شبکه های بیان ژنی برای آنالیز داده ماکروواری DNA

حکیمه زالی[2,1]، مصطفی رضایی طاویرانی[*3]، جعفر سلیمیان[4]، غلامرضا اولاد[4]، صیاد بسطامی نژاد[5]

1) کمیته تحقیقات دانشجوی، دانشگاه علوم پزشکی شهید بهشتی
2) دانشکده پیراپزشکی ، دانشگاه علوم پزشکی شهید بهشتی
3) مرکز تحقیقات پروتئومیکس ، دانشکده پیراپزشکی ، دانشگاه علوم پزشکی شهید بهشتی
4) مرکز تحقیقات میکروبیولوژی، دانشگاه علوم پزشکی بقیه الله(عج)
5) مرکز تحقیقات میکروبشناسی بالینی ، دانشکده پزشکی ، دانشگاه علوم پزشکی ایلام

تاریخ دریافت: **تاریخ پذیرش:**

چکیده

بر خلاف دیدگاه تقلیل گرایانه بیولوژی کلاسیک ، رویکرد کل نگر در بیولوژی با انفجار در پیشرفت تکنیک هایی با تکنولوژی بالا و تولید حجم عظیم داده خود را نشان داده است . اکنون چالش بیولوژیست ها کشف روش های تحلیل این داده ها است تا بتوانند در رسیدن به درک سیستم پویای پیچیده حیات کمک کنند. در بین تکنولوژی های بسیار پیشرفته اخیر که بیشترعمومی هستند میکروواری DNA از مشهورترین آن ها است . میکروواری،سطوح بیانی هزاران ژن را به طور همزمان مورد بررسی قرار داده و یک تصویر کلی از فعالیت رونویسی سلول د رشرایط چندگانه فراهم می کند. میکروواری دری را به روی قلمروی جدید جستجوهای بیولوژیکی شامل توضیح ژنها در فرایندهای ویژه مانند چرخه سلولی ، رشد و پیشرفت سلولی، ارزیابی اثر اختلالات شیمیایی و ژنتیکی و شناسایی ژنهای مرتبط با انواع بیماری ها فراهم نموده است. حجم خالص داده تولید شده به وسیله مطالعات میکروواری نیاز به پیشرفت ابزارهای کامپیوتری آنالیز آماری پیشرفته است .در این مطالعه به مرور روش های آماری که برپایه تئوری گراف است پراخته شده است. ساختن شبکه های بیانی ژن (GCN)، ادغامGCN با دیگر داده ها، آنالیز GCN ها و کاربرد GCN برای مطالعه سرطان به طور کامل مورد بررسی قرار گرفت. در نهایت می توان گفت که مطالعه ژنوم از طریق میکروواری شامل نمونه های بیشتر و در طیف وسیعی از گونه ها امکان پذیر است و در آینده کاربرد روش های متا آنالیز برای ادغام حجم بزرگی از داده ها مانند هم ترازی شبکه ای احتمالا کمک به روشن ساختن مشابهت و تفاوت بین طیف وسیعی از گونه ها ،بافت ها و حالتهای بیماری خواهد شد.

واژه های کلیدی: میکروواری ، شبکه بیان ژن، آنالیز شبکه

***نویسنده مسئول:** مرکز تحقیقات پروتئومیکس ، دانشکده پیراپزشکی ، دانشگاه علوم پزشکی شهید بهشتی

Email: rezaei.tavirani@ibb.ut.ac.ir

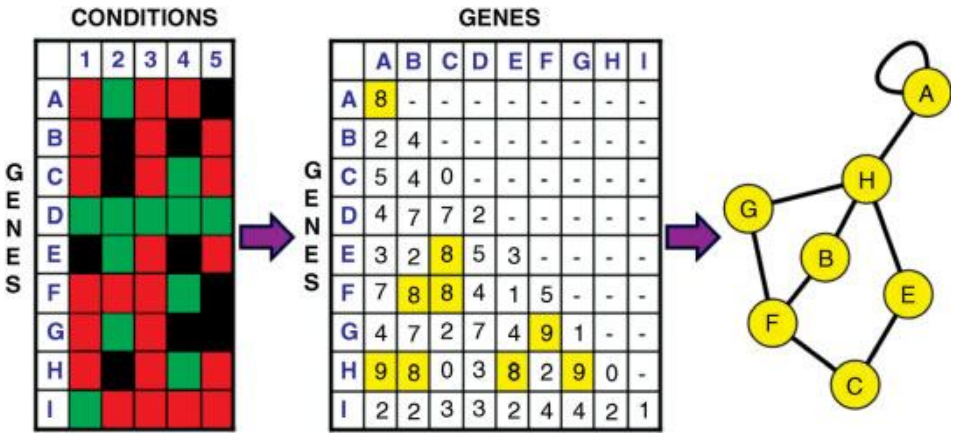
مقدمه

فلسفه بیولوژی کلاسیک براساس دیدگاه تقلیل گرایانه است و برای سالها محققان برای موفقیت این رویکرد به مطالعه و پژوهش پرداختند.اما با توجه به پیشرفت سریع و فوری تکنولوژی های توالی یابی شواهد زیادی نشان داده اند که عملکرد های سلولی به ندرت ایزولاسیون انجام می دهند و بر خلاف آن ، اغلب شامل اینترکشن مربوط به تعداد زیادی از اجزاء و ترکیبات سلولی است که یکی از چالش هایی که اکنون بیولوژیست ها با آن روبرو هستند رسیدن به درک این سیستم پویای پیچیده است.

سالها اخیر انفجاری در پیشرفت تکنیک هایی با تکنولوژی بالا برای دستیابی و نشان دادن جنبه های مختلف فعالیت ژن به وجود آمده است. اکنون با استفاده از این تکنولوژی های جدید، شناسایی ارتباطات جدید بین ژنها با قدرت تفکیک پذیری بالاتر نسبت به گذشته ممکن ساخته است. برای مثال خیلی زود این امکان وجود خواهد داشت که نقشه کل مجموعه اینترکشن های پروتئین برای هرارگانیسم نیز مشخص شود. دسترسی این مجموعه داده وسیع ژنوم یک فرصت بی نظیر برای کشف ویژگیهای سلولی جدید از منظر سیستمی می دهد و توانایی دانشمندان را در پیش بینی صحیح عملکرد ژن در حجم های وسیع افزایش می دهد. دربین تکنولوژی های بسیارپیشرفته اخیر که بیشترعمومی هستند میکروآری DNA از مشهورترین آن ها است . میکروآری،سطوح بیانی هزاران ژن را به طور همزمان مورد بررسی قرارداده و یک تصویرکلی از فعالیت رونویسی سلول درشرایط چندگانه فراهم می کند.قبل از پیشرفت این تکنیک تعیین زمانی و مکانی یک ژن که رونویسی می شود یک تلاش زمانبر و دشوار بود . میکروآری دری را به روی قلمروی جدید جستجوی بیولوژیکی شامل توضیح ژنها در فرایندهای ویژه مانند چرخه سلولی و رشد و پیشرفت، ارزیابی اثر اختلالات شیمیایی و ژنتیکی و شناسایی ژنهای مرتبط با انواع بیماری ها فراهم نمود.

حجم خالص داده تولید شده به وسیله مطالعات میکروآری نیاز به پیشرفت ابزارهای کامپیوتری آنالیز آماری پیشرفته است.در این مطالعه به مرور یک زیرکلاس از این روش ها که برپایه تئوری گراف است پراخته شده است.

به طورخاص مطالعات وتکنیک های بحث شده در این جا به مقایسه الگوهای بیان ژن برروی یک نمره مشابهت و جفت که بیش از یک آستانه است به ایجاد یک شبکه که ایجاد پایه ای برای آنالیز بعدی متصل می شود(شکل 1,11) شبکه هایی که با این روش از داده میکروآری ایجاد می شوند اغلب به عنوان شبکه های بیانی ژن gene coexpression networks(GCNs) مطرح می شود و کمک می کنند به شناسایی ژنهایی که الگوهای بیان مشابهی رو در بین شرایط تجربی متفاوت نشان می دهند.ژنهایی که در چنین الگوهایی را بازی می کنند اغلب به وسیله برنامه تنظیم رونویسی مشابهی در عملکرد ویژه ای از سلول کنترل می شوند یا عضوی از یک مسیر مشابه هستند و یا با هم تشکیل کمپلکس پروتئینی می دهند.



رونویسی ژن

سلول یک سیستم خیلی پیچیده است که بایدبه انواعی از تغییرات شرایط داخلی و خارجی به منظور عملکردهای بی شمارش پاسخ دهد.به طور باورنکردنی تقریبا هم اطلاعات لازم جهت حفظ این سیستم در DNA ژنومی کد شده است.واحد بنیادی اطلاعات در DNA ژن است اگر چه تعدادی از ژنهابه طورثابتی کمتر یا بیشتر رونویسی می شوند، بیشتر برطبق نیاز رونویسی انجام می شود و تنظیم رونویسی یک ژن شامل یک پاسخ خوش صدا ست که به طیف وسیعی از شرایط داخل سلولی و خارج سلولی پاسخ می دهد. رمزگشایی کد تنظیمی ژنوم(زمان و چگونگی خاموش وروشن شدن ژنها)یک هدف اصلی بیولوژی و تکنیک میکرو اری است که با اندازه گیری سطوح بیان ژنها تحت شرایط مختلف سلولی به این هدف کمک می کند.

میکرواریDNA

یک آزمایش میکرواری سطوح بیانی هزاران ژن تحت یک شرایط تجربی را مورد بررسی قرارمی دهد.بیشترمطالعات شامل آزمایشات میکرواری چندگانه ، یک طیفی از شرایط رو پوشش می دهند.برای مثال،مطالعات میکرواری اغلب سطوح بیانی ژنها درشرایط رشد مختلف در مراحل پیشرفت و تکامل متفاوت و یا نمونه های بافتی بیمار و سالم را مورد مقایسه قرار می دهند. چندین صفحه آزمایش مختلف برای اجرای آزمایشات میکرواری به وجود آمده است و شامل پروپ یا DNA ژنومی است که در یک آزمایش میکرواری نمونه های ازDNA مکمل یا RNA مکمل به پروپ هیبرید می شودو فراوانی هر تداعی در نمونه به وسیله آشکارسازی اتوماتیک فلورسانت کمی سازی می شود.یک پروفایل بیان ژنی اغلب بیانگر یک بردار از مقدار عددی است و معمولا با یک تصویر heatmap(شکل 1,11چپ) نشان داده می شوند که پروپ هایی با بیان بالا به رنگ قرمز و پروپ هایی با بیان پایین به رنگ سبز مشخص می شوند.

بیشتر مطالعات شامل هردو تکرار بیولوژیکی و تکرار آزمایشگاهی است. تکرارها معمولا به یک نمره متوسط از پروپ ها تبدیل می شود و سطوح بیانی جهت گرفتن گزارشی از اغتشات ذاتی و تنوع میکرواری ها مورد نرمال سازی قرار می گیرد.مقالات زیادی در نرمال سازی داده های میکرواری وجود دارد . جهت مطالعه بیشتر و گرفتن اطلاعات در این مورد به رفرنس های 9و10 مراجعه شود. معمولاداده نرمان سازی شده با مقادیر نزدیک صفرنشان دهنده این است که هیچ تغییری در سطح بیان پروپ ها صورت نگرفته است در حالی که اعداد مثبت و منفی به ترتیب بیانگر افزایش و کاهش سطح بیان است.بعد از نرمال سازی فیلتر بیشتری قرار داده می شود مانند برداشتن پروپ هایی که سطوح بیان آنها در شرایط متنوع معنی دار نیستند.

مطالعات متنوعی از میکرواری گزارش شده است که بر روی لیست ژنی با بیان متفاوت، آنالیزهای آماری مانند آنالیز واریانس (ANOVA) آنالیز enrichmentمجموعه ژنی (GSEA)یا LIMMAگزارش نموده اند . تفاوت بیان ژن با تفاوت معنی دار سطوح بیان در بین دو یا تعداد بیشتری از شرایط مختلف دیده می شود.برای مثال،بیان متفاوت به طورمتوالی جهت شناسایی ژنهایی که دارای بیان بالایی در یک حالت بیماری خاص (درمقابل نرمال و شرایط غیر بیماری) هستند ،استفاده می شود در این مطالعه با تمرکز بر روش های مبتنی برشبکه به جای شناسایی روابط بر اساس بیان ژن به بررسی پرداخته می شود. ارتباطات برپایه سطوح بیان همزمان باعث تسخیرتمایل کلی برای یک جفت یا گروهی از ژنها که دارای سطوح بیان مشابه در آن شرایط هستند ایجاد می کند.(سطح بیان بالا دریک مجموعه از شرایط و پایین در گروهی دیگر)

شبکه ها

میکرواری ایجاد حجم عظیمی از داده می کند که نیاز به روش های پیچیده ای برای داده کاوی ، تفسیر و بیان دوباره است. یک ساختارداده عمومی که برای این کار به وجود آمد شبکه است. مزیت روشهای بر پایه شبکه قدرتشان ،در سازماندهی حجم عظیم داده است که با دسترسی به روش های تئوری گراف ،داده کاوی و تفسیر تصویری به آسانی میسر می گردد.شبکه ها یک راه طبیعی در تفسیرها ی مدل بین ژنها که

متشکل از با گره های ژنی نماینده و لبه هایی است که شامل انواع مختلفی از اینترکشن های متنوع است که از منابع داده مختلف استنتاج شده است . شبکه برای انواع وسیعی از مسائل بیولوژی مانند نقش اینترکشن پروتئین ها، کشف جایگاه اتصال فاکتور رونویسی و مدل سازی اینترکشن های ژنتیکی به کاربرده می شود.بعلاوه چندین ابزار برای تصویرسازی وآنالیز شبکه های بیولوژی مانند VisAnt، cytoscape و tYNAپیشرفت کرده است.برای مرور برنامه های شبکه بیولوژیکی، به رفرنس 23مراجعه شود. در ادامه برروی برنامه های شبکه متمرکز شده تا ارتباطات بیان ژن های استنتاج شده از آزمایشات میکروارای DNA مورد بررسی قرارداده شود.

ساختن شبکه بیان ژن (GCN)

GCN به جفت ژنهایی (گره هایی) اتلاق می شود که به طور معنی داری تحت شرایط (آزمایش میکروارای) در یک مطالعه باهم بیان می شوند. اولین مرحله ایجاد یکGCN نمره همه جفت بردارهای ژنی است و مرحله دوم انتخاب یک نمره آستانه ای است و همه جفت ژنهای که نمره های آن ها از این عدد تجاوز می کند به هم وصل می شوند. نتیجه GCN می تواند هم بدون وزن (اینترکشن های دوتایی است :حضور یا غایب)یا وزن دار(مانند احتمالات نماینده احتمال یک ارتباط داده شده) باشد. دراین فصل طرحها نمره دهی متفاوتی مورد بحث قرار خواهندگرفت وروشهایی برای انتخاب یک آستانه نمره دهی مرورخواهدشد.

به طورعمومی شبکه ها می توانند هم به صورت جهت دار و هم بدون جهت باشد.در اینجا برروی شبکه های بدون جهت که بیانگر ارتباط دوطرفه بیانی است اما لزوما بیانگر علیت نیست تمرکز خواهد شد.چندین روش برای استنباط شبکه های تنظیمی جهت دار از داده ها بیانی تولیده شده از ژنهایی که بیانشان متوقف شده یا دارای بیان چند برابر شده اند به وجود آمده است. روش های شبکه ای جهت دار به خوبی توسط Markowetz و Spang (28) مرور شده است .

قالب و فرمت داده و نحوه نمایش داده

یک مطالعه بیانی شامل آزمایش های میکروارای چندگانه است عمکرد هر آزمایش یک شرایط سلولی خاص را نشان می دهد .نتیجه داده ها اغلب به صورت ماتریس بیان می شود که ردیف ها پروپ ها (ژنها)را نشان می دهند و ستون ها بیانگر شرایط آزمایش هستند .

$$M = \begin{bmatrix} M_{1,1} & M_{1,2} & \cdots & M_{1,N} \\ M_{2,1} & M_{2,2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ M_{G,1} & M_{G,2} & \cdots & M_{G,N} \end{bmatrix}$$

دراین جا ماتریس M شامل Gردیف و Nستون است که به ترتیب بیانگرتعداد ژنها و تعداد شرایط هستند.هربخش در ماتریس مسئول سطح بیان یک ژن در یک شرایط است.هرردیف نماینده سطح بیان یک ژن درهمه شرایط آزمایش است و به عنوان بردار ژن برمی گردد(تصویر1,11چپ)که اغلب برای شناسایی گروههای ژنی و شرایطی که الگوهای مشابهی را نشان می دهند خوشه بندی می شود.

(بخش 1,3,5,11را ببینید)ماتریس بیانی به عنوان داده و برای همه آنالیز های بعدی شبکه به کارگرفته می شود که اولین مرحله بامحاسبه نمره بیانی جفت ژن انجام می شود.

محاسبه نمره های جفت ژن

چندین روش برای مقایسه پروفایل های بیانی جفت ژنهاوجود دارد.دراین بخش اول برروی چهارروش نمره دهی عمومی شامل فاصله اقلیدسی، اطلاعات دوطرفه،ضریب همبستگی پیرسون وهمبستگی رتبه بندی اسپیرمن تمرکز می شود. اگر چه روش های دیگری پیشنهاد شده است،بیشترمطالعات تا امروز یکی از 4روش شرح داده شده در اینجا را استفاده می کنند.هر روش نمره دهی یکی ازجنبه های داده رابه خوبی بررسی می کند،فاصله اقلیدسی فاصله هندسی بین دو بردار را اندازه می گیرد و برای هر دو ویژگی بزرگی و جهت بردار ژنی محاسبه صورت می گیرد.اطلاعات دوطرفه سطوح بیانی یک ژن چه مقدارباعث کاهش غیرمطمئن سطوح بیان دیگری می

شود را محاسبه می کند.ضریب همبستگی پیرسون تمایل دوبردار درافزایش یا کاهش با هم اندازه گیری می کند و کیفیت ارتباط تشابه و مطابقت کلی آنها را به ما می گوید.همبستگی رتبه بندی اسپیرمن با همبستگی پیرسون یکسان است،اما برروی رتبه(rank) بیان به جای عدد(value) بیان عمل می کند و بنابراین به طور قدرتمندی برای نشان دادنoutlierها مناسب است. هرروش مزایا و معایب مختص به خود را برای شناسایی ارتباط بیانی دارد .

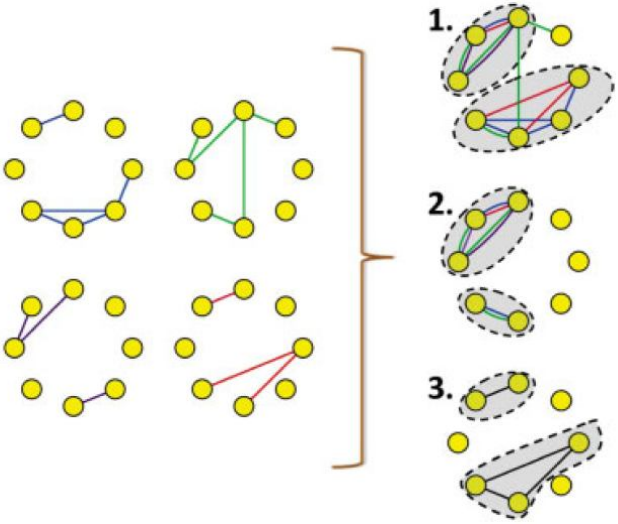
انتخاب یک آستانه

وقتی یک مقیاس نمره دهی انتخاب می شود و همه جفت ژنها نمره دهی می شوند،مرحله بعد انتخاب یک نمره آستانه و ایجاد یک GCN برای اتصال همه جفت ژنها با نمره هایی است که ازاین آستانه تجاوز می کنند.روش های آستانه سازی ساده همراه با انتخاب یکcutoff نمره دهی است.این روش وقتی که یک تفسیر واضح از طرح نمردهی در دسترس باشد می تواند مفید باشد.برای مثال،چندین مطالعه از ضریب همبستگی پیرسون با cutoff 50/ که مسئول همبستگی میانه است یا cutoff 80/ مسئول همبستگی قوی و بالا است. یک جایگزین برای انتخاب یک آستانه ردیفی،انتخاب یکcutoff که مسئول سطح معنی داری مورد نظر است.تبدیل z فیشر تعیین یک تخمین معنی داری از ضریب همبستگی پیرسون را بر پایه تعداد شرایط درآن محاسبه می شود.چنین استراتژی می تواند به این دلیل مفید باشد که تعداد شرایط مورد آزمایش در مطالعه میکرواری می تواند به طور قابل توجهی متنوع باشد برای مثال یک مطالعه شامل بخش های مجزای ازیک ژن ویژه که ممکن است شامل تنهایک مشت از شرایط باشد درحالی که یک مطالعه کلینیکی ممکن است شامل صدها نمونه بیمارباشد.به طور مستقیم ،یک همبستگی پیرسون9/ که بیشتر از 4شرایط را محاسبه نموده باید نمره بدتری ازیک همبستگی9/ باشد که بیشتر از 100 شرایط را محاسبه نمودهمانطور که مطرح بالای همبستگی می تواند در طی مقدارکمی از شرایط با شانس بدست می آید.تبدیلz فیشر معنی داری بالاتری از ژنهابا همبستگی قویتری در تعداد بیشتری از نمونه ها تعیین می کند(38). آستانه اپتمم ساختار شبکه ناشناخته است در این مورد آستانه ها بر پایه جایگشت(permutation) می تواند مفید باشد که یک تخمین معنی داری بر پایه تجربی نیز بدست آورد چنین تخمینی به وسیله مقایسه یک ماتریس داده بیانی جایگشت تصادفی بدست آمده است روشهای برپایه جایگشتی برای انتخابcutoff شبکه میانی در چندین مطالعه مورد استفاده قرارگرفته است(39و31). روش های دیگری نیز برای انتخاب یک cutoffبرای ایجاد شبکه میانی پیشنهاد شده است.یکی از معمولی ترین آنها انتخاب یک آستانه برپایه دانش بیولوژیکی قبلی است(37و40-41)برای مثال اگریک فرآیند بیولوژیکی ویژه ای تحت آزمایش قرار گیرد یک آستانه می تواند مانند آنچه در اکثریت ژنها ی تشکیل دهنده آنها که به وسیله اتصالات بیانی متصل شده اند انتخاب شود .متناوبا یک آستانه می تواند به طور محکمی بر پایه توپولوژی شبکه اصولی انتخاب شود.برای مثال تعداد ارتباطات به طور تدریجی می تواند کاهش یابد تا ضریب خوشه بندی شبکه با رجوع به رونوشت تصادفیآن بهینه شود. روش های دیگر نیز بر پایه تئوری ماتریس تصادفی پیشنهاد شده است انتخاب یک استانه همبستگی از حداقل چگالی شبکه بدست می آید.

ادغام GCNها با دیگر داده ها

همانطور که تکنولوژی میکرواری تکامل می یابد روش ها متا آنالیز تلاش دارد تا ویژگیهای معمول در مجموعه داده را بیابد که اغلب ایجاد پیش بینی قوی تری نسبت به آنها که بر اساس مجموعه داده های تنها ساخته می شوند ایجاد می کند. مطالعات اخیر و تکنولوژی هایی که در تکنیک متا آنالیز جهت ترکیب مجموعه دادهای بیولوژیک چند گانه در قالب یک شبکه معرفی می شوند. چنین روش هایی می تواند به طور وسیعی به سه رویکرد اولیه طبقه بندی شود(شکل 2,11) . درررویکرد اول،شبکه ترکیبی از واحد اتصالاتی که از مجموعه داده جداگانه گرفته شده است(شکل 2,11(1)) بدست می آید. روش های برپایه این رویکرد حساسیت بالا را با هزینه اختصاصیت (specificity)بدست می آورند. به طور عمومی رویکرد

بر پایه اتحاد(union) وقتی که برای مجموعه داده میکروواری جداگانه مورد استفاده قرار می گیرد،شبکه حاصله برای اینکه استفاده کاربردی داشته باشد احتمالا دارای اتصالات خیلی چگالی خواهد بود. رویکرد دوم محدود به اتصالاتی می شود که به وسیله انواع اینترکشن های چندگانه حمایت می شود(شکل 2,11(2)).برطبق مطالعات اخیر بر پایه روشهای احتمالی ادغام شبکه تمرکز دارد این روش ها منابع داده



ادغام مجموعه داده های بیانی چندگانه

نتایج مطالعات متعدد نشان داده است که ژنهایی که در شرایط چندگانه هم بیان هستند احتمالا عضو مسیر مشابهی هستند یا جزئی ا کمپلس پروتئینی محسوب می گردند.اما شناسایی ارتباطات تنظیم ژنها با هم از الگوهای بیانی یک آزمایش و احد با حضور مسیرهای overlapو شلوغی در داده پیچیده می شود.بنابراین اگر چه میکروواری یک دیدگاه کلی با ارزش از الگوهای بیان ژن ارائه می کند پیش بینی های اختصاصی اغلب با یک نسبتی از مثبت کاذب بالا همراه است.مطالعات اخیر مزایای افزایش را برابری با لا بودن حجم داده های میکروواری ارائه نموده که با شناسایی ژنهایی که در شرایط چندگانه در مطالعات چندگانه دارای تنظیم همزمان هستند را نشان می دهد.چنین استراتژی هایی به یک حرکت افزایش یافته اخیر به سمت فهرست کردن سیستماتیک مجموعه داده ها بیانی در پایگاه داده هایی مانند GEO، Array Express وStandford Microarray Database کمک کرده است.بیشتر پایگاه داده های میکروواری نیاز مصنوعی که بر اساس توانایی شان در پیش بینی ویژگی های مختلف (مانند عملکرد به اشتراک گذاشته شده) ویک آستانه بر اساس ترکیب وزنی انتخاب می کند(شکل 2,11(3)) چنین روش هایی دارای مزیت تفاوت ذاتی در کیفیت بین انواع داده ها ی متنوع هستند اخیرا دیده شده که دارای یک موج محبوبیت هستند.

به فرمت های استاندارد شده ای برای بیان شرایط آزمایش هستند که استفاده از روش های اتوماتیک را برای شناسایی مطالعاتی که شرایط خاصی را مورد آنالیز قرار داده اند هموار می سازد. انواع ترکیب نتایج حاصل مجموعه داده های بیانی چندگانه شامل: ادغام داده در یک گونه ،ادغام داده در بین گونه ،ادغام منابع داده هتروژن توسط روش هایی که در چهار چوبی برپایه شبکه است.

روش ها ی تقاطعی (intersection) و اتحادی (union)

یک کلاس از روش های تجمعی شبکه شامل اتحاد ارتباطی در شبکه می شود(شکل 2,11(1)) ترکیب انواع مختلف اینترکشن در یک شبکه اتحادی به این روش اجازه تعیین ارتباطات بین انواع داده های ترکیب کننده می دهد.اگرچه تجمع شبکه ساده از نظر مفهومی ثابت شده که یک روش قدرتمنداست .مطالعات آغازین تجمع شبکه جفت ژنهای را شناسایی نمود که ارتباط آنها از منابع داده چندگانه مانند پروتئین پروتئین،بیان ژن و اینترکشن های هم فنوتیپ (co phenotype) ژنی بدست آمده بود.یک استراتژی

جایگزین یافت زیر شبکه هایی است که با انواع اینترکشن های چندگانه که لزوما مشترک (overlap) نبودند حمایت می شود.چنین رویکردهایی دارای مزیت مجموعه داده هایی است که ایجاد ارتوگونال اماتکمیلی مانند اینترکشن های ژنتیکی سنتزی کشنده و اینترکشن های پروتئینی می کند. مطالعات دیگری از تجمع شبکه جهت تعیین ارتباطات پیچیده بین انواع اینترکشن های مختلف و همچنین تشخیص و خصوصیت یابی حالتهای بیماری انسانی استفاده کرده اند.

روش های احتمالی

اگر چه به طور واضح روش های تجمع ساده در بالا شرح داده شده مطالعات اخیر برروی روش های پیشرفته تری برای ادغام منابع داده در شبکه احتمالی ساده تمرکز کرده اند.شبکه های احتمالی می توانند به عنوان نماینده یک مشابهت عملکردی متوسط در انواع بافت های مختلف،مراحل پیشرفت و شرایط مختلف در نظر گرفته شوند،بنابراین شبکه نهایی می تواند همه آمیزش های ممکن را در طیف وسیعی از حالتهای مختلف بیان نماید.چندین مزیت با گرفتن یک دید احتمالی از آمیزش ژنی بدست می آید که شامل ترکیب و تلفیق مطلقی از عدم اطمینان و شک و توانایی محاسبه برای Pleiotropy ژنی(زمانی اتفاق می افتد که یک ژن صفات فنوتیپی چندگانه را تحت تاثیر قرار می دهد) است.

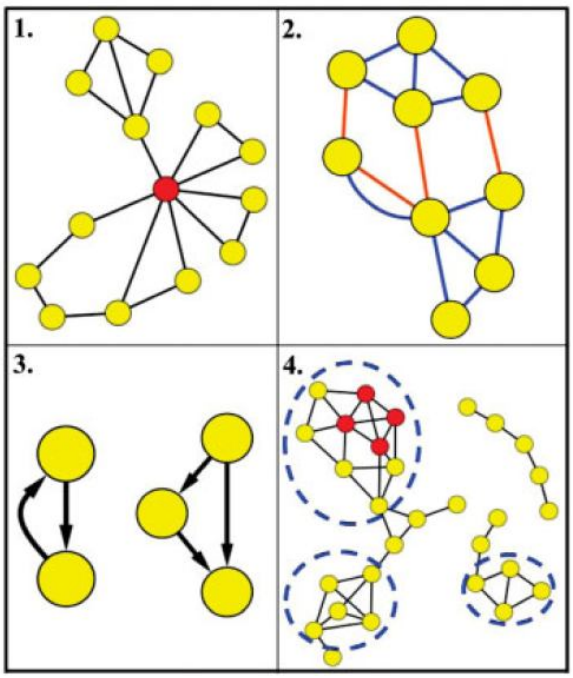
روش احتمالی برای ادغام داده ها به منابع داده قبل از ترکیب آنها به صورت یک شبکه واحد در فرآیندی که تفاوت در صحت را محاسبه می کند وزن می دهد.

برای تایید کردن چهار شبکه نشان داده شده در شکل 2,11 را ملاحظه کنید شبکه ای با رنگ قرمز بیانگر سطح بالای اطمینان اینترکشن هاست،شبکه آبی رنگ بیانگر اطمینان متوسط اینترکشن هاست و شبکه سبز و بنفش بیانگر اینترکشن با اطمینان پایین است.یک روش احتمالی می تواند همه این شدت ها ی اینترکشن را یکپارچه سازد و نتیجتا یک اینترکشن قرمز برای حضور در شبکه نهایی کافی است(شکل 2,11(3)).برعکس سایر انواع اینترکشن ها برای حضور در شبکه نهایی کافی نیستند و تنها زمانی می توانند به کار گرفته شوندکه یک شدت متوسط (آبی رنگ) توسط حداقل یک دیگر از اینترکشن های دیگر حمایت شود(شکل2,1(3)) با این طریق روش های احتمالی یک توافق منطقی بین رویکردهای اتحادی (شکل 2,11(1)) و رویکرد های تقاطعی(شکل2,11(2)) بیان می کنند.

یک موضوع مهم در ترکیب منابع مختلف از اطلاعات در یک قالب احتمالی، چگونگی وزن دهی به انواع مختلف داده است.بیشتر روش ها به هر مجموعه داده ای که براساس توانائیشان در به دام اندازی ارتباطات عملکردی شناخته شده است وزن دهی می کنند.(برای مثال ژنهای شرکت کننده در فرآیندهای بیولوژیکی مشابه شناخته می شوند.) یک قالب عمومی شبکه بیضین یک مدل گرافی در اتصال پراکندگی های احتمالی چند متغیره است که مستقل شرطی بین متغیر ها را به دام می اندازد.به علت توانایی شان در به کاربردن noise و شرح فرآیندهای احتمالی پیچیده شبکه بیضین به طور گسترده ای برای ساختن شبکه های بیانی به کاربرده می شود مطالعات شبکه های بیضین برای ادغام منابع داده ای چندگانه در یک شبکه عملکردی منفرد در مخمر و دیگر گونه های یوکاریوتی به کارگرفته شد(97).

آنالیز GCNها

پس از ساختن GCN داده کاوی از هزاران اینترکشن بر تکنیک هایی استوار است که بر پایه ایده های تئوری گراف هستند. در این اینجا سه دسته عمومی از روشهای پیش بینی های بیولوژیک از شبکه های بیان ژن مورد بحث قرارمی گیرد.در بخش اول شناسایی ژنها با بسیاری از الگوها ی اینترکشن که اغلب به قطب های شبکه(hub) برمی گردد(شکل 3,11(1)) سپس روش های کشف موتیف های شبکه که مسئول ساختارهای زیرگرافی مانند مثلث های اینترکشن یا لوپ های پس خورو پیش خورهستند



در انتهای بخش رویکردهایی برای آشکارسازی زیر نواحی یک شبکه(اغلب به عنوان مدول های ژنی) که در شکل3,11.(4) نشان داده شده است پرداخته می شود.

hub های شبکه

اکثر شبکه های بیولوژیکی از نوع شبکه قابل گسترش (Scale-free network)هستند بدین معنی که در شبکه، بیشتر ژنها دارای ارتباطاتی با تنها چند hub ژنی (گره ای دارای بیشترین ارتباطات ژنی)هستند.این نوع پراکندگی که به ندرت در شبکه های تصادفی(Random network) ظاهر می شود گفته می شود که ایجاد خواص سودمند برای سیستم کلی می کند که شامل قدرت در شکست و عدم موفقیت گره ژن تصادفی است و یک فاصله متوسط کوتاهتر بین هر دو گره ایجاد می کند . ژنهای hub به منظور اهمیت شان در ساختار کلی شبکه ،در بسیاری از شبکه ها بیولوژی نشان داده که یک گرایش افزایش یافته به حضور دارند که ضروری به نظر می رسد و این ژن ها دارای ارتولوگ در دیگر ارگانیسم هستند که در حین تکامل خیلی حفاظت شده است و در بیماریهای پیچیده نیز به عنوان یک شاخص دیده می شوند. به دلیل خصوصیات گفته شده به طورخاصی در GCNها حضور دارند(39و34).

موتیف های شبکه

یکی از راههای تعیین خصوصیتات کلی یک شبکه ، امتحان کردن الگوهای اینترکشنی زیرگراف های محلی آن شبکه است. ساختار زیرگراف محلی که یک شبکه غنی از آنهاست و مقایسه شده با همتاهای تصادفی آن به عنوان موتیف های شبکه معروف هستند.موتیف ها ی شبکه می تواند شامل هر تعداد گره با ساده ترین الگو از سه گره متصل به شکل یک مثلث باشد.برای مثال مثلث آبی رنگ درشکل 3,11(2) می تواند بیانگر یک موتیف شبکه باشد که دارای فرکانس بالایی در شبکه هستند در حالی که مثلث های نارنجی این گونه نیستند. موتیف های شبکه بیانگر بلوک های ساختار ی ساده ای از سیستم های پیچیده هستند و بنابراین ممکن است بینشی از اصول طراحی ساختاری شبکه ارائه کنند.یک مزیت موتیف های شبکه این است که آنها می توانند جهت تعیین نقش های فرض شده ژنها تنها بر اساس قرارگیری آنها در توپو لوژی شبکه محلی مورد استفاده قرارگیرند.

برخی موتیف های شبکه مستقیما قابل تفسیر از لحاظ بیولوژیکی هستند مانند لوپ های پس خور و پیش خور که در شکل 3,11(3) دیده می شود.هردو موتیف نقش مهمی در تنظیم رونویسی ژن بازی می کنند و افزایش بیانشان در شبکه های تنظیم رونویسی در گونه های چند اشتقاقی دلالت بیشتری بر ارتباط بیولوژیکی شان است.مطالعات اخیر به جستجوی چنین

موتیف های اینترکشنی سه راهی در GCNپرداخته اند.کشف موتیف های شبکه در شبکه های ادغامی بینشی از ارتباط بین انواع مختلف اینترکشن می دهد در حالی که ظهور رتبه(order) بیشتر موضوع های(theme) شبکه شامل وقوع چندگانه موتیف های شبکه نیز در شبکه های بیولوژی غنی است.

مدول های ژنی

اغلب هدف از یک آزمایش میکرواری تعیین گروه هایی از ژنها با سطوح بیان مشابه تحت شرایط است.مدول های ژنی شناخته شده در این روش ممکن است بر برنامه های تنظیمی متمایزی دلالت کنند و اغلب به خوبی مسیرها و کمپلس های پروتئینی را نشان می دهند.برای مثال قوی ترین سیگنال در مطالعه ماکرواری اغلب کلاسترینگ (گروه بندی با هم براساس پروفایل بیانی مشابه) ژنهایی که زیر واحدهای ریبوزومی را کد می کنند ،است. شواهد یک طبیعت مدولار را پیشنهاد می کند که سازماندهی سیستم های سلولی را به عهده دارد و مطالعات متعددی نیز به جستجو و مطالعه فرآیندهای سلولی در سطح مدول های ژنی بر خلاف مطالعه تک ژنی در یک زمان پرداختند(152-150) . یک طیف وسیعی از روش ها برای شناسایی مدول های ژنی با استفاده از داده های میکرواری به کارگرفته شده است. از این روش ها گروهی که به طور مستقیم روی ماتریس داده میکرواری عمل می کنند(شناسایی مدول های ژنی با روش خوشه بندی بردار ژنی) و دسته ای که GCN را به عنوان ورودی می پذیرند(روش هایی برپایه شبکه) شامل می شود.

یکی از روش های خوشه بندی بردار ژنی ،خوشه بندی سلسله مراتبی است. در این نوع خوشه بندی ، ارتباط بین ژنها به وسیله یک درخت نشان داده می شود که طول شاخه ها انعکاسی از درجه شباهت بین بردارهای ژنی است که به وسیله روش نمره دهی مانند ضریب همبستگی پیرسون بدست می آید.خوشه ها از درخت تشکیل شده اند که با گروه بندی تعدادی از نزدیکترین جفت گره ها و جایگزینی آنها بایک گره منفرد که بیانگر میانگین مجموعه است.دیگر روش های عمومی خوشه بندی شامل خوشه بندی k-means است یک روش تفکیک کننده جزیی است که جمع فاصله میانه خوشه را کوچک می کند و نقشه ها خود سازمان داده که بیانگر خوشه ها به عنوان گره ها در یک گرافی است که به یک فضای با ابعاد بیشتر (درابتدا به صورت تصادفی) نقشه بندی می شود و به طور تکراری تنظیم می شود تا داده ها سازگار شوند.روش هایی که ژنها را به خوشه ها از یک زیرمجموعه از شرایط شناسایی می کند به روش ها دو کلاسترینگ (biclustering) یا کلاسترینگ دو طرفه است.چنین روش هایی در به دام اندازی ژنهایی که تحت یک شرایط خاص هم بیان هستند موثر است و قادر به شناسایی ژنهایی که در چند مسیر حضور دارند نیز هستند(155,153). روشهای خوشه بندی بردار ژنی که در بالا گفته شد به عنوان روش های بدون ناظر دسته بندی شوند.روش های تحت نظارت یک کلاس جدا هستند که اطلاعات بیولوژی قبلی مانند انواع اینترکشن ها و مسیرهای بیولوژیک شناخته شده در آن شرکت دارند. این روش ها یک راه مفید در ترکیب داده های میکرواری با دیگر اطلاعات هستند،اما آنها همچنین می توانند باعث محدود شدن کشف های جدید با تمرکز بر روی ارتباطات بیولوژیکی پایه گذاری شده موجود شوند.بیشتر اینکه روش های تحت نظارت به طور بالایی وابسته به کیفیت داده هایی هستند که آموزش دیده اند. با این حال روش های کلاسترنیگ تحت نظارت ثابت شده که برای کشف ارتباطات جدید بین ژنها و فرایند های بیولوژیک شناخته شده مفید هستند.

از دسته دوم روش های شناسایی مدول های ژنی بر اساس روش های شبکه ای بود که روش های خوشه بندی شبکه ای است که در آن مجموعه ژنهایی که دارای چگالی ارتباطات بیشتر نسبت به یکدیگر هستند نسبت به آنها که با شانس ارتباط دارند شناسایی می کند و یک چگالی ارتباط کلی از شبکه ارائه می نماید. زیر شبکه هایی با ارتباطات چگال بیانگر گروهی از ژنها هستند که دارای ارتباط دوطرفه هستند و به طور محکمی با هم بیان می شوند و بنابراین احتمالا مدول های ژنی هستند که باهم تنظیم می شوند. یک مزیت بزرگ رویکرد خوشه بندی شبکه ای توانایی آن

در تحمل داده ها از دست داده شده است.یعنی برای همه اعضا یک جدول ژنی لازم نیست که در همه شرایط که آنها دارای بیان مشابه هستند تست شوند و ایجاد آن زیر مجموعه هایی از ژنها که به اندازه کافی چگال به طور معنی درای ارتباط برقرار کرده اند علاوه بر آن به جای این حقیقت که بسیاری از ژنها در چند فرآیند شرکت می کنند بیشتر الگوریتم خوشه بندی شبکه ای غیر استاندارد تنها ژنها را به یک خوشه ارجاع می دهند.روشهای شبکه ای اجازه اشتراک بین گروههای هم تنظیم ژنی را می دهند.

ترکیبات ارتباطی

ساده ترین راه گروه بندی یک شبکه به زیر شبکه ها،شناسایی ترکیبات ارتباطی (اتصالی) آنهاست.یک ترکیب ارتباطی یک مجموعه ماکزییم از گره هایی است که هر جفت گره می تواند به حداقل یک مسیر برسد. به طور کلی سه ترکیب ارتباطی در شبکه وجود دارد که در شکل 3,11(4) (چپ،بالاراست،پایین راست) مشخص شده است. یک مزیت ترکیبات ارتباطی این است که به راحتی می توان آنها را از طریق یک جستجوی ردیفی یا جستجوی عمیق درجه اول (depth- first search) شناسایی شود. اما اکثر شبکه های بیولوژیکی به اندازه کافی مدولار برای ترکیبات ارتباطی نیستند که در عمل بتواند با بسیاری از آنها که تنها شامل یک ترکیب ارتباطی بزرگ است مفید واقع شود. به عنوان یک نتیجه،ترکیبات ارتباطی اغلب می توانند به زیر شبکه ها ی بیشتر شکسته شود (برای مثال سمت چپ ترین ترکیب ارتباطی در شکل 3,11(4) می تواند به دوترکیب ارتباطی که با دایره با خط غیر ممتد به طور محکمی شکسته شده است).یک عیب دیگر ترکیبات ارتباطی این است که یک مسیر بین یک مجموعه از گره ها نمی تواند یک درجه بالایی از چگالی ارتباطی را ضمانت نمایند که یک نمونه شاهد برای ترکیب ارتباطی در سمت راست بالای تصویر 3,11(4) نشان داده شده است.

دسته های شبکه

یک رویکرد جایگزین که یک چگالی بالای ارتباطی را تضمین می کند شناسایی دسته های شبکه است. یک دسته یک مجموعه از گره هایی است که به طور کاملی به یکدیگر متصل شده اند.برای مثال گره های قرمز در شکل3,11(4) نماینده یک دسته از سایز4 رانشان می دهد.با کمک تعریف یک دسته دارای kگره دارای k*(k-1)/2 ارتباط (همه ترکیبات ممکن جفت گره ها) است. می توان دید که یک دسته از سایزk (k دسته) نیز شامل دسته های چندگانه ای از سایز-k-1,k 2,...,1 است که الگوریتم ها برای شناسایی دسته ها ی حداکثر اغلب در عمل به کار گرفته می شود. مشکل شناسایی همه دسته های بیشین در یک شبکه از نظر محاسباتی گران است بنابراین روش های پیشرفته مانند الگوریتم انشعاب و تجدید(شاخه و کران) دقیق توسط Cliquer (165)به کارگرفته شده یا روش استخراج تحت پوشش راس به وسیله Voy و همکارانش (166) باید برای شناسایی هایشان به کارگرفته شود.

حاشیه نویسی مدول های ژنی

به علت طبیعت مدولار شبکه های سلولی، امروز استفاده از مدول های ژنی برای تعیین عملکرد ژنی به طور وسیعی مورد توجه قرارمی گیرند.زیر شبکه های شناسایی شده در شبکه ها اغلب جهت تخصیص ژنها به یک عملکردی بیولوژیکی مفروض مورد استفاده قرار می گیرند. عملکرد بیولوژیک ژنهای حاشیه نویسی نشده یک مدول را بر اساس اصل گناه از سوی انجمن(guilt by association) با توجه به حضور ژنهایی که عملکرد شناخته شده دارند تعیین می شود. براین منظور اغلب از پایگاه داده هستی شناسی های کنترل شده ای مانند GO، KEGG یاGen MAPP مورد استفاده قرارمی گیرد.

GCN برای مطالعه سرطان

سرطان یک بیماری است که تنظیم سلولی به هم خورده و از دست رفته است و بسیاری از محققینی که از مطالعات آنالیز میکروآری استفاده کرده اند تلاش دارند تا کمک کنند به حل مکانیسم های پیچیده ای که در انتقال از حالتهای سرطانی به نرمال وجود دارد(187-185). بیشتر مطالعات برروی شناسایی ژنهایی با بیان متفاوت مانند ژنهایی که دارای بیان بالایی در حالتهای توموری در مقایسه با بافت نرمال هستند.یک رویکرد جایگزین شناسایی ژنهایی است که

به طور متفاوتی هم بیان هستند(مانند یک گروه از ژنها که باهم در بافت نرمال بیان می شوند اما در بافت توموری نیستند) که ممکن است نمایانگر برنامه تنظیمی باشد که در سرطان تخریب شده یا متوقف شده است.کارهای اخیر بر روی شناسایی مسیرهای تمرکز کرده اند که الگوهای بیان سرطانی ناسازگاری را نشان می دهند .پایگاه داده Oncomine (202) شامل رویکردهای متفاوتی از داده های هزاران میکروواری است که در طیف گسترده ای از انواع سرطان ها ،مراحل پیشرفت و زمینه های بیماری ، مورد آزمایش قرارداده است. رویکردهای شبکه ای در تعدادی از مطالعات ارائه شده است که کمک می کند به شناسایی ژنها و گروه هایی از ژنها که باعث تبدیل بافت نرمال به یک مرحله بدخیمی می شود.قالب های شبکه ای به طور ویژه ای برای این کار جذاب هستند و آنها را قادر به کشف تغییرات ارتباطات شبکه رونویسی در حالتهای بیماری می شود. هدف نهایی این مطالعات تعیین بیومارکرها و هدف های درمانی خاصی برای هر زیر گونه از سرطان است .برخلاف موفقیت مطالعاتی که در بالا ذکر شد،سرطان یک بیماری خیلی پیچیده است و در حال حاضر هیچ تست کلینیکی تائید شده ای بر پایه کاربردهای میکروواری وجود ندارد. در آینده تلاش های جمعی در حجم بالا مانند اطلس ژنوم سرطان می تواند این قول را بدهد که درک از پایه های سلولی سرطان افزایش یابد.

نتیجه گیری

دراین مطالعه خلاصه ای از روشهای شبکه ای برای آنالیز داده بیان ژن تولیده شده به وسیله میکروواری DNA ارائه گردید.GCN ها ایجاد یک وسیله مستقیم برای تجسم و مطالعه اینترکشن های پیچیده بین هزاران ژن در یک زمان فراهم می کنند.قالب شبکه برای ادغام با دیگر منابع داده خیلی مناسب است و کاربرد الگوریتم ها بر پایه تئوری گراف را قادر می سازد .آنالیز GCN دیدگاه هایی از خصوصیات کلی سیستم ها ارائه می کند و این در حالی است که همزمان ایجاد پایه ای برای تنظیم ژنی خاص و فرضیه عملکرد می کند.

GCN ها یک روش راحت مفید برای آنالیز داده میکروواری هستند.اما این مهم است که اهمیت امتحان داده بیان اوریجیال تحت شناسایی گروه ها ی ژنی مورد نظر تائید شود، همچنان که ژنهایی که می توانند در همبستگی قوی با یک تنوعی از دلایل غیر ضروری و زائد تظاهر یابند.برای مثال اگر شرایط آزمایش بسیار مشابه ،مورد آزمایش قرارگیرد،ممکن است این انتظار را ایجاد کند که یک بخش قابل توجهی از ژنوم ممکن است الگوهای بیان مشابه ای را نشان دهند.بیشتر اینکه ژنها با سطوح بیان که با هم در یک شرایط منفرد حداکثر هستند و در غیر اینصورت دارای بیان پایینی هستند دارای یک همبستگی بسیار قوی هستند. با امتحان داده بیانی اوریجینال (مانند شکل یک heatmap) چنین مواردی می تواند به راحتی از پیش بینی های مربوطه بیولوژیکی جدا شود.

در بلوغ تکنولوژی های بسیار پیشرفته نیاز به روش های محاسباتی نیز همچنان رو به افزایش خواهد بود.اگر چه تکنولوژی های توالی یابی جدید مانند 454،Solexa وSolid ،فرصتهای بی نظیر را پیشنهاد می کنند ،آنها همچنین تولید میزان های شوک برانگیزی از داده هستند .همچنین مطالعات میکروواری شامل نمونه های بیشتر(مانند گروههای بیمار بزرگتر) و یک طیف وسیعی از گونه ها را پوشش می دهد.در آینده کاربرد روش های متا آنالیز برای ادغام حجم بزرگ داده مانند هم ترازی (alignment) شبکه احتمالا کمک به روشن ساختن مشابهت و تفاوت بین طیف وسیعی از گونه ها ،بافت ها و حالتهای بیماری خواهد کرد.

سپاسگزاری

این مطالعه حاصل طرح تحقیقاتی با کد -1-1391 159-8981 است که توسط کمیته پژوهشی دانشجویان دانشگاه علوم پزشکی شهید بهشتی تصویب گردیده است .از مرکز تحقیقات پروتئومیکس دانشگاه علوم پزشکی شهید بهشتی که در اجرای پروژه همکاری داشتند تشکر می گردد.

References

1-Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA, 1996; 93: 10614-9.

2-Dudoit S, Yang YH, Callow MJ, Speed TP. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. 2000; Technical report # 578. Available from: http://www.stat.Berkeley.EDU/users/terry/zarray/Html/papersindex.html.

3-Satagopan JM, Panageas KS. Tutorial in biostatistics, a tatistical perspective on gene expression data analysis. Stat Med 2003; 22:481–99.

4-Golub TR, Slonim DK, Tamayo P, Gaasenbeek CHM, Mesirov JP, Coller H, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999; 286:531-7.

5-Leukemia, in Mosby's Medical, Nursing &Allied Health Dictionary. 4th ed. Mosby-YearBook:Inc;1994.P. 903.

6-Matutes E. T-cell prolymphocytic leukemia, a rare variant of mature post-thymic T-cell leukemias, has distinct clinical and laboratory characteristics and a poor prognosis. Cancer Cont J 1998; 5 :14-9.

7-Valbuena JR, Herling M, Admirand JH, Padula A, Jones D, Medeiros LJ. T-cell prolymphocytic leukemia involving extramedullary sites. Am J Clin Pathol 2005;123: 456-64.

8-Ross JA, Kasum CM, Davies SM, Jacobs DR, Folsom AR, Potter JD. Diet and risk of leukemia in the Iowa Women's Health Study. Cancer Epidemiol Biomarkers Prev 2002;11: 777-81.

9- Kinzler KW, Vogelstein B. Genetic basis of human cancer. McGraw-Hill

10-Wiernik, Peter H. Adult leukemias. New York: BC Decker; 2001. P. 3–15.

11-Robinette MS, Cotter S, Van de W. Quick look series in veterinary medicine: Hematology. Teton New Media 2001; 105.

12-Kong SW, Pu WT, Park PJ. A multivariate approach for integrating genome-wide expression data and biological knowledge. Bioinformatics 2006; 22:2373-780.

13- Liu Q, Dinu I, Adewale AJ, Potter JD, Yasui Y. Comparative evaluation of gene-set analysis methods. Bioinformatics 2007;8:431.

14- Goeman JJ, Bühlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics 2007; 23:980-7.

15- Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004; 5: R80.

16- Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:3-6.

17-Wei Huang D, Sherman BT, Richard AL. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Prot 2009; 4: 4-8.

18-Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503-11.

19-Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. J Am Statistic Assoc 2002; 97: 75-9.

20- Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. Nature Genet 1999; 21:10-4.

21- Habeck M. DNA microarray technology to revolutionise cancer treatment. Lancet Oncol 2001; 2:5.

22-Nguyen DV, Arpat AB, Wang N, Carroll RJ. DNA microarray experiments: biological and technological aspects. Biometrics 2002; 58:701-17.

Gene Expression Networks to Analysis DNA Microarray Data

Zali H[1,2] , Rezaei Tavirani M[3], Salimian J[4], Aolad G.R[4], BasataminejadS[5]*

(Received: Accepted:)

Abstract

Unlike the reductionist views of classical biology, holistic approach in biology is shown with an explosion in the development of high-tech techniques to produce large amounts of data. Now the challenge for biologists is to discover ways to analyze this data in order to ability to support understanding the complex dynamic systems of life. In recent advanced technologies that are more public, DNA microarray are most famous. Microarray simultaneously examines expression levels of thousands of genes and provides a snapshot of the transcriptional activity of the cells in multiple conditions. Microarray have provided chance , especially in search of new territory, to describe the genes involved in biological processes such as cell cycle, growth and development of cell, assessment of chemical and genetic disorders and to identify genes associated with various diseases. Sheer volume of data produced by microarray studies need to develop advanced statistical analysis computer tools. In this study has been reviewed statistical methods based on graph theory. Construction the gene expression network (GCN), GCN integration with other data, GCN analysis and application GCN for cancer research fully described. Finally we can say that study of the genome through microarray includes more samples and is possible in wide range of the species and in future applications of meta-analysis methods to integrate large amounts of data such as network alignment may help to clarify the similarities and differences between a wide range of species, tissues and disease states.

Keywords: Microarray, Gene expression network, Network analysis

1. Students' Research Committee, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. Faculties of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3. Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
4. Microbiology Research Center, Baghyatollah University of Medical Sciences, Tehran, Iran
5. Dept of Clinical Microbiology, Faculty of Medicine, Ilam University of Medical Sciences, Ilam, Iran
**(corresponding author)*