی افراد برتر برای صفت یا صفات اقتصادی مهم در جمعیت‌ها می‌باشد. بنابراین، بالاچار می‌بایست با نگرش زیست‌شناسی سامانه‌ای روابط بین ژنی را با نگرش گسترده‌تر و مجموعه‌وار در مورد برهم‌کنش ژن‌ها (روابط علی و معلولی) و با استفاده از شبکه‌های تنظیمی ژنی مورد بررسی قرار داد. استخراج روابط تنظیمی در لایه‌های مختلف اومیکس، اگر چنانچه در قالب برنامه‌های اصلاح نژاد آینده‌گرا صورت گیرد، شاید بتواند در بهبود ژنتیکی صفات مؤثر واقع شود. بنابراین لزوم آشنایی متخصصین اصلاح نژاد با روش‌های زیست‌شناسی سامانه‌های

- Alexandre, P.A., Kogelman, L.J., Santana, M.H., Passarelli, D., Pulz L.H., Fantinato-Neto, P. *et al.* (2015). Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. *BMC genomics*. 16: 1.
- Bar, D., Grohn, Y.T., Bennt, G., Gonzalez, R.N., Hertl, J.A., Schulte, H.F. *et al.* (2008). The cost of generic clinical mastitis in dairy cows as estimated by using dynamic programming. *Journal of Dairy Science*, 91: 2205- 2214.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, 136: 215-33.
- Bolstad, B.M., Collin, F., Brettschneider, J., Simpson, K., Cope, L., Irizarry, R.A. *et al.* (2005). Quality Assessment of Affymetrix GeneChip Data." In Gentleman R, Carey V, Huber W, Irizarry R and Dudoit S (eds.), *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, chapter 3, pp. 33–47. Springer, New York.
- Cánovas, A., Reverter, A., DeAtley, K.L., Ashley, R.L., Colgrave, M.L., Fortes, M.R., *et al.* (2014). Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PloS one*. 9: e102551.
- Davis, S. and Meltzer, P. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*. 14: 1846–1847.
- Fortes, M.R., Nguyen, L.T., Neto, L.R.P., Reverter, A., Moore, S.S., Lehnert, S.A. *et al.* (2016). Polymorphisms and genes associated with puberty in heifers. *Theriogenology*. 86(1): 333-339.
- Gámez, J.A., Mateo, J.L. and Puerta, J.M. (2011). Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood. *Data Mining and Knowledge Discovery*. 22:106-148.
- Gene expression omnibus (GEO). (2017) Available at <https://www.ncbi.nlm.nih.gov/geo/>. NCBI, USA.

محاسباتی که برای استخراج رابطه علی- معلولی بین سازوکارهای زیست مولکولی توسعه یافته‌اند، ضروری به نظر می‌رسد.

نتیجه گیری

نتایج خوشه‌بندی و هستی‌شناسی گروه‌های ژنی مؤثر بر بیماری ورم پستان در بافت غدد پستانی گاو شیری نشان داد که بیشتر ژن‌های متفاوت بیان شده در دسته‌های ژنی مقاومت به بیماری‌ها، پاسخ به محرک‌ها و التهاب قرار می‌گیرند. در این پژوهش تعداد ۵ ژن هاب شناسایی شدند (S100A12، CCL2، BCL2A، AOX1 و MGP). همچنین نتایج حاصل نشان داد ژن‌های AOX1 و BCL2A، MGP به‌عنوان ژن‌های تنظیمی ژنوم گاو در پاسخ به بیماری ورم پستان ناشی از باکتری‌های *E. coli* و استافیلوکوکوس اورئوس در گاوهای شیری نقش دارند. اطلاعات حاصله از جدول حاشیه‌نویسی عملکردی و هستی‌شناسی ژنی نشان داد ژن‌های تنظیمی نقشی همانند ژن‌های متفاوت بیان‌شده (پاسخ-های التهابی، پاسخ به محرک‌ها و مرگ برنامه‌ریزی‌شده سلولی) در بروز فنوتیپی بیماری ورم پستان در گاو شیری دارند.

منابع

- شریفی، س.، پاکدل، ع. و ابراهیمی، ا. (۱۳۹۶). فراتحلیل (متا-آنالیز) داده‌های بیان ژن بافت پستان آلوده شده با باکتری اشریشیاکلی در گاوهای شیری. *مجله علوم دامی ایران*، دوره ۴۸، شماره ۳، صص ۳۵۲-۳۴۳.
- Abad, S. V., Alijani, S., Kia, H. D., Zali, H., Karim, S. K., and Pashaie, M. B. (2015). Bioinformatics analysis of *E. coli* causing mastitis in Holstein dairy cattle by using microarray data. *Koomesh*, 17(1).
- Adomas, A., Heller G., Olson, A., Osborne, J., Karlsson, M., Nahalkova, J. *et al.* (2008). Comparative analysis of transcript abundance in *pinus sylvestris* after challenge with a saprotrophic, pathogenic or mutualistic fungus. *Physiology*. 28: 885-897.

- Gholami, A. and Siahkoohi, H.R. (2009). A two-step wavelet-based regularization for linear inversion of geophysical data. *Geophysical Prospecting*, 57: 847-862.
- Grisart, B., Coppeters, W., Farnir, F., Karim, L., Ford, C., Berzi, P. *et al.* (2002). Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome research*, 12(2): 222-231.
- Gronlund, U., Hulten, C., Eckersall, P.D., Hogarth, C., Waller, K.P., (2003). Haptoglobin and serum amyloid A in milk and serum during acute and chronic experimentally induced *Staphylococcus aureus* mastitis. *Journal of Dairy Research*, 70: 379-386.
- Hajibemani, A., Sharifiyazdi, H., Mirzaei, A., and Ghasrodashti, A. R. (2012). Characterization of single nucleotide polymorphism in the 5'-untranslated region (5'-UTR) of Lactoferrin gene and its association with reproductive parameters and uterine infection in dairy cattle. In Veterinary Research Forum (Vol. 3, No. 1, p. 37). Faculty of Veterinary Medicine, Urmia University, Urmia, Iran.
- Heckerman, D. (1997). Bayesian networks for data mining. *Data Mining and Knowledge Discovery*. 1: 79-119.
- Horvath, S. (2011). *Weighted network analysis: applications in genomics and systems biology*. Springer Science & Business Media, Berlin, Germany. pp: 90-149.
- Huang, D.W., Sherman, B.T, Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1):1-13.
- Jorgensen, H. B., Buitenhuis, B., Rontved, C. M., Jiang, L., Ingvarsten, K. L. and Sorensen, P. (2012). Transcriptional profiling of the bovine hepatic response to experimentally induced *E. coli* mastitis. *Physiol Genomics*, 44: 595-606.
- Kauffmann, A., Gentleman, R. and Huber, W. (2009). arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 25(3): 415-6.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002). The human genome browser at UCSC. *Genome research*, 12(6): 996-1006.
- Kgwatalala, P. M., Ibeagha-Awemu, E. M., Hayes, J. F., and Zhao, X. (2009). Stearoyl-CoA desaturase 1 3' UTR SNPs and their influence on milk fatty acid composition of Canadian Holstein cows. *Journal of Animal Breeding and Genetics*, 126(5): 394-403.
- Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 9: 559.
- Li, K., Chen, Y., Li, W., He, J. and Xue, Y. (2018). Improved gene expression programming to solve the inverse problem for ordinary differential equations. *Swarm and Evolutionary Computation*, 38: 231-239.
- Li, Q. L., Zhang, Z. F., Xia, P., Wang, Y. J., Wu, Z. Y., Jia, Y. H. *et al.* (2015). A SNP in the 3'-UTR of HSF1 in dairy cattle affects binding of target bta-miR-484. *Genet Mol Res*, 14(4): 12746-55.
- Moran, B., Butler, S.T., Moore, S.G., MacHugh, D. and Creevey, C.J. (2015). Differential gene expression in the endometrium reveals cytoskeletal and immunological genes in lactating dairy cows genetically divergent for fertility traits. *Reproduction, Fertility and Development*. 29(2): 274-82.
- Mungloo-Dilmohamud, Z., Jaufeerally-Fakim, Y. and Peña-Reyes, C. (2017). A Meta-Review of Feature Selection Techniques in the Context of Microarray Data. In: Proceedings of International Conference on Bioinformatics and Biomedical Engineering. Switzerland, Springer, Cham, 33-49.

جدول ۱ پیوست- تعداد آزمایش‌های ریز آرایه DNA انجام گرفته به ازای هر بافت در هر GSE در گونه گاو

تخمندان		رحم		کبد		غدد پستانی		ماهیچه اسکلتی	
تعداد	سری آزمایش	تعداد	سری آزمایش	تعداد	سری آزمایش	تعداد	سری آزمایش	تعداد*	سری آزمایش
۱۲	GSE11495	۱۸	GSE11835	۳۶	GSE10695	۱۰	GSE15019	۱۵	GSE19586
۱۶	GSE18290	۱۱	GSE1414	۱۸	GSE19696	۱۰	GSE15020	۶	GSE21782
۱۸	GSE21005	۱۱	GSE15544	۱۹	GSE23894	۱۰	GSE15022	۱۲	GSE37127
۳۰	GSE23348	۱۲	GSE16880	۱۲	GSE24596	۳۰	GSE15025	۶۶	GSE42932
۱۰	GSE23449	۱۸	GSE17921	۵۴	GSE25319	۱۲	GSE18768	۹	GSE49992
۱۲	GSE24477	۱۲	GSE21049	۲۸	GSE45857	۴۹	GSE24217	۱۴	GSE52145
۱۹	GSE39589	۱۶	GSE22590	۱۰	GSE87391	۸۸	GSE24560	۲۰	GSE75347
۱۵	GSE42535	۶	GSE24936			۵۴	GSE25319	۶	GSE75348
۳	GSE42838	۱۲	GSE27817			۳۰	GSE25413	۱۰	GSE9256
۱۶	GSE44704	۳۳	GSE30694			۱۴	GSE27380		
۱۸	GSE46996	۷۹	GSE33030			۸	GSE27840		
۱۹	GSE49505	۱۶	GSE33496			۳۶	GSE27851		
۶	GSE65269	۱۲	GSE36079			۲۴	GSE32186		
۶	GSE69094	۱۲	GSE36080			۲۸	GSE33720		
۱۹	GSE89279	۱۲	GSE38512			۳۰	GSE50685		
		۶	GSE39796			۱۴	GSE65059		
		۱۲	GSE40101			۷	GSE9344		
		۹	GSE53457						
		۸	GSE74987						
		۱۴	GSE75750						
فراوانی GSE به ازای هر بافت									
٪۱۷		٪۲۶		٪۱۴		٪۳۵		٪۸	

*تعداد نمونه مورد استفاده در هر سری آزمایشی (GSE)

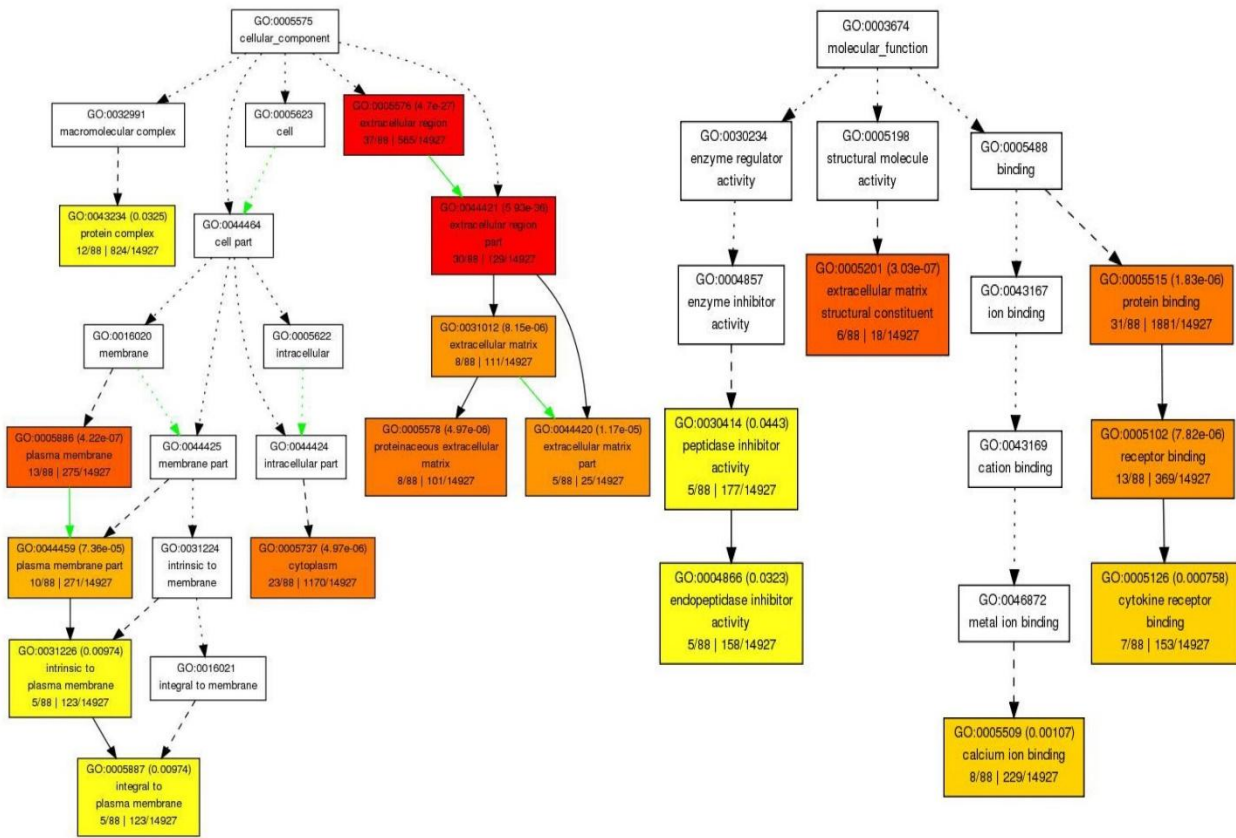
جدول ۲ پیوست - فهرستی از ابزارهای برخط برای استنباط شبکه ژنی

ابزار	دسته	مدل	روش	تارنما
BNW	پیوسته	شبکه بیزی	تحلیل بیزی	http://bnw.genenetwork.org/BNW/sourcecodes/home.php
javabayes	پیوسته	شبکه بیزی	تحلیل بیزی	https://www.cs.cmu.edu/~javabayes/
SEBINI	پیوسته	شبکه بیزی	تئوری اطلاعات	http://www.emsl.pnl.gov/NIT/NIT.gov
LibB	پیوسته	شبکه بیزی	تحلیل بیزی	http://www.cs.huji.ac.il/labs/compbio/libb/
BANJO	پیوسته	شبکه بیزی	تحلیل بیزی	http://www.cs.duke.edu/~amink/software/
Bayes Server	پیوسته	شبکه بیزی	تحلیل بیزی	https://www.bayesserver.com/docs/introduction/bayesian-networks
BayAnet	پیوسته	شبکه بیزی	تحلیل بیزی	http://www.cs.man.ac.uk/~gbrown/bayes_nets/#

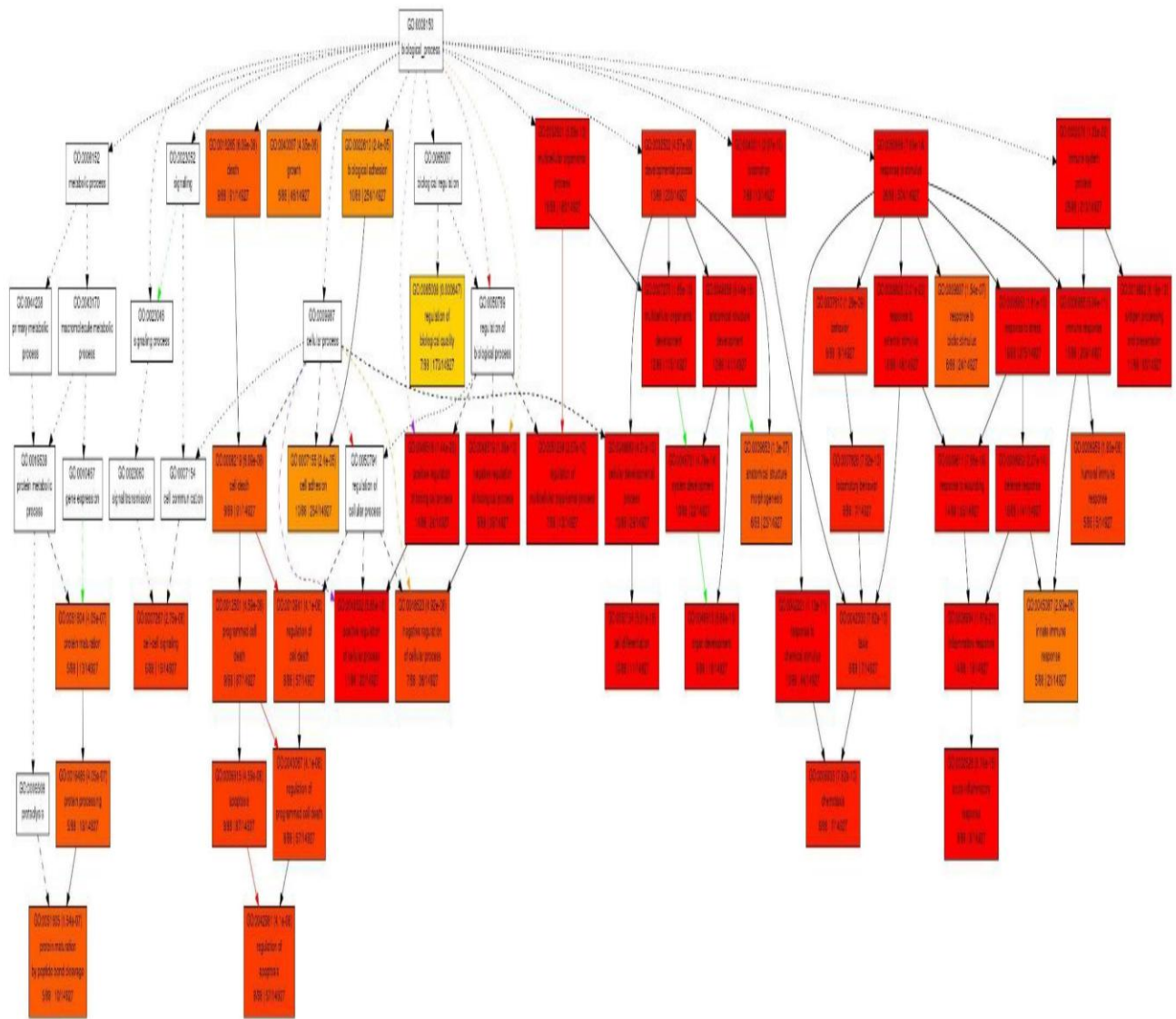
جدول ۳ پیوست - خصوصیات بسته‌های نرم افزاری موجود در CRAN جهت استخراج شبکه تنظیمی به روش بیزی

bnlearn	bnstruct	catnet	deal	pcalg	gRbase	آیتم
						داده های گسسته
	○	○				داده های پیوسته
○	○	○		○	○	داده های مختلط
		○	○			الگوریتم های محدودیت گرا
				○	○	الگوریتم های نمره گرا
		○	○	○	○	الگوریتم های هیبرید
	○		○	○		دستکاری ساختار
	○					برآورد پارامتر
			○	○	○	پیش بینی

deal توانایی یادگیری ساختار شبکه به همراه یادگیری فراسنجه آن به روش بیزی برای داده‌های گسسته، پیوسته و مختلط را داراست. در این بسته نرم افزاری چگالی پسین شبکه به عنوان تابع نمره و از شروع کننده‌های مجدد تصادفی استفاده می‌کند. Pcalg نرم‌افزاری فراگیر برای اجرای الگوریتم PC است و اختصاصاً برای اندازه‌گیری اثرات تصادفی طراحی شده است. این نرم افزار علاوه بر توانایی کار با داده‌های گسسته و پیوسته، قادر به محاسبه اثرات متغیرهای پنهان شبکه نیز هست. Catnet با هدف کار بر روی داده‌های گسسته و شبکه‌های بیزی پایا یا ثابت از دیدگاه فراوانی گراها توسعه داده شده است. یادگیری ساختار در دو گام صورت می‌پذیرد. ابتدا، نظم‌دهی گره‌های گراف با استفاده از چسپاندن شبیه سازی انجام می‌شود، متناوباً، یک نوع ترتیب بندی سفارشی گره‌ها توسط کاربر نیز قابل اعمال است. همچنین توانایی یادگیری فراسنجه‌های شبکه و پیش بینی از دیگر قابلیت‌های این بسته نرم افزاری است. بسته gRbase در طبقه بندی یادگیری فراسنجه‌های شبکه قرار می‌گیرند. آنها تحت فرض گسسته بودن تمامی متغیرها، با دستکاری فراسنجه‌های شبکه توانایی استنباط شبکه و پیش بینی را نیز دارند. bnstruct علاوه بر یادگیری ساختار شبکه با استفاده از الگوریتم‌های محدودیت گرا، نمره گرا و هیبرید، قادر به یادگیری پارامترهای شبکه از داده‌های گمشده را نیز داراست. bnlearn برای اجرای تعدادی از الگوریتم‌ها، جهت استنباط در مورد ساختارهای بیزی کاربرد دارد. این مجموعه قابلیت کار بر روی سه نوع داده نرمال، گسسته و شرطی پیوسته را داراست. این مجموعه نرم افزاری علاوه بر استنباط شبکه، قادر به یادگیری فراسنجه‌های شبکه و یادگیری ساختار شبکه نیز می‌باشد.



شکل ۱ پیوست- نتایج غنی سازی ژن های متفاوت بیان شده در عبارات مؤلفه های سلولی و عملکرد مولکولی در بافت پستانی



شکل ۲ پیوست- نتایج غنی سازی ژن های متفاوت بیان شده در عبارت فرآیندهای زیستی در بافت غدد پستانی