دامستیک؛ دوره ۲۲، شماره ۲ (شماره ۲۳ پیاپی): ۱۳-۵

نشریهٔ علمی- ترویجی (حرفه‌ای)

دامِستیک

انجمن علمی- دانشجویی گروه مهندسی علوم دامی دانشگاه تهران؛ پاییز ۱۴۰۱

https://domesticsj.ut.ac.ir/article_90439.html

مقاله علمی- ترویجی

روش تجزیه و تحلیل شبکهٔ هم‌بیانی ژنی وزن‌دار و کاربرد آن در ژنتیک و اصلاح نژاد دام و طیور

وحید دهقانیان ریحان[۱]، مصطفی صادقی[۲*]، فرزاد غفوری[۱]

[۱] دانشجوی دکتری تخصصی ژنتیک و اصلاح نژاد دام و طیور، گروه مهندسی علوم دامی، دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، البرز، ایران

[۲] دانشیار ژنتیک و اصلاح نژاد دام و طیور، گروه مهندسی علوم دامی، دانشکدگان کشاورزی و منابع طبیعی دانشگاه تهران، کرج، البرز، ایران

https://doi.org/10.22059/domesticsj.2022.344763.1098

چکیده

در میان انواع مختلف شبکه‌ها، شبکه‌های هم‌بیانی ژنی بیشترین انعطاف‌پذیری را برای بررسی صفات مختلفی همچون صفات عملکردی و تولیدمثلی، بیماری‌ها و غیره دارند. هم‌بیان بودن ژن‌ها عموماً به همبستگی بین ژن‌ها در سطوح رونوشت اشاره دارد، همچنین از طرفی دیگر می‌تواند در تمام مقیاس‌های بیولوژیکی (مانند پروتئین‌ها، متابولیت‌ها و یا به صورت ترکیبی بین رونوشت‌ها، پروتئین‌ها و متابولیت‌ها) برای مطالعهٔ روابط همبستگی بین ژن‌ها استفاده شود. شبکه‌های هم‌بیانی تا حدی به دلیل امکان بهره‌گیری از فناوری‌هایی مانند ریزآرایه‌ها، RNA-Seq و طیف‌سنجی جرمی محبوب شده‌اند، چرا که امکان بررسی واسطه‌های مولکولی در مقیاس‌های بیولوژیکی مختلف به روشی ساده و در تعداد نسبتاً زیادی نمونه را فراهم می‌کند. همچنین، با استفاده از این روش اندازه‌گیری معنی‌داری بیان همزمان ژن‌ها از نظر بیولوژیکی در انواع سلول‌های خاص امکان‌پذیر است. به عنوان یک مقایسه، بیشتر شبکه‌های تعاملی پروتئین با پروتئین (PPI) صرفاً نشان‌دهندهٔ تعاملات کلی بین ژن‌ها می‌باشند که اشاره‌ای به نوع سلول و بخش زمانی بیان ژن‌ها ندارند، در حالی که شبکه‌های هم‌بیانی ژنی را می‌توان با استفاده از داده‌های به دست آمده از انواع سلول‌های خاص از افراد مختلف (بالادست و پایین‌دست در مورد یک صفت فنوتیپی مانند افراد با باروری بالا و پایین) و در سراسر مراحل رشد بازسازی کرد. یکی از پرکاربردترین الگوریتم‌ها برای ساخت شبکه‌های هم‌بیانی ژنی، تجزیه و تحلیل شبکهٔ هم‌بیانی ژنی وزن‌دار (WGCNA) است که به دلیل استفادهٔ گسترده از آن در بسیاری از مطالعات هم‌بیانی، توضیح نحوهٔ عملکرد آن آموزنده خواهد بود. بنابراین شناسایی ماژول‌ها، ژن‌ها و مسیرهای متابولیکی- سیگنالینگ مرتبط با صفات مختلف مورد مطالعه با استفاده از روش WGCNA، ممکن است بینش جدیدی در رابطه با مکانیسم‌های مولکولی را نشان دهند. در واقع هدف از این مطالعه، ارائه توضیحات اجمالی در رابطه با روش تجزیه و تحلیل شبکهٔ هم‌بیانی ژنی وزن‌دار و کاربرد آن در ژنتیک و اصلاح نژاد دام و طیور است.

کلمات کلیدی: WGCNA، ترانسکریپتوم، سیستم بیولوژی، شبکه‌های هم‌بیانی ژنی

***نویسنده مسئول:** sadeghimos@ut.ac.ir

بخش: ژنتیک و اصلاح‌نژاد دام و طیور **دبیر تخصصی:** دکتر معصومه ناصرخیل

تاریخ دریافت: ۱۴۰۱/۰۴/۰۵ **تاریخ بازنگری:** ۱۴۰۱/۰۴/۲۵ **تاریخ پذیرش:** ۱۴۰۱/۰۵/۰۶ **تاریخ انتشار آنلاین:** ۱۴۰۱/۰۹/۱۵

رفرنس‌دهی: دهقانیان ریحان، و.، صادقی، م.، غفوری، ف. روش تجزیه و تحلیل شبکهٔ هم‌بیانی ژنی وزن‌دار و کاربرد آن در ژنتیک و اصلاح نژاد دام و طیور. *علمی- ترویجی (حرفه‌ای) دامِستیک*، ۱۴۰۱؛ ۲۲(۲): ۱۳-۵.

مقدمه

با توجه به توسعه روز افزون فناوری‌های دیجیتال با توان بالا و ثبت اطلاعات ژنومی و فنوتیپی و تولید آبرداده‌ها (Big data) و به دنبال آن نیاز به ذخیره، بازیابی و تجزیه و تحلیل داده‌های حاصل از مطالعات متعدد، باعث پیدایش علوم جدیدی مانند بیوانفورماتیک، تکنیک‌های هوش مصنوعی، یادگیری ماشینی، شبکه عصبی مصنوعی و سایر تکنولوژی‌های نوین شد (Ghafouri *et al.*, 2020). در این میان، بیوانفورماتیک یک علم میان رشته‌ای است که داده‌های مختلف زیستی (شامل DNA، RNA، پروتئین و غیره) را توسط تئوری‌های آماری با الگوریتم‌ها و برنامه‌های کامپیوتری، ذخیره، بازیابی و تجزیه و تحلیل می‌کند. فعالیت‌های بیوانفورماتیک به طور کلی به چهار بخش کلی شامل ژنومیکس (Genomics)،ترانسکریپتومیکس (Transcriptomics)، پروتئومیکس (Proteomics) و متابولومیکس (Metabolomics) دسته‌بندی می‌شوند. سیستم بیولوژی یا زیست‌شناسی سامانه‌ای نیز رویکردی از علم بیوانفورماتیک است که در سال‌های اخیر با گسترش داده‌های پُربرونداد بیش از پیش مورد توجه قرار گرفته است و تحقیقات بیشتری در جهت مطالعات سیستماتیک در زمینهٔ سلول‌ها، اندام‌ها و موجودات زنده، به ویژه فرآیندهای سلولی مانند برهمکنش‌های مولکولی، ارتباطات بین سلولی، تقسیم سلولی، هموستازی و سازگاری‌های محیطی در این زمینه صورت می‌گیرد. موضوعات اصلی مرتبط با سیستم بیولوژی شامل آنالیز شبکه‌های مختلف است که مهم‌ترین این شبکه‌ها شامل شبکه‌های ژنی، شبکه‌های تعاملی پروتئین- پروتئین (PPI)، شبکه‌های متابولیسمی و شبکه‌های سیگنالینگ می‌باشند (باباعباسی، ۱۳۹۵).

در سال‌های اخیر، تجزیه و تحلیل شبکه هم‌بیانی ژنی وزن‌دار (WGCNA: Weighted Gene Co-expression Network Analysis) به عنوان یک چارچوب برای پیاده‌سازی رویکرد سیستم بیولوژی مطرح شده است (Zhao *et al.*, 2020; Zhang *et al.*, 2022). هدف اصلی از ارائه این تکنیک تجزیه و تحلیل رفتار هم‌بیانی مجموعه‌ای از ژن‌ها به جای بررسی انفرادی آن‌ها بوده است. در این روش به جای ارتباط چندین هزار ژن با یک متغیر یا یک صفت، همبستگی چند گروه ژنی (ماژول) با متغیر یا صفت خاص مورد نظر برای ساخت شبکهٔ هم‌بیانی وزن‌دار بررسی می‌شود (درزی و همکاران، ۱۴۰۰). در واقع هدف از ارائه این مطالعه، تشریح روش کار، بیان مزایا و معایب روش WGCNA نسبت به سایر شبکه‌ها و به ویژه شبکهٔ غیر وزن‌دار و همچنین بررسی کاربردهای این روش در زمینه‌های مختلف علمی به ویژه در رشتهٔ مهندسی علوم دامی و زمینهٔ تخصصی ژنتیک و اصلاح‌نژاد دام و طیور است.

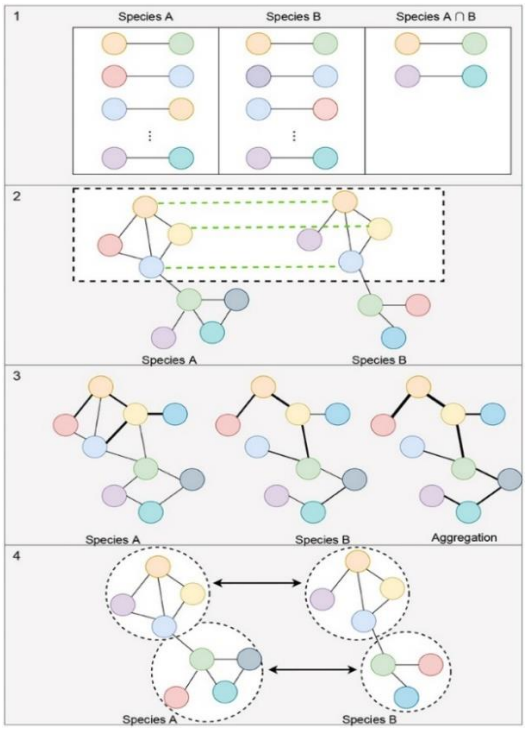
تجزیه و تحلیل شبکهٔ هم‌بیانی ژنی وزن‌دار (WGCNA)

تجزیه و تحلیل شبکهٔ همبستگی وزن‌دار که به عنوان شبکه هم‌بیانی ژنی وزن‌دار نیز شناخته می‌شود، یک روش داده کاوی پُرکاربرد به ویژه برای مطالعه شبکه‌های بیولوژیکی با همبستگی جفتی بین متغیرها است. روش WGCNA توسط استیو هوروات (Steve Horvath) و همکاران (۲۰۰۸)، استاد ژنتیک انسانی در دانشگاه کالیفرنیا توسعه داده شد (Ovens *et al.*, 2021). تئوری تجزیه و تحلیل شبکهٔ هم‌بیانی ژنی وزن‌دار بر مقایسهٔ همزمان بیان تعداد بسیار بالای ژن‌ها تحت عنوان ماژول و با توجه به تعریف اثر تعاملی ژن‌های هاب درون هر ماژول با یکدیگر (مقایسهٔ ماژول‌های هم‌بیان) با صفت مورد مطالعه (مقایسهٔ ماژول‌های هم‌بیان با صفت) استوار است؛ همچنین امکان مقایسهٔ توپولوژی شبکه‌های مختلف (بررسی تفاوت شبکه) را نیز فراهم می‌سازد (Horvath, 2011). اصول کلی روش WGCNA در چهار مرحله کلی شکل می‌گیرد: اول، ساخت شبکهٔ هم‌بیانی ژنی که به صورت ریاضی به وسیلهٔ یک ماتریس همجواری (Adjacency matrix) نمایش داده می‌شود که نشان‌دهندهٔ شباهت هم‌بیانی بین یک جفت ژن است. دوم، شناسایی ماژول‌ها؛ به این صورت که روش WGCNA از خوشه‌بندی سلسله مراتبی برای شناسایی ماژول‌ها استفاده می‌کند. برای اندازه‌گیری تفاوت بین خوشه‌ها، WGCNA از یک رویکرد عدم تشابه همپوشانی توپولوژیکی جهت دسته‌بندی ماژول‌ها استفاده می‌کند. سوم، مرتبط ساختن ماژول‌ها به فنوتیپ؛ چندین روش را می‌توان برای اندازه‌گیری ارتباط یک ماژول با یک صفت فنوتیپی استفاده کرد. به عنوان مثال می‌توان از ME (Module Eigengene) که به عنوان اولین جزء اصلی یک ماژول و یا MS (Multiple Sclerosis) که به عنوان میانگین اهمیت ژن تعریف می‌شود، استفاده نمود. چهارم، مطالعهٔ روابط بین ماژول‌ها که با استفاده از دو رویکرد kME_i و kIM_i است که در نهایت برای شناسایی مهمترین ژن‌های هاب در بین ماژول‌ها کاربرد دارد، صورت می‌گیرد (Zhang and Horvath, 2005; Li *et al.*, 2018). در واقع در روش WGCNA فرض بر این است که شبکه های ژنتیکی از معیار توپولوژی بدون مقیاس تبعیت می‌کنند. روش WGCNA به جای تقسیم‌بندی بیان مشترک ژن (متصل = ۱، غیر متصل = ۰)، از یک آستانه نرم برای تعیین وزن لبه‌های جفت ژن‌ها استفاده می‌کند؛ در این راستا

اثبات شده است که نتایج قوی‌تری نسبت به شبکه‌های غیر وزن‌دار ارائه می‌دهد. یک آستانهٔ نرم مناسب، شبکه هم‌بیانی حاصل را به یک شبکه بدون مقیاس نزدیک‌تر می‌کند. همچنین این روش به جای مرتبط کردن ژن‌های فردی به فنوتیپ، بر رابطهٔ بین چند ماژول و صفت یا صفات مورد مطالعه تمرکز می‌کند، به گونه‌ای که تا حد زیادی مشکلات رایج در مورد تست‌های متعدد در تجزیه و تحلیل داده‌های ریزآرایه را کاهش می‌دهد. بنابراین می‌توان عنوان کرد که روش WGCNA به طور گسترده در تجزیه و تحلیل داده‌های ترانسکریپتومی استفاده می‌شود به گونه‌ای که در آن نمونه‌ها مستقل از یکدیگر فرض می‌شوند (Fuller *et al.*, 2007; Li *et al.*, 2018).

مزایای شبکهٔ هم‌بیانی ژنی وزن‌دار نسبت به شبکهٔ هم‌بیانی ژنی غیر وزن‌دار

- در شبکه‌های هم‌بیانی از دو نوع حد آستانه برای تعیین وجود یا عدم وجود همبستگی بین دو ژن استفاده می‌شود؛ حد آستانهٔ سخت (در شبکه‌های غیر وزن‌دار) و نرم (در شبکه وزن‌دار). برای تعیین وضعیت همبستگی ژن‌ها در شبکه‌های هم‌بیانی ژنی غیر وزن‌دار از حد آستانهٔ سخت استفاده می‌شود، در این روش اطلاعات هم‌بیانی به صورت دوتایی یا باینری می‌باشند که منجر به از دست رفتن اطلاعات زیادی می‌شود. در مقابل در شبکه‌های هم‌بیانی ژنی وزن‌دار از حد آستانهٔ نرم استفاده می‌شود که نتایج قوی‌تر و دقیق‌تری را ارائه می‌دهد؛ به گونه‌ای که ماهیت پیوستهٔ اطلاعات همبستگی نیز حفظ می‌شود (Zhang and Horvath, 2005).
- در شبکه‌های هم‌بیانی وزن‌دار تفسیر هندسی (شبکه) بر اساس تفسیر زاویه‌ای همبستگی راحت‌تر است، در حالی که در شبکه‌های هم‌بیانی غیر وزن‌دار به این شکل نیست (Horvath and Dong, 2008).
- در شبکه‌های هم‌بیانی وزن‌دار اطلاعات آماری به دست آمدهٔ شبکه را می‌توان برای بهبود روش‌های استاندارد داده کاوی مانند تجزیه و تحلیل خوشه‌ای استفاده کرد، زیرا معیارهای (عدم) شباهت اغلب می‌توانند به شبکه‌های وزن‌دار تبدیل شوند (Horvath and Dong, 2008; Oldham *et al.*, 2012).
- شبکه‌های هم‌بیانی وزن‌دار نتایج آماری قوی‌تری در رابطه با ماژول‌های ژنی حفاظت شده ارائه می‌دهند؛ به گونه‌ای که می‌توانند برای ارزیابی نسبت به شرایط مختلف مورد استفاده قرار گیرند. همچنین، امکان مطالعهٔ تفاوت‌ها و شباهت‌ها بین ساختار ماژول‌های شبکه‌های مختلف را نیز فراهم می‌کند (Langfelder *et al.*, 2011؛ شکل ۱).
- شبکه‌های هم‌بیانی وزن‌دار را اغلب می‌توان با شبکه‌های "قابل فاکتورسازی" (Factorizable) تقریب زد (Dong and Horvath, 2007; Horvath and Dong, 2008). دستیابی به چنین تقریبی اغلب برای شبکه‌های پراکنده و بدون وزن دشوار است. بنابراین، شبکه‌های هم‌بیانی وزن‌دار امکان پارامترسازی محدود (با توجه به ماژول‌ها و وضعیت عضویت ژن‌ها در ماژول‌ها) را فراهم می‌کنند؛ به عبارتی شبکهٔ هم‌بیانی وزن‌دار می‌تواند تجزیه شود و یا به یک شبکه‌های ساده‌تری دسته‌بندی و تقریب شود (Horvath, 2011; Ranola *et al.*, 2013; Guillen-Gamez and Migallón, 2018).



شکل ۱- مقایسه برای شباهت شبکه‌های هم‌بیانی ژنی دو گونهٔ مختلف. (۱) اندازه‌گیری شباهت‌های توپولوژیکی بین شبکه‌ها: همان‌طور که در تصویر نشان داده شده است، شامل شناسایی جفت ژن‌های حفظ‌شده یا مقایسه‌های زیرشبکه‌ای پیچیده‌تر مانند شمارش تعداد زیرشبکه‌های کوچک است. (۲) نمایش پیوندهای ارتولوگ (با پیوندهای سبز بین شبکه‌ها نشان داده شده است) برای شناسایی ماژول‌های حفاظت شده ژن‌ها. (۳) نمونه‌ای از مقایسهٔ وزن لبه‌ها یا تجمیع شبکه که در آن گره‌های شبکه به طور خودکار با هم تطبیق داده می‌شوند و وزن لبه‌ها برای به دست آوردن مقدار شباهت بین شبکه‌ها جمع می‌شوند. (۴) تشخیص ماژول که در آن هر ماژول را می‌توان با استفاده از آمار اتصال شبکه و چگالی مقایسه کرد (Ovens *et al.*, 2021).

محدودیت‌های شبکهٔ هم‌بیانی ژنی وزن‌دار

در حالی که روش WGCNA یک ابزار قدرتمند تجزیه و تحلیل برای ریزآرایه است، کاربران باید از محدودیت‌های آن نیز آگاه باشند. اول اینکه روش WGCNA فرض می‌کند که داده‌های ریزآرایه به درستی از قبل پردازش و نرمال شده‌اند که برای نرمال‌سازی داده‌های بیانی، چندین تابع R در بسته‌های Bioconductor پیاده سازی شده است (Gentleman *et al.*, 2005). دوم، نتایج WGCNA مشابه بسیاری از روش‌های داده‌کاوی دیگر، می‌تواند به دلیل آلودگی‌های بافت یا طراحی ضعیف آزمایش، اریب یا نامعتبر باشد. سوم، اگر چه چندین روش تشخیص ماژول هم‌بیان در بستهٔ نرم‌افزاری R پیاده‌سازی شده است، امّا اینکه کدام روش بهترین روش می‌باشد را ارائه نمی‌دهد (Fuller *et al.*, 2011).

تئوری روش تجزیه و تحلیل شبکهٔ هم‌بیانی ژنی وزن‌دار

یکی از پرکاربردترین تکنیک‌های مورد استفاده برای تشخیص ماژول در شبکه‌های هم‌بیانی ژنی (GCN: Gene Co-expression Network)، تحلیل شبکه‌ای هم‌بیانی ژنی وزن‌دار است. اگر چه WGCNA در سال ۲۰۰۸ ابداع گردید، امّا هنوز هم برای شناسایی ماژول‌های مهم و ژن‌های هاب مرتبط با بیماری‌ها (Allen *et al.*, 2018; Swarup *et al.*, 2019)، مسیرهای بیولوژیکی (Silva-Vignato *et al.*, 2019) و رشد و نمو (Spadafora *et al.*, 2018) مورد استفاده قرار می‌گیرد. فرآیند تشکیل یک شبکهٔ هم‌بیان ژنی وزن‌دار به این صورت است که پس از نرمال‌سازی داده‌های بیانی ژنی (که به فرمت ریزآرایه می‌باشند)، ابتدا ضریب همبستگی پیرسون که می‌تواند به صورت بی‌علامت (Unsigned) یا علامت‌دار (Signed) (به ترتیب معادلات (۱) و (۲)) باشد، محاسبه می‌شود. البته روش‌های دیگری نیز برای محاسبهٔ ضریب همبستگی وجود دارد؛ با این حال معمولاً از فرمول ضریب همبستگی پیرسون استفاده می‌شود.

$$S_{ij}^{unsigned} = |cor(x_i,x_j)| \qquad \text{معادلهٔ (۱)}$$

$$S_{ij}^{singned} = 0.5 + 0.5cor(x_i,x_j) \qquad \text{معادلهٔ (۲)}$$

منظور از محاسبهٔ ضریب همبستگی اندازه‌گیری میزان شباهت (Similarity value) بین دو مؤلفهٔ x_i و x_j است که به ترتیب پروفایل بیانی مربوط به ژن‌های i و j می‌باشند. در محاسبهٔ ضریب همبستگی از طریق معادلهٔ (۱)، صرفاً استفاده از قدر مطلق همبستگی ممکن است اطلاعات بیولوژیکی مبهمی در اختیار بگذارد، زیرا هیچ تمایزی بین سرکوب بیان، فعال‌سازی و یا تحریک بیان ژن وجود ندارد. این در حالی است که در معادلهٔ (۲) میزان شباهت بین دو ژن نشان‌دهندهٔ علامت همبستگی پروفایل بیانی دو ژن نیز می‌باشد. بنابراین، محاسبهٔ میزان شباهت علامت‌دار و بدون علامت در نحوهٔ برخورد با ژن‌های همبسته منفی متفاوت است؛ ژن‌های با همبستگی منفی بالا (نزدیک به ۱-) شباهت کمی در یک شبکه علامت‌دار دارند (با توجه به معادلهٔ ۲ نزدیک به صفر)، امّا در یک شبکه بدون علامت از شباهت زیادی برخوردار هستند (با توجه به معادلهٔ ۱ نزدیک به ۱). با این حال، دامنهٔ نمرهٔ شباهت (S_{ij}) برآورد شدهٔ پروفایل بیانی دو ژن از طریق هر دو معادله، بین صفر و ۱ می‌باشد (Fuller *et al.*, 2011).

از مقادیر محاسبه شدهٔ فوق، برای ساخت ماتریس شباهت یا ماتریس همبستگی (Similarity Matrix or Correlation Matrix) استفاده می شود (معادلهٔ ۳).

$$\boldsymbol{S} = [s_{ij}] \qquad \text{معادلهٔ (۳)}$$

در مرحلهٔ بعد مؤلفه‌های ماتریس همبستگی از طریق تابع مجاورت و با توان β به ماتریس مجاورت یا همجواری (Adjacency Matrix) تبدیل می‌شود. منظور از تابع مجاورت تابعی است که برای اندازه‌گیری کمی قدرت همبستگی بین هر جفت ژن مورد استفاده قرار می‌گیرد (معادلهٔ ۴ و ۵ به ترتیب نشان دهندهٔ تابع مجاورت و ماتریس همجواری است). ماتریس A بر اساس تعیین حد آستانه برای ماتریس S تعریف می‌شود. از آستانهٔ سخت (دو قطبی کردن یا تقسیم به دو بخش) در شبکه‌های هم‌بیانی ژنی غیر وزن‌دار استفاده می‌شود؛ بدین صورت که اگر $s_{ij} > \tau$ همجواری برابر با ۱ دارد و در غیر این صورت برابر صفر خواهد بود. از آنجایی که آستانهٔ سخت اتصالات ژن را به صورت دوتایی کدگذاری می‌کند، می‌تواند نسبت به انتخاب آستانه حساس باشد و منجر به از دست رفتن اطلاعات هم‌بیانی شود. ماهیت پیوستهٔ اطلاعات هم‌بیانی را می‌توان با استفاده از آستانهٔ نرم حفظ کرد که منجر به یک شبکهٔ وزن‌دار می‌شود (Zhang and Horvath, 2005; Fuller *et al.*, 2011).

$$a_{ij} = (s_{ij})^{\beta} \qquad \text{معادلهٔ (۴)}$$

$$A = [a_{ij}] \qquad \text{معادلهٔ (۵)}$$

توان β پارامتری برای تعیین آستانهٔ نرم است. پیش فرض پارامتر β در شبکه‌های هم‌بیانی بدون علامت و علامت‌دار به ترتیب برابر با ۶ و ۱۲ می‌باشد. میزان β باید به گونه‌ای تعیین

گردد تا مؤلفهٔ مقیاس آزاد در شبکه (Scale-free network) ایجاد شود؛ شبکهٔ بی‌مقیاس یا شبکهٔ مقیاس آزاد (Scale-free network) شبکه‌ای است که توزیع درجهٔ آن تحت تبدیل مقیاس، بدون تغییر باقی بماند. به عبارت دیگر، اگر با چند برابر کردن متغیر توزیع درجه، شکل توزیع تغییری نکند، گفته می‌شود که شبکه بی‌مقیاس است. امّا برای ایجاد این خاصیت در شبکه، مقدار β با استفاده از معیار توپولوژی مقیاس آزاد تعیین می‌شود که عبارت است از کوچک‌ترین مقدار β، که به طورتقریبی توپولوژی مقیاس آزاد در شبکه ایجاد می‌شود. در این راستا با توجه به معادلهٔ ۶ می‌توان نتیجه‌گیری کرد که شبکهٔ همجواری وزنی با میزان شباهت هم‌بیانی ژن‌ها به صورت خطی در مقیاس لگاریتمی مرتبط است. علاوه بر این مقدار بالای β، ضرایب همبستگی بالا را به ضرایب همجواری بالا تبدیل می‌کند؛ در حالی که مقادیر کم را به سمت صفر سوق می‌دهد. به علت این که روش آستانهٔ نرم اعمال شده بر روی یک ماتریس همبستگی زوجی (منظور یک جفت است) منجر به ماتریس مجاورت وزنی می‌شود، به عنوان تجزیه و تحلیل شبکهٔ هم‌بیان ژنی وزن‌دار نام برده می‌شود (Zhang and Horvath, 2005; Fuller *et al.*, 2011).

$$\log(a_{ij}) = \beta \log(s_{ij})$$ معادلهٔ (۶)

یک گام مهم در چهارچوب روش WGCNA استفاده از معیار هم‌پوشانی توپولوژیکی (TOM: Topological Overlap Matrix) جهت خوشه‌بندی ژن‌ها در گروه‌های ژنی یا ماژول‌ها و شناسایی ماژول‌های با میزان هم‌بیانی بیشینه می‌باشد. با استفاده از TOM می‌توان به طور همزمان میزان شباهت بیانی یک جفت ژن و تمام ژن‌های متصل به آن جفت ژن را محاسبه کرد. بدین منظور مؤلفه‌های ماتریس همجواری به مقادیر جدیدی با استفاده از تابع TOM-Similarity (معادلهٔ ۷) در ماتریس هم‌پوشانی توپولوژیکی تبدیل می‌شوند. طبق قرارداد حداکثر میزان شباهت بیانی دو ژن برابر با ۱ و حداقل صفر می‌باشد. ماتریس TOM یک ماتریس متقارن است. بسیاری از روش‌های خوشه‌بندی استاندارد جهت خوشه‌بندی به معیار عدم تشابه نیاز دارند، که در آن ژن‌های با بیان همزمان تفاوت کمی دارند. بنابراین، معیار عدم تشابه هم‌پوشانی توپولوژیکی به عنوان ورودی در خوشه‌بندی سلسله مراتبی پیوند متوسط استفاده می‌شود. سپس ماژول‌ها به عنوان شاخه‌های درخت خوشه‌ای تعریف می‌شوند (شکل ۲). بدین منظور، روش برش شاخهٔ پویا (Dynamic branch cutting) و یا روش برش درخت پویا (Dynamic tree cutting) به کار گرفته می‌شود (Langfelder *et al.*, 2007; Fuller *et al.*, 2011). این روش تشخیص ماژول در بسیاری از مطالعات مفید و کاربردی بوده است (Oldham *et al.*, 2008; Presson *et al.*, 2008; Weston *et al.*, 2008) (معادلهٔ ۸).

$$TOM_{ij} = \frac{\sum_{u\neq i} a_{iu}a_{uj}+a_{ij}}{min\,(k_i,k_j)+1-a_{ij}}$$ معادلهٔ (۷)

$$k_i = \sum_{u\neq i} a_{ui}$$ معادلهٔ (۸)

$$dissTOM_{ij} = 1 - TOM_{ij} = 1 - \frac{\sum_{u\neq i} a_{iu}\,a_{uj} + a_{ij}}{min\,(k_i,k_j) + 1 - a_{ij}}$$

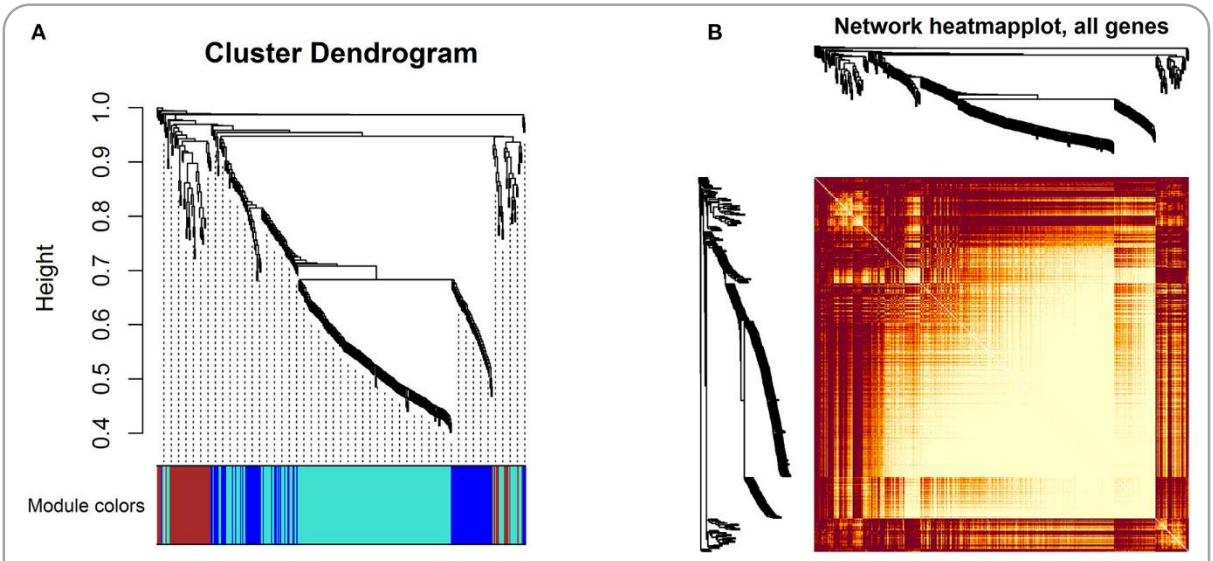
در بیشتر موارد در صورتی که بتوان پروفایل‌های بیانی همه ژن‌های در یک ماژول را با استفاده از یک پروفایل بیانی منفرد داشت، مفید خواهد بود. بدین منظور، ژن ویژهٔ ماژول (ME: Module Eigengene) به عنوان اولین مؤلفهٔ اصلی پروفایل‌های بیانی استاندارد یک ماژول معین تعریف می‌شود (Langfelder *et al.*, 2007; Horvath and Dong, 2008). ME را می‌توان میانگین وزنی بیان ژن‌های ماژول در نظر گرفت. پس از شناسایی ماژول‌ها مهم‌ترین مسئله، برقراری ارتباط بین ماژول‌ها با صفت مورد مطالعه و سپس مطالعهٔ روابط بین ماژول‌ها می‌باشد. برای مطالعهٔ ارتباط ماژول‌ها با صفت فنوتیپی از دو روش تعیین ضریب همبستگی ME ماژول مربوط با صفت مورد نظر و یا آزمون معنی‌داری ماژول (MS: Multiple Sclerosis) که عبارت است از میانگین معنی‌داری ژنی (GS: Gene Significance) کل ژن‌های موجود در داخل ماژول مربوط استفاده می‌شود. در واقع GS یک گره به عنوان همبستگی بین گره و صفت فنوتیپی تعریف می‌شود. ماژول‌هایی که بیشترین سطح معنی‌داری را در صفت مورد نظر دارند، ممکن است مسیرهای مرتبط با صفت فنوتیپی مورد نظر را نشان دهند (Fuller *et al.*, 2011; Li *et al.*, 2018).

علاوه بر این، MEها نشانگرهای زیستی قوی را تعریف می‌کنند (Foroushani *et al.*, 2017) و می‌توانند به عنوان ساختارهایی در مدل‌های پیچیده یادگیری ماشینی مانند شبکه‌های بیزی استفاده شوند (Agrahari *et al.*, 2018). برای مطالعهٔ روابط بین ماژول‌ها می‌توان شبکه‌های هم‌بیانی را بین شبکه‌های MEها (Module eigengene networks)، یعنی شبکه‌هایی که گره‌های آن‌ها ماژول هستند، بازسازی کرد. مطالعهٔ روابط ماژول‌ها می‌تواند به یافتن این که کدام ماژول‌ها با صفت مورد نظر بسیار مرتبط هستند، کمک کند. به گونه‌ای که برای شناسایی ژن‌های هاب در داخل یک ماژول مشخص، می‌توان از دو نوع معیار اتصال استفاده کرد. معیار اتصال اول که بر پایهٔ ME ($k_{ME,i}$) شناخته می‌شود، بر اساس همبستگی هر ژن با ME

مربوطه تعریف شده است (معادلهٔ ۹). معیار اتصال دوم که معیار اتصال درون ماژولی ($k_{IM,i}$) نامیده می‌شود، بر اساس مجموع مجاورت‌ها با توجه به ژن‌های ماژول تعریف می‌شود (معادلهٔ ۱۰). در عمل، این دو معیار معادل هم هستند. گره‌هایی که بیشترین تعداد لبه‌ها را در داخل ماژول‌های مهم و معنی‌دار دارند، به عنوان محرک‌های (Driver) کلیدی و مهم هستند، زیرا عملکرد این ژن بر همه ژن‌های متصل تأثیر می‌گذارد (Langfelder *et al.*, 2007; Horvath and Dong, 2008; Langfelder *et al.*, 2011). اگر یک ژن ویژه (E^q) را در ماژول q در نظر داشته باشیم، آنگاه معیار اندازه‌گیری اتصال i-اُمین ژن در داخل آن ماژول برابر با معادلهٔ ۹ خواهد بود. دامنهٔ معیار عضویت ماژول ($MM^q(i)$) مقادیر $[-1,1]$ می‌باشد. اگر $MM^q(i)$ نزدیک به صفر باشد، i-اُمین ژن بخشی از ماژول q نیست. از طرف دیگر، اگر $MM^q(i)$ نزدیک به ۱ یا ۱- باشد، به شدت به ژن‌های ماژول q متصل است. در مورد معیار اتصال داخل ماژولی نیز میزان اتصال یا بیان همزمان i-اُمین ژن را با توجه به ژن‌های یک ماژول خاص اندازه‌گیری می‌کنند (Fuller *et al.*, 2011; معادلهٔ ۱۰).

$$k^q_{ME,i} = MM^q(i) = cor(x_i, E^q) \qquad \text{معادلهٔ (۹)}$$

$$k_{IM,i} = \sum_{u \in Module} a_{ui} \qquad \text{معادلهٔ (۱۰)}$$



شکل ۲– WGCNA: (A) دندروگرام خوشه‌های ژن‌های مشترک. شاخه‌ها و نوارهای رنگی ماژول خاص را نشان می‌دهند. (B) پلات TOM، رنگ روشن نماد همپوشانی کم است و رنگ قرمز به تدریج تیره‌تر شده است نشان دهندهٔ همپوشانی بالاتر بین ژن‌های مشترک است. بلوک‌های رنگ‌های تیره‌تر در امتداد خط مورب مربوط به ماژول‌ها هستند (Ghahramani *et al.*, 2021).

کاربرد تکنیک WGCNA

این تکنیک به دلیل کاربرد گسترده در بسیاری از علوم زیستی با مجموعه داده‌های با ابعاد و توان بالا، بیشتر در برنامه‌های ژنومی و ترانسکریپتومی مورد استفاده قرار می‌گیرد. روش WGCNA به طور گسترده برای تجزیه و تحلیل داده‌های بیانی ژن استفاده شده است (Horvath *et al.*, 2006; Langfelder *et al.*, 2013; Hartung *et al.*, 2018). به طور کلی از روش WGCNA می‌توان به عنوان یک تکنیک کاهش داده، برای خوشه‌بندی (خوشه‌بندی گامتی)، تکنیک غربال‌گری (برای مثال غربالگری ژن)، تکنیک یکپارچه‌سازی داده‌های مکمل ژنومی (بر اساس همبستگی وزنی بین متغیرهای کمی) و همچنین یک تکنیک داده کاوی یاد کرد. این روش همچنین با انتخاب ژن‌های هاب مشترک در ماژول‌ها، امکان پرداختن به تکنیک‌های متاآنالیز مبتنی بر شبکه را نیز فراهم می‌کند (Horvath, 2011; Langfelder *et al.*, 2013). روش WGCNA کاربرد گسترده‌ای در علوم زیستی و به ویژه در مورد بیماری‌ها (مطالعهٔ مکانیسم‌های بیماری، نوع بیماری)، مطالعه در مورد بافت سلولی (مطالعهٔ عملکرد طبیعی بافت‌ها جهت شناسایی پاتوژنز بیماری‌ها)، تحقیقات دارویی (شناخت اثرات داروها بر بدن)، تحقیقات تکاملی (در مورد تکامل ژن‌ها و یا اشتقاق گونه‌ها)، حاشیه‌نویسی عملکرد ژن‌ها و همچنین در علوم اعصاب (مانند تصویربرداری مغز (MRI)) دارد (Guillen-Gamez and Migallón, 2018). در این راستا روش WGCNA در سال‌های اخیر در زمینهٔ رشته مهندسی علوم دامی جهت شناسایی ژن‌ها و مسیرهای متابولیکی و سیگنالینگ مؤثر بر صفات مورد مطالعه به کار گرفته شده است. در مطالعهٔ وو و همکاران (۲۰۲۰) برای

شناسایی ژن‌های هاب در رشد فولیکول‌های مو در مرحلهٔ جنینی در بز کرکی مغولستان به روش WGCNA، تعداد ۱۰ ماژول شناسایی شد که از این تعداد یک ماژول با تعداد کل ۳,۱۶۶ ژن، با توجه به تجزیه و تحلیل الگوی بیانی ژنی، به عنوان یک ماژول خاص معرفی گردید. در مجموع ۵۸۴ ژن هاب کاندید در این ماژول با توجه به ضرایب همبستگی بین ژن‌ها، ME و اتصال ژن‌ها انتخاب و مورد بررسی قرار گرفت که در نهایت با استفاده حاشیه-نویسی ژن، ژن *WNT10A* به عنوان یک ژن کلیدی در رشد و بلوغ فولیکول‌های مو در پوست در بزهای کرکی بومی مغولستان در مرحلهٔ جنینی شناسایی شد. در تحقیقی دیگر، فرهادیان و همکاران (۲۰۲۱) از ترکیب متاآنالیز و WGCNA به عنوان رویکردی برای شناسایی ژن‌های مؤثر بر روی تولید شیر استفاده کردند که در نهایت دو مسیر متابولیکی Ubiquitin-dependent ERAD و Chaperone cofactor-dependent protein refolding به ترتیب در مقایسهٔ مراحل قبل از پیک تولید شیر نسبت به پیک و همچنین پیک تولید شیر نسبت به مرحلهٔ بعد از آن به عنوان مسیرهای متابولیکی معنی‌دار شناسایی شدند. در ادامه در روش WGCNA نیز پنج ماژول عملکردی مهم مربوط به فرآیند شیردهی شناسایی شده بود که در نهایت ژن‌های *GJA1*، *AP2A2* و *NPAS3* به عنوان ژن‌های هاب کاندید در ماژول‌های شناسایی شده معرفی شدند.

در مطالعهٔ بختیاری زاده و همکاران (۲۰۲۰) که بر روی بیماری ورم‌پستان به عنوان یک بیماری پیچیدهٔ چند عاملی انجام شده بود، با استفاده از ۱۰ نمونهٔ RNA-Seq (شامل ۵ نمونه از شیر دام مبتلا به ورم‌پستان و ۵ نمونه از شیر دام سالم) و با تکیه بر روش WGCNA، ۲۵ ماژول شناسایی شده بود که از نظر بیولوژیکی با التهاب، پاسخ ایمنی و ایجاد ورم پستان مرتبط بودند. آن‌ها گزارش کردند که از میان ژن‌های هاب شناسایی شده، ۲۵۰ ژن در شبکه‌های هم‌بیانی و هم در شبکه‌های PPI نقش‌های مهمی در پاسخ ایمنی یا مسیرهای التهابی دارند. همچنین در مطالعهٔ قهرمانی و همکاران (۲۰۲۱) نیز بر روی بیماری ورم‌پستان با استفاده از رویکرد ترکیبی متاآنالیز و WGCNA بر روی داده‌های RNA-Seq و ریزآرایه کار کرده بودند. در نتایج این مطالعه، مسیرهای سیگنالینگ پراکسیزم، گیرندهٔ شبیه *NOD*، *IL-17* و *TNF* به عنوان مسیرهای معنی‌دار در ورم‌پستان معرفی شده بودند. علاوه بر این، ژن‌های *PRDX5*، *RAB5C*، *ACTN4*، *SLC25A16*، *MAPK6*، *CD53*، *NCKAP1L*، *ARHGEF2*، *COL9A1* و *PTPRC* نیز به عنوان ژن‌های هاب در ماژول‌های عملکردی شناسایی شدند که در ادامه از الگوریتم‌های وزن‌دهی صفات (Attribute weighting algorithms) در یادگیری ماشینی نظارتی (Supervise Machine-Learning) جهت بهینه‌سازی مدل‌های پیش‌بینی و استخراج زیرمجموعه‌ای از ویژگی‌های ورودی (ژن‌ها) با حذف مواردی که حاوی اطلاعات کمی یا بدون اطلاعات بودند، استفاده گردید.

نتیجه‌گیری کلی و اُفق دید آیندهٔ روش WGCNA در ژنتیک و اصلاح نژاد دام و طیور

روش تجزیه و تحلیل شبکهٔ هم‌بیانی ژنی وزن‌دار بینش جدیدی را در مورد شبکه‌های تنظیمی پیچیدهٔ فرآیندهای بیولوژیکی ارائه می‌دهد و با توجه به قدرت تجزیه و تحلیل این روش می‌توان ژن‌های کاندید جدیدی را که ممکن است در راهبردهای اصلاح‌نژادی آینده برای بهینه‌سازی مدل‌های پیش‌بینی ارزش اصلاحی حیوانات مفید باشد، پیشنهاد کند. همچنین، این روش در ترکیب با تکنیک‌های شناسایی ژنی دیگر همچون متاآنالیز، امکان دستیابی به تجزیه و تحلیل‌هایی با وضوح بالاتر را فراهم می‌کند که می‌تواند مهم‌ترین ژن‌ها و مسیرهای متابولیکی- سیگنالینگ عملکردی را که ممکن است نشانگرهای زیستی قوی‌تری برای پیش‌بینی بهتر صفات فنوتیپی ارائه دهد. بنابراین، روش تجزیه و تحلیل شبکهٔ هم‌بیانی ژنی وزن‌دار به عنوان یک تکنیک بالقوه در شناسایی نشانگرهای زیستی معنی‌دارتر برای مطالعات آتی، می‌تواند در ژنتیک و اصلاح‌نژاد دام و طیور حائز اهمیت باشد.

منابع

باباعباسی، ب. (۱۳۹۵). "*بیوانفورماتیک سلولی و مولکولی*." پژوهشگاه رویان، پژوهشکده زیست‌شناسی و علوم پزشکی تولیدمثل جهاد دانشگاهی، مرکز تحقیقات پزشکی تولیدمثل، گروه ژنتیک، تهران، ایران: ۹-۳۱۸.

درزی، م.، گرگین، س.، مجیدزاده، ک.، و اسمعیلی، ر. (۱۴۰۰). "شناسایی ژن‌های مرتبط با پیش‌آگهی در سرطان پستان Her2-enriched با استفاده از تجزیه و تحلیل شبکه هم‌بیانی ژنی." *فصلنامه بیماری‌های پستان ایران*، ۱۴(۱)، ۴۹-۶۳.

Agrahari, R., Foroushani, A., Docking, T.R., Chang, L., Duns, G., and et al. (2018). "Applications of Bayesian network models in predicting types of hematological malignancies." *Scientific Reports*, 8(1), 1-12.

Allen, M., Wang, X., Burgess, J.D., Watzlawik, J., Serie, D.J., and et al. (2018). "Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases." *Alzheimer's & Dementia*, 14(3), 352-366.

Bakhtiarizadeh, M.R., Mirzaei, S., Norouzi, M., Sheybani, N., and Vafaei Sadi, M.S. (2020). "Identification of gene modules and hub genes involved in mastitis development using a systems biology approach." *Frontiers in Genetics*, 11, 722.

reveals novel transcription factors associated with bisphenol A dose-response." *Frontiers in Genetics*, 9, 508.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., and et al. (2008). "Functional organization of the transcriptome in human brain." *Nature Neuroscience*, 11(11), 1271-1282.

Oldham, M.C., Langfelder, P., and Horvath, S. (2012). "Network methods for describing sample relationships in genomic datasets: application to Huntington's disease." *BMC Systems Biology*, 6(1), 1-18.

Ovens, K., Eames, B.F., and McQuillan, I. (2021). "Comparative analyses of gene co-expression networks: Implementations and applications in the study of evolution." *Frontiers in Genetics*, 12.

Presson, A.P., Sobel, E.M., Papp, J.C., Suarez, C.J., Whistler, T., and et al. (2008). "Integrated weighted gene co-expression network analysis with an application to chronic fatigue syndrome." *BMC Systems Biology*, 2(1), 1-21.

Ranola, J.M., Langfelder, P., Lange, K., and Horvath, S. (2013). "Cluster and propensity based approximation of a network." *BMC Systems Biology*, 7(1), 1-20.

Silva-Vignato, B., Coutinho, L.L., Poleti, M.D., Cesar, A.S., Moncau, C.T., and et al. (2019). "Gene co-expression networks associated with carcass traits reveal new pathways for muscle and fat deposition in Nelore cattle." *BMC Genomics*, 20(1), 1-13.

Spadafora, R., Lu, J., Khetani, R.S., Zhang, C., Iberg, A., and et al. (2018). "Lung-resident mesenchymal stromal cells reveal transcriptional dynamics of lung development in preterm infants." *American Journal of Respiratory and Critical Care Medicine*, 198(7), 961-964.

Swarup, V., Hinz, F.I., Rexach, J.E., Noguchi, K.I., Toyoshiba, H., and et al. (2019). "Identification of evolutionarily conserved gene networks mediating neurodegenerative dementia." *Nature Medicine*, 25(1), 152-164.

Weston, D.J., Gunter, L.E., Rogers, A., and Wullschleger, S.D. (2008). "Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants." *BMC Systems Biology*, 2(1), 1-17.

Wu, Z., Hai, E., Di, Z., Ma, R., Shang, F., and et al. (2020). "Using WGCNA (weighted gene co-expression network analysis) to identify the hub genes of skin hair follicle development in fetus stage of Inner Mongolia Cashmere goat." *PloS One*, 15(12), e0243507.

Zhang, B., and Horvath, S. (2005). "A general framework for weighted gene co-expression network analysis." *Statistical Applications in Genetics and Molecular Biology*, 4(1). 17

Zhang, P., Li, Q., Wu, Y., Zhang, Y., Zhang, B., and et al. (2022). "Identification of candidate genes that specifically regulate subcutaneous and intramuscular fat deposition using transcriptomic and proteomic profiles in Dingyuan pigs." *Scientific Reports*, 12(1), 1-13.

Zhao, X., Wang, C., Wang, Y., Zhou, L., Hu, H., and et al. (2020). "Weighted gene co-expression network analysis reveals potential candidate genes affecting drip loss in pork." *Animal Genetics*, 51(6), 855-865.

Publisher Note
Animal Science Students Scientific Association, Campus of Agriculture and Natural Resources at the University of Tehran

Submit Your Manuscript:
https://domesticsj.ut.ac.ir/contacts?_action=loginForm

Dong, J., and Horvath, S. (2007). "Understanding network concepts in modules." *BMC Systems Biology*, 1(1), 1-20.

Farhadian, M., Rafat, S.A., Panahi, B., and Mayack, C. (2021). "Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process." *Scientific Reports*, 11(1), 1-15.

Foroushani, A., Agrahari, R., Docking, R., Chang, L., Duns, G., and et al. (2017). "Large-scale gene network analysis reveals the significance of extracellular matrix pathway and homeobox genes in acute myeloid leukemia: an introduction to the Pigengene package and its applications." *BMC Medical Genomics*, 10(1), 1-15.

Fuller, T., Langfelder, P., Presson, A., and Horvath, S. (2011). "Review of weighted gene coexpression network analysis." *In Handbook of Statistical Bioinformatics*. Springer, Berlin, Heidelberg, 369-388.

Fuller, T.F., Ghazalpour, A., Aten, J.E., Drake, T.A., Lusis, A.J., and et al. (2007). "Weighted gene coexpression network analysis strategies applied to mouse weight." *Mammalian Genome*, 18(6), 463-472.

Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A., and Dudoit, S. (Eds.). (2005). "*Bioinformatics and computational biology solutions using R and Bioconductor*." New York: Springer, 1 (0).

Ghafouri, F., Mehrabani Yeganeh, H., and Mohamadian Jeshvaghani, S. (2020). "Big data and the role of high-throughput technologies in livestock and poultry breeding". *Professional Journal of Domestic*, 20(1), 34-40.

Ghahramani, N., Shodja, J., Rafat, S.A., Panahi, B., and Hasanpur, K. (2021). "Integrative systems biology analysis elucidates mastitis disease underlying functional modules in dairy cattle". *Frontiers in Genetics*, 12.

Horvath, S. (2011). "*Weighted network analysis: applications in genomics and systems biology*." Springer Science & Business Media.

Horvath, S., and Dong, J. (2008). "Geometric interpretation of gene coexpression network analysis." *PLoS Computational Biology*, 4(8), e1000117.

Horvath, S., Zhang, B., Carlson, M., Lu, K.V., Zhu, S., and et al. (2006). "Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target." *Proceedings of the National Academy of Sciences*, 103(46), 17402-17407.

Langfelder, P., Luo, R., Oldham, M.C., and Horvath, S. (2011). "Is my network module preserved and reproducible?." *PLoS Computational Biology*, 7(1), e1001057.

Langfelder, P., Mischel, P.S., and Horvath, S. (2013). "When is hub gene selection better than standard meta-analysis?." *PloS One*, 8(4), e61505.

Langfelder, P., Zhang, B., and Horvath, S. (2008). "Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R." *Bioinformatics*, 24(5), 719-720.

Li, J., Zhou, D., Qiu, W., Shi, Y., Yang, J.J., and et al. (2018). "Application of weighted gene co-expression network analysis for data from paired design." *Scientific Reports*, 8(1), 1-8.

Liu, W., Li, L., Ye, H., and Tu, W. (2017). "Weighted gene co-expression network analysis in biomedicine research." *Sheng wu Gong Cheng xue Bao= Chinese Journal of Biotechnology*, 33(11), 1791-1801.

Maertens, A., Tran, V., Kleensang, A., and Hartung, T. (2018). "Weighted gene correlation network analysis (WGCNA)

Domestic; 22 (2): 5-13

Scientific-Extensional (Professional) Journal

Domestic

Animal Science Students Scientific Association at the University of Tehran, Autumn 2022

https://domesticsj.ut.ac.ir/article_90439.html

Scientific-Extensional Article

Method of weighted gene co-expression network analysis and its application in animal and poultry breeding and genetics

Vahid Dehghanian Reyhan[1], Mostafa Sadeghi[2*] and Farzad Ghafouri[1]

[1] Ph.D. Student of Animal and Poultry Breeding & Genetics, Department of Animal Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Alborz, Iran

[2] Associate Professor of Animal Breeding and Genetics, Department of Animal Science, College of Agriculture and Natural Resources, University of Tehran, Karaj, Alborz, Iran

https://doi.org/10.22059/domesticsj.2022.344763.1098

Abstract

Among the different types of networks, gene co-expression networks have the most flexibility to study different traits such as functional and reproductive traits, diseases, etc. Gene co-expression generally refers to the correlation between genes at the levels of transcripts; on the other hand, it can also be used at all biological scales (such as proteins, metabolites, or in combination between transcripts, proteins, and metabolites) to study the correlation relationships between genes. Co-expression networks have become popular in part because of the use of technologies such as microarrays, RNA-Seq, and mass spectrometry, as they allow the study of molecular mediators at different biological scales in a simple and relatively large number of samples. In addition, using this method, it is possible to measure the biological co-expression of genes simultaneously in specific cell types. By comparison, most protein-protein interaction (PPI) networks merely indicate general interactions between genes that do not refer to cell type and gene expression time, while gene expression networks can be reconstructed using data obtained from specific cell types from different individuals (upstream and downstream about a phenotypic trait such as individuals with high- and low-fertility) and throughout the developmental stages. One of the most widely used algorithms for constructing gene co-expression networks is weighted gene co-expression network analysis (WGCNA) due to its widespread use in many co-expression studies that will be instructive to explain how it works. As a result of identifying modules, genes, and metabolic-signaling pathways associated with various studied traits using the WGCNA method, they may show new insights into molecular mechanisms. Consequently, the purpose of this study was to provide a brief description of the method of weighted gene co-expression network analysis and its application in animal and poultry breeding and genetics.

Keyword(s): WGCNA, Transcriptome, Biological system, Gene co-expression networks

***Corresponding Author E-mail:** sadeghimos@ut.ac.ir

Section: Animal and Poultry Breeding & Genetics **Associate Editor:** Dr. Masoumeh Naserkheil

Received: 26 Jun 2022 **Revised:** 16 Jul 2022 **Accepted:** 28 Jul 2022 **Published online:** 06 Dec 2022

Citation: Dehghanian Reyhan, V., Sadeghi, M., Ghafouri, F. Method of weighted gene co-expression network analysis and its application in animal and poultry breeding and genetics. *Professional Journal of Domestic*, 2022; 22(2): 5-13.