

مروری نوین بر فناوری ریز آرایه: مبانی تئوریک، ساختار داده خام، تحلیل‌های آماری و تفسیر نتایج

آرش جوانمرد^{۱*}، کریم حسن پور^۱، محمد سلیمان اختیاری^۲، آیه سادات صدر^۳، زهرا رودباری^۴، پویا مطیع نوپور^۵،

فرزاد غفوری^۶

شماره صفحات

۵-۳۰

(۱) استادیار ژنتیک و اصلاح نژاد دام، گروه مهندسی علوم دامی، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران.

(۲) دانش آموخته کارشناسی ارشد، بیوشیمی بالینی، دانشکده دامپزشکی، دانشگاه تبریز، تبریز، ایران.

(۳) استادیار ژنتیک و اصلاح نژاد دام، پژوهشکده آبی پروری آب‌های جنوب کشور، مؤسسه تحقیقات علوم شیلاتی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، اهواز، ایران.

(۴) استادیار ژنتیک و اصلاح نژاد دام، گروه مهندسی علوم دامی، دانشکده کشاورزی، دانشگاه جیرفت، جیرفت، ایران.

(۵) دانشجوی دکتری تخصصی، گروه بیوتکنولوژی، دانشکده کشاورزی، دانشگاه تربیت مدرس، تهران، ایران.

(۶) دانشجوی دکتری تخصصی ژنتیک و اصلاح نژاد دام و طیور، گروه مهندسی علوم دامی، پردیس کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.

*نویسنده مسئول: arash_707@yahoo.com

تاریخ پذیرش: ۱۴۰۱/۰۵/۱۰

تاریخ دریافت: ۱۴۰۱/۰۲/۰۶

چکیده

دستیابی به فناوری‌هایی که بتواند رفتارهای ژن‌ها در سطح رونوشت سلولی و اثر متقابل بین آن‌ها را به تصویر بکشد، مطلوبیت ویژه‌ای دارد. در همین راستا، فناوری ریزآرایه، بر مبنای کشفیات سابق واکنش پلی‌مراز کمی سازی شده (qPCR)^۱ و فناوری هیبریداسیون ساترن بلاتینگ^۲ می‌باشد که از همان قوانین آما، در مقیاس بزرگ‌تر برای اندازه‌گیری میزان بیان چندین هزار ژن استفاده می‌کند. در واقع، روشی مبتنی بر هیبریداسیون^۳ در ترکیب با فناوری نانو و استفاده از رنگ‌های فلورسنت نشان‌دار شده می‌باشد. این فناوری بر مبنای تراشه و صفحه شیشه‌ای است که توالی‌های شناساگر خاص در یک نظم دو بُعدی به تعداد چند ده هزار واحد در کنار هم قرار گرفته و به‌طور هم‌زمان شناسایی الگوی بیان افتراقی چندین هزار توالی خاص یا مناطق کد شونده را عهده‌دار می‌باشد. پیش‌آگاهی از جزئیات، معماری ریزآرایه و پشت‌صحنه مورد استفاده برای این تکنیک بر روی تجزیه و تحلیل داده‌های خام تأثیر مستقیم دارد. روش‌های پیش‌پردازش داده‌ها و آزمون‌های نرمال‌سازی آن‌ها، استفاده از روش تجزیه مؤلفه‌های اصلی، رسم نمودار حرارتی، خوشه‌بندی و رسم دندروگرام تست همبستگی داده و روش پیشرفته TSN از جمله روش‌های پیشرو هستند. همچنین، چالش مقایسه میانگین چندگانه، وجود خطای نوع یک و مفهوم شاخص چند برابر شدن (Fold change) نیز به‌طور اجمالی مورد بحث قرار می‌گیرد. در مطالعه مروری حاضر، سعی شده است که مبانی تئوریک، ساختار داده‌های خام، تحلیل‌های آماری و تفسیر نتایج فناوری ریزآرایه مورد بحث قرار گیرد و به‌طور مستقل قادر باشد مسیر داده تا تفسیر و همچنین چالش‌های موجود را با اطلاعاتی که داده می‌شود به خوبی پشت سر قرار دهد.

کلمات کلیدی: پیش‌پردازش داده‌ها، تحلیل آماری، ترانسکریپتوم، ریزآرایه و نرمال‌سازی داده‌ها.

¹ Quantitative PCR

² Southern blotting

³ Hybridation

⁴ Clustering

مقدمه

بیان ژن فرآیندی است که در آن اطلاعات موجود در توالی ژنوم به محصولات همچون، RNA غیرکدکننده، اسید آمینه و پروتئین، تبدیل می‌شوند. تنظیم ژن در واقع فرآیند نظارتی است که به سلول در بافت مربوطه اجازه می‌دهد پروتئین‌های مورد نیاز را در زمان مناسب و به‌مقدار لازم تولید کند. فرآیند، بیان ژن از مهم‌ترین اصول در تکامل و تمایز صفات یوکاریوت‌ها می‌باشد. بنابراین، در یک سلول، الگوی بیان شده چندین هزاران ژن خاص تعیین می‌کند که آن سلول چه وظیفه‌ای را می‌تواند در بدن انجام دهد. در واقع، بیان ژن، یک فرآیند دقیق تنظیم شده است که به سلول اجازه می‌دهد در محیط‌های مختلف از خود واکنش مناسبی را نشان دهد (Kulkarni et al., 2019). این فرآیند، را به‌عنوان کلید "روشن و یا خاموش کردن ژن" نیز معرفی می‌کنند که هدف از آن کنترل زمان ساخت پروتئین‌ها، کنترل حجم یا میزان تولید پروتئین‌های مورد نیاز بدن و همچنین، افزایش یا کاهش بیان آن‌ها است. بررسی سطح رونوشت سلولی^۵ به درک چگونگی عملکرد مجموعه ژنی با یک‌دیگر در ایجاد مسیرهای متابولیکی و تنظیمی و همچنین پیام‌رسانی درون سلولی کمک می‌کند؛ از طرف دیگر، الگوهای هم‌بیانی و ژن‌های دارای اثر متقابل را آشکار می‌سازد که این باعث تعیین عملکرد ژن‌ها می‌گردد. سلول، یک سیستم پیچیده است که باید به انواعی از تغییرات شرایط داخلی و خارجی به‌منظور عملکردهای بی‌شمار خود پاسخ دهد (Asyali et al., 2006). تنظیم رونویسی یک ژن، شامل مجموعه پاسخی است که به طیف وسیعی از شرایط داخل سلولی و خارج سلولی پاسخ می‌دهد. رمزگشایی پروفایل سطح رونوشت سلولی (زمان و محل بافت و همچنین چگونگی خاموش و روشن شدن ژن‌ها)، یک هدف اصلی بیولوژیکی است که با اندازه‌گیری سطح بیان ژن‌ها تحت شرایط مختلف سلولی به این هدف کمک می‌کند. علم مطالعه سطح رونوشت سلولی یکی از زمینه‌های پیشرفته در زمینه علوم بعد از ژنوم است. در واقع، سطح رونوشت سلولی به مجموعه نسخه‌های RNA در نوع خاصی از یک سلول یا بافت در شرایط ویژه‌ای از رشد و یا تحت شرایط فیزیولوژیکی خاص اطلاق می‌شود که شامل RNAهای پیام بر^۶ RNAهای انتقال دهنده^۷ RNA ریبوزومی و سایر RNAهای غیر کدکننده^۸ می‌شوند (Kulkarni et al., 2019).

در واقع، یک سیستم حیاتی، حاکی از وجود یک مسیر برای جریان انتقال اطلاعات ژنتیکی است، به این صورت که DNA باعث، ساخت RNA شده و متعاقباً، ترجمه RNA منجر به تولید پروتئین می‌شود. انطباق این مسیر انتقال اطلاعات ژنتیکی بر فرآیند تمایز سلولی رمزهای جالبی را از پدیده حیات در یک سیستم حیاتی، (انسان، دام، گیاه و حتی میکروارگانیسم‌های ابتدایی) رمزگشایی می‌کند. به‌عنوان مثال، در یک موجود پیشرفته مانند انسان یا دام، قرار گرفتن سلول‌های یکسان در یک بافت و گاهاً، در بافت‌های مختلف نسبت به هم با وظایف کاملاً اختصاصی همراه است که در مجموع، یک سیستم حیاتی را سرپا نگه

⁵ Gene Expression

⁶ Transcriptome

⁷ Messenger RNA

⁸ Transfer RNA

⁹ Non-coding RNA

می‌دارد و اینکه چه علتی باعث شده است که از یک سلول تخم به تنهایی، با محتوای ژنوم نیمی از والد نر و نیمی از والد ماده بتواند خاصیت ایجاد یک موجود کامل را داشته باشد و در دوران جنینی به سوی یک بافت خاص از پرده‌های جنینی تمایز یابد (شکل ۱) (Andrews, 2004). به نظر می‌رسد که، یکی از دلایلی که سلول‌های اختصاصی هر بافت در یک موجود وظیفه متفاوتی برعهده می‌گیرند و ساختار عملکردی اختصاصی پیدا می‌کنند، وقوع جهش‌های مختلف در ژنوم بافت‌های مختلف باشد (Kulkarni *et al.*, 2019). اما، امکان وقوع جهش‌های زیاد به علت برهم زدن ساختار فراکتال کل مجموع بعید به نظر می‌رسد و نظریه غالب این است که پروفایل بیان ژن در بافت‌های مختلف متفاوت می‌باشد.

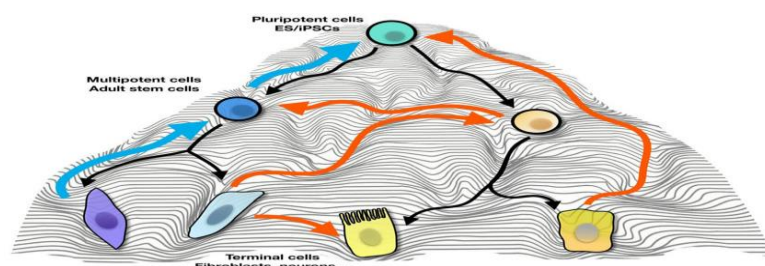


شکل ۱: نمایی کلی از فرآیند تمایز سلولی از سلول‌های مورولا و بلاستوسیت (به وجود آمدن سلول‌ها و بافت‌های مختلف از یک سلول) (Hardy, 1999; Hansen and Tribulo, 2019)

Figure 1- General view of the process of cell differentiation from merula cells and blastocyst (the formation of different cells and tissues from a cell) (Hardy, 1999; Hansen and Tribulo, 2019)

مدل پیشنهادی ولینگتون^۱ برای تمایز سلولی و نقش فرآیندهای اپی ژنتیک

در خصوص تمایز سلولی و فرآیندهای جنینی، جنین شناس معروف بنام پرفسور ولینگتون، مدلی را برای تحلیل ایجاد سلول‌ها با عملکرد اختصاصی در بافت‌های یک موجود بیان کرد که طی این مدل جهش در ژنوم نمی‌تواند عامل اصلی تمایز باشد؛ بلکه، بسیاری از سلول‌های یک ارگانیسم حاوی ترکیب ژنوم با ساختار یکسان می‌باشد. اما، پدیده‌های اپی ژنتیک و پروفایل بیان متفاوت در دو سلول متفاوت متعلق به دو بافت اختصاصی و فرآیند اپی ژنتیکی متفاوت در سطح هیستون‌ها می‌تواند یک عامل منطقی برای توجیه این تفاوت‌ها باشد (شکل ۲) (Nelson, 1992).



شکل ۲: مدل پیشنهادی برای تمایز و تأثیر اپی ژنتیک و بیان افتراقی ژن‌ها (Botchkarev *et al.*, 2012)
Figure 2- The proposed model for differentiation and epigenetic effect and differential expression of genes (Botchkarev *et al.*, 2012)

¹ Merula cells and blastocyst ⁰

¹ Waddington ¹

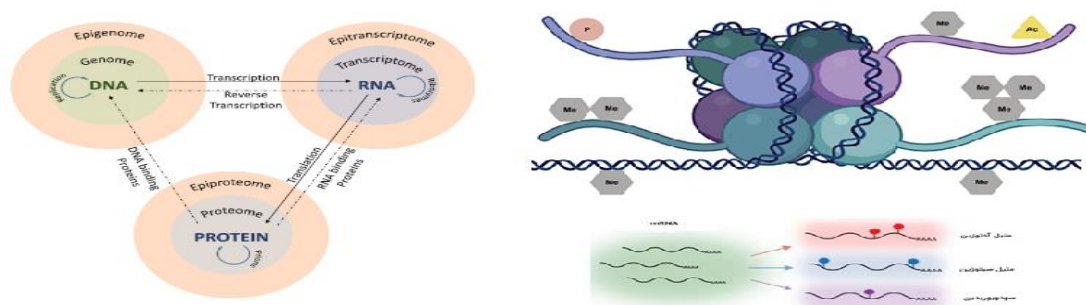
در این مدل پیشنهادی، بیان شده است که اگر یک سلول را مثل یک گویچه دوار در نظر گرفت که از یک مسیر با شیب خاص از بالا رها سازی شده است، این گویچه می‌تواند، به تصادف در پایین شیب به چند راهی برسد و هر چه در پایین سطح شیب‌دار به چند راهی‌های منشعب‌تر برسد، به‌علت بیان متفاوت در سطح ژنوم می‌تواند عملکرد متمایزتری از سلول‌های بنیانگذار و اجدادی اولیه پیدا کند. در واقع، فرآیندهای اپی‌ژنتیک و تغییراتی که در هیستون‌ها اتفاق می‌افتد به طور هدفمند قادر است از این خاصیت توتی پوتنسی (قدرت تولید و ایجاد یک ارگانیسم از یک جزء آن) سلول‌های جنینی اولیه استفاده کند و بر این اساس، هر سلول از سه لایه جنینی را به ساخت بافت‌های تخصصی ویژه و با عملکرد از پیش تعیین شده هدایت می‌کند (Cooper, 1990).

تصویر سازی کلی از ساختار ژنوم پستانداران

ژنوم پستانداران، شامل ۲۰ تا ۲۵ هزار ژن است که در شماره کروموزومی متفاوت مستقر هستند. این ساختار ژنی و در مجموعه سلول‌های مختلف و بافت‌های بدن تشابه بالایی دارند. اما، الگوی بیان افتراقی و پروفایل ترانسکرپتوم و همچنین اثر متقابل بین ژن‌ها در دو بافت متفاوت تا حدودی می‌تواند متفاوت باشد.

هیستون‌ها و نقش اپی ژنتیک بر تنظیم بیان ژن در یک بافت اختصاصی

استیلاسیون پروتئین‌های هیستونی یکی از مهم‌ترین فرآیندهای اپی‌ژنتیکی است که به‌منظور تنظیم بیان ژن‌ها رخ می‌دهد. کروماتین به واسطه اتصال گروه استیل به دنباله هیستونی نوکلئوزوم‌های خود، رشته DNA را در دسترس فاکتورهای رونویسی و دیگر پروتئین‌های تنظیم کننده بیان ژن قرار می‌دهد (شکل ۳) (Mazzone *et al.*, 2019).



شکل ۳: نقش هیستون‌ها و اپی ژنتیک در روشن و خاموش شدن بیان ژن (Singh *et al.*, 2019)

Figure 3- The role of histones and epigenetics in turning gene expression on and off (Singh *et al.*, 2019)

جزئیات تغییرات اپی‌ژنتیک شامل مکانیسم‌های مولکولی همچون استیلاسیون، متیلاسیون، فسفوریلاسیون و ADP ریبولوزیلاسیون است که مهم‌ترین آن‌ها استیلاسیون می‌باشد. به یاد داشته باشیم که فعال شدن یا غیر فعال شدن فرایند بیان ژن‌ها پستانداران، توسط آنزیم‌ها تغییردهنده مبتنی بر بازسازی کروماتین، هیستونی و یا کمپلکس تغییر دهنده کروماتین وابسته

به ADP صورت می‌گیرد. فاکتورهای مانند انتقال‌دهنده استیل به اسید آمینه لیزین موجود در پروتئین (KATs) و فاکتورهای دیالاستیل‌ه کردن هیستون‌ها (HDACs) که استیل‌اسیون برگشت‌پذیر شناخته شده‌ترین و هیستون‌های را کاتالیز می‌کنند، قابل فهم‌ترین آنزیم‌های تغییر دهنده هیستونی هستند که در ارتباط با مکانیسم‌های جذب و تنظیم عناصر ژن‌ها و نقش آن‌ها در رونویسی هستند. فاکتورهای فعال‌کننده رونویسی به‌طور هم‌زمان با KATs فعال می‌شوند و در حالی که فاکتورهای سرکوب‌سازی رونویسی با فعالیت HDACs شروع به کار می‌کنند (Singh *et al.*, 2019).

باید توجه داشت که در بین ۳۰ الی ۳۳ هزار ژن تشکیل دهنده ژنوم یک موجود، در بحث تمایز، علاوه بر اینکه ساختار ژنوم و جهش‌ها در سطح بافت‌های مختلف و کل سلول‌ها یک ساختار محافظت شده است، بیان ژن و الگوی بیان هم در کل مجموعه از یک ساختار فراکتال و حفاظت شده تبعیت می‌کند و تفاوت‌ها به تعداد محدود ژن ختم می‌شود. در واقع اثرات متقابل و هم‌بیانی بین ژن‌ها است که بیان افتراقی در تعداد محدودی از ژن‌ها را ایجاد می‌کند و آن نیز منجر به تخصص در عملکرد یک بافت می‌شود. به‌عنوان یک مدل ساده، برای تصویر سازی این مفاهیم، اگر در ساده‌ترین حالت دو ژن A و B را در نظر بگیریم چهار رابطه بین این دو ژن متصور می‌شویم که یا هر دو اثر سینرژیک یا هم‌بیانی دو حالت (کم و زیاد) یا آنتاگونیسم دو حالتی یکی کم دیگری زیاد دارد و این حالت را اگر به ۳۰ الی ۳۳ هزار تعمیم دهیم می‌توانیم داستان کل و تصویر سازی در این خصوص را تجسم داشته باشیم (Retis-Resendiz *et al.*, 2021).

در واقع، غلظت RNA در هر بافت می‌تواند با بافت دیگر، به‌علت تفاوت در اپی‌ژنتیک و نواحی هتروکروماتینی و یوکروماتینی متفاوت باشد، یعنی، به بیان ساده‌تر ۲۵ هزار ژن مستقر در بافت تولیدمثلی با ۲۵ هزار تا ژن در بافت روده و دستگاه گوارش می‌تواند الگوی متفاوتی را نشان دهند و با دو نوع مبانی روبرو خواهیم بود که یک نوع تفاوت ذاتی در پروفایل بیان در بین بافت‌ها می‌باشد و یک نوع دیگر تأثیر ایجادکننده تفاوت بیان عوامل محیطی همچون بیماری، تغذیه و استرس بر تغییر این الگوی بیان و تغییر ژن‌های خاص می‌باشد (Yu and Ren, 2017). به عبارت دیگر عوامل خارجی و محرک‌های محیطی می‌توانند باعث تغییر در مکانیسم بیان و ایجاد حالتی همچون بیان کم یا زیاد و یا حتی عدم بیان شوند (Zhou *et al.*, 2011). پس از دیدگاه یک متخصص اصلاح دام یا گیاه بررسی تأثیر محرک‌هایی مانند بیماری، استرس‌های محیطی از جمله شوری، کم آبی و خشک‌سالی بر ترانسکریپتوم، متیل‌اسیون و آسیلاسیون هیستون‌ها می‌تواند بیومارکرهایی را در حیوانات مقاوم یا حساس و یا در انسان‌ها میان انسان‌های سالم و بیمار (مانند سرطان یا یک بیماری با طیف شیوع بالا) معرفی کند (Hon *et al.*, 2009).

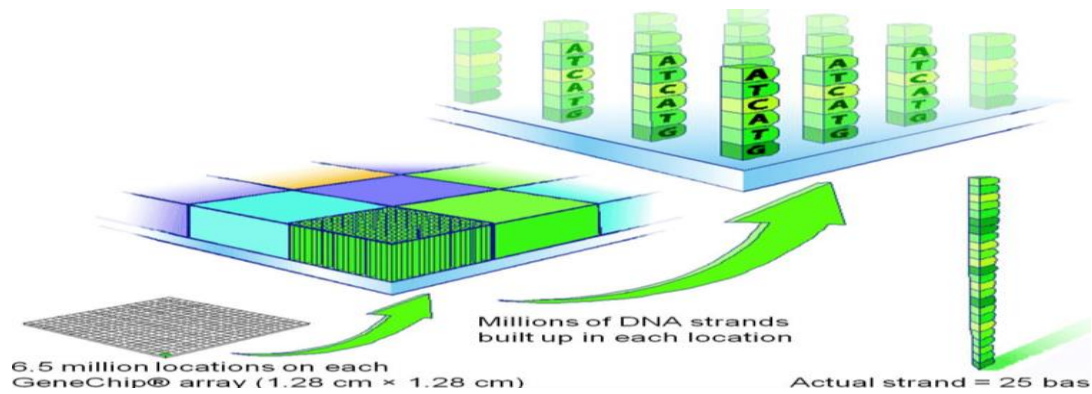
مروری بر فناوری qPCR

رمزگشایی از معماری ژنتیکی و سطح رونوشت برداری سلولی و یا سیستم حیاتی یک فرد که برتری در عملکرد در یک زمینه اختصاصی دارد، می‌تواند دارای اهمیت باشد. به‌عنوان مثال، یک فرد برتر نسبت به سایر افراد که در مواجهه با بیماری تحت تأثیر قرار نمی‌گیرد و دارای سیستم ایمنی است. در واقع، این فرد به محرک‌های محیطی مخرب عکس‌العمل قوی نشان می‌دهد و

می‌توان از اهمیت و سیمای مولکولی ژنوم و نسخه‌برداری آن و نقاط تنظیمی این موضوع پرده‌برداری کرد. لذا، ابزارهای مولکولی که بتواند اندازه‌گیری غلظت RNA را در یک بافت اختصاصی کمی‌سازی کند، ایده‌آل خواهد بود. پس ساده‌ترین حالت ممکن اینست که علت عملکرد خوب اختصاصی یک بافت در انجام وظیفه خود را به‌میزان غلظت، تعداد نسخه‌برداری و سرعت نسخه‌برداری آن مرتبط دانست و اگر بتوان میزان غلظت RNA را اندازه گرفت، در واقع به هدف اصلی دست یافته‌ایم (Hong *et al.*, 2021). رفتار بیانی یک ژن در شرایط مختلف می‌تواند تغییر کند و در میان ژن‌های ساختاری که همیشه یک بیان بدون واریانس دارند (ژن‌های مرجع)، بعضی ژن‌های کاندیدا تحت تیمارهای خاص و تغذیه افتراق معنی‌داری دارند و در روش‌های اولیه کمی‌سازی RNA، چون موانعی طبیعی همچون طول عمر کم و ناپایداری، تک رشته بودن، وجود باز اوراسیل و نواحی کدکننده و غیرکدکننده وجود دارند و رنگ‌های فلئورسنت نیز به تک رشته اتصال پیدا نمی‌کنند، ناچار به پیاده‌سازی روشی تحت عنوان نسخه‌برداری معکوس و تولید cDNA می‌شویم که در مرحله بعد بتواند ساب‌گرین و پروب‌ها را با این ساختار اثر متقابل دهد. مطمئناً وجود RNA زیاد در یک بافت می‌تواند منجر به تولید cDNA و به‌دنبال آن تولید رنگ بیش‌تر و ارتفاع گرفتن زودتر از موعد منحنی سیکل آستانه یا Cycle of threshold (Ct) شود. در واقع وجود سیکل آستانه کم رابطه مستقیم با بیان بالا دارد و ژن‌های مرجع هم وسیله‌ای برای نرمال‌سازی داده‌های خام می‌باشند. در خروجی‌های این روش پایه، چون تعداد ژن‌های محدودی وجود دارد، عملاً، داده خام همان شکل عددی سیکل آستانه می‌باشد که وارد محاسبات آماری t تست و شاهدهی و موردی می‌شود. اما، یک ژن هیچ وقت نمی‌تواند معرف و حلال مشکل سیمای کلی ترانسکریپتوم باشد و نیاز به روش‌های پیشرفته و مبتنی بر نانو همواره احساس می‌گردد. ترانسکریپتوم بیش از این که رصد عملکرد انفرادی یک ژن خاص باشد، نگاه به یک سیستم و اجرای پیچیده در این سیستم ساده است (Laikhter, 2015).

فناوری ریزآرایه

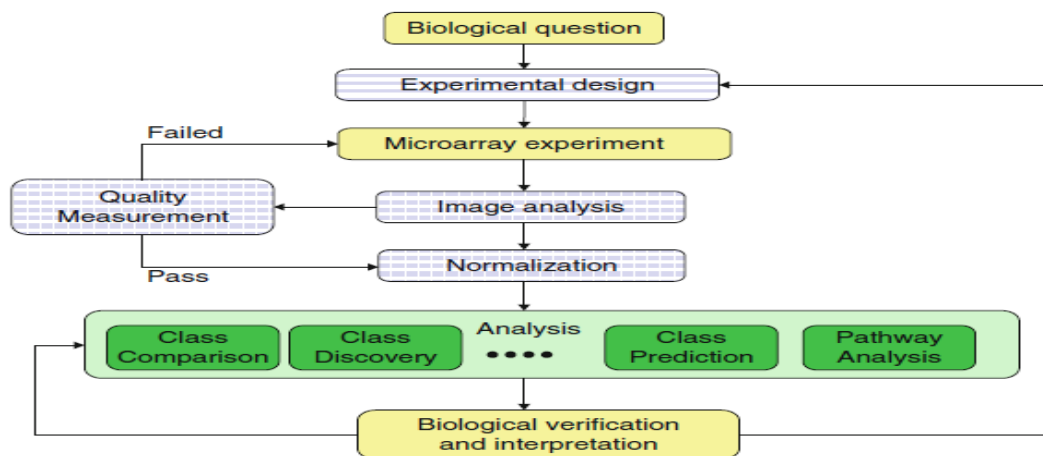
همانطور که در مباحث قبل اشاره شد، در روش RT_PCR بررسی بیان نسبی بر اساس سیکل آستانه است و نگاه انفرادی به یک سیستم پیچیده با ابعاد یک تک ژن نمی‌تواند نگاهی درست و قضاوتی عادلانه دربر داشته باشد. در همین راستا، فناوری ریزآرایه بر مبنای قوانین qPCR و فناوری هیبریداسیون ساترن بلاتینگ می‌باشد که در مقیاس بزرگ‌تر برای اندازه‌گیری میزان بیان تعداد متعددی ژن مورد استفاده قرار می‌گیرد. همچنین این روش مبتنی بر هیبراسیون در ترکیب با فناوری نانو و استفاده از رنگ‌های فلئورسنت نشان‌دار است. این فناوری بر مبنای چیپ و صفحات شیشه‌ای (اسلایدهای میکروسکوپی) است که تعداد چند ده هزار از توالی‌های شناساگر خاص در یک نظم دو بعدی کنار هم قرار گرفته‌اند و به‌طور هم‌زمان شناسایی الگوی بیان افتراقی چندین هزار توالی خاص یا مناطق کدکننده را عهده‌دار است. در این روش جایگزین، در یک میکروچیپ سیلیکونی $1/5 \times 1/5$ پیکسل‌های مختلف وجود دارد و هر پیکسل شامل میلیون‌ها پروب است که هر پروب بیانگر یک ژن بوده و هر پیکسل با پیکسل بعدی توالی متفاوتی دارد. تعداد پیکسل بسته به نوع چیپ و گونه مورد مطالعه می‌تواند، متفاوت باشد (Heller, 2002).



شکل ۴: نمایی کلی از اصول حاکم بر ریزآرایه و جزئیات آن. در خرد کردن باید cDNA

در حد و حدود طول پروبها باشد (Russell *et al.*, 2008).

Figure 4- An overview of the principles governing the microarray and its details. In chopping, the cDNA should be about the length of the probes (Russell *et al.*, 2008).



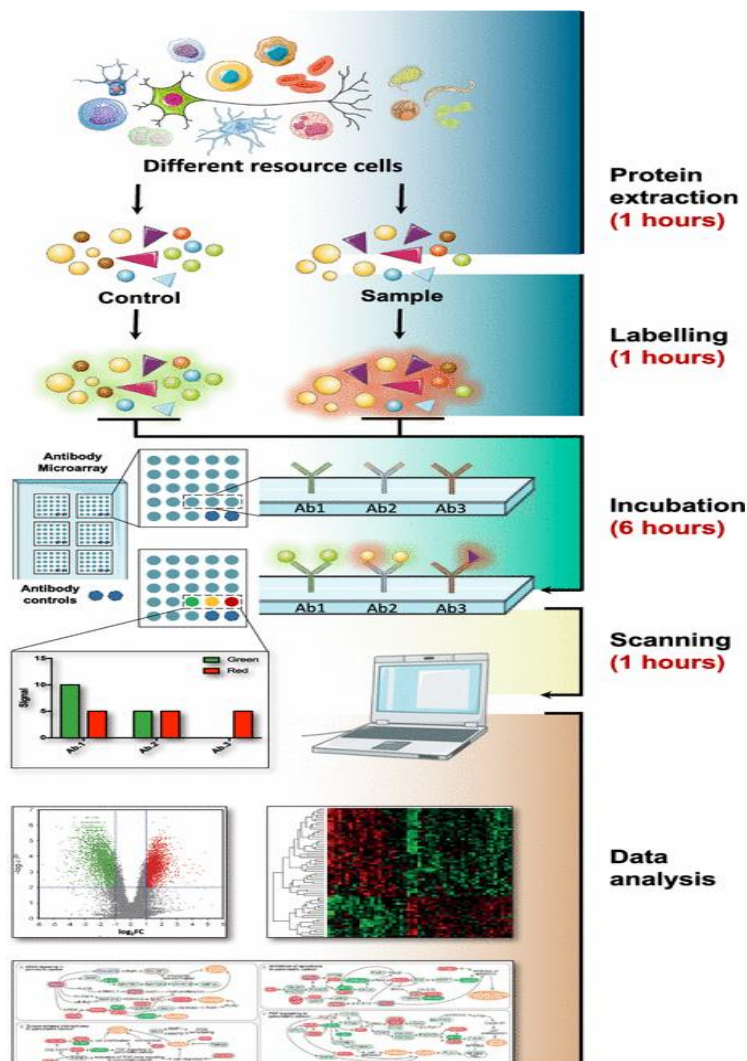
شکل ۵: خلاصه‌ای از کلیات طراحی یک مطالعه بیان ژن و بررسی الگوهای افتراقی معنی‌دار مبتنی بر ریزآرایه (Heller, 2002)

Figure 5- Summary of the general design of a gene expression study and investigation of significant differential patterns based on microarray (Heller, 2002)

نحوه هیبریداسیون و تشخیص در سطح بیان ژن

روش‌های ریزآرایه را می‌توان به روش دو کاناله (قدیمی) و تک کاناله (امروزی) دسته‌بندی نمود. تفاوت این دو روش در رابطه با این است که آیا نمونه بیمار و سالم با یک رنگ متفاوت (در دو کاناله) یا با یک رنگ یکسان (تک کاناله) برچسب می‌خورد، متمایز می‌شوند (Ehrenreich, 2006). در روش امروزی یا تک کاناله، نمونه‌های سالم و سرطانی با هم یک رنگ مشترک فلورسنت می‌گیرند. این روش اگر چه سرعت پردازش کمی دارد، اما دقت آن بالا است و حتی با تکنولوژی‌هایی مانند RNA-Seq هم می‌تواند در رقابت باشد. در روش دو کاناله ابتدا از نمونه‌های سلول سرطانی و سالم استخراج RNA انجام می‌شود و سپس تبدیل به cDNA دو رشته صورت می‌گیرد. در مرحله بعد سپس رشته واسرشت سازی شده و به رشته‌های یکسانی تقسیم می‌شوند. در روش دو کاناله از دو رنگ مجزا استفاده می‌شود. به عنوان مثال، به محتوای سلول سالم رنگ سبز ۵۲۳ نانومتر و به محتوای سلول سرطانی رنگ قرمز ۶۵۳ نانومتر اختصاص داده می‌شود. سپس، محلول‌ها روی یک میکروچیپ فلوسل ریخته شده

و یک شبانه روز گذشت میزان لازم است. سپس، هیبریداسیون بین محتوای نمونه‌ها و پروب‌های ژنی از قبل قرار داده شده روی چیپ با لیزر تصویربرداری می‌شود.



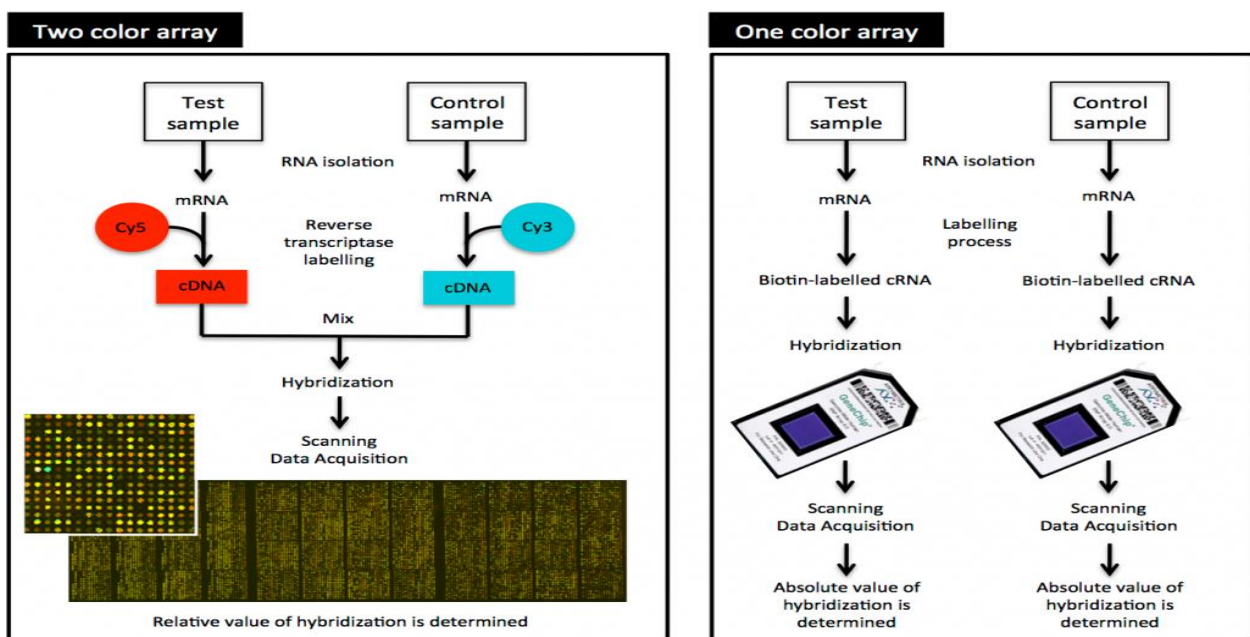
شکل ۶: نمای کلی از خروجی‌های یک مطالعه بیان ژن مبتنی بر ریزآرایه (از ابتدا تا پایان راه) (Ramakrishnan *et al.*, 2005)

Figure 6 - Overview of the outputs of a microarray-based gene expression study (from start to finish)

(Ramakrishnan *et al.*, 2005)

در این حالت، هر چه مقدار RNA بیشتر باشد، هیبریداسیون بیشتر و پیکسل بیش‌تری رنگ می‌گیرد و در مرحله بعد، بر اساس پردازش تصویر، نتایج کمی‌سازی می‌شود. درای مرحله تفسیر به راحتی انجام می‌شود به گونه‌ای که هر چه هیبریداسیون بیشتر باشد، سینگال دریافتی بیشتر خواهد بود و بنابراین، سطح بیان نیز بالاتر می‌باشد. روش دو کاناله هر چند زودتر جواب می‌دهد، اما، احتمال بروز خطا در آن بیشتر است، چون لیزر وقتی نمونه‌ها را با دو رنگ مجزای رنگ‌آمیزی شده تشخیص می‌دهد، ممکن است خوانش آن با خطا مواجه شود. در این پازل در واقع ژنی که در هیچ کدام یک از سلول‌ها بیان نشده باشد با رنگ سیاه، ژنی که در هر دو سلول بیان شده باشد با رنگ زرد، ژنی که تنها در سلول‌های سالم بیان شده باشد با رنگ سبز و ژنی که در سلول سرطانی بیان شده باشد با رنگ قرمز نشان داده خواهد شد. رنگ سیاه در واقع مخلوطی از سبز و قرمز است و نشان

می‌دهد که محصولات سلول سالم و بیمار هر دو به یک اندازه به پروب‌های آن ژن‌های ساختاری متصل شده است. دقت شود که قطعات خرد شده می‌تواند به مکان نادرست اتصال پیدا کند و اختصاصی عمل نکند. در این حالت، فرآیند شستشو انجام می‌شود که تا حدودی اتصالات غیر اختصاصی را حذف می‌کند، اما در هر صورت همیشه این احتمال به صفر نمی‌رسد و باز غیر اختصاصی بودن یک مورد اجتناب ناپذیر است و در نتیجه، باید رویهٔ پیش پردازش و تصحیح پیش زمینه داده‌های میکروآرئی روی آن اعمال شود (Heller, 2002; Ehrenreich, 2006). در فناوری ریزآرایه، معماری چیپ و فلسفه تولید و پیدایش آن خیره‌کننده می‌باشد. به گونه‌ای که در صفحه فلوسل نانویی، برای یک ژن در موقعیت جغرافیایی و موقعیت‌های مکانی مختلف، کاوشگر خاص آن ژن به‌طور مکرر ریخته شده است، بنابراین موضوع تکرار تکنیکی رعایت شود. اگر بنا به دلایلی بافر هیبریداسیون به‌طور یکنواخت روی صفحه پخش نشود، حباب تشکیل شود و یا صفحه شیب نسبی پیدا کند و به بعضی پیکسل‌ها بیش‌تر مواد برسد، در آن صورت نتایج بیان ژن به‌علت هیبریداسیون کاذب نشان داده خواهد شد. بنابراین به‌طور خلاصه در بحث ریزآرایه دو رویه پردازش و تصحیح پیش زمینه داده‌های میکروآرئی و بحث تکرار بیولوژیکی مطرح است (Botchkarev *et al.*, 2012).



شکل ۷: تفاوت تکنیک‌های ریزآرایه دوکاناله (قدیمی) با میکروآرئی تک کاناله (جدید) و رنگ‌آمیزی آن‌ها (Hall *et al.*, 2007)

Figure 7- Difference between two-channel microarray techniques (old) and single-channel microarray (new) and their staining (Hall *et al.*, 2007)

موضوع دیگری که در بحث ریزآرایه دارای اهمیت است، وجود حداقل ۳۰ پروب برای یک ژن در موقعیت‌های مختلف در سطح پروب می‌باشد و در نهایت، برای این ژن با این حجم تکرار و جواب در سلول‌های سالم و بیمار میانگین گرفته می‌شود و این میانگین‌گیری جلوی وقوع بسیاری از خطاهای ناشی از وجود حباب و عدم یکنواختی صفحه فلوسل را می‌گیرد. دقت شود

¹ Washing 3

¹ Background correction 4

که این پروب‌ها بر اساس نواحی مختلف (نه یک ناحیه یکسان) از ژن کاندید طراحی شده است و به بحث ایزوفورم‌ها (وجود چندین آگرون و اینترون متفاوت برای یک ژن) نیز توجه می‌گردد. پیرایش متناوب غیرمعمول در سلول‌های سرطانی، شایع بوده و ایزوفورم‌های حاصل می‌توانند به‌عنوان بیومارکر و یا هدف درمانی برای طراحی دارو مورد استفاده قرار گیرند. ایزوفورم‌های ژن، mRNAهایی هستند که از یک مکان تولید می‌شوند، اما از نظر مکان‌های شروع رونویسی (TSS)، توالی‌های DNA کدکننده پروتئین (CDS) و یا مناطق ترجمه نشده (UTRs) متفاوت هستند و به طور بالقوه عملکرد ژن را تغییر می‌دهند (Angenendt, 2005).

ماهیت داده خام در مطالعات ریزارایه و مقایسه آن با داده‌های qPCR

داشتن اطلاعات در رابطه با ماهیت داده خام در یک مطالعه، کمک بزرگی به تجزیه و تحلیل‌های آن دارد. در آزمایش‌های بیان ژن به روش qPCR که از فرمول‌های Livak و Pfaffl method) چون پایه ریاضی فرمول عدد ۲ می‌باشد، در هر سیکل واکنش زنجیره پلی‌مرز، رشته DNA به‌طور تئوریک دو برابر می‌شود را تصحیح و مطالب را همپوشانی می‌کنند، زیرا، الگوی بیان روند یکنواختی دارد و مقایسه حداکثر ۵ الی ۶ ژن با یک ژن کنترل در دو بافت است. در این حالت به‌طور مستقیم داده ورودی قابل ارزیابی با روش‌های مناسب آماری است. در داده‌های ریزارایه چون بحث مقایسه چندین هزار ژن به‌طور هم‌زمان است، نیازمند دیدگاه‌های خاص متفاوت به تناسب داده وجود دارد و رفتار ژن‌های مختلف در دو بافت درگیر و سالم قطعاً پیچیده می‌باشد و تصویرسازی کل داستان در یک قاب قطعاً سخت و مشکل خواهد بود. لذا، اولین خاصیت دیتای خام ریزارایه غیریکنواختی داده‌ها^{۱۵} می‌باشد، یعنی تعداد زیاد ژن با بیان متفاوت و در مقیاس گسترده و طیف وسیع، پس اولین تفاوت این است که در داده‌های ریزارایه از سیستم لگاریتم‌گیری استفاده می‌گردد (Morey et al., 2006).

یکی از دلایل منطقی که در داده‌های ریزارایه سیستم لگاریتم‌گیری پیاده می‌شود این است که ارزش خام و مطلق مد نظر نیست و تغییرات و واریانس مقیاس‌سنجی کمی می‌شود. از طرف دیگر، در ریزارایه چون داده پربرونداد^{۱۶} است، همه زوایای تغییرات جلوی چشم ظاهر نمی‌شوند و بسیاری از تغییرات بیان کم خواهد بود، اما اصل مهم در فنوتیپ پنهان می‌ماند و در مقابل ژن‌های با بیان زیاد، ژن‌های دارای بیان کم جلوه نمی‌کنند در نتیجه لگاریتم‌گیری داده خام در تکنیک ریزارایه سعی می‌کند این موارد را مدیریت کند. سیستم لگاریتم‌گیری قضاوت کاربر و محقق را نسبت به تغییرات تمام ژن‌ها نزدیک به واقعیت می‌کند. از سوی دیگر این سیستم منجر می‌شود تصحیح بر مبنای اکثر ژن‌هایی که ساختاری هستند و واریانس در تمام سلول‌های سالم و بیمار یکسانی دارند، صورت پذیرد (Popovici et al., 2009).

1 Heterogeneity of data 5
1 High throughput data 6

لگاریتم مورد استفاده در آنالیز داده‌های خام ریزآرایه لگاریتم بر مبنای ۲ می‌باشد و از لحاظ ریاضی لگاریتم بر مبنای دو خواص و ویژگی‌های جالبی دارد؛ از جمله این که حداقل لگاریتم بر پایه ۲ صفر است و توابع تعریف شده بر این اساس می‌توانند ضرب را به جمع و تقسیم را به تفریق تبدیل کنند. آن چه بیش‌تر بر آن تاکید می‌شود خروجی دیتای مستقیم از ریزآرایه ۱۰۰،۱۰۰،۱۰۰۰ می‌باشد، در حالی که برای ما تغییرات و واریانس مهم است و عدد مطلق نقش آفرین نیست. در واقع ممکن است در تحلیل هم‌زمان ژن‌ها، ژن‌هایی باشند که تفاوت بیان یک واحدی در آن مهم‌تر از چند ده برابری تفاوت بیان در ژن‌های دیگر است. به‌طور کلی لگاریتمی کردن محدوده داده را از طول موج صفر تا ۴۰۰۰۰ به ۱ الی ۱۶ تقلیل می‌دهد و در یک نمودار کنار هم کشیدن الگو بیان ژن‌ها در مقایسه ۱ الی ۱۶ بهتر از مقیاس صفر تا سه هزار است (Dittami *et al.*, 2013).

انجام مرحله پیش پردازش یا تعیین کیفیت داده‌های خام

با توجه به این که اساس ریزآرایه، هیبریداسیون در محلول بافر و هیبریداسیون بین خردشده‌های cDNA و پروب‌های مستقر می‌باشد، اتصال زمانی مطلوب است که اختصاصی باشد، یعنی پروب باید اختصاصی و بی‌همتا باشد و همچنین GC پروب‌های مختلف باید انرژی یکسانی را برای اتصال داشته باشند. پس GC پروب‌ها باید یکسان باشد. در عمل این پیش‌بینی‌های تئوری ممکن است درست درنیایند و بنا به دلایل متعددی داده‌ها نویز یا خطا داشته باشند. در این مرحله، باید پیش‌پردازش داده‌های خام انجام شود. علل مختلفی سبب ایجاد این نویزها می‌شود که از جمله آن‌ها می‌توان به کمیت و کیفیت RNA، تولید cDNA اشتباه، تفاوت در اتصال و هیبریداسیون ناقص، کافی نبودن سلول در موارد تومور مغزی و آلودگی در حین استخراج RNA اشاره کرد (Silver *et al.*, 2009).

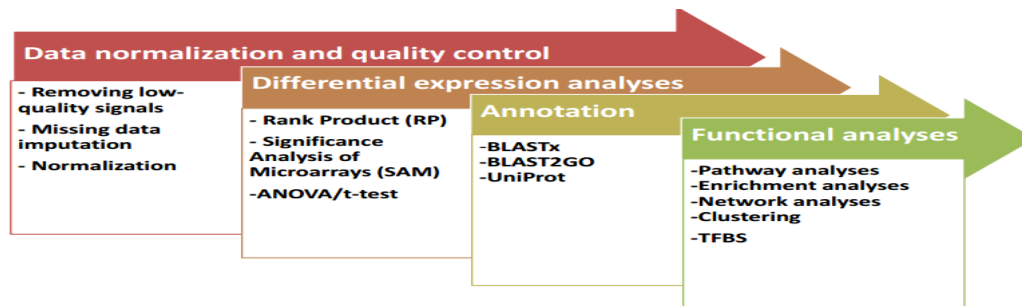
مکانیسم انجام پیش پردازش یا تعیین کیفیت داده‌های خام

در معماری ریزآرایه به‌طور عمده از طراحی جالبی استفاده شده است. در این طراحی یک سری پروب مخفی استفاده می‌شود که نه مکمل دارند و نه مشابهی برای آن‌ها جهت هیبریداسیون وجود دارد؛ بلکه توالی‌های خامی هستند که اصلاً به گونه مورد مطالعه حتی ارتباط دور هم ندارد. این موارد خود می‌توانند به نتایجی که سیاه مطلق نیستند، یک نوع معنی بخشی مجدد دهند. لازم به ذکر است که در سلول‌های سرطانی و بیمار دقیقاً کیفیت و کمیت نمونه‌های اولیه به هم کاملاً شبیه باشند تا در اسکنر دستگاه نتایج کاذب نشان ندهند. در واقع پیش شرط آنالیز این است که کیفیت مطلوب و یکسانی از دو نمونه به دستگاه معرفی شود. پس عملاً باید یک سری روش آماری به کار گرفته شود تا کیفیت داده‌ها و اعتبار اولیه داده‌ها سنجیده شود که این مورد براساس نرمال‌سازی و کنترل کیفیت داده‌های ریزآرایه می‌باشد (Edward, 2003).

روش‌های آماری تست کیفیت داده ریزآرایه

اهمیت مسئله کیفیت داده موضوعی مهم و ضروری برای هر پروفایل داده است، چرا که تضمین سلامت و صحت داده‌ها هنگام ترکیب و یکپارچگی آن‌ها، از الزام‌های اساسی مدیریت کیفیت داده می‌باشد. پژوهشگران حوزه کیفیت

در تلاش هستند که موضوع و اهداف مورد نظر در اندازه‌گیری کیفیت داده را توسعه ببخشند (Pollock., 2012). در تکنیک ریزآرایه داده‌های تولیدی با استفاده از تست‌های استفاده از روش تجزیه به مؤلفه‌های اصلی، رسم نمودار حرارتی^{۱۷} کلاسترینگ و رسم دندروگرام، تست همبستگی داده، TSNE و تعیین کیفیت می‌شوند (Russell *et al.*, 2008).



شکل ۸: مراحل تجزیه و تحلیل داده‌های ریزآرایه برای رسیدن به تفسیر نهایی

و نتیجه‌گیری (Silver *et al.*, 2009) برگرفته از اسلایدهای دکتر اسلاملو

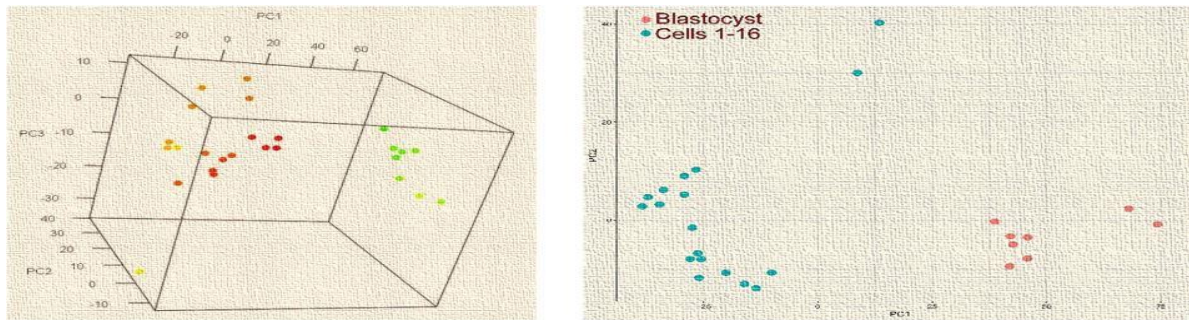
Figure 8 - Steps of microarray data analysis to reach final interpretation and conclusion (Silver *et al.*, 2009)

روش تجزیه به مؤلفه‌های اصلیدر شروع آنالیز، انتظار این است که اختلاف واریانس درون گروهی در حداقل مقدار ممکن باشد و به عنوان مثال، بین سلول تیمار و کنترل، انتظار اصلی تفاوت در رفتارهای ژن‌ها می‌باشد و همچنین روش تجزیه به مؤلفه‌های اصلی سعی دارد تا با کاهش ابعاد در وهله اول ثابت کند که سلول‌های سالم و سرطانی در دو گروه مجزا قرار بگیرند و در واقع، از این تکنیک می‌توان دو استفاده کرد. یک مورد گروه‌بندی نمونه‌ها است (بررسی کیفیت دیتا) و دوم گروه‌بندی بر اساس تعداد ژن‌ها است که به شناسایی بیومارکرهای ژنی منجر خواهد شد. کنترل کیفیت در مقایسات فرد به فرد، فرد به ژن و ژن به ژن می‌تواند منجر شود. دقت شود که در خروجی‌های اولیه ریزآرایه از بین ۳۰ الی ۳۳ هزار ژن انتظار این است که تصویر بیان در سلول‌های سرطانی و سالم بین کلیه افراد نرمال باشند، چون که بسیاری از ژن‌ها، ژن‌های ساختاری هستند و تنها تعداد محدودی از ژن‌ها با حداکثر فراکتال و پایداری در الگو دارای تغییر هستند و بخش اعظم ژن‌ها یکسان هستند. از طرف دیگر، تجزیه به مؤلفه‌های اصلی نشان‌دهنده کیفیت کلی داده‌ها است و همچنین به پاسخ‌های سودمندی محقق را سوق می‌دهد که تفاوت بیانی معنی‌دار ژن‌ها در سلول تخم تا مرحله ۱۶ سلولی و بلاستوسیست در توسعه جنینی گاو کدام ژن‌ها بیشترین تفاوت در بیان را دارند و معنی‌دار هستند (Khaje Ghiassi *et al.*, 2022).

نمودار حرارتی و نحوه تفسیر خروجی‌ها

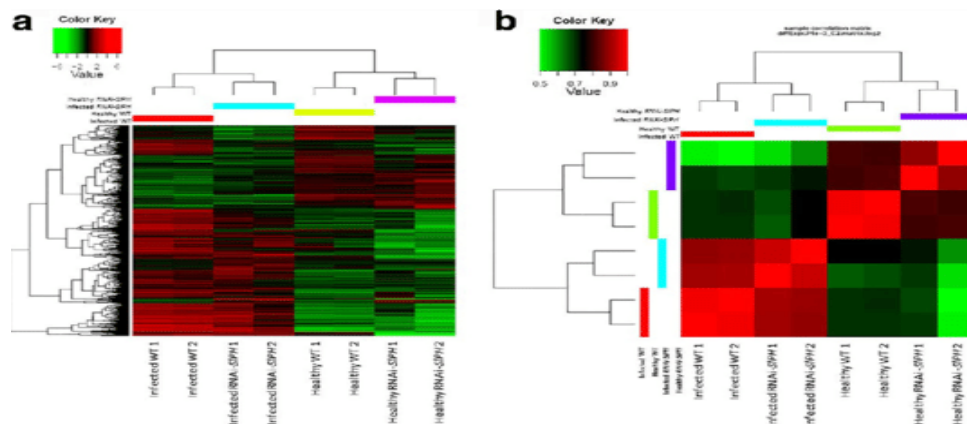
نمودار Heatmap از اهمیت خاصی برخوردار است و گروه‌بندی بین ژن با ژن، فرد با ژن و فرد با فرد در این خصوص مطرح می‌باشد و می‌تواند همبستگی بیومارکر در سلول‌های سالم و به طور جدا در سلول‌های بیمار و همچنین هم‌بیانی یا Co-expression ژن‌ها را برای محقق مشخص کند (Barad *et al.*, 2017).

¹ Heat map



شکل ۹: تصویر روش تجزیه مؤلفه‌های اصلی (PCA) داده ریزآرایه سلول تخم تا مرحله ۱۶ سلولی و بلاستوسیست در توسعه جنینی گاو. نمودار حاکی از کیفیت و دسته‌بندی مناسب داده‌ها می‌باشد و داده‌ها به نرمال‌سازی نیازی ندارند (Khaje Ghiassi *et al.*, 2022).

Figure 9 - Principal component analysis (PCA) image of the bovine embryonic cell to a 16-cell and blastocyst stage of microarray data. The graph indicates the quality and proper classification of the data and the data do not need normalization (Khaje Ghiassi *et al.*, 2022).

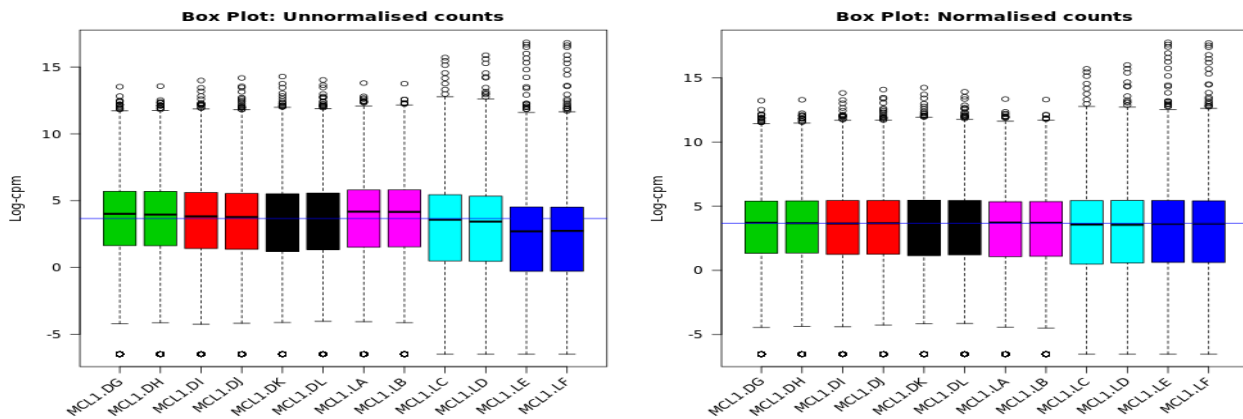


شکل ۱۰: نمودار همبستگی پیرسون بین سطوح بیان ژن‌ها در نمونه‌های مختلف (Barad *et al.*, 2017).
Figure 10- Heat map of Pearson correlation between expression levels of genes in different samples (Barad *et al.*, 2017).

نمودار بارپلات و نحوه تفسیر خروجی‌ها

همان‌طور بیکه در شکل ۱۱ دیده می‌شود میانگین، میانه، چارک‌های اول، دوم، سوم و چهارم مهم هستند. میانگین زمانی کاربرد دارد که داده‌ها نرمال باشند. از طرفی اگر داده‌ها نرمال نباشند، از میانه استفاده می‌شود. در واقع نرمال نبودن داده‌ها نشان می‌دهد که سلول‌ها شبیه نیستند، دیتا افتراقی نمی‌شود و داده‌ها دارای نویز یا خطا هستند و یا تکرارهای بیولوژیکی کنار هم قرار نمی‌گیرد. چارک‌ها تصویر مناسبی از توزیع داده‌ها به ارائه می‌دهند (Brennan *et al.*, 2018). پس علت تعیین کیفیت داده‌ها و پیش‌پردازش داده‌های به جلوگیری از وقایع تصادفی و کنترل نشده، یا حذف نمونه اشتباهی خاص در جامعه به گروه‌های سالم یا بیمار و یا وجود خطا برمی‌گردد. در واقع هدف از کنترل کیفیت داده‌ها، حذف واریانس داخل گروه و افزایش مشاهده احتمالی واریانس بین دو گروه می‌باشد. عدم مشاهده کیفیت در نمونه‌های اولیه اطلاع از این دارد که باید از روش‌های دیگری استفاده شود تا این که داده‌ها نرمال شوند؛ به گونه‌ای که این نرمال کردن می‌تواند مشکلات تکنیکی در زمان انجام کار رخ دهد، مانند

تشکیل حباب یا شیب در صفحه فلورسل. نرمال کردن به داده‌های خام و اولیه آسیب نمی‌رساند، چون این روش تغییرات اولیه موجود را از بین نمی‌برد. بلکه در نرمال کردن تفاوت‌های فردی و خطای ناشی از هیبریداسیون و شستشو از بین می‌روند. از جمله روش‌های نرمال کردن داده‌های بیان استفاده از روش Quantile است. جهت نرمال کردن داده‌های غیر نرمال ریزآرایه از روش‌هایی مانند سینگنال‌های کلی هیبریداسیون، استفاده از ژن‌های کنترل داخلی و قرار دادن mRNA خارجی استفاده می‌شود. برای تست آماری همبستگی داده‌های خام هم معمولاً از آزمون فیشر (F-test) استفاده می‌گردد.



شکل ۱۱: تصویری از داده‌های اولیه بیان ژن نرمال و غیرنرمال توسط بار چارت (Brennan *et al.*, 2018)

Figure 11- An image of the primary data of normal and abnormal gene expression by bar chart (Brennan *et al.*, 2018)

مفهوم لگاریتم بر پایه دو شاخص Fold change

پس از اتمام بخش آزمایشگاهی ریزآرایه و ایجاد داده‌های خام با دو سطح داده روبرو خواهیم بود. در داده‌های خام در واقع هدف پیاده‌سازی یک آزمون t (T-test) معمولی برای خروجی‌های با ماهیت کمی است، پس در واقع از نمونه‌های بیولوژیکی در سلول‌های سالم یک میانگین و از داده‌های متعلق به بیمار نیز یک میانگین گرفته می‌شود. اما براساس آن‌ها هم، مقیاس به مقایسه لگاریتم در پایه ۲ برده می‌شود و مانند آزمون t با یک سطح احتمال داده‌ها مقایسه می‌شوند. کار اگر چه ساده به نظر می‌رسد، ولی در عمل چالش‌هایی به‌علت ماهیت خاص داده وجود دارد. در آزمون t اگر حتی میانگین‌ها برابر باشند ولی واریانس‌ها متفاوت، می‌تواند معنی‌داری و تفاوت‌های موجود را بیرون بیاورد و در واقع بحث ریزآرایه هم منطبق با این خاصیت آزمون t می‌باشد.

لگاریتم مبنای دو (میانگین بیان ژن در سلول بیمار - میانگین بیان ژن در سلول سالم) اگر مثبت باشد، در سلول بیمار ژن افزایش بیان داشته و اگر منفی باشد در سلول بیمار ژن کاهش بیان داشته است و با مرتب کردن داده‌ها می‌توان ژن‌های افتراقی با درجه معنی‌داری را تشخیص داد. مفهوم شاخص LogFold change در واقع از تقسیم لگاریتم بیان ژن در سلول بیمار بر لگاریتم بیان ژن در سلول سالم حاصل می‌شود و یا برعکس، چون به حالت مقایسه سالم به بیمار یا بیمار به سالم می‌تواند

بستگی داشته باشد (Khaje Ghiassi *et al.*, 2022).

مشکل و چالش مقایسات میانگین چندگانه

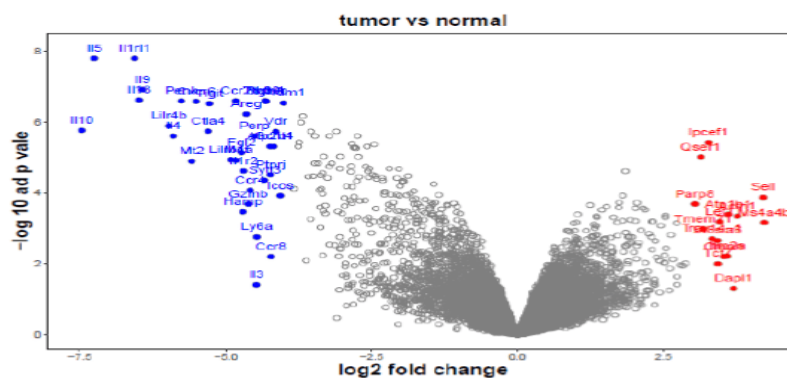
در مقایسات آماری و طرح آزمایشات در واقع شانس وقوع خطا ($1-\alpha$) و عدم خطا ($1-(1-\alpha/5)$) می باشد و زمانی که از آزمون مقایسات میانگین چندگانه استفاده می شود، مجبور به تصحیح p-value خواهیم بود و تصحیح P باعث کاهش خطای نوع یک می شود.

	فرض صفر درست	فرض صفر غلط
رد فرض صفر	خطای نوع اول α	توان آزمون $1-\beta$
پذیرش فرض صفر	خطایی رخ نداده $1-\alpha$	خطای نوع دوم β

شکل ۱۲: انواع خطا در آزمایشات آماری (Russell et al., 2008)

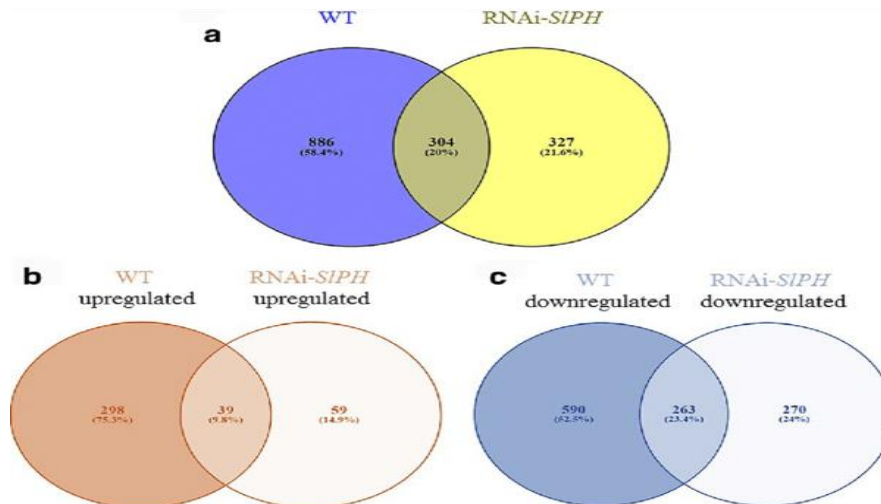
Figure 12- Types of errors in statistical tests (Russell et al., 2008)

در واقع با این استراتژی جلوی وقوع false positive و false negative گرفته می شود، یعنی تفاوت بیان بین سلول سرطانی و سالم وجود ندارد و به صورت اشتباه داده ها معرفی شده اند و یا تفاوت بین نمونه های سرطانی و سالم وجود داشت و به اشتباه متوجه نشده این امر نشده ام. این همان مفهوم False discovery rate می باشد که نشان دهنده دقت می باشد. در صورتی که در خروجی تجزیه و تحلیل های LogFold change برابر $0/5$ شده باشد، از لحاظ آماری یعنی 2 به توان $0/5$ یا رادیکال دو می باشد که برابر $1/4$ خواهد شد. پس در واقع p_value adjusted واریانس را در کنار میانگین به صورت همزمان نشان می دهد و ژن هایی که میانگین یکسان ولی واریانس کوچک متفاوت دارند را بیرون می کشد. در آنالیزها خروجی هایی هستند که LogFold change بالای یک یا کوچک تر از یک باشد. اگر در مواردی p_value غیر معنی دار باشد، اما $\log_{2}FC$ بالاتر یا کمتر از یک باشد، بیان کننده تعداد نمونه کم، نویز زیاد و یا واریانس درون گروه زیاد است (Popovici et al., 2009). آزمون های آماری مانند FDR از آزمون آسان گیر و آزمون بونفرونی از جمله آزمون های سختگیرانه می باشند (Retis-Resendiz et al., 2021).



فیلتراسیون ژن‌هایی که در بافت‌های مختلف رفتار یکسان یا غیریکسان دارند

در این مرحله از آنالیز سعی می‌شود که از لیست کلی بیان ژن‌های افتراقی به دسته‌بندی‌های سازماندهی شده‌ای دست یافت که در آن ژن‌های با بیان اختصاصی در یک بافت (Tissue specific gene expression pattern) مشخص شود، به گونه‌ای که ژن‌های دارای بیان معنی‌دار متفاوت و مشترک میان بافت‌های مورد مطالعه قابل تشخیص و دسته‌بندی باشند (Hall *et al.*, 2007).

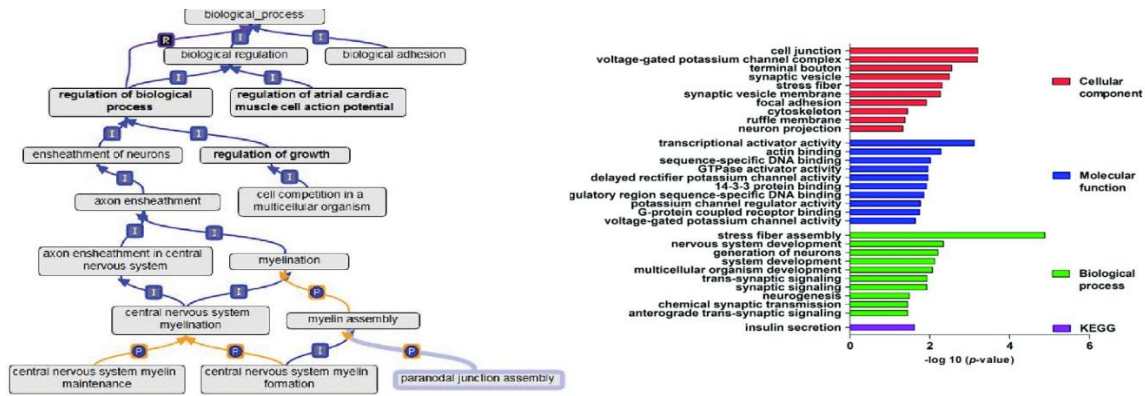


شکل ۱۴: نمودار ون. نشان‌دهنده تعداد ژن‌های دارای معنی‌دار برای هر بافت و همچنین ژن‌های مشترک میان دو یا چند بافت مورد مطالعه (Barad *et al.*, 2017).

Figure 14- Venn diagram. Present the number of genes with significance for each tissue as well as common genes between two or more tissues studied (Barad *et al.*, 2017).

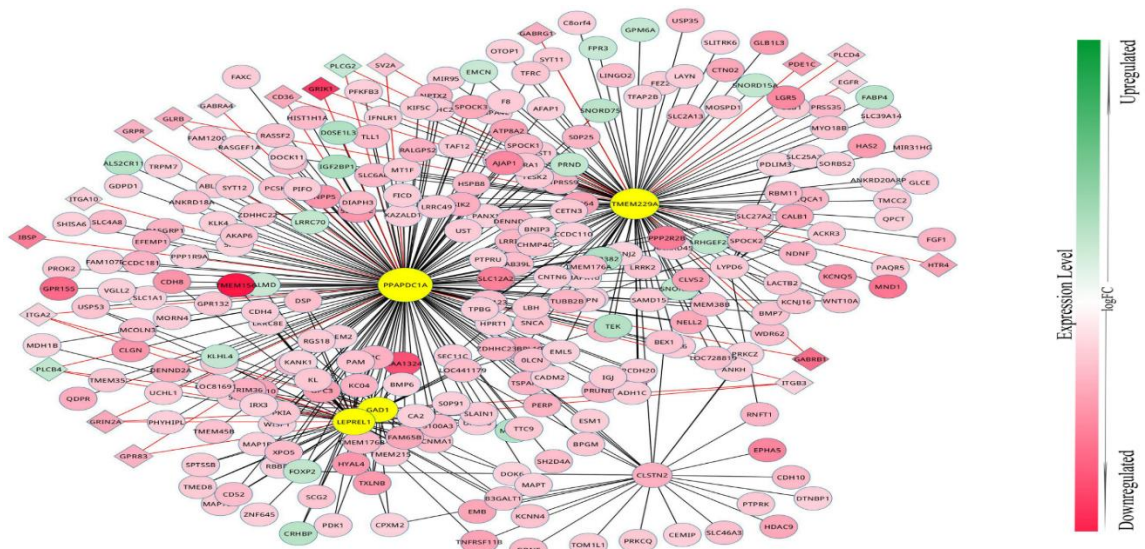
آنالیزهای Downstream، هستی‌شناسی ژن‌ها

پس از مشخص شدن لیست ژن‌های هاب با بیان افتراقی معنی‌دار بین نمونه‌های سالم و بیمار موضوع در این زمان خاتمه پیدا نمی‌کند و با توجه به اثرات متقابلی که بین ژن‌های مختلف معنی‌دار وجود دارد؛ رویکرد زیست‌شناسی سامانه‌ای، ارتباط بین ژن‌ها و هم‌بیانی و همچنین کشف ژن‌های بزرگ اثر صورت می‌گیرد. در این مرحله ژن‌های با الگوی بیان افتراقی (مهمترین ژن‌های کلیدی و تأثیرگذار) در چرخه‌های سلولی مرتبط با آن صفت اولیه مورد بررسی مشخص می‌شوند و همچنین تعاملات بین پروتئینی و فعالیت‌های بیولوژیکی آن‌ها در سلول مشخص می‌شود. در واقع این روش‌ها یک نوع فیلتراسیون هوشمندانه برای رسیدن از یک لیست ژنی گسترده به یک لیست ژنی مؤثر و با بالاترین درجه تأثیرگذاری می‌باشد (Ramakrishnan *et al.*, 2005). در مرحله بعد بررسی و آنالیز مسیرهای متابولیکی و سیگنالینگ مرتبط با ژن‌های بزرگ اثر انجام و بازسازی شبکه اثرات متقابل پروتئین- پروتئین نیز صورت می‌گیرد. با توجه به مسیرهای معنی‌دار شناسایی شده می‌توان به این نتیجه رسید که ژن‌های بزرگ اثر استخراج شده در کدام مسیرها فعال هستند و چه وظایف خاصی در بافت‌های مورد مطالعه و بدن برعهده دارند.



شکل ۱۵: روش‌های آماری مورد نیاز جهت آنالیز و شناسایی گروه‌های ژنی با بیان افتراقی معنی‌دار در مقایسه نمونه‌های سالم و بیمار با هم. سمت چپ: روندنمای مراحل مختلف آنالیز. سمت راست: مسیرهای متابولیکی (در سه دسته) و سیگنالی مرتبط با ژن‌های دارای تفاوت بیانی معنی‌دار (Russell *et al.*, 2008)

Figure 15- Statistical methods required to analyze and identify gene groups with significant differential expression in comparing healthy and diseased samples together. Left side: flow chart of different stages of analysis. Right: metabolic pathways (in three categories) and signaling associated with genes with significant expression differences (Russell *et al.*, 2008).



شکل ۱۶: بازسازی شبکه اثر متقابل پروتئین- پروتئین یا شبکه ژنی جهت شناسایی گروه‌های ژنی (ماژول‌ها) و یا ژن‌های بزرگ اثر با بیان افتراقی معنی‌دار در مقایسه نمونه‌های سالم و بیمار با هم (Russell *et al.*, 2008).

Figure 16- Reconstruction of protein-protein interaction network or gene network to identify gene groups (modules) or large effect genes with significant differential expression comparing healthy and diseased samples together (Russell *et al.*, 2008)

آنالیز WGCNA

اخیراً، یک بسته تحلیلی نرم‌افزار R به نام تجزیه و تحلیل شبکه هم‌بیانی وزن‌دار (WGCNA) توسعه یافته است و برای مجموعه داده‌های ریزآرایه یا RNA-Seq با توان عملیاتی بالا دارای کاربرد می‌باشد، زیرا بینش‌هایی در سطح سیستم، حساسیت بالا به فراوانی کم یا تغییرات کوچک تاشوی ژن‌ها را ارائه می‌دهد. بدون از دست دادن اطلاعات، این رویکرد اخیراً برای تجزیه و

تحلیل داده‌های پروتئومی و متابولومیک نیز مورد استفاده قرار گرفته است. با این حال، با توجه به این واقعیت که پوشش کم فناوری‌های تحلیلی پروتئومی و متابولومیک فعلی، باعث می‌شود که قالب مجموعه داده‌ها اغلب ناقص باشد. این روش نیاز به اصلاح دارد تا از آن برای تفسیر بیولوژیکی معنی‌دار بتوان استفاده کرد که این امر با آنالیزهای WGCNA که با کلاستر بندی مجموع ژنی این نقص رفع گردیده است.

تجزیه و تحلیل شبکه همبستگی وزن‌دار^۱ که با عنوان تحلیل شبکه ژن‌های هم بیان وزن‌دار^۲ (WGCNA) شناخته می‌شود، یک روش داده کاوی پرکاربرد است که به ویژه برای مطالعه شبکه‌های بیولوژیکی، براساس همبستگی بین زوج متغیرها مورد استفاده قرار می‌گیرد. با این روش ماژول‌ها (خوشه‌ها)، هاب ژن‌های درون ماژول‌ها^۱ و رئوس شبکه را با توجه به عضویت آن‌ها در هر ماژول تعریف می‌کنند. به کمک این روش روابط بین ماژول‌های هم بیان را نیز مطالعه کرده و توپولوژی شبکه‌های مختلف را مقایسه می‌کنند (تحلیل تفاقی شبکه)^۲. WGCNA را می‌توان به عنوان یک تکنیک تقلیل داده^۳ یا به عنوان یک روش خوشه‌بندی (خوشه بندی فازی)^۴ یا حتی به عنوان یک روش انتخاب ویژگی و همچنین به عنوان چارچوبی به منظور یکپارچه‌سازی داده‌های ژنومی استفاده کرد. با وجود این که WGCNA از مبانی تکنیک‌های سنتی کاوش داده‌ها استفاده می‌کند، زبان شبکه بصری و چارچوب تحلیل آن فراتر از هر تکنیک تحلیل استاندارد دیگری است. از آنجایی که در این تکنیک از روش آنالیز شبکه استفاده می‌شود، لذا آن را در زمره روش‌های بیولوژی سیستم‌های ژنتیکی تقسیم‌بندی می‌کنند.

گام مهم در آنالیز مبتنی بر ماژول در شبکه‌ها خوشه بندی ژن‌ها در درون ماژول‌ها می‌باشد که برای این منظور از معیار نزدیکی شبکه^۵ استفاده می‌گردد. به عبارت دیگر، اگر دو ژن مقادیر پراکسیمیتی بالایی داشته باشند ارتباط محکم‌تری با هم دارند. در روش WGCNA از معیار هم‌پوشانی توپولوژیک^۶ برای محاسبه پراکسیمیتی استفاده شده است. در این معیار میزان ارتباط دو ژن با همدیگر و با سایر ژن‌ها استفاده می‌شود و همین معیار برای خوشه‌بندی ژن‌ها نیز مورد استفاده می‌گردد. در همین مبنا می‌توان هر ماژول را شاخه‌های یک خوشه در نظر گرفت که با رهیافت Dynamic branch cutting به دست آمده است.

مطالعات بسیار زیادی نیز از زمان معرفی این تکنیک، توسط محققان حوزه ژنتیک انجام شده‌است. به گونه‌ای که از این روش در شناسایی ژن‌های مهم تأثیرگذار بر بیماری‌ها یا صفات مختلف استفاده شده است. در واقع WGCNA یک روش مکمل برای مطالعات کلینیکی یا آزمایشگاهی است تا محققان به صورت هدف‌دار برای مطالعه پروفاایل ژنی بیماری‌ها یا صفات در

¹ Weighted correlation network analysis

² Weighted gene co-expression network analysis

² Intramodular hubs ¹

² Differential network analysis ²

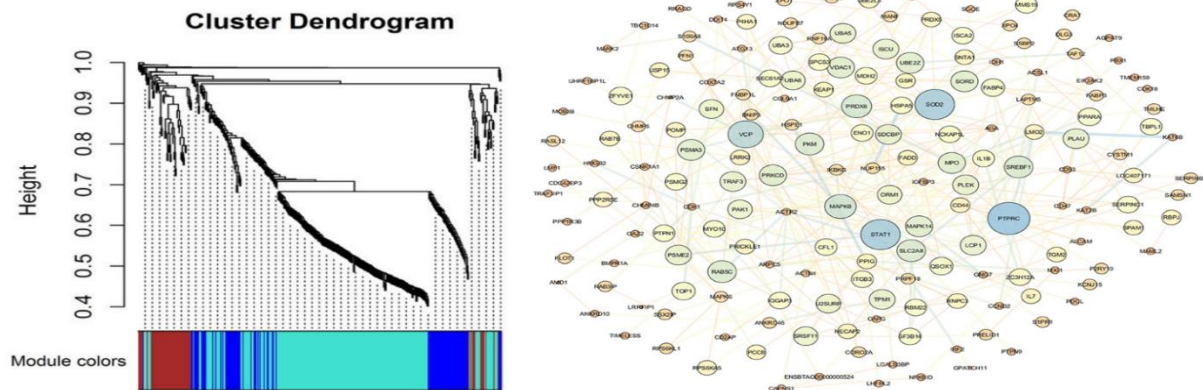
² Data reduction ³

² Fuzzy clustering ⁴

² Network proximity measure ⁵

² Topological overlap measure ⁶

موجودات زنده اقدام نمایند. این رو در تلفیق با سایر روش‌ها همچون یادگیری ماشین یا بهره‌گیری از الگوریتم‌های داده کاوی می‌تواند اطلاعات بسیار زیاد و کاربردی را به دنیای بیولوژی معرفی نماید. شبکه بازسازی شده حاصل از این روش را می‌توان برای استفاده در نرم افزارهای دیگری همچون Gephi یا Cytoscape استخراج نموده و برای آنالیزهای بیشتر از آن‌ها بهره برد.



شکل ۱۷: خروجی‌های ناشی از تجزیه و تحلیل شبکه‌های ژنی هم‌بیانی وزن دار (WGCNA) (Ghahramani *et al.*, 2021)
Figure 17- Outputs from weighted co-expression gene network analysis (WGCNA) (Ghahramani *et al.*, 2021)

معرفی پایگاه‌های مورد نیاز برای تحلیل‌های پسین داده‌های ریزآرایه

پایگاه داده زیستی KEGG

این پایگاه، پس از دریافت ژن‌های معنی‌دار استخراج شده از آزمایش‌های بیان ژن یا توالی‌یابی کل ژنوم، مسیرهای متابولیکی-سیگنالینگ مرتبط معنی‌دار را با توجه به لیست ژنی دریافت شده معرفی می‌نماید و همچنین در آن تعدادی عملکرد سلول و ارگانیسم‌ها نیز به صورت جزئی به تصویر کشیده خواهد شد. در هر خروجی شامل یکسری واکنش بین بیومولکول‌های مختلف می‌باشد که بیان‌کننده اطلاعاتی جامع درباره ژن‌ها و پروتئین‌های دخیل در آن است. در این پایگاه داده می‌توان مسیرهای مختلف را با هم مقایسه و تحلیل کرد. هدف این پایگاه داده در واقع، سازماندهی اطلاعات موجود در مورد ارتباط بین مارکر مولکول‌های زیستی به ویژه مسیرهای بیوشیمیایی، مسیرهای تنظیمی و فرآیندهای زیستی است.

پایگاه داده زیستی STRING

اثرات متقابل بین پروتئین‌ها را اغلب می‌توان از ارتباط ژنومی بین ژن‌هایی که آن‌ها را رمزگذاری می‌کنند استنباط کرد. پایگاه داده STRING یک منبع جهانی محاسبه شده برای کاوش و تجزیه و تحلیل این ارتباطات است. STRING شامل یک چارچوب امتیازدهی منحصر به فرد براساس معیارهای انواع مختلف ارتباط در برابر یک مجموعه مرجع مشترک است که در یک امتیاز اطمینان واحد در هر پیش‌بینی ادغام شده است. نمایش گرافیکی شبکه اثرات متقابل پروتئینی استنباط شده و وزن‌دار، نمای سطح بالایی از اثرات متقابل عملکردی را فراهم می‌کند که تجزیه و تحلیل مدولار بودن در فرآیندهای بیولوژیکی را تسهیل

می‌کند. این پایگاه داده به طور مداوم به روزرسانی می‌گردد. پایگاه داده STRING یکی از چندین منبع آنلاین است که به شبکه‌های ارتباط پروتئینی در سطح ارگانسیم اختصاص داده شده است. چارچوب‌های رایج موجود در این پایگاه شامل هستی‌شناسی ژن که شامل عملکرد مولکولی، فرآیند بیولوژیکی و اجزای سلولی می‌باشند و همچنین شامل مسیرهای KEGG، Reactome، Pfam و SMART نیز است.

پایگاه داده زیستی DAVID

هدف تمام ابزارهای موجود در منابع بیوانفورماتیک DAVID ارائه تفسیر عملکردی از فهرست‌های بزرگی از ژن‌های مشتق شده از مطالعات ژنومی است. پایگاه دانش توسعه یافته DAVID، امروزه تقریباً تمام منابع بیوانفورماتیک عمومی شناخته شده و اصلی را یکپارچه می‌کند. ده‌ها میلیون شناسه ژن/پروتئین متنوع و اصطلاحات حاشیه‌نویسی از انواع پایگاه داده‌های عمومی بیوانفورماتیک در این پایگاه داده موجود است. برای هر فهرست ژنی آپلود شده، اکنون نه تنها تجزیه و تحلیل غنی‌سازی ژنی معمولی، بلکه ابزارها و عملکردهای جدیدی را نیز ارائه می‌سازد. این پایگاه داده شامل مجموعه‌ای کامل از ابزار شناسایی برای فهم عملکرد بیولوژیکی ژن‌ها می‌باشد. همچنین با اتصال به پایگاه داده‌های دیگر مانند GO، Wikipathways، KEGG و Reactome به بررسی عملکردی و مسیرهای متابولیکی ژن‌های معنی‌دار معرفی شده به آن می‌پردازد.

پایگاه داده زیستی Gene Ontology

پروژه هستی‌شناسی ژن (GO) (<http://www.geneontology.org/>) واژگان و طبقه‌بندی‌های ساختار یافته و کنترل شده‌ای را ارائه می‌دهد که چندین حوزه از زیست‌شناسی مولکولی و سلولی را پوشش می‌دهد و به طور رایگان برای استفاده جامعه در حاشیه‌نویسی ژن‌ها در دسترس است. پایگاه داده GO واژگان و حاشیه‌نویسی‌های ارائه شده را یکپارچه می‌کند و دسترسی کامل به این اطلاعات را در قالب‌های مختلف فراهم می‌سازد. در واقع، هستی‌شناسی ژن^۷ به صورت خلاصه GO یک پروژه اطلاعاتی ساختار یافته و قابل پردازش از عملکرد ژن‌ها و محصولات ژنتیکی را ارائه می‌دهد. امروزه داده‌های پروژه GO به طور گسترده در علوم زیستی استفاده می‌شوند و محتوای آن هم از نظر کیفیت و هم از نظر کمیت مدام در حال تغییر است. آنتولوژی ژن (GO) یک لغت نامه برای توصیف محصولات ژن است و تلاشی برای هماهنگ کردن توصیف محصولات ژن‌ها در پایگاه‌های داده‌ای مختلف می‌باشد. هستی‌شناسی شامل سه زیر آنتولوژی است که شامل فرآیند بیولوژیکی (BP)^۸، عملکرد مولکولی^۹ (MF) و اجزاء سلولی^{۱۰} (CC) می‌باشد (Balakrishnan *et al.*, 2013). فرآیند بیولوژیکی نشان دهنده فعالیت بیولوژیکی است که ژن در آن شرکت می‌کند مانند سیکل سلولی. عملکرد مولکولی بیان‌کننده فعالیت‌هایی است که در سطح سلول رخ می‌دهند و اجزای

2	Gene Ontology	7
2	Biological process	8
2	Molecular function	9
3	Cellular Component	0

سلولی نیز تعیین می‌کند که محصول ژن در کدام قسمت سلول یافت می‌شود (Yavari *et al.*, 2008).

پایگاه داده زیستی Revigo

نتایج آزمایش‌های بیولوژیکی با کارایی بالا معمولاً با آزمایش‌های آماری برای دسته‌های عملکردی ژن غنی شده تعریف شده توسط هستی‌شناسی ژن (GO) تفسیر می‌شوند. فهرست‌های حاصل از اصطلاحات GO ممکن است حجیم و بسیار زائد باشند و بنابراین تفسیر آن دشوار خواهد بود. REVIGO یک وب سرور است که فهرست‌های طولانی و نامفهومی از اصطلاحات GO را با یافتن زیرمجموعه‌ای نماینده از اصطلاحات با استفاده از یک الگوریتم خوشه‌بندی ساده که بر معیارهای تشابه معنایی متکی است، خلاصه می‌کند. علاوه بر این، REVIGO این مجموعه اصطلاح غیر زائد GO را به روش‌های متعددی برای کمک به تفسیر تجسم می‌کند: مقیاس‌گذاری چند بُعدی و تجسم‌های مبتنی بر نمودار به دقت زیربخش‌ها و روابط معنایی در داده‌ها را ارائه می‌کنند، در حالی که نقشه‌های درختی و آبرهای برجسب نیز به عنوان نماهای جایگزین ارائه می‌شوند. REVIGO به صورت رایگان در <http://revigo.irb.hr> در دسترس است.

پایگاه زیستی Biomart

Biomart یک پایگاه داده از اسامی و اطلاعات ژن‌های گونه‌های مختلف و نیز مسیرها و داده‌های زیستی دیگر از آن‌ها است. Biomart یک بسته جدید Bioconductor است که منابع داده BioMart را با نرم‌افزار تجزیه و تحلیل داده‌ها در Bioconductor ادغام می‌کند. این بسته می‌تواند طیف وسیعی از شناسه‌های ژن یا محصول ژنی (مانند شناسه‌های کاوشگر Entrez-Gene و Affymetrix) را با اطلاعاتی مانند نماد ژن، مختصات کروموزومی، هستی‌شناسی ژن و حاشیه‌نویسی OMIM حاشیه‌نویسی کند. علاوه بر این، Biomart بازیابی توالی‌های ژنومی و اطلاعات پلی‌مورفیسم تک نوکلئوتیدی را امکان‌پذیر می‌کند، که می‌تواند در تجزیه و تحلیل داده‌ها استفاده شود. بازیابی سریع و به‌روز داده‌ها امکان‌پذیر است، زیرا بسته درخواست‌های مستقیم SQL را در پایگاه‌های داده Bio-Mart (به عنوان مثال Ensembl) اجرا می‌کند. بسته Biomart یکپارچه‌سازی دقیق از پایگاه‌های داده BioMart بزرگ، عمومی یا محلی نصب شده با تجزیه و تحلیل داده‌ها در Bioconductor فراهم می‌کند که یک محیط قدرتمند برای داده کاوی بیولوژیکی ایجاد می‌کند.

پایگاه زیستی PANTHER

پایگاه داده PANTHER برای تجزیه و تحلیل بازده توالی‌های پروتئین طراحی شده است. یکی از ویژگی‌های کلیدی، هستی‌شناسی ساده عملکرد پروتئین است که امکان مرور پایگاه داده توسط توابع بیولوژیکی را فراهم می‌سازد. مدل‌های آماری (مدل‌های مارکوف پنهان یا HMM) از هر یک از این گروه‌ها ساخته می‌شوند. مزیت این رویکرد این است که توالی‌های جدید را می‌توان به صورت خودکار طبقه‌بندی کرد. برای اطمینان از طبقه‌بندی عملکردی دقیق، HMMها نه تنها برای خانواده‌ها، بلکه برای زیرخانواده‌های عملکردی نیز ساخته می‌شوند. هم‌ترازی‌های توالی چندگانه و درختان فیلوژنتیک برای هر خانواده در

دسترس هستند. نسخه فعلی PANTHER شامل توالی‌هایی از همه ارگانیس‌ها در پایگاه داده پروتئین GenBank است و HMMها برای طبقه‌بندی محصولات ژنی در کل ژنوم انسان و مگس سرکه *Melanogaster* استفاده شده است.

پایگاه زیستی TAIR

منبع اطلاعاتی آرابیدوپسیس (TAIR؛ <http://arabidopsis.org>) یک منبع وب جامع از بیولوژی آرابیدوپسیس برای دانشمندان گیاهی است. TAIR اطلاعات مربوط به ژن‌ها، پروتئین‌ها، عملکرد ژن، بیان ژن ارتولوگ‌ها، فنوتیپ‌های جهش یافته، مواد بیولوژیکی مانند کلون‌ها و ذخایر دانه، نشانگرهای ژنتیکی، نقشه‌های ژنتیکی، سازمان‌دهی ژنوم، تصاویر گیاهان جهش یافته، محلی‌سازی‌های زیر سلولی پروتئین را مدیریت و ادغام می‌کند. این واحد در درجه اول بر روی برخی از روش‌های اساسی برای جستجو، مرور، تجسم و تجزیه و تحلیل اطلاعات در مورد ژن‌ها و ژنوم آرابیدوپسیس تمرکز دارد.

پایگاه زیستی Enrichr

تجزیه و تحلیل غنی‌سازی، یک روش محبوب برای تجزیه و تحلیل مجموعه‌های ژنی تولید شده توسط آزمایش‌های گسترده ژنوم است. Enrichr در حال حاضر شامل مجموعه بزرگی از مجموعه‌های ژنی متنوع است که برای تجزیه و تحلیل و دانلود در دسترس است. ویژگی‌های جدیدی به Enrichr اضافه شده‌اند که شامل توانایی ارسال مجموعه‌های فازی، آپلود فایل‌های BED است. به طور کلی، Enrichr یک منبع جامع برای مجموعه‌های ژنی و موتور جستجویی است که دانش بیولوژیکی را برای اکتشافات بیولوژیکی بیشتر جمع‌آوری می‌کند. Enrichr به صورت رایگان در <http://amp.pharm.mssm.edu/Enrichr> در دسترس است.

جدول ۱: مروری کوتاه بر مطالعات میکروآرای در حوزه علوم دامی در داخل کشور و خلاصه خروجی های بدست آمده

Table 1. Mini review on microarray research in animal science reports and brief outputs and results

نویسنده و تاریخ	شماره دسترسی	گونه مورد مطالعه	صفت	تعداد نمونه	جزئیات نمونه	ژنهای مهم کلیدی	هستی شناسی ژنها
شریفی و همکاران (۱۳۹۶)	GSE15205 GSE25217 GSE24560 GSE25413 GSE32186 GSE50685	گاو	ورم پستان	۱۲۷	۵۴ سالم ۷۳ الوده به E.coli	TNF-alpha سیتوکینین	ایمنی، التهاب، پروتئولیز سلول، مرگ یاخته ای، بیان ژن اران ا بلی مرز
گلزار و همکاران (۱۳۹۹)	GSE38345	گاو	تکامل سلول تخم	۳۶	۶ تیمار با ۶ تکرار زمانی	MRPS10, MRPS18A, MRPL16 MRPL17 MRPL22, ATP5B ATP5C1 KPNA1 RGS2 TESC KRTB FN KRT1B	انهدام میتوکندری، ترجمه ژنهای میتوکندری، ATP سنتتاز
پیمان خواجه غیائی و همکاران (۱۴۰۱)	GSE18290 GSE24094	گاو	توسعه جنین گاو	۲۸	۱۸ رویان ۸ پلاستوسیت	FBX5 RNR123 CTNBL CUL1 PA264	مسیر آپاتوز، تقسیم میتوز، سنتز استروئید، متابولیسم اسید فولیک، سنتز ATP اسپلایسوزوم یو بیکوتینین پروتئولیز بیوزنژ سیستم ایمنی پیری سلول
رضایی سینگی و همکاران (۱۳۹۹)	GSE23592	مرغ	استرس گرمایی	۱۸	۶ مغز ۶ کبد ۶ ماهیچه	RPGR UBASH38 ALCAM RPL28 SAC3 BOLLA1 BTBD9 FN1 SOD2 HMOY1 PCK2 STAT3 HSPA1A IL6 IFIH1 PAR14 IL1B ICAMI CCL5 MYC KRAS HSPA8 ATP58 JAK1 FOXA2	پاسخ به استرس مسیرهای متابولیکی سلولی، بیوزنژ، آناتومی سختاری
مهدی نژاد و همکاران (۱۴۰۰)	GSE33667	پرندگان حیات وحش	مهاجرت	۲۲	۱۶ مهاجر		
شبه ناصری (۱۳۹۹)	GSM628120 GSM628145 GSM628137 GSM628146 GSM628152	گاو	ماهیه اسکلتی	۱۳	از تولد تا ۳۰ ماهگی		در مسیرهای تولید انسولین WNT IGF1
وجدی و همکاران (۱۳۹۴)	GSE47599	گاو	ورم پستان	۳۲	ورم پستان		فرایندهای سیستم ایمنی، پاسخهای سیستم ایمنی، فعال شدن نوترفیلها سیتوکینین ها
اسکندرنسب و همکاران (۱۴۰۰)	GSE33680	گاو	تولید شیر	۸ عدد	۴ پرشیر ۴ کم شیر		متابولیسم گلیکولپتید IL6 EGFR1

نتیجه‌گیری کلی

بیان ژن یک فرآیند دقیق تنظیم شده است که به سلول اجازه می‌دهد در محیط‌های مختلف از خود واکنش نشان دهد. این فرآیند را به عنوان کلید روشن و یا خاموش کردن ژن نیز معرفی می‌کنند که هدف از آن کنترل زمان ساخت پروتئین‌ها، کنترل حجم یا میزان تولید پروتئین‌های مورد نیاز بدن و افزایش یا کاهش بیان آن‌ها است. رمزگشایی از معماری ژنتیکی و ترانسکریپتوم و یا سیستم بیولوژیکی یک فرد که برتری در عملکرد در یک زمینه اختصاصی دارد، می‌تواند دارای اهمیت باشد. به عنوان مثال، یک فرد نسبت به سایر افراد که در مواجهه با بیماری تحت تأثیر قرار نمی‌گیرد می‌تواند دارای سیستم ایمنی برتری باشد و این برتری می‌تواند ناشی از افزایش و یا کاهش بیان تعدادی از ژن‌ها در بدن باشد. در واقع این فرد به محرک‌های محیطی مخرب عکس‌العمل قوی نشان می‌دهد و می‌توان از اهمیت و سیمای مولکولی ژنوم و نسخه‌برداری آن و نقاط تنظیمی این موضوع پرده‌برداری کرد. لذا، ابزارهای مولکولی که بتوانند اندازه‌گیری غلظت RNA را در یک بافت اختصاصی کمی‌سازی کنند، ایده‌آل خواهد بود. فناوری ریزآرایه بر مبنای قوانین qPCR و فناوری هیبریداسیون سادرن بلاتینگ می‌باشد که در مقیاس بزرگ‌تر برای اندازه‌گیری میزان بیان تعداد متعددی ژن مورد استفاده قرار می‌گیرد. همچنین این روش مبتنی بر هیبراسیون در ترکیب با فناوری نانو و استفاده از رنگ‌های فلئورسنت نشان‌دار است. این فناوری به طور همزمان شناسایی الگوی بیان افتراقی چندین هزار توالی خاص یا مناطق کدکننده را عهده‌دار است. روش‌های پیش‌پردازش داده‌ها و آزمون‌های نرمال‌سازی آن‌ها، استفاده از روش تجزیه مؤلفه‌های اصلی، رسم نمودار حرارتی، کلاسترینگ و رسم دندروگرام تست همبستگی داده و روش پیشرفته TSN از جمله روش‌های پیشرو هستند. همچنین، چالش مقایسه میانگین چندگانه، وجود خطای نوع یک و مفهوم Fold change نیز به طور اجمالی مورد بحث قرار گرفته است. در مطالعهٔ مروری حاضر، مبانی تئوریک، ساختار داده‌های خام، تحلیل‌های آماری و تفسیر نتایج فناوری ریزآرایه مورد بحث قرار گرفت و به‌طور مستقل سعی شد از داده خام تا تفسیر بررسی شود و همچنین چالش‌های موجود را با اطلاعاتی که داده می‌شود به خوبی پشت سر قرار داد.

منابع:

- Andrews, R. M. (2004). Patterns of embryonic development. Reptilian incubation: environment, evolution and behaviour, 75-102.
- Aneh Shebh Naseri, F., Bahreini Behzadi, M., Roudbari, Z., Eskandarynasab, S. (2020). Gene expression analysis of skeletal muscle to identify genes that influence bovine growth process. *Veterinary Researches & Biological Products*, 2020; 33(4): 123-131.
- Angenendt, P. (2005). Progress in protein and antibody microarray technology. *Drug Discovery Today*, 10(7), 503-511.
- Asyali, M. H., Colak, D., Demirkaya, O., and Inan, M. S. (2006). Gene expression profile classification: a review. *Current Bioinformatics*, 1(1): 55-73.
- Balakrishnan, R., Harris, M. A., Huntley, R., Van Auken, K., and Cherry, J. M. (2013). A guide to best practices for Gene Ontology (GO) manual annotation. Database, 2013.
- Barad, S., Sela, N., Dubey, A.K., Kumar, D., Luria, N., Ment, D., Cohen, S., Schaffer, A.A. and Prusky, D. (2017). Differential gene expression in tomato fruit and *Colletotrichum gloeosporioides* during colonization of the RNAi-SIPH tomato line with reduced fruit acidity and higher pH. *BMC Genomics*, 18(1): 1-21.

- Botchkarev, V. A., Gdula, M. R., Mardaryev, A. N., Sharov, A. A., and Fessing, M. Y. (2012). Epigenetic regulation of gene expression in keratinocytes. *Journal of Investigative Dermatology*, 132(11): 2505-2521.
- Brennan, G.P., Vitsios, D.M., Casey, S., Looney, A.M., Hallberg, B., Henshall, D.C., Boylan, G.B., Murray, D.M. and Mooney, C. (2018). RNA-sequencing analysis of umbilical cord plasma microRNAs from healthy newborns. *PloS One*, 13(12), e0207952.
- Cooper, R. (1990). S. Natusch and G. Swainson. William Swainson FRS, FLS, &c. The anatomy of a nineteenth century naturalist. Third (revised) edition. Privately published (S. Natusch, 46 Owhiro Bay Parade), Wellington: 1987. Pp [6], 184; illustrated. Price: not stated. ISBN 0-473-00569-7.
- Dittami, S. M., Hostyeva, V., Egge, E. S., Kegel, J. U., Eikrem, W., and Edvardsen, B. (2013). Seasonal dynamics of harmful algae in outer Oslofjorden monitored by microarray, qPCR, and microscopy. *Environmental Science and Pollution Research*, 20(10): 6719-6732.
- Edwards, D. (2003). Non-linear normalization and background correction in one-channel cDNA microarray studies. *Bioinformatics*, 19(7): 825-833.
- Ehrenreich, A. (2006). DNA microarray technology for the microbiologist: an overview. *Applied Microbiology and Biotechnology*, 73(2): 255-273.
- Eslandarynasab S, Roudbari Z, Mohammadabadi M. (2021). Transcriptome Analysis of Mammary Gland Tissue to Identify Major Genes Involved in the Milk Production. *Rap*, 12 (32) :160-166
- Farhadi, G., Roshanfekr, H., Fayazi, J., Nazari, M., Behdani, E. (2020). Identification of Mitochondrial Genes Affecting the Development of Oocyte Under FSH Hormone by Using a Microarray Data in Cow. *Iranian Journal of Animal Science Research*, 12(2): 200-201.
- 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.
- Hall, D. A., Ptacek, J., and Snyder, M. (2007). Protein microarray technology. *Mechanisms of Ageing and Development*, 128(1): 161-167.
- Hansen, P. J., and Tribulo, P. (2019). Regulation of present and future development by maternal regulatory signals acting on the embryo during the morula to blastocyst transition—insights from the cow. *Biology of Reproduction*, 101(3): 526-537.
- Hardy, K. (1999). Apoptosis in the human embryo. *Reviews of Reproduction*, 4: 125-134.
- Heller, M. J. (2002). DNA microarray technology: devices, systems, and applications. *Annual Review of Biomedical Engineering*, 4(1): 129-153.
- Hon, G. C., Hawkins, R. D., and Ren, B. (2009). Predictive chromatin signatures in the mammalian genome. *Human Molecular Genetics*, 18(R2): R195-R201.
- Hong, L. Z., Zhou, L., Zou, R., Khoo, C. M., Chew, A. L. S., Chin, C. L., and Shih, S. J. (2021). Systematic evaluation of multiple qPCR platforms, NanoString and miRNA-Seq for microRNA biomarker discovery in human biofluids. *Scientific Reports*, 11(1): 1-11.
- Khaje Ghiassi, P., Amin Afshar, M., Ghaderi Zefrehei, M., Emam jome kashan, N., and Honarvar, M. (2022). 'Bioinformatics analysis of gene sets enrichment between bovine embryonic cleavage and development blastocysts using DNA microarray data', *Iranian Journal of Animal Science*, 53(1): 53-68. (In Farsi)
- Khaje Ghiassi, P., Amin Afshar, M., Ghaderi Zefrehei, M., Emam jome kashan, N., Honarvar, M. (2022). Bioinformatics analysis of gene sets enrichment between bovine embryonic cleavage and development blastocysts using DNA microarray data. *Iranian Journal of animal Science*, 53(1): 53-68.
- Kulkarni, A., Anderson, A. G., Merullo, D. P., and Konopka, G. (2019). Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Current Opinion in Biotechnology*, 58: 129-136.
- Laikhter, A. (2015). Overview of qPCR molecular probes. *Current Protocols Essential Laboratory Techniques*, 11(1): 10-5.
- Mahdavi Nezhad, Z., Rajabi Maham, H., Ramezani, M., Sharifi Zarchi, A. (2021). Differential Expression analysis of microarray data for three migratory and one non-migratory bird as candidates of migratory and non-migratory species. *Journal of Animal Environment*, 13(3): 57-68.
- Mazzone, R., Zwergel, C., Artico, M., Taurone, S., Ralli, M., Greco, A., and Mai, A. (2019). The emerging role of epigenetics in human autoimmune disorders. *Clinical Epigenetics*, 11(1): 1-15.
- Morey, J. S., Ryan, J. C., and Van Dolah, F. M. (2006). Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. *Biological Procedures Online*, 8(1): 175-193.
- Nelson, E. (1992). Shepherd, W. and Cook, W. The Botanic Garden, Wellington. A New Zealand history 1840–1987. Millwood Press, Wellington: 1988. Pp 396; illustrated. Price: none stated. ISBN: 0-908582-79-X.
- Pollock, S. E. (2012). Data quality rules in the analytic health repository. The University of Utah.
- Popovici, V., Goldstein, D. R., Antonov, J., Jaggi, R., Delorenzi, M., and Wirapati, P. (2009). Selecting control genes for RT-QPCR using public microarray data. *BMC Bioinformatics*, 10(1): 1-10.
- Ramakrishnan, R., Bao, P., and Müller, U. R. (2005). High sensitivity expression profiling. In *Microarray Technology and Its Applications* (pp. 229-250). Springer, Berlin, Heidelberg.

Retis-Resendiz, A. M., González-García, I. N., León-Juárez, M., Camacho-Arroyo, I., Cerbón, M., and Vázquez-Martínez, E. R. (2021). The role of epigenetic mechanisms in the regulation of gene expression in the cyclical endometrium. *Clinical Epigenetics*, 13(1): 1-23.

Rezaei Sinaki, M., Sadeghi, M., Bahrami, A., Moradi Shahrabak, M. (2020). Identification of genes, biological pathways and signaling affecting heat stress with microarray data sets in poultry. *Iranian Journal of animal Science*, 2020; 51(3): 243-251.

Russell, S., Meadows, L. A., and Russell, R. R. (2008). *Microarray technology in practice*. Academic Press.

Sharifi, S., Pakdel, A., Ebrahimie, E. (2017). Meta-analysis of transcriptomic data of mammary gland infected by *Escherichia coli* Bacteria in dairy cows. *Iranian Journal of animal Science*, 48(3): 343-352.

Silver, J. D., Ritchie, M. E., and Smyth, G. K. (2009). Microarray background correction: maximum likelihood estimation for the normal-exponential convolution. *Biostatistics*, 10(2): 352-363.

Singh, S., Sonkar, S. K., Sonkar, G. K., and Mahdi, A. A. (2019). Diabetic kidney disease: a systematic review on the role of epigenetics as diagnostic and prognostic marker. *Diabetes/Metabolism Research and Reviews*, 35(5), e3155.

Vajdi Hokamabad, S., Alijani, S., Daghighkia, H., Zali, H., Khodakarim, S., Begdeli Pashaei, M. (2015). Bioinformatics analysis of *E. coli* causing mastitis in Holstein dairy cattle by using microarray data. *Koomesh*, 17 (1) :214-223.

Yavari, F., Towhidkhah, F., Gharibzadeh, S., Khanteymooori, A. R., and Homayounpour, M. M. (2008). Modeling large-scale gene regulatory networks using gene ontology-based clustering and dynamic bayesian networks. In 2008 2nd International Conference on Bioinformatics and Biomedical Engineering (pp. 297-300). IEEE.

Yu, M., and Ren, B. (2017). The three-dimensional organization of mammalian genomes. *Annual review of cell and developmental biology*, 33, 265.

Zhou, V. W., Goren, A., and Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nature Reviews Genetics*, 12(1): 7-18.