

مدل سازی شبکه ژنی و کاوشگری در برخی از بافت های گونه گاو با استفاده از داده های ریزآرایه DNA

امین مرتضوی^۱، امیر رشیدی^{۲*}، مصطفی قادری-زفrehی^۳، پرهام مرادی^۴ و محمد رزم کبیر^۵

۱، ۲ و ۵. دانشجوی دکتری، استاد و دانشیار، ژنتیک و اصلاح دام، گروه علوم دامی، دانشگاه کردستان، سنندج، ایران

۳. استادیار زیست سامانه های محاسباتی، گروه علوم دامی دانشگاه یاسوج، یاسوج، ایران

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

(تاریخ دریافت: ۱۳۹۶/۸/۲۴ - تاریخ پذیرش: ۱۳۹۷/۵/۱)

چکیده

هدف از این پژوهش، شناسایی ژن ها و کاوشگرهای هاب ناشی از ایجاد شبکه های بیزی در داده های بیان ژن و کاوشگری در بافت های مختلف گونه گاو با استفاده از داده های ریزآرایه DNA بود. با استفاده از داده های خام بیان ژن و کاوشگری در هر بافت، ژن ها و کاوشگرهایی که عامل بیشترین میزان واریانس بیان ژن بودند، شناسایی شده و سپس شبکه بیزی، برای آن ها برآش داده شد. ژن ها و کاوشگرهای هاب با استفاده از نسبت فراسنجه های درجه درون و بیرون شناسایی شدند. اندازه پوشش مارکفی در شبکه های مختلف متفاوت بود که به احتمال بیانگر وجود زیرساختارهای ساختارشناختی (توپولوژیکی) برای شبکه های یادشده در بافت های مختلف است. در روش شبکه ژن محور، در بافت ماهیچه CBR1، LOC788826، در بافت پستانی NID2، COL5A2، LOC616942، FXYD3، در بافت کبدی LOC100132279، MGC127133، MBOAT2، CLDN2، ANKRD1، در بافت رحم JGFBP1، DGAT2، CKMT1، ISG15، CKMT1 و در بافت تخمدان LOC286871 و INHBA به عنوان ژن های هاب شناسایی شدند. همچنین در روش ایجاد شبکه کاوشگر محور، مشخص شد که دو ژن JSP.1 و BOLA-DQB در بافت های کبد و رحم فعال هستند. ولی این ژن ها در روش ژن محور در ایجاد شبکه ژنی به عنوان ژن هاب استخراج نشدند.

واژه های کلیدی: توپولوژی، شبکه بیزی، هاب.

Gene-based and probe-based network modeling in some bovine tissues using DNA microarray data

Amin Mortazavi¹, Amir Rashidi^{2*}, Mostafa Ghaderi-Zefrehi³, Parham Moradi⁴ and Mohammad Razm Kabir⁵

1, 2, 5. Ph.D. Candidate, Professor and Associate Professor, Department of Animal Science, University of Kurdistan, Sanandaj, Iran

3. Assistant Professor, Department of Animal Science, Yasuj University, Yasuj, Iran

4. Associate Professor, Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran

(Received: Nov. 15, 2017 - Accepted: Jul. 23, 2018)

ABSTRACT

The aim of this study was to extract genes and probes hub based on Bayesian networks on probe and gene transcriptomic data over different tissues in Bovine species. Using raw probe and gene transcriptomic data in each tissue, genes and probes with the highest expression variances were extracted and fitted to Bayesian networks. The hub genes and hub probes were identified using the ratio of in-and-out degree. The size of the Markov Blanket was different in different networks. This might be indicative of existing of substructures topology of aforementioned network on different tissues. Using gene based network, in muscle (CBR1 and LOC788826); in mammary of (NID2, COL5A2, LOC616942 and FXYD3); in liver (LOC100132279; MGC127133; MBOAT2; CLDN2; ANKRD1; JGFBP1; DGAT2; CKMT1, ISG15, CKMT1), and in the ovary (LOC286871 and INHBA) were extracted as hub genes. It was shown that the hubs were different in different tissues. These results can be used for more accurate bioassays of each tissue. Using probe based network, two genes BOLA-DQB and JSP.1 would have function in liver and uterus tissues. The results of this study can reduce the gap between phonemics-and Genomics distance on investigated tissues in bovine species.

Keywords: Bayesian network, hub, topology.

* Corresponding author E-mail: arashidi@uok.ac.ir

مقدمه

پیچیدگی خاصی دارند، زیرا به دلیل چالش شمار کم نمونه‌ها در برابر شمار زیاد ژن‌ها اغلب منجر به یافته‌های چندگانه مثبت و منفی دروغین می‌شوند. این موضوع می‌تواند کار یادگیری ساختار شبکه‌ی بی‌زی^۵ و استنباط شبکه^۶ در زیست سامانه‌ها را مشکل کند (Elo *et al.*, 2005). به‌ویژه هنگامی که شبکه‌های بی‌زی برای تحلیل داده‌های بیان ژن (Friedman, 2004)، تعامل‌های پروتئین-پروتئین (Sachs *et al.*, 2005) و یکپارچه کردن داده‌های ژنتیکی ناهمگون (Chang & McGeachie, 2011) به‌کار می‌روند. در بسیاری از موارد، یادگیری شبکه‌های بی‌زی با استفاده از داده‌های با حجم بزرگ (چندین هزار متغیر) بدون در نظر گرفتن فرضیه‌های محدودکننده روی ساختار گراف‌های حلقه‌ای جهت‌دار با ماهیت توزیع‌های محلی^۷ امکان‌پذیر نیست (Scutari, 2014). بهینه‌سازی یا کاهش شمار آزمون‌های استقلال شرطی و امتیازدهی‌های شبکه یکی از راه‌حل‌هایی است که برای رفع این مشکل پیشنهاد شده است (Scutari, 2014). در این میان الگوریتم‌های نمره‌گرا از جمله تپه-نوردی^۸ کارایی خود را در مدل‌سازی شبکه‌های بی‌زی به‌خوبی نشان داده‌اند (Scutari, 2014). همچنین در این الگوریتم یافتن نقطه‌ی بهینه‌ی اصلی^۹ با توجه به امکان وجود چندین نقطه‌ی بهینه‌ی محلی^{۱۰} کار مشکلی است، به‌همین دلیل نوع خاصی از این الگوریتم با نام الگوریتم تپه‌نوردی با نقطه‌ی آغاز تصادفی^{۱۱} ایجاد شد که با توجه به وجود گزینه‌ها و حالت‌های مختلف در اجرای آن (شمار آغاز تصادفی، نقطه‌ی آغاز، شمار والد به ازای هر گره و ...)، یافتن بهینه‌ترین حالت در مدل‌سازی شبکه را با توجه به ساختار داده‌های موجود چالش‌برانگیز کرده است. به همین دلیل در این پژوهش سعی شد، برای آزمون مدل‌ها و حالت‌های مختلف الگوریتم تپه‌نوردی روی داده‌ی بیان ژنی بافت‌های مختلف از شبکه‌ی بی‌زی استفاده شود. همچنین تأثیر این حالت‌ها بر فراسنجه‌های هر شبکه‌ی برازش‌شده بحث و بررسی شد. از سوی دیگر توجه به

فناوری ریزآرایه DNA امکان بررسی بیان همزمان هزاران ژن را در یک بافت یا یاخته فراهم کرده است. هدف بیشتر آزمایش‌های ریزآرایه DNA به دست آوردن فهرستی از ژن‌های متفاوت بیان شده در یک شرایط خاص است. پیش از ظهور فناوری توالی‌یابی RNA-seq، یکی از گسترده‌ترین پلات‌فرم‌های مورد استفاده در پژوهش‌های رونوشت‌برداری (ترنسکریپتومی)، آرایه‌های افی‌متریکس^۱ بودند که به‌طور معمول این آرایه‌ها متشکل از ویفر کوارتزی ۵۰۰ هزار دئوکسی‌الیگونوکلوئوتید ۲۵ نوکلئوتیدی متصل‌به‌هم هستند که کاوشگر^۲ نامیده می‌شوند (Stalderi & Harrison, 2007). ترکیب کاوشگرها با هم یک مجموعه کاوشگری^۳ را ایجاد خواهد کرد. هر مجموعه کاوشگری به‌طور معمول معرف یک ژن است. شبکه‌های ژنی می‌توانند روی داده‌های کاوشگری و یا روی داده‌های بیان ژنی ایجاد شوند. مدل‌سازی شبکه‌های ژنی با استفاده از داده‌های بیان ژن (ژن یا کاوشگر محور) می‌تواند نگرشی عمیق نسبت به فرآیندهای زیستی در سطح شبکه را ارائه کند (Liu *et al.*, 2016). بسیاری از صفات مهم اقتصادی و بیماری‌ها در اصلاح نژاد دام حاصل اثر متقابل و پیچیده ژن-ژن و ژن-محیط هستند. امروزه در حوزه ژنتیک و اصلاح نژاد دام، برای استخراج شبکه‌ی بیان ژن از داده‌های بیان ژن یا بررسی‌های پویا ژنگانی (ژنومی)^۴ استفاده شده است (Alexandre *et al.*, 2015; Fortes *et al.*, 2012; Girard Komolka *et al.*, 2015; Kogelman *et al.*, 2014; Verardo *et al.*, 2016; Ramayo-Caldas *et al.*, 2016; Weber *et al.*, 2016). شماری از پژوهشگران کاربرد شبکه‌ی بی‌زی را برای شناسایی رابطه‌های علت و معلولی بین ژن‌ها بررسی کرده‌اند (Friedman, 2004; Ghaderi-Zefrehei *et al.*, 2015). با این حال، در حوزه ژنتیک انسانی و به‌ویژه ژنگان‌شناسی (ژنومیکس) از شبکه‌های بی‌زی استفاده فراوانی شده است (Hageman *et al.*, 2011; Malovini *et al.*, 2009; Sherif *et al.*, 2015; Stingo *et al.*, 2015).

5. Structural network learning

6. Network Inference

7. Local distribution

8. Hill climbing

9. Global optimum

10. Local optimum

11. Random-restart hill climbing

1. Affymetrix

2. Probe

3. Probe set

4. GWAS

استفاده کرد. در نهایت کاوشگرهای هاب را می‌توان منبع جدیدی از SNPهای عملکردی به شمار آورد. از سوی دیگر، تاکنون مقایسه‌ای بین دو روش بازسازی ساختار شبکه (استفاده از داده‌های بیان ژن یا کاوشگر) در چندین بافت مختلف در گونه گاوی انجام نگرفته است. بنابراین هدف از این پژوهش، استخراج ساختارشناختی بیان هر بافت، یافتن بهینه‌ترین حالت از الگوریتم تپه نوردی برای مدل‌سازی شبکه بیزی ژن محور و کاوشگر محور و مقایسه ژن‌ها و کاوشگرهای تنظیمی مشترک در بافت‌های مرتبط با صفات اقتصادی مهم در گونه گاو است.

مواد و روش‌ها

در این پژوهش از داده‌های پنج آزمایش ریزآرایه DNA مربوط به پنج بافت متفاوت (ماهچه‌های کالبدی یا اسکلتی: GSE75347، غدد پستانی: GSE24560، کبد: GSE25319، رحم: GSE33030 و تخمدان: GSE23348) گونه گاو استفاده شد. داده‌ها به صورت Soft فایل از پایگاه GEO استخراج شدند (<https://www.ncbi.nlm.nih.gov/geo>)، بنابراین لزومی به انجام عملیات پیش‌پردازش^{۱۱} داده‌های ریزآرایه DNA نبود. در این راستا با جستجوی ترکیبی واژگان کلیدی در گونه گاو، شمار آزمایش‌ها برای هر بافت مشخص شد. آن دسته از آزمایش‌هایی که دارای بیشترین آرایه (نمونه) و حاصل پلت فرم یکسان (GPL1221) بودند، استفاده شدند. همچنین سعی شد آزمایش‌هایی انتخاب شوند که تا حد ممکن شرایط آزمایشی و تیماری همسان با هم داشته باشند. اطلاعات کامل این آزمایش‌ها در جدول ۱ ارائه شده است. با در دست داشتن جدول حاشیه‌نویسی^{۱۲}، نگاشت کاوشگرها روی نام ژن‌ها انجام شد. برای انجام این نگاشت، از معیار، انحراف معیار میانه^{۱۳} استفاده شد. در مرحله بعد، برای جلوگیری از کاهش تأثیر نوبز در مدل‌سازی شبکه بیزی ژن محور و کاوشگر محور، تلاش شد ۱ درصد ژن‌ها و کاوشگرهایی که بیشترین

ویژگی‌های ساختار شناختی (توپولوژیکی) شبکه همچون پوشش مارکفی^۱، عامل شاخه‌ساز^۲، اندازه همسایگی^۳، درجه^۴، نزدیکی^۵، بینابینی^۶، درون به مرکز^۷ و برون به مرکز^۸ می‌تواند در شناخت ژن‌ها و کاوشگرهای هاب تأثیر بسزایی داشته باشند. بر پایه نظریه و مفهوم در یک شبکه نگاره‌ای (گرافیکی)، پوشش مارکفی به والدین، همسر و فرزندان یک ژن اطلاق می‌شود (Nagarajan *et al.*, 2013). استخراج پوشش مارکفی برای شبکه‌های مقیاس کوچک یکی از روش‌های مدل‌سازی شبکه است. از مفهوم‌های دیگر در شبکه بیزی می‌توان به اندازه همسایگی اشاره کرد که معرف ژن‌های مجاور یک ژن خاص است که در شناسایی ژن‌های با کارکرد همسان در شبکه کمک شایانی می‌کند. همچنین مفهوم عامل شاخه‌ساز نیز نمایانگر شمار ژن‌هایی است که می‌تواند توسط یک ژن خاص تحت تأثیر قرار بگیرند که یکی از عامل‌های ایجاد ویژگی دسته‌های ژنی^۹ و تنک‌کنندگی^{۱۰} در شبکه است (Nagarajan *et al.*, 2013). کاربرد این نوع الگوریتم و روش‌های استخراج ارتباط بین ژن‌ها می‌تواند سبب شناسایی رابطه علت- معلولی شود. با توجه به این موضوع که در حوزه ژنتیک و اصلاح نژاد، مدل‌سازی شبکه‌های بیزی بیان ژن و استخراج رابطه‌های علت- معلولی بین ژن‌ها در بافت‌های مختلف کمتر مورد توجه قرار گرفته است، نتایج به‌دست‌آمده از کاربرد این نوع الگوریتم سبب کاهش فاصله بین ژنومیکس و فنومیکس در این حوزه شده و سبب آسانگری در مدل‌سازی ساختار شبکه به دو صورت ژن محور و کاوشگر محور شده است. یک هدف عمده از مدل‌سازی شبکه بیزی کاوشگر محور، استخراج کاوشگرها است. کاوشگر، توالی الیگونوکلئوتیدی کوچک از cDNA هستند. لذا می‌توان SNPهای این قطعه‌ها را به‌طور غیرمستقیم استخراج و در برنامه‌های اصلاح نژاد

1. Markov Blanket
2. Branching Factor
3. Neighborhood Size
4. Degree
5. Closeness
6. Betweenness
7. Centrality
8. Eccentricity
9. Modularity
10. Sparsity

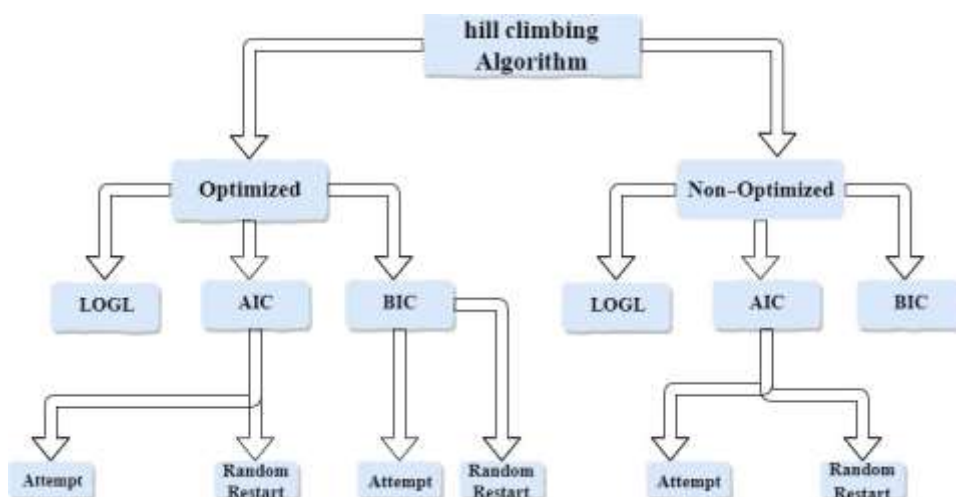
11. Preprocess
12. Annotation table
13. Median absolute deviation

داده‌های بیان ژن بافت ماهیچه کالبدی، به بررسی میزان کفایت تکرارهای بوت استرپ^۸ در ۵ تکرار مختلف (۱۰، ۲۰، ۵۰، ۸۰ و ۱۰۰) و برای هر کدام از تکرارها، توان و آستانه در دو حالت ۶۵ درصد و ۸۵ درصد توان^۹ و برای^{۱۰} یال (میزان توان ارتباط بین ژن‌ها) در نظر گرفته شد. به منظور بررسی تأثیر (نبود) بوت استرپ و همچنین عامل‌های محدودکننده (انتخاب اتصال‌هایی با بیشترین توان در آستانه‌ای معین) روی فراسنجه‌های شبکه‌ی بیزی برازش داده شده، در مقایسه با اجرای خود الگوریتم به‌تنهایی، سه مدل استفاده شد (استفاده از خود الگوریتم، الگوریتم+بوت استرپ و الگوریتم+بوت استرپ+فراسنجه‌های محدودکننده). جزئیات همه‌ی حالت‌های اجرای الگوریتم‌ها در زیرنویس جدول ۲ قابل مشاهده است. در شکل ۲ به صورت کلی حالت‌های شمار تکرارهای بوت استرپ و آستانه روی داده‌های بیان ژنی بافت ماهیچه کالبدی نشان داده شده است. فراسنجه‌های شبکه برای همه‌ی ۵ بافت مورد بررسی و هم در هر دو حالت اجرای شبکه‌ی ژن محور و کاوشگر محور برازش به‌دست آمدند. ژن‌ها و کاوشگرهای هاب با استفاده از ماتریس کناری^{۱۱} و نسبت فراسنجه‌ی درجه‌ی درون^{۱۲} به درجه‌ی بیرون^{۱۳} ناشی از شبکه‌های ژنی مربوطه به‌دست آمدند. فراسنجه‌های ساختارشناختی شبکه نیز با استفاده از بسته نرم‌افزاری igraph در محیط R برآورد شدند (Csardi & Nepusz, 2006). برای استخراج اطلاعات حاشیه‌نویسی ژن‌های شناسایی‌شده در ژنگان (ژنوم) گو‌ی گاو (نام RefSeq، PubMed و Entrez ID، شمار جفت باز و محل ژن روی کروموزوم) از سایت UCSC Genome Browser و از Assembly (Bos_taurus_UMD_3.1.1/bosTau8) استفاده شد (Kent et al., 2002). همچنین برای گردآوری اطلاعات مورد نیاز در مورد عملکرد زیستی ژن‌های شناسایی‌شده در ژنگان گاو از NCBI و <http://www.genecards.org/> استفاده شد (Rebhan et al., 1997).

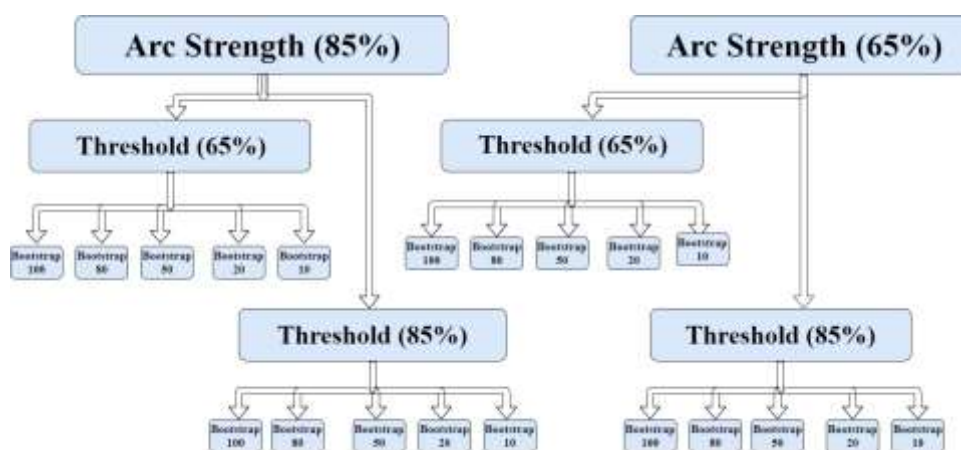
واریانس را در داده‌های بیان ژن ایجاد کرده بودند، انتخاب شوند (Langfelder & Horvath, 2008). سپس این‌ها به‌عنوان کاوشگر و ژن‌های متفاوت بیان شده^۱ شناخته و به‌عنوان ورودی شبکه‌ی بیزی استفاده شدند. برای استخراج شبکه و فراسنجه‌های شبکه از بسته نرم‌افزاری bnlearn در محیط R و برای برازش شبکه رویدادهای ژن محور و کاوشگر محور از الگوریتم تپه‌نوردی استفاده شد (Nagarajan et al., 2013) تا شبکه‌ی به‌دست‌آمده بیشترین همخوانی را با مجموعه داده‌ها داشته باشد. در کل این نوع الگوریتم‌ها، میزان همخوانی هر ساختار ممکنه شبکه را با مجموعه داده ارزیابی می‌کنند به‌گونه‌ای که هیچ ساختار ممکنه بیشترین همخوانی را نداشته باشد. با توجه به وجود معیارهای متفاوت برازش شبکه در الگوریتم تپه‌نوردی همچون تابع‌های امتیاز متفاوت (معیار اطلاعات بیزی^۲، معیار اطلاعات آکایک^۳، معادل درجیلت بیزی^۴ و لگاریتم درست‌نمایی^۵)، (نبود) بود گزینه بهینه‌سازی^۶، نقاط آغاز تصادفی و شمار تلاش‌ها^۷ و از سوی دیگر به دلیل ارزیابی درستی و اطمینان از نتایج به‌دست‌آمده، شبکه‌های بیزی برازش داده شده در ۱۲ حالت مختلف بررسی شدند. حالت‌های مختلف برازش شبکه تنها روی داده‌های بافت ماهیچه کالبدی بررسی و به علت حجم بالای نتایج از گنجانیدن نتایج دیگر بافت‌ها خودداری شد. یکی از معیارهای انتخاب حالت بهینه، استفاده از تابع‌های امتیازدهی است. به‌طوری‌که هر تابع امتیازدهی که کمترین امتیاز را در مدل‌سازی یا بازسازی ساختار شبکه داشته باشد، به‌عنوان بهترین حالت مدل‌سازی شناخته می‌شود. همچنین در انتخاب حالت بهینه شاخص‌های دیگری همچون زمان اجرای الگوریتم نیز مدنظر بود. در شکل ۱ حالت‌های مختلف الگوریتم تپه‌نوردی اجرا شده روی داده‌های بیان ژنی بافت ماهیچه کالبدی معرفی شده است. پس از انتخاب بهترین حالت از بین ۱۲ حالت موجود برای اجرای الگوریتم روی

8. Bootstrap
9. Arc strength
10. Arc direction
11. Adjacency matrix
12. In degree
13. Out degree

1. Differentially Expressed Genes or DE's
2. Bayesian Information Criterion or BIC
3. Akaike Information Criterion or AIC
4. Bayesian Gaussian Equivalent or BGE
5. Logarithm Likelihood or LogL
6. Optimization
7. Attempt



شکل ۱- بررسی حالت‌ها مختلف الگوریتم تپه نوردی اجرا شده بر روی داده‌های بیان ژنی بافت ماهیچه کالبدی
Figure 1- Examining different scenarios of hill climbing algorithm implemented on skeletal muscle tissue expression data



شکل ۲- بررسی حالت‌ها مختلف بوت‌استرپ الگوریتم بر روی داده‌های بیان ژنی بافت ماهیچه کالبدی
Figure 2. Examining different parameters in Bootstrap algorithm in skeletal muscle tissue expression data

نسبت به دو معیار امتیازی دیگر در همه حالت‌ها به خود اختصاص داد. نتایج به دست آمده نشان داد، استفاده از گزینه بهینه‌سازی به دلیل کاهش شمار آزمون‌های مورد استفاده تنها سبب کاهش زمان اجرای الگوریتم شده و هیچ‌گونه تأثیری بر فراسنجه‌های شبکه برازش داده شده ندارد. لازم به یادآوری است استفاده از گزینه بهینه‌سازی بیشتر برای دسته الگوریتم‌های محدودیت‌گرا کاربرد دارد. از سوی دیگر با توجه به رابطه‌های علی- معلولی (تأثیرپذیری و تأثیرگذاری) بین ژن‌ها نمی‌توان شمار ارتباط‌های بین ژن‌ها، به‌ویژه در مورد ژن‌های هاب با بیشترین درجه تأثیرگذاری و تأثیرپذیری را به صورت محدود در نظر داشت. بدین منظور سعی شد که به ازای هر ژن بیشترین شمار والد در نظر گرفته شود. از سوی

نتایج و بحث

ارزیابی و انتخاب بهینه‌ترین حالت مدل‌سازی شبکه
نتایج به دست آمده از اجرای ۱۲ حالت مختلف الگوریتم تپه‌نوردی در جدول ۲ ارائه شده است. در واقع حالت‌های ۱ تا ۶ نمایانگر استفاده از الگوریتم تپه‌نوردی و حالت‌های ۷ تا ۱۲ بیانگر کاربرد الگوریتم تپه‌نوردی با نقطه آغاز تصادفی است. تفاوت بین این دو حالت از الگوریتم، در معلوم بودن شمار تکرارها، شمار نقاط آغاز تصادفی، شمار تلاش‌ها برای تعیین جهت اتصال‌ها و شمار والدین به ازای هر ژن است.

با توجه به انتخاب مدل با کمترین میزان امتیاز برای مدل‌سازی و از سوی دیگر با مقایسه مقدار برآوردشده برای توابع امتیازدهی، معیار BIC کمترین مقادیر را

یال‌های استخراج‌شده در شبکه‌های برازش داده شده درصد‌های متفاوتی از توان و جهت یال داشتند. از سوی دیگر میانگین پوشش مارکفی و میانگین اندازه‌ی همسایگی نیز با افزایش توان یال کاهش یافت. چنین حالتی نیز با افزایش بوت استرپ‌ها مشاهده شد. با توجه به این شواهد و همچنین مدت‌زمان اجرای الگوریتم، برای همه‌ی داده‌ها مورد بررسی، شمار بوت استرپ‌ها ۵۰ و بیشترین مقادیر توان یال ۸۵ درصد در نظر گرفته شد. بنابراین از یک‌سو با افزایش توان و جهت یال‌ها و از سوی دیگر با افزایش شمار تکرار بوت‌استرپ‌ها، بسیاری از اتصال‌های ضعیف بین ژن‌ها برداشته‌شده و اطمینان به‌درستی نتایج و شبکه‌ی برازش یافته بیشتر می‌شود. کاهش میزان فراسنجه‌های پوشش مارکفی و اندازه‌ی همسایگی بر نتیجه‌ی افزایش توان و برای اتصال‌ها در آستانه‌ی بالاتر همگی نیز مؤید همین مطلب بودند.

دیگر، با توجه به نظریه‌ی الگوریتم تپه‌نوردی و همچنین امکان و احتمال زیاد شناسایی نقاط بهینه‌ی محلی به‌جای نقاط بهینه‌ی اصلی در زمان بهینه‌سازی تابع هدف و نیز با توجه به حالت‌های جهان واقعی، حالتی که بیشترین شمار تکرار برای آغاز دوباره‌ی روند بهینه‌سازی و همچنین بیشترین شمار تلاش برای وجود یا تعیین جهت بهینه‌ی اتصال‌های بین ژن‌ها را داشته باشد، مطلوب است. درنهایت با توجه به نتایج به‌دست‌آمده جدول ۲ و لحاظ همه‌ی نکات یادشده، حالت ۹ با معیار امتیاز BIC، بیشترین شمار والد به ازای هر ژن، بیشترین شمار تکرار و آغاز دوباره‌ی تصادفی و بیشترین شمار تلاش برای کاهش/حذف یا معکوس کردن برای یال‌ها در هر بار آغاز تصادفی و در نظر گرفتن حالت بهینه استفاده شد. هم‌چنین نتایج به‌دست آمده از جدول ۳ نشان داد که با افزایش تکرار بوت‌استرپ، شمار یال‌های شبکه‌ی بی‌زی استخراجی، کاهش یافتند. اگرچه بسیاری از

جدول ۱. آزمایش‌های ریزآرایه‌ی DNA بافت‌های مختلف گونه‌ی گاو مورد استفاده در این پژوهش

Table 1. DNA Microarray data from various bovine tissues used in this research

Experiment	Platform	Tissue	Title	Array
GSE23348	GPL2112	Ovary	Prostaglandin F2 α -induced changes in transcriptome of bovine mid cycle versus early corpus luteum	30
GSE33030	GPL2112	Uterus	Effects of pregnancy and progesterone supplementation on endometrial gene expression in cattle	79
GSE75347	GPL2112	Muscle	Comparative analysis of Musculus longissimus dorsi expression of Holstein x Charolais F2 cattle (SEGFAM) with high and low intramuscular fat (IMF) content	20
GSE25319	GPL2112	Liver	Gene transcription profiles showing early systemic response to E. coli (k2bh2) infection in bovine liver	54
GSE24560	GPL2112	Mammary	Comparative Expression Profiling of E. coli and S. aureus inoculated primary Mammary Gland Cells sampled from Cows with different genetic Predisposition for Somatic Cell Score	88

جدول ۲. حالت‌های مختلف الگوریتم تپه‌نوردی اجراشده روی داده‌های بیان ژنی بافت ماهیچه‌ی کالبدی

Table 2. Different sets of hill climbing algorithms tested on skeletal muscle tissue

Model	N	E	MB	NH	BF	Score Values			T	
						BGE	AIC	BIC		
1	121	1497	81.79	24.74	12.37	-3742	15046	1420	16664	190980
2	121	1497	81.79	24.74	12.37	-3742	15046	1420	16664	22186560
3	121	1801	84.18	29.77	14.88	-3876	18388	17431	20310	26615160
4	121	1801	84.18	29.77	14.88	-3879	18388	17431	20310	227700
5	121	2007	84.18	33.17	16.59	-3973	18805	17746	20933	250260
6	121	2007	84.18	33.17	16.59	-3973	18805	17746	20933	29286840
7	121	1801	84.18	29.77	14.88	-3876	18388	17431	20310	26615160
8	121	1497	81.79	24.74	12.37	-3742	15046	1420	16664	22186560
9	121	1497	81.79	24.74	12.37	-3742	15046	1420	16664	190980
10	121	1497	81.79	24.74	12.37	-3742	15046	1420	16664	190980
11	121	1497	81.79	24.74	12.37	-3742	15046	1420	16664	191220
12	121	1801	81.79	29.77	14.88	-3876	18456	17499	20378	227700

N: شمار گره‌ها، E: شمار یال‌ها، MB: پوشش مارکفی، NH: اندازه همسایگی، BF: عامل شاخه‌ساز، OP: بهینه‌سازی، SC: امتیاز، T: تعداد تست‌ها

N: Number of node, E: Number of Edge, MB: Markov Blanket, NH: Neighborhood size, BF: Branching factor, OP: optimization, SC: score, T: number of test used.

1: Optimized=True, Score=BIC, 2: Optimized=False, Score=BIC, 3: Optimized=False, Score=AIC, 4: Optimized=True, Score=AIC, 5: Optimized=True, Score=Log Likelihood, 6: Optimized=False, Score=Log Likelihood, 7: Optimized=False, Score=AIC, restart=0, perturb=1, 8: Optimized=False, Score=BIC, restart=0, perturb=1, 9: Optimized=True, Score=BIC, restart= Inf, perturb= Inf, 10: Optimized=True, Score=BIC, restart=0, perturb=Inf, 11: Optimized=True, Score=BIC, restart=1, perturb=1, 12: Optimized=True, Score=AIC, restart=1, perturb=1.

جدول ۳. ساختار شبکه بیزی برازش داده شده برای بافت ماهیچه کالبدی در ترکیب‌های مختلف الگوریتم تپه‌نوردی

Table 3. Fitted structural Bayesian network for skeletal muscle in different combinations of hill climbing algorithm

		No. of Arcs					Average Markov Blanket					Average of Neighborhood Size				
		Bootstrap Replication					Bootstrap Replication					Bootstrap Replication				
A	T	10	20	50	80	100	10	20	50	80	100	10	20	50	80	100
65	65	70	43	46	35	41	1.2	0.74	0.78	0.58	0.68	1.1	0.71	0.76	0.58	0.68
65	85	30	20	16	16	15	0.5	0.33	0.26	0.26	0.25	0.5	0.33	0.26	0.26	0.25
85	65	63	42	37	38	38	1.1	0.71	0.63	0.64	0.63	1	0.69	0.61	0.63	0.63
85	85	18	15	16	13	15	0.3	0.25	0.26	0.21	0.25	0.3	0.25	0.26	0.21	0.25

A: Arc direction and strength, T: Threshold

A: قدرت و جهت یال‌ها، T: آستانه

جدول ۴. برآورد فراسنجه‌های شبکه بیزی در بافت‌ها و در حالت‌های متفاوت (کاوشگر محور درون پراتنز)

Table 4. Gene-based (and probe-based) estimation of bayesian network parameters in different tissues and conditions

T*	M*	N*	E*	MB*	NH*	BF*
Muscle	1	109(121)	1194(1423)	75.6(82.4)	21.9(23.6)	10.9(11.8)
	2	109(121)	12(16)	0.22(0.26)	0.22(0.2)	0.11(0.1)
	3	23(36)	13(19)	1.13(1.06)	1.13(1.06)	0.57(0.54)
Mammary	1	98(121)	736(903)	47(50)	15(14)	7.51(7.46)
	2	98(121)	64(71)	1.6(1.4)	1.3(1.1)	0.65(0.59)
	3	65(90)	60(71)	2.3(1.8)	1.8(1.5)	0.9(0.7)
Liver	1	102(121)	672(887)	61.7(64.8)	13.1(14.6)	6.5(7.3)
	2	102(121)	39(52)	0.84(0.98)	0.76(0.86)	0.38(0.43)
	3	102(79)	672(54)	61.7(1.52)	13.1(1.3)	6.59(0.6)
Uterus	1	89(121)	388(577)	25.3(32.4)	8.72(9.54)	4.36(4.7)
	2	89(121)	40(50)	1.01(0.86)	1.90(0.83)	0.45(0.41)
	3	53(80)	39(40)	1.66(1.27)	1.47(1.2)	0.74(0.6)
Ovary	1	100(121)	952(661)	81.5(89)	19.4(17.6)	9.5(8.81)
	2	100(121)	16(26)	0.32(0.45)	0.32(0.43)	0.16(0.21)
	3	26(36)	14(21)	1.08(1.11)	1.08(1.11)	0.54(0.55)

* T: بافت، M: مدل، N: شمار گره‌ها، E: شمار یال‌ها، MB: پوشش مارکفی، NH: اندازه همسایگی، BF: عامل شاخه‌ساز

مدل ۱: الگوریتم تپه‌نوردی، مدل ۲: الگوریتم تپه‌نوردی + بوت استرپ، مدل ۳: الگوریتم تپه‌نوردی + بوت استرپ + قدرت یال + آستانه = ۸۵٪

* T: Tissue, M: Model, N= Number of Node, E= Number of Edge, MB= Average of Markov Blanket Size, NH=Average of Neighborhood Size, BF= Average of Branching Factor

Model1: Hill Climbing Algorithm, 2: Hill Climbing+ Bootstrap, 3: Hill Climbing+ Bootstrap+ Arc Strength+ Threshold= 85% .

شده باشد.

در حالت استفاده از الگوریتم تپه‌نوردی (حالت ۱) بیشترین شمار ژن (۱۰۹) در بافت ماهیچه کالبدی مشاهده شد، ولی با در نظر گرفتن فراسنجه‌های محدودکننده در حالت ۳، بیشترین شمار ژن (۱۰۲) در بافت کبد مشاهده شد. احتمال دارد این نتایج نمایانگر تفاوت فضای ساختارشناختی رونویسی در بافت‌های مورد نظر باشد. میزان برآورد شده برای فراسنجه پوشش مارکفی، همسایگی و شاخه‌ساز برای همه بافت‌ها در حالت ۱ به‌طور معنی‌دار و غیرمنتظره- ای بالاتر از دیگر حالت‌ها بودند. مقادیر برآورد شده برای فراسنجه پوشش مارکفی در حالت‌های ۲ و ۳ نزدیک به هم بودند، با این حال برای حالت ۲ مقدار کمتری از این فراسنجه مشاهده شد که این احتمال دارد به دلیل استفاده از عامل‌های محدودکننده برای توان یال‌ها و حذف آن‌ها باشد. در حالت ۱ بیشترین میزان میانگین پوشش مارکفی بعد از بافت کبد

مدل‌سازی ساختار شبکه بیزی در بافت‌های مورد

بررسی

مقدار برآورد فراسنجه‌های شبکه بیزی ژن محور و کاوشگر محور برای ۵ بافت مورد بررسی در جدول ۴ نشان داده شده است. اعداد درون پراتنز معرف میزان برآورد فراسنجه‌ها در حالت شبکه بیزی کاوشگر محور هستند. همان‌طور که در جدول ۴ قابل مشاهده است در حالت ۱ شمار گره‌ها و یال‌ها به میزان قابل‌توجهی بالاتر از دیگر حالت‌ها است که سبب شده میزان فراسنجه‌های پوشش مارکفی، شاخه و همسایگی نسبت به دو حالت دیگر بیشتر برآورد شود. همچنین با توجه به جدول ۴ و در حالت‌های ۱ و ۲ شمار گره‌ها در همه بافت‌ها با هم برابر بود، درحالی‌که شمار یال‌ها برای حالت‌ها دو و سه تا حدودی برابر بودند. این پدیده احتمال دارد به دلیل استفاده از حالت‌های محدودکننده در حالت ۳ (میزان توان یال و آستانه) و بوت‌استرپ برای بازسازی شبکه در حالت ۲ ایجاد

برآورد فراسنجه‌های ساختارشناختی در بافت‌های

مورد بررسی

شماری از فراسنجه‌های ساختارشناختی مهم به‌دست‌آمده از اجرای شبکه‌ی بیزی در داده‌های بیان ژن بافت‌های مورد بررسی به‌صورت جداگانه در جدول ۵ نشان داده شده است. از این فراسنجه‌ها می‌توان برای شناسایی ژن‌ها در بافت‌های مورد بررسی استفاده کرد. شماری از این فراسنجه‌ها با توجه به حالت‌های تنظیمی و جهت‌دار درون و بیرون شونده از آن‌ها مانند درجه، نزدیکی، خارج از مرکز و توان، حالت‌های مختلفی دارند. شماری دیگر از فراسنجه‌های ساختارشناختی همچون ضریب خوشه‌بندی^۱ و فاصله^۲ به دلیل ناچیز بودن از محاسبه حذف شدند. ناچیز بودن ضریب خوشه‌بندی این حقیقت را آشکار کرد که در کل امکان ایجاد دسته‌های ژنی^۳ شبکه‌های مختص هر بافت کم است. به‌طور عمده تغییر دامنه مقدار این فراسنجه بین ۰ و ۱ است. هراندازه مقدار این فراسنجه به سمت عدد ۱ میل کند، نشان‌دهنده ایجاد ماژول در شبکه است. چون در این پژوهش میزان ضریب خوشه‌بندی به‌دست‌آمده ۰ بود، بنابراین بر نبود زمینه تشکیل ماژول دلالت دارد.

بیشترین (۰/۰۰۰۲) و کمترین (۰/۰۰۰۰۶) مقدار فراسنجه بینابینی ژن محور به‌ترتیب برای بافت ماهیچه کالبدی و تخمدان مشاهده شد. فراسنجه بینابینی ژن محور به‌دست‌آمده از این پژوهش معیاری از میزان اتصال‌های بین گره‌های درون شبکه است (Csardi & Nepusz, 2006)، اگرچه در بافت‌های ماهیچه و رحم بیشترین (۰/۰۳) و کمترین (۰/۰۱) مقدار خود را داشتند.

فراسنجه نزدیکی^۴ که به‌صورت تقاطع میانگین کوتاه‌ترین راه‌هایی که به یک گره ختم می‌شود یا شمار مرحله‌های مورد نیاز برای دستیابی به دیگر گره‌ها از یک گره خاص تعریف می‌شود، دو مؤلفه نزدیکی از بیرون و درون دارد (Csardi & Nepusz, 2006). این فراسنجه در بافت تخمدان و کبد بیشترین

(۶۱)، برای بافت تخمدان (۸۱) مشاهده شد درحالی‌که با انجام بوت‌استرپ یکی از کمترین مقادیر پوشش مارکفی مربوط به بافت ماهیچه کالبدی (۰/۲۲) در حالت (۱) و تخمدان (۱/۰۸) در حالت (۳) بود. بیشترین میزان میانگین معیار پوشش مارکفی در حالت سوم مربوط به بافت غدد پستانی (۲/۳۱) است. در حالت ۱ بیشترین (۲۱/۹۱) و کمترین (۸/۷۲) مقدار همسایگی به ترتیب مربوط به بافت‌های ماهیچه کالبدی و رحم بودند. در حالت ۲ و ۳ نیز کمترین مقدار فراسنجه همسایگی مربوط به بافت‌های ماهیچه کالبدی (۰/۲۲) و بافت تخمدان (۱/۰۸) مشاهده شد. بیشترین میزان این معیار نیز به ترتیب برای حالت‌های ۲ و ۳ برای بافت‌های رحم (۱/۹۰) و پستان (۱/۸۵) برآورد شد. برای فراسنجه شاخه‌ساز نیز در حالت ۱ مقادیر فراتر از حد انتظار و میزان این فراسنجه در حالت ۲ کمتر از میزان برآورد شده در مقایسه با حالت ۳ در همه بافت‌ها بود. بیشترین مقدار فراسنجه شاخه‌ساز در حالت‌های ۱، ۲ و ۳ به ترتیب برای بافت‌های ماهیچه کالبدی (۱۰/۹۵)، غده پستانی (۰/۶۵) و غده پستانی (۰/۹۲) مشاهده شد. حال آنکه کمترین مقدار از این فراسنجه برای بافت‌های رحم (۴/۳۶)، بافت ماهیچه کالبدی (۰/۱۱) و بافت تخمدان (۰/۵۴) مشاهده شد. با توجه به ماهیت ترکیب فراسنجه‌های مختلف الگوریتم تپه‌نوردی (خود الگوریتم، بوت‌استرپ و بوت‌استرپ با محدودیت در توان یال و آستانه برای استخراج میانگین شبکه‌های بوت‌استرپی) ژن‌های متفاوتی شناسایی شدند. در این بررسی از حالت ۳ (بوت‌استرپ به همراه محدودیت در توان یال و آستانه شبکه‌های نمونه‌گیری شده به ۰/۸۵) پس از بررسی نتایج و به‌عنوان کامل‌ترین و مطمئن‌ترین حالت در استخراج ساختار شبکه بیزی انتخاب شد. این نتایج نشان داد، ساختار ساختارشناختی شبکه به‌دست‌آمده در بافت‌های مختلف می‌تواند توسط ترکیب‌های حالت‌های مختلف فراسنجه‌های الگوریتم تپه‌نوردی تحت تأثیر واقع شود.

1. Cluster coefficient
2. Distance
3. Modularity
4. Closeness

همان ژن‌های با بیشترین درجه تفاوت بیانی نبودند. در نتیجه می‌بایست در هنگام تجزیه و تحلیل نتایج به این نکته مهم توجه داشت. در بهترین حالت باید با در نظر گرفتن این مورد، برای شناسایی ژن‌های متفاوت بیان شده و ژن‌های هاب، طراحی آزمایش ریزآرایه DNA برای هر ۵ بافت مختلف، انجام گیرد. اگرچه با توجه به این موارد، انتظار می‌رود در تحلیل نتایج و تعمیم ژن‌های هاب شناسایی شده به موارد کلی‌تر باید کمی با احتیاط عمل کرد.

برای ژن LOC78882 در بافت ماهیچه کالبدی، COL5A2 در بافت پستانی، MBOAT2 در بافت کبد و LOC286871 در بافت تخمدان هیچ‌گونه اطلاعاتی در ژنگان گاو به دست نیامد. بیشترین شمار ژن‌های هاب برای بافت کبد شناسایی شدند. با توجه به نقش اصلی کبد در سوخت‌وسازی بدن و به‌عنوان پایه‌ای برای تبدیل فرایندهای سوخت‌وسازی قابل توجه است. سپس بافت پستانی با چهار ژن هاب شناسایی شده که این ژن‌ها در ساختار پروتئین‌های غشایی و نقل و انتقال‌های یونی نقش مهمی را بر عهده دارند.

در بافت رحم نیز ژن‌های شناسایی شده نقش مهمی در ساخت (سنتر) چربی، تبدیل انرژی، رشد و مقاومت ویروسی داشتند. ژن‌های شناسایی شده در بافت تخمدان نقش مهمی در رشد بدن، تولید مانند و تولید شیر بر عهده دارند. نتایج به‌دست‌آمده از این پژوهش نشان داد، در بافت ماهیچه کالبدی ژن‌ها در تبدیل انرژی در چرخه‌های بیوشیمیایی نقش فعالی دارند.

در جدول ۷ ژن‌های منطبق بر کاوشگرهای هاب استخراج شده و به تفکیک بافت‌ها ارائه شده است. با توجه به جدول ۷، زیاد بودن شمار کاوشگرها به نسبت ژن‌ها در هر بافت در فناوری افی‌متریکس قابل‌درک است. با توجه به فناوری خاص تولید داده‌های ریزآرایه DNA از مجموعه کاوشگرهای تراشه‌های افی‌متریکس، مشکل در نگاشت این کاوشگرها به برخی ژن‌ها وجود دارند از جمله بعضی کاوشگرها به یک هدف خاص دورگ (هیبرید) نمی‌شوند و یا به نواحی چندگانه دورگ می‌شوند (Liu et al., 2016). افزون بر این، در حاشیه‌نویسی‌های اصلی افی‌متریکس اغلب بیان یک ژن توسط چندین مجموعه کاوشگر با مقادیر

(۰/۳) و کمترین (۰/۰۱) مقدار را داشتند. فراسنجه بیرون از مرکز نیز که به‌صورت بیشترین طول نامحدود و کوتاه‌ترین مسیر بین یک گره و دیگر گره‌ها در شبکه تعریف می‌شود (Csardi & Nepusz, 2006) در بافت رحم بیشترین (۱/۹۵) و در بافت ماهیچه کمترین مقدار خود را (۰/۹۸) داشت. فراسنجه توان^۱ که به‌صورت مجموع وزن‌های یال‌های کناری برای هر گره تعریف می‌شود (Csardi & Nepusz, 2006) بیشترین مقدار خود در بافت ماهیچه (۱/۶) و کمترین آن در بافت تخمدان (۱/۲) را داشت.

فراسنجه‌های شبکه بیزی کاوشگر محور شبکه‌ای و ساختارشناختی برای بافت‌های مورد بررسی در جدول ۵ ارائه شده است. مقادیر کاوشگر محور برای فراسنجه‌های ضریب خوشه، ضریب خوشه وزن‌دار، فاصله به درون و فاصله به بیرون نیز همانند شبکه ژن محور بسیار ناچیز و نزدیک به ۰ بود. بیشترین مقدار کاوشگر محور برای فراسنجه بینابینی بافت‌های ماهیچه و کبد (۰/۰۰۰۳) و کمترین آن (۰/۰۰۰۰۶) مشاهده شد. بافت تخمدان و پستانی بیشترین (۰/۰۳۶) و کمترین (۰/۲۴) میزان اتصال‌ها را داشتند، که این بیانگر اهمیت به‌نسبت زیاد بین ژن‌های موجود در بافت تخمدان نسبت به دیگر بافت‌ها است. قابل توجه است که میزان یال‌های وارد شونده و خارج شونده با هم مساوی بودند. بافت تخمدان همچنین بیشترین مقدار از فراسنجه نزدیکی (۰/۰۳) و کمترین مقدار از فراسنجه توان (۱/۱) را داشت. کمترین مقدار فراسنجه نزدیکی (۰/۱۸۷) در بافت کبد مشاهده شد. بافت‌های رحم و ماهیچه کالبدی به ترتیب بیشترین (۱/۹) و کمترین (۰/۹۸) مقدار از فراسنجه خارج از مرکز را داشتند. بیشترین مقدار (۱/۶) از فراسنجه توان برای بافت ماهیچه محاسبه شد.

شناسایی ژن‌های هاب در بافت‌های متفاوت

محل قرارگیری ژن‌های هاب شناسایی شده و کاوشگرهای معادل آن‌ها روی کروموزوم‌ها در بافت‌های مورد بررسی در جدول ۶ ارائه شده است. نتایج به‌دست‌آمده نشان دادند، ژن‌های هاب در اصل

موارد یادشده، نتایج به دست آمده از جدول نامبرده نمایانگر حالت ها متفاوتی همچون نگاشت درست کاوشگر به ژن، عدم نگاشت برای برخی کاوشگرها، نگاشت کاوشگر همسان بر ژن مورد نظر در دو بافت متفاوت و نگاشت دو کاوشگر روی یک ژن بود.

بسیار متفاوت اندازه گیری می شود، اما نتایج آن بسیار کم در سطح ژن خلاصه می شود (Milchevskaya *et al.*, 2017). این واقعه می تواند خطاهایی را در تحلیل های رونویسی متکی به ریزآرایه DNA ایجاد کند، اگرچه این مشکل در فناوری RNA-Seq وجود ندارد. با توجه به

جدول ۵- فراسنجه های شبکه ژن محور (و کاوشگر محور) برای ۵ بافت مورد بررسی
Table 5- Gene-based (and probe-based) topological parameters of 5 studied tissues

Parameters	Tissue				
	Skeletal muscle	Mammary	Liver	Uterus	Ovary
Betweenness	0.0002(0.00002)	0.00008(0.00008)	0.00006(0.00003)	0.0001(0.00004)	0.00006(0.00001)
Edge betweenness	2.773(1.111)	1.675(2.467)	1.538(1.173)	1.935(1.745)	1.222(1.08)
In degree	0.013(0.015)	0.012(0.008)	0.012(0.014)	0.015(0.008)	0.018(0.013)
Out degree	0.013(0.015)	0.012(0.008)	0.012(0.014)	0.015(0.008)	0.018(0.013)
Degree	0.027(0.030)	0.024(0.017)	0.025(0.028)	0.031(0.016)	0.036(0.026)
In closeness	0.016(0.029)	0.018(0.011)	0.018(0.024)	0.027(0.012)	0.031(0.023)
Out closeness	0.016(0.029)	0.018(0.011)	0.181(0.024)	0.022(0.012)	0.031(0.023)
Closeness	0.018(0.029)	0.019(0.012)	0.018(0.025)	0.023(0.013)	0.032(0.023)
In eccentricity	1.095(0.542)	0.745(1.032)	0.75(0.609)	0.911(0.797)	0.593(0.568)
Out eccentricity	0.984(0.542)	0.781(0.956)	0.732(0.536)	0.755(0.784)	0.593(0.5)
Eccentricity	0.984(1.057)	1.709(4.714)	1.732(1.292)	1.955(1.708)	1.312(1.272)
Alpha centrality	4.698(1.542)	1.890(2.252)	1.892(1.609)	2(1.873)	1.625(1.590)
In strength	2.460(0.514)	0.672(0.780)	0.696(0.560)	0.688(0.645)	0.562(0.568)
Out strength	0.841(0.514)	0.672(0.780)	0.696(0.560)	0.688(0.645)	0.562(0.568)
Strength	1.682(1.028)	1.345(1.560)	1.392(1.121)	1.377(1.29)	1.125(1.136)
Sub graph centrality	1(1)	1(1)	1(1)	1(1)	1(1)

جدول ۶- کارکرد ژن های شناسایی شده به تفکیک بافت های مورد بررسی

Table 6- Characteristics of discovered hub genes for 5 studied tissues-gene based

	PubMed ID	Location	Length (bp)	Probe
Muscle	CBR1: Carbonyl Reductase 1	chr1:150,054,221-150,064,637	10417	Bt.5304.1.S1_at
	NID2 Nidogen-2 Precursor	chr10:44894656-44986659	92004	Bt.13542.1.S1_at, Bt.278.1.S1_at, Bt.8124.1.S2_at, Bt.9360.1.S1_at
mammary	LOC616942 MHC class I A-like precursor	chr23:27842253-27846497	4245	Bt.3213.1.A1_at, Bt.5395.1.S1_a_at, Bt.16087.1.S1_at, Bt.16628.2.A1_at
	COL5A2 Collagen Alpha-2(V) Chain Preproprotein	chr2:189031915-189179879	147965	Bt.27856.1.S1_at, Bt.29821.1.S1_at, Bt.357.1.S1_at, Bt.4670.1.S1_at
	FXYP Domain-Containing Ion Transport Regulator 3 Precursor	chr18:46033469-46039482	6014	Bt.24211.2.S1_at, Bt.4209.2.S1_at, Bt.5324.1.S1_s_a, Bt.8804.1.S1_at
	MBOAT2	chr11:88418175-88450929	32755	Bt.16592.1.A1_at, Bt.17110.1.A1_at, Bt.17516.1.A1_at, Bt.9549.1.S1_at
Liver	ANKRD1 Ankyrin Repeat Domain-Containing Protein	chr26:12571209-12580580	9372	Bt.4594.1.S1_at, Bt.5324.1.S1_s_at, Bt.6376.1.A1_at, Bt.12961.1.A1_at
	CLDN2 Claudin-2	chrX:60134285-60136956	2672	Bt.13106.1.A1_at, Bt.16250.2.S1_at, Bt.28443.1.S1_at, Bt.3805.1.S1_at
	MGC127133 beta-Hydroxysteroid Dehydrogenase-Like	chr1:150108438-150110508	2071	Bt.13855.2.S1_at, Bt.468.1.S1_at, Bt.13796.1.S1_at, Bt.11702.1.A1_at, Bt.17729.1.A1_at,
Uterus	IGFBP: Insulin-like Growth Factor-Binding Protein 1 Precursor	chr4:76720425-76725309	4885	Bt.190.1.A1_at, Bt.21934.1.S1_a_at, Bt.26011.1.A1_at, Bt.26921.1.A1_a, Bt.4594.1.S1_at, Bt.569.1.S1_at, Bt.8027.1.S1_at, Bt.25270.1.A1_a_at
	Diacylglycerol O-Acyltransferase 2	chr15:55940757-55973229	32473	
	CKMT1 Creatine Kinase U-Type Mitochondrial Precursor	chr21:55877337-55882400	5064	
	ISG15 exon2 Ubiquitin-Like Protein ISG15	chr16:52714627-52715654	1028	
Ovary	INHBA Inhibin Beta A Chain Precursor	chr4:79986254-79997756	11503	Bt.15511.1.A1_at, Bt.18639.1.A1_at, Bt.4751.1.S1_a_at

جدول ۷- نگاشت فهرست ژن‌های منطبق بر کاوشگرهای هاب مستخرج به تفکیک بافت‌ها

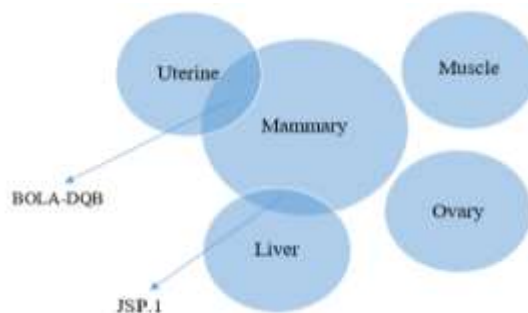
Table 7- Genes Match to Hub Probes found for Each Tissue Separately

Tissue	Probe set	Gene
Muscle	Bt.5304.1.S1_at	GSTM3
Mammary	Bt.13542.1.S1_at, Bt.278.1.S1_at , Bt.8124.1.S2_at, Bt.9360.1.S1_at, Bt.3213.1.A1_at, Bt.5395.1.S1_a_at, Bt.16087.1.S1_at, Bt.16628.2.A1_at, Bt.24211.2.S1_at, Bt.4209.2.S1_at, Bt.5324.1.S1_s_at , Bt.8804.1.S1_at, Bt.27856.1.S1_at, Bt.29821.1.S1_at, Bt.357.1.S1_at, Bt.4670.1.S1_at, t.12304.1.S1_at, Bt.15788.1.S1_a_at, Bt.3805.1.S1_at , Bt.9966.1.S1_at	ADAM12, ASPN, C3, CFB, CHRDL1, COL1A2, COL5A2, ISG15, JSP.1, JSP.1 , LOC100128178, LOC507402, LOC616942, NELL2, NID2, S100A12, S100A8, TGFB1, VCAN
Liver	Bt.16592.1.A1_at, Bt.17110.1.A1_at, Bt.17516.1.A1_at, Bt.9549.1.S1_at , Bt.4594.1.S1_at , Bt.5324.1.S1_s_at , Bt.6376.1.A1_at, Bt.12961.1.A1_at, Bt.13106.1.A1_at, Bt.16250.2.S1_at, Bt.28443.1.S1_at, Bt.3805.1.S1_at	BOLA-DQB , FMO5, HAO1, JSP.1, JSP.1 , MCF2, MGC13057, MICAL2, PRLR, SLC10A1, SLC01B3
Uterus	Bt.13855.2.S1_at, Bt.468.1.S1_at, Bt.13796.1.S1_at, Bt.11702.1.A1_at, Bt.17729.1.A1_at , Bt.190.1.A1_at, Bt.21934.1.S1_a_at , Bt.25270.1.A1_a_at , Bt.26011.1.A1_at , Bt.26921.1.A1_at, Bt.4594.1.S1_at , Bt.569.1.S1_at , Bt.8027.1.S1_at	atADAMDEC1, ANPEP, BOLA-DQB , CFTR, IGFBP1, LOC783399, MAP2, MAP2 , MMD2, SLC27A6
Ovary	Bt.15511.1.A1_at, Bt.18639.1.A1_at, Bt.4751.1.S1_a_at	FABP3, FBXO32, HLA-DQA1

حروف قرمز معرف عدم نگاشت کاوشگر، حروف سبز و آبی معرف نگاشت یک ژن در چندین بافت و حروف بنفش معرف نگاشت دو کاوشگر به یک ژن می‌باشد. The red words indicate the non-mapping genes for the probe. The green and blue words indicate the mapping of an identical probe to the target gene in the two specified tissue, and the purple words represent the mapping of two probes to a gene.

است. شمار ژن‌های هاب استخراجی در بافت‌های مختلف متفاوت بود. همچنین شمار کاوشگرهای استخراجی بافت‌های مختلف نیز یکسان نبودند. با در نظر گرفتن شرایط آزمایشی، بافت‌های رحم و کبد در این پژوهش از جمله بافت‌هایی بودند که بیشترین شمار ژن هاب را داشتند. اما نبود ژن‌های همسان هاب در بافت‌های مختلف از نتایج فراوان این پژوهش بود. شمار زیادی کاوشگرهای هاب در هر بافت به جز ماهیچه شناسایی شدند. کاوشگرهای استخراجی در بافت‌های مختلف همسان نبودند. بنابراین احتمال دارد این واقعه نمایانگر وجود تفاوت در فرآیند تنظیم بیانی در بافت‌های مختلف باشد. با این حال می‌بایست به شرایط متفاوت آزمایشی و تیمارهای مختلف داده‌های بیان ژنی هر آزمایش برای هر بافت نیز توجه خاصی داشت و در تعمیم نتایج با احتیاط بیشتری رفتار کرد. یکی از علل استخراج شبکه بیزی کاوشگر محور به همراه شبکه ژن محور، استخراج هاب کاوشگرهایی بود که با شماره شناسه آن‌ها بتوان به استخراج ویژگی‌های یک ژن اقدام کرد. در صورتی که کاوشگرهایی استخراج شوند که در بافتی نقش هاب را داشته باشند، به احتمال بیانگر این واقعیت است که شاید حامل یک نشانگر عملکردی یا یک SNP باشند. لذا به جای انتخاب ژن، می‌توان به انتخاب کاوشگر همت گماشت.

در جدول ۸ ژن‌های هاب شناسایی شده از نگاشت کاوشگرها به همراه اطلاعات کلی در مورد نام ژنی آن‌ها در حالت‌های مختلف (Entrez ID, PubMed) ارائه شده است. همان‌طور که مشاهده می‌شود، شمار ژن‌ها به ازای هر بافت در روش نگاشت ژنی به نسبت روش ژن محور (جدول ۶) خیلی بالاتر بود و همه ژن‌های متفاوت از روش ژن محور به‌استثنای بافت پستانی (LOC616942, NID2 و COL5A2) و بافت رحم (IGFBP1) تشخیص داده شدند. این نتیجه معرف یکسان بودن و سازگاری جزئی کاوشگرها با نتایج ژن محور است. در بافت‌های پستان و کبد دو کاوشگر Bt.3805.1.S1_at و Bt.5324.1.S1_s_at پس از نگاشت به ژن‌های یکسان (JSP.1) منتهی شده و نتایج نشان داد، ژن یادشده در هر دو بافت حضور دارد. همچنین ژن BOLA-DQB پس از نگاشت در بافت‌های کبد و رحم ظاهر شد. این واقعه نشان‌دهنده بیان ژن در هر دو بافت است. همچنین در بافت رحم دو کاوشگر (Bt.569.1.S1_at و Bt.21934.1.S1_a_at) به یک ژن (MAP2) منتهی شدند. ژن ISG15 در هر دو بافت رحم (داده ژن محور) و پستانی (داده کاوشگر محور) ظاهر شد. در شکل ۳ وجود ژن‌های مشترک بین بافت‌های مورد بررسی در قالب نمودار ون ترسیم شده



شکل ۳. نمودار ون از ژن های مشترک بین ۵ بافت مورد بررسی پس از نگاشت کاوشگرهای هاب به ژن ها
Figure 3. Venn diagram of common genes between the 5 studied tissues after the mapping of hub probes to genes

جدول ۸. نام علمی ژن های شناسایی شده کاوشگر محور به تفکیک بافت های مورد بررسی
Table 8. Characteristics of the Discovered Hub Genes for 5 Studied Tissues- Probe Based

Tissue	Pubmed ID	Enterez ID	RefSeq ID	Location	Length (bp)	Product PubMed Name
Muscle	GSTM3	615507	NM_001046560.1	chr3:33768050-33770824	2775	Glutathione S-Transferase Mu 3
	ADAM12	407229	NM_001001156.1	chr26:45848826-46238138	389313	Disintegrin And Metalloproteinase Domain-Containing Protein 12 Precursor
	ASPN	507990	NM_001034309.2	chr8:85499331-85524670	25340	Asporin Precursor
	C3	280677	NM_001040469.2	chr7:18989627-19025593	35967	Complement C3 Preproprotein
	CFB	514076	NM_001040526.1	chr23:27225266-27231292	6027	Complement Factor B Precursor
Mammary	CHRD1	100336905	NM_001205187.1	chrX:63688443-63744897	56455	Chordin-Like Protein 1
	COL1A2	282188	NM_174520.2	chr4:11624470-11661163	36694	Collagen Alpha-2(I) Chain Precursor
	ISG15	281871	NM_174366.1	chr16:52714627-52715654	1028	
	JSP.1	407173	NM_001040498.1	chr23:28469733-28473374	3642	MHC Class I JSP.1 Precursor
	LOC616942	616942	NM_001206459.1	chr23:27842253-27846497	4245	Major Histocompatibility Complex Class I A-Like Precursor
	NELL2	524622	NM_001102084.1	chr5:35657329-36042698	385370	Protein Kinase C-Binding Protein NELL2 Precursor
	NID2	521854	NM_001102065.1	chr10:44894656-44986659	92004	Nidogen-2 Precursor
	S100A12	282467	NM_174651.3	chr3:17164142-17165266	1125	Protein S100-A12
	S100A8	616818	NM_001113725.2	chr3:17146918-17147899	982	Protein S100-A8
	TGFB1	539596	NM_001205402.1	chr7:49062892-49095978	33087	Transforming Growth Factor-Beta-Induced Protein Ig-H3 Precursor
Liver	VCAN	282662	NM_181035.2	chr7:85666058-85782896	116839	Versican Core Protein Precursor
	BOLA-DQB	282495	NM_001012676.3	chr23:25375270-25388621	13352	Major Histocompatibility Complex Class II DQ Beta Precursor
	FMO5	788719	NM_001101304.1	chr3:22571922-22615026	43105	Dimethylaniline Monooxygenase [N-Oxide-Forming]
	HAO1	533957	NM_001079781.2	chr13:51150852-51173199	22348	Hydroxyacid Oxidase 1
	JSP.1	407173	NM_001040498.1	chr23:28469733-28473374	3642	MHC Class I JSP.1 Precursor
	PRLR	281422	NM_001039726.2	chr20:39073246-39137480	64235	Prolactin Receptor Long Form Precursor
	SLC10A1	532890	NM_001046339.1	chr10:81867933-81890829	22897	Sodium/Bile Acid Cotransporter
	SLCO1B3	404160	NM_205804.2	chr5:89323086-89391585	68500	Solute Carrier Organic Anion Transporter Family Member 1B3
	BOLA-DQB	282495	NM_001012676.3	chr23:25375270-25388621	13352	Major Histocompatibility Complex Class II DQ Beta Precursor
	IGFBP1	282259	NM_174554.3	chr4:76720425-76725309	4885	Insulin-Like Growth Factor-Binding Protein 1 Precursor
Uterus	MAP2	281294	NM_001205807.1	chr2:97843887-97991789	147903	Microtubule-Associated Protein 2
	MMD2	525334	NM_001075870.2	chr25:39609810-39672870	63061	Monocyte To Macrophage Differentiation Factor 2
	FABP3	281758	NM_174313.2	chr2:122723225-122783830	60606	Fatty Acid-Binding Protein Heart
Ovary	FBXO32	513776	NM_001046155.1	chr13:82227724-82228041	318	F-Box Only Protein 32

نتیجه‌گیری

این پژوهش نشان داد، ژن‌های متفاوت بیان شده در هر بافت به‌حتم نمی‌توانند ژن‌های هاب باشند. ژن‌های هاب شناسایی شده به شیوه ژن محور بسیار متفاوت از ژن‌های شناسایی شده به شیوه کاوشگر محور بودند، با این حال در روش ژن محور برای هر بافت ژن‌های به‌کلی متفاوتی شناسایی شدند. درحالی‌که نتایج روش کاوشگر محور بیانگر وجود ژن‌های مشترک در بافت‌های متفاوت بود (پستان و کبد ژن JSP.1 و در بافت کبد و رحم ژن BOLA-DQB). بنابراین نمی‌توان نتایج مدل‌سازی ژن محور را با کاوشگر محور همسان در نظر گرفت.

سپاسگزاری

از همکاری سازنده آقای Marco Scutari پژوهشگر مؤسسه ژنتیک UCL لندن و برنامه‌نویس بسته نرم‌افزاری bnlearn در ارائه پیشنهادها و نظرهای سازنده در زمینه اجرا و انتخاب بهینه‌ترین حالت الگوریتم تپه‌نوردی، تشکر و قدردانی می‌گردد.

در این پژوهش از نگرش علوم رایانه‌ای و یادگیری ماشین به بررسی رابطه‌های بین ژنی در هر بافت با ارزیابی کارایی حالت‌های برازش شبکه بیزی به کمک الگوریتم تپه‌نوردی پرداخته شد. نتایج نشان داد، استفاده از سه حالت متفاوت ترکیب فراسنجه‌های الگوریتم تپه‌نوردی (استفاده از الگوریتم به‌تنهایی، نبود) استفاده از بوت‌استرپ و بوت‌استرپ به همراه محدودیت در انتخاب آستانه و توان و برای یال) می‌تواند تأثیر متفاوتی روی شمار و نوع ژن‌ها و کاوشگرهای شناسایی شده داشته باشد. با این حال تأیید درستی نتایج این پژوهش و ژن‌های هاب شناسایی شده در بافت‌های مختلف، مستلزم تکرار نتایج همسان در دیگر بررسی‌ها است. بنابراین پیش از اعتبارسنجی نتایج این تحقیق، نمی‌توان آن را به‌غیر از داده‌های مورد تجزیه در این پژوهش تعمیم داد. استفاده از بررسی‌های شبیه‌سازی ضروری به نظر می‌رسد. نتایج به‌دست آمده از

REFERENCES

- Alexandre, P. A., Kogelman, L. J., Santana, M. H., Passarelli, D., Pulz, L. H., Fantinato-Neto, P., Silva, P. L., Leme, P. R., Strefezzi, R. F. & Coutinho, L. L. (2015). Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. *BMC Genomics*, 16(1), 1073.
- Chang, H.-H. & Mcgeachie, M. (2011). Phenotype prediction by integrative network analysis of SNP and gene expression microarrays. In: *Proceedings of Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 30 Aug - 03 Sep 2011, Boston Marriott Copley Place Hotel, Boston, MA, USA, pp. 6849-6852.
- Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. *Inter Journal, Complex Systems*, 1695(5), 1-9.
- Elo, L. L., Lahti, L., Skottman, H., Kylänemi, M., Lahesmaa, R. & Aittokallio, T. (2005). Integrating probe-level expression changes across generations of Affymetrix arrays. *Nucleic Acids Research*, 33(22), e193-e193.
- Fortes, M., Snelling, W., Reverter, A., Nagaraj, S., Lehnert, S., Hawken, R., Deatley, K., Peters, S., Silver, G. & Rincon, G. (2012). Gene network analyses of first service conception in Brangus heifers: use of genome and trait associations, hypothalamic-transcriptome information, and transcription factors. *Journal of Animal Science*, 90(9), 2894-2906.
- Friedman, N. (2004). Inferring cellular networks using probabilistic graphical models. *Science*, 303(5659), 799-805.
- Gene Expression Omnibus. (2017). NCBI: National Center for Biotechnology Information, Retrieved November 14, 2017, from <https://www.ncbi.nlm.nih.gov/geo/>
- Ghaderi-Zefrehei, M., Dolatabady, M. & Rowghani, E. (2015). Simple gene regulatory network of immune system candidate genes in dairy cattle. *Research Opinions in Animal and Veterinary Sciences*, 5(12), 499-506.
- Girard, A., Dufort, I., Douville, G. & Sirard, M. A. (2015). Global gene expression in granulosa cells of growing, plateau and atretic dominant follicles in cattle. *Reproductive Biology and Endocrinology*, 13(1), 17.
- Hageman, R. S., Leduc, M. S., Korstanje, R., Paigen, B. & Churchill, G. A. (2011). A Bayesian framework for inference of the genotype-phenotype map for segregating populations. *Genetics*, 187(4), 1163-1170.

11. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, 12(6), 996-1006.
12. Kogelman, L. J., Cirera, S., Zhernakova, D. V., Fredholm, M., Franke, L. & Kadarmideen, H. N. (2014). Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. *BMC Medical Genomics*, 7(1), 57.
13. Komolka, K., Ponsuksili, S., Albrecht, E., Kühn, C., Wimmers, K. & Maak, S. (2016). Gene expression profile of Musculus longissimus dorsi in bulls of a Charolais × Holstein F 2-cross with divergent intramuscular fat content. *Genomics Data*, 7131-133.
14. Langfelder, P. & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.
15. Liu, F., Zhang, S.-W., Guo, W.-F., Wei, Z.-G. & Chen, L. (2016). Inference of gene regulatory network based on local bayesian networks. *PLoS Computational Biology*, 12(8), e1005024.
16. Malovini, A., Nuzzo, A., Ferrazzi, F., Puca, A. A. & Bellazzi, R. (2009). Phenotype forecasting with SNPs data through gene-based Bayesian networks. *BMC Bioinformatics*, 10(2), S7.
17. Milchevskaya, V., Tödt, G. & Gibson, T. J. (2017). A Tool to Build Up-To-Date Gene Annotations for Affymetrix Microarrays. *Genomics and Computational Biology*, 3(2), 38.
18. Nagarajan, R., Scutari, M. & Lèbre, S. (2013). *Bayesian networks in R*. Springer.
19. Ramayo-Caldas, Y., Fortes, M. R. S., Hudson, N. J., Porto-Neto, L. R., Bolormaa, S., Barendse, W., Kelly, M., Moore, S. S., Goddard, M. E. & Lehnert, S. A. (2014). A marker-derived gene network reveals the regulatory role of ,and in intramuscular fat deposition of beef cattle. *Journal of Animal Science*, 92(7), 2832-2845.
20. Rebhan, M., Chalifa-Caspi, V., Prilusky, J. & Lancet, D. (1997). GeneCards: integrating information about genes, proteins and diseases. *Trends in Genetics*, 163(4), 13.
21. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A. & Nolan, G. P. (2005). Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721), 523-529
22. Scutari, M. (2014). Bayesian network constraint-based structure learning algorithms: Parallel and optimised implementations in the bnlearn r package. *arXiv preprint arXiv:1406.7648*.
23. Sherif, F. F., Zayed, N. & Fakhr, M. (2015). Discovering Alzheimer genetic biomarkers using Bayesian networks. *Advances in bioinformatics*, 2015.
24. Stalteri, M. A. & Harrison, A. P. (2007). Interpretation of multiple probes sets mapping to the same gene in Affymetrix GeneChips. *BMC Bioinformatics*, 8(1), 13.
25. Stingo, F. C., Swartz, M. D. & Vannucci, M. (2015). A Bayesian approach to identify genes and gene-level SNP aggregates in a genetic analysis of cancer data. *Statistics and Its Interface*, 8(2), 137-151.
26. Verardo, L., Lopes, M., Wijga, S., Madsen, O., Silva, F., Groenen, M., Knol, E., Lopes, P. & Guimarães, S. (2016). After genome-wide association studies: Gene networks elucidating candidate genes divergences for number of teats across two pig populations. *Journal of Animal Science*, 94(4), 1446-1458.
27. Weber, K. L., Welly, B. T., Van Eenennaam, A. L., Young, A. E., Porto-Neto, L. R., Reverter, A. & Rincon, G. (2016). Identification of gene networks for residual feed intake in Angus cattle using genomic prediction and RNA-seq. *PloS One*, 11(3), e0152274.