مطالعه روش های آنالیز شبکه جهت تفسیر فنوتیپ های پیچیده در شبکه های بیولوژیک

حکیمه زالی[۱]، مصطفی رضایی طاویرانی[۲*]،خدیجه حیدربیگی[۳]، مهدی شهریاری نور[٤]

۱)دانشکده پیراپزشکی ، دانشگاه علوم پزشکی شهید بهشتی

۲)مرکز تحقیقات پروتئومیکس ، دانشکده پیراپزشکی ، دانشگاه علوم پزشکی شهید بهشتی

۳)گروه بافت شناسی دانشکده پزشکی دانشگاه علوم پزشکی ایلام

٤)گروه میکروبیولوژی، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات گیلان

تاریخ دریافت: ۹۱/۱۰/۱۵ **تاریخ پذیرش: ۹۱/۱۲/۱۱**

چکیده

آنالیز شبکه ژنی بخش مهمی از مطالعات بیولوژی سیستم ها است. در مقایسه با مطالعات سنتی ژنوتیپ / فنوتیپ که با تمرکز بر ایجاد روابط بین ژنهای منفرد و ویژگی مورد نظر است ، آنالیز شبکه ما را قادر می سازد تا به مشاهده تمام ژن ها با هم پرداخته که به نوبه خود عملکرد بیولوژیکی درست را نشان می دهد. آنالیز شبکه همچنین کمک به استنتاج اطلاعاتی سودمند از شبکه می کند و در کشف فرآیندهای بیولوژیکی از یک شبکه نیز یاری می رساند. در این مطالعه روش های اصلی و برنامه های کاربردی در آنالیز شبکه به سه موضوع اصلی جهت تفسیر فنوتیپ های پیچیده می پردازند. نخستین جنبه ،شناسایی اهمیت هر گره در شبکه است که مهمترین یا ضروری ترین ژن شبکه، همچنین کم اهمیت ترین و قابل چشم پوشی ترین ژن تعیین می شود. جنبه دیگر، شناسایی ژن های عملکردی تر در کل شبکه است که با اندازه گیری ارتباطات مستقیم یک ژن با سایر ژن ها و توجه به کل ارتباطات شبکه بدست می آید. شناسایی مسیرها یا جریان ها در شبکه ها با کمک ژنهای شناخته شده به عنوان ورودی و خروجی شبکه جنبه نهایی مورد بحث در آنالیز شبکه است.اگر چه این روش ها دارای مزایای زیادی است بیولوژی شبکه هنوز با بسیاری از چالش ها روبرو است بنابراین بیشتر روش ها در هم ادغام شده تا ابزاری مهم را برای آنالیز شبکه ایجاد نمایند .تسلط بر این روش ها ضروری به نظر می رسد اما برای درک بیولوژی کافی نیست. بنابراین مهمترین مسئله انجام سوالی درست است تا ابزارهای آنالیز شبکه مناسب انتخاب شود و جهت ارزیابی نتیجه آنالیز نیز آزمایشات تجربی صورت گیرد. در نهایت می توان گفت بیولوژی شبکه و بیولوژی ملکولی دارای هدف اصلی مشابه هستند و آن درک بهتر فرآیند های بیولوژیکی و کشف مکانیسم بیمارهای انسانی است.

واژه های کلیدی: گره، لبه، شبکه بیان ژن، آنالیز شبکه

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

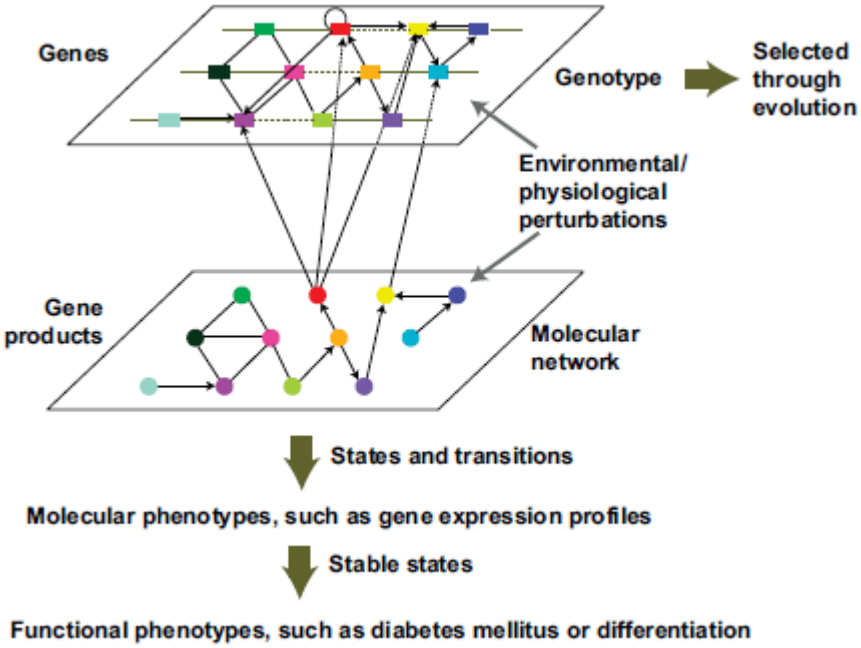
Email: rezaei.tavirani@ibb.ut.ac.ir

مقدمه

در سالهای اخیر انفجاری در پیشرفت تکنیک هایی با تکنولوژی بالا برای دستیابی و نشان دادن جنبه های مختلف فعالیت ژن به وجود آمده است.استفاده از این تکنولوژی های جدید،اکنون جهت شناسایی ارتباطات جدید بین ژنها با 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 هر ارگانیسم 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 ها در پیش بینی صحیح عملکرد ژن در حجم های وسیع افزایش می دهد. ایجاد حجم عظیمی از داده نیاز به روش های پیچیده ای برای داده کاوی ، تفسیر و بیان دوباره است. یک ساختار داده 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 آسانی میسر می گردد.. شبکه ها یک راه طبیعی در تفسیرها ی مدل بین ژنها که متشکل از گره های ژنی 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 های رونویسی و مدل سازی اینترکشن های ژنتیکی به کاربرده می شود[۱].

بر خلاف نظریه مندلی که رابطه یک ژن یک فنوتیپ را مطرح نمود C.H.Waddington در سال ۱۹۵۷ چشم انداز اپیژنتیکی را جهت نشان دادن اثرات چند ژنی و شبکه ای ژن ها در متابولیسم سلولی مطرح نمود. با توجه به دانش کنونی ما، متابولیسم سلولی در مدل Waddington می تواند به صورت شبکه های مولکولی گسترش یابد که می تواند حالت پایداری از وضعیت مورد نظر را در قالب یک شبکه ارائه نماید. چنین حالت پایداری و انتقال از یک حالت به حالت دیگر به طور محاسباتی از طریق شبکه های شبیه سازی شده(۲-4) مورد تجزیه و تحلیل قرار می گیرند و ارزیابی تجربی با بررسی پروفیل های بیان ژن در انتقال های تکثیر / تمایز ،آشفتگی های جهش ژنی و یا تنش های زیست محیطی و یا فیزیکی مورد توجه قرار می گیرد[۵، ۶].انتقال از یک حالت پایدار به حالت دیگر معمولا به فنوتیپ های پیچیده مربوط می شود که می تواند هر دو جنبه فیزیولوژیکی و پاتولوژیکی را در بر دارد مانند آنچه در دیابت شیرین و یا تکثیر سرطانی دیده می شود (شکل ۱) [۷]. عملکرد ژنی ایزوله نمی شود، بنابراین نمی توان عملکرد را به صورت جداگانه مطالعه نمود. نه تنها عملکرد محصولات ژن های منفرد، بلکه خود اینترکشن ژن ها با یکدیگر است که به طور فزاینده برای موفقیت موجودات زنده بالاتر مهم است و مزیت انتخابی ژن ها و شبکه برگرفته از آنها را تعیین می کند.

آنالیز شبکه چه کمکی به ما می کند؟ دراینجا شبکه ی ژنی ارائه می شود که به وسیله ی آزمایش مورد ارزیابی قرارگرفته است.چه اطلاعاتی رامی توان ازاین شبکه به دست آورد؟ چگونه می توان فرایند بیولوژیکی را با کمک یک شبکه فهمید؟به طوراساسی به سه موضوع پرداخته خواهد شد.سنتی ترین جنبه ،شناسایی اهمیت هر گره در شبکه است(برای مثال کدام ژن مهمترین یا ضروری ترین ژن است ،کدام ژن کم اهمیت ترین یا قابل چشم پوشی می باشد.). جنبه دیگر، شناسایی اینکه کدام ژن ها عمل کردنی تر نسبت به بقیه در کل شبکه هستند و این نه تنها با اندازه گیری ارتباطات مستقیم بلکه با توجه به کل ارتباطات شبکه بدست می آید. از این طریق ما قادر هستیم تا ارتباطات عملکردی بین همه ی ژن ها را به وسیله شبکه اینترکشن پروتئین پروتئین و دیگر انواع شبکه های معتبرتجربی را بنا نهیم.



شکل۱ فنوتیپ پیچیده توسط حالت ثابت از شبکه مولکولی تعیین می شود. یک شبکه مولکولی توسط شبکه ژنتیکی کد گذاری می شود. تاثیر متقابل مولکول ها در شبکه و همچنین اینترکش هایشان با محیط و نشانه های رشد و نمو، تعیین حالت پایداری از شبکه می کند، که در نهایت تعیین فنوتیپ هایی می کند که انعکاسی از سیستم است(۷).

بیشترمطالعات اخیر برروی شناسایی مسیرها یا جریان هایی که ازطریق شبکه هایی با ورودی و خروجی ژنی شناخته شده تمرکز می شود. این روش ها می توانند ژن های حد وسط ناشناخته را شناسایی نمایند و همچنین مشخص کنند کدام ژن ها مهمترین ژن در این فرآیند های بیولو ژیک هستند.همه ی این جنبه های مختلف می تواند به خوبی جهت شناسایی بیماری های انسانی در سطوح مختلف و با دیدگاه های مختلف به خدمت گرفته شوند .ما با بحث پیرامون این سه جنبه با جزئیاتی شامل تعدادی از روش های مربوطه شروع خواهیم نمود.

یک شبکه N شامل یک مجموعه ای V(N) از رأس ها (یا گره ها)به همراه یک مجموعه ای ازE(N) لبه (لینک) که جفت رأس های مختلف را به هم وصل می کند.معمو لاً ،گره ها بیان کننده ژنها یا پروتئین ها هستند در حالی که لینک ها اینترکشن ها را نشان می دهند.

زمانی یک شبکه N ، یک شبکه وزن دار محسوب می شود که هر لینک آن دارای عددی است که بیانگر شدت آن لبه است. معمولا وزن لبه حاکی از تضمین فعل و انفعالات آزمایشات بیولوژیک است.

یک شبکه N را زمانی می توان شبکه جهت دار نامید که همه لبه های آن جهت دار باشد و شبکه N را شبکه بدون جهت می نامند که هیچیک از لبه های آن جهت دار نباشد.معمولا شبکه های سیگنالینگ و شبکه های تنظیم رونویسی شبکه های جهت دارند که جهت ها بیانگر انتقال سیگنال یا تنظیم رونویسی هستند.

برای هر شبکه N و هر رأس خاص v در V(N) ،تعداد رأس ها v’ در V(N) که به طور مستقیم به v وصل شده درجه v نامیده می شود.

به طورخاص برای هر شبکه جهت دارN و هر رأس ویژه v درV(N) تعداد رأس ها v’ درV(N) که به طورمستقیم به v به وسیله یک لبه روبه داخل وصل می شود،درجه ای v نامیده می شود و تعداد رأس ها v’ درV(N) که به طور مستقیم به v به وسیله یک

لبه روبه خارج از v وصل میشود(لبه هایی که از v خارج می شوند) درجه خارجی از v نامیده می شود.

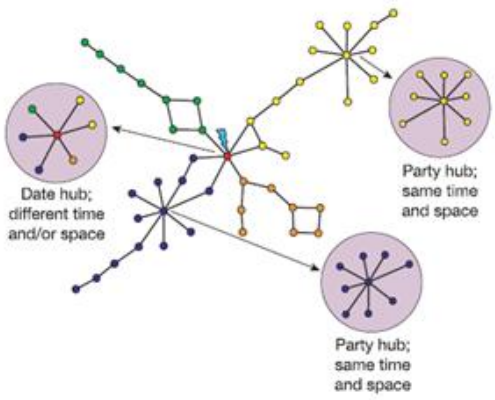
کمترین تعداد لبه های که باید بین رأس v و رأس دیگر v' در یک شبکه N عبور کند به کوتاهترین طول مسیر بین v و v' نامیده می شود. برای هرشبکه ارتباطی N،متوسط کوتاهترین طول مسیر بین هر جفت رأس به طول مسیر مشخصه (characteristics path length (CPL)) مشهوراست.

شناسایی ژنهای مهم براساس توپولوژی شبکه

شناسایی ژنهای مهم در فرآیندهای بیولوژیکی یکی از شایعترین و مهمترین جنبه ها در همه انواع مطالعات بیولوژی است .در شبکه ها بیولوژیکی ،ایده اولیه رسیدن به این هدف است که تاثیر بر شبکه یا تخریب شبکه از طریق دست کاری ژنهای خاص محاسبه نمایند.اگر برداشتن یک ژن از یک شبکه منجر به تغییرات و اثرات کوچکی شود،این ژن درحفظ عملکرد صحیح شبکه بیولوژیک باید کم اهمیت باشد . در مقابل اگر منجر به فروپاشی یا اثر بزرگی در شبکه شود مانند تقسیم کل شبکه به دوزیر شبکه،این ژن احتمالا یک نقش ضروری و مهم در فرآیند بیولوژی ایفا می کند. این فرضیه توسط داده تجربی که بیانگر حضور ژن هایی با نفوذ و تاثیر بالاتر در شبکه هستند اثبات شده است که می توانند به عنوان ژنهای کشنده تر محسوب گردند و در طی تکامل نیز حفاظت شده اند و دارای نقش اصلی در حفظ عملکرد بیولوژیکی در شبکه هستند. جهت ارزیابی اهمیت ژنها ،اندازه گیری ها متنوعی می تواند در زمینه های مختلف مورد استفاده قرارگیرد(۸-۱۱).

درجه

مهترین توجه بصری زمانی است که لبه های بیشتر که برداشته می شود آسیب بیشتری به شبکه وارد می شود.بنابراین ،ژنهایی با درجه بیشتر به عنوان قطب(HUB) شبکه شناخته می شود که خیلی مهم هستند. شواهد نشان داده که دست کاری ها در هاب منجر به افزایش قابل ملاحظه ای درCPL در شبکه بیولوژیکی نسبت به دست کاریهای تصادفی می شود. علاوه بر این سایراطلاعات مانند داده های بیان ژن جهت یافتن Party hubs, date hubs که بیانگر تفاوت عملکردهای بیولوژیک هستند می تواند مورد استفاده قرار گیرد(۱۲).با توجه به شکل ۲ مشاهده می شود که Party hub گره ای در شبکه است که به طور همزمان با همه راس های شریک خود ارتباط دارد و یک عملکرد ویژه ای را درون مدول به اجرا در می گذارد در حالی که date hub با شریک های متفاوت در زمان و مکان های متفاوتی اینترکشن دارد و در واقع عامل اتصال فرآیند های بیولوژیک به یکدیگر است. Party hubها دارای بیان همزمان بالایی با شریک های خود هستند در حالی که date hub ها دارای بیان همزمان کمتری با شریک های خود دارند.



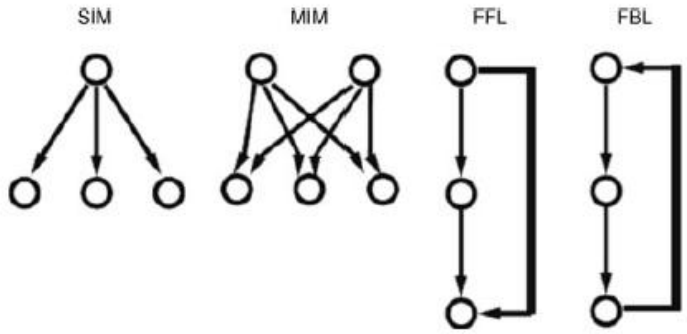
شکل ۲- گره های Party hub و date hub در شبکه را نشان می دهد.

Betweenness

مرکزیت یا اتصال به یک شبکه می تواند به وسیله CPL اندازه گیری شود. در شبکه های بیولوژیکی CPL نشان دهنده سرعت انتقال سیگنال یا سرعت پاسخ بیولوژیکی است.بنابراین ملاحظه دیگر اهمیت یک ژن، تغییرات CPL است وقتی که ژن دست کاری شود. این تغییرات می تواند به طورمستقیم به وسیله محاسبه دوباره CPL زمانی که هر ژن از شبکه برداشته می شود یا به طور غیر مستقیم با استفاده از Betweenne های هر ژن انجام شود. Betweenne یک رأس v به عنوان تعداد کوتاهترین مسیر های عبوری از آن است که بر تعداد همه کوتاهترین مسیرها تقسیم می شود. در مقایسه با Betweenne ، محاسبه دوباره CPL دقیق تر است اما زمان بر است. درحقیقت ،یک همبستگی بسیار زیادی بین نتایج محاسبه دوباره CPL واندازه گیری های Betweenne وجود دارد،بنابراین اساسا اندازه گیری Betweenne های یک ژن جهت دیدن تاثیر آن برروی CPL کافی است. همچنین به آسانی می توانیم ببینیم که یک ژن با Betweenne بالا لازم نیست که یک hub یا دارای درجه خیلی بالایی باشد،اما در مشاهده همه مجموعه ژنها ، همبستگی Betweenne و درجه مشاهده می شود..

موتیف های شبکه

در مقایسه با دو اندازه گیری قبلی که در انواع شبکه ها می تواند به کارگرفته شود، موتیف های شبکه ای در شبکه های جهت دار ،مانند شبکه های تنظیم رونویسی یا شبکه های هدف فاکتور رونویسی بکار گرفته می شوند.موتیف های شبکه ای می تواند به عنوان بلوکهای اولیه جهت تشکیل کل شبکه در نظر گرفته شود و نشان داده شده که در نگهداری استحکام ،اغتشاش و دست کاری، پاسخ های سریع و انتقال سیگنال دقیق می تواند مهم باشد. بنابراین ژن هایی که در موتیف های مختلف شبکه شرکت دارند باید خیلی مهم باشد و شمارش موتیف های شبکه ای یک سنجش برای ارزیابی اهمیت ژنها می شود(۱۳-۱۵) .در طرح ۱ چندین موتیف شبکه معرفی می شود.



طرح ۱ -چندین موتیف معمول شبکه.

موتیف های تک ورودی(SIM): یک گروه از گره ها که بوسیله ی یک گره منفرد بدون هیچ نوع تنظیم دیگری تنظیم می شود.

موتیف های چند ورودی(MIM) :یک گروه ازگره ها که گروه های دیگری از گره ها را با هم دیگر تنظیم می کند.

لوپ های پیش خور(FFL): یک گره، گره دیگری را تنظیم می کند وسپس این دو گره، گره سوم را با همدیگر تنظیم می کنند.

لو پ های پس خور(FBL که به عنوان لوپ های چندجزئی (MCL) نیز شناخته می شوند): یک گره بالا دست به وسیله یک گره پایین دست تنظیم می شود.

در شبکه های بیو لوژیکی ،ژنها در MIMوSIM ها معمولا گردن باریک بطری شبکه را تعیین می کنند که بیان می کنند که احتمالا حذف یا موتاسیون این ژن ها باعث تاثیرات کشنده می شود.FFLها وFBLها می تواند کنترل دقیق و پاسخ سریع را فراهم سازند که به طور دقیق در فرآیند های بیولوژیکی و پاسخ ها مورد نیاز است. موتیف های شبکه به آن چه در بالا گفته شد محدود نمی شود اما همه موتیف هایی هستند که معنی بیولوژیکی برای آنها اثبات شده است. جستجو درانواع مختلف موتیف های شبکه، ما را قادر به یافتن ژنهای مهم برای عملکردهای معینی که مورد نظر ما است خواهد نمود.

ساختار سلسه مراتبی

در شبکه های انتقال سیگنال یا شبکه های تنظیمی رونویسی، ژنها می توانند به چندین لایه تقسیم شوند و سیگنالها از بالا به پایین جریان می یابد. این نوع از ساختار را ساختار سلسه مراتبی می نامند. به غیر از درجه و موتیف های شبکه،ژنها در لایه های مختلف یا گره هایی با مبدا مختلف (تنظیم شده به وسیله این ژن ها) می توانند اطلاعاتی را از روی فرآیند های بیولوژیکی شناخته شده فراهم نمایند(۱۶).

این آنالیز انجام شده بر پایه ی توپولوژی شبکه به طور گسترده ای جهت شناسایی ژن های مهم در مطالعات چند گانه در گونه های مختلف مورد استفاده قرار می گیرد .با این حال برخی دیگر از هشدارها در همه این سنجش های ما در کنار حقایقی که برپایه ی ملاحظات مختلف صورت می گیرد باید اعلام گردد.اول اینکه سخت است که به تاثیر ترکیبی ژن ها توجه نشود، مانند زمانی که حتی یکی از دو ژن با ارتباطات بسیار مشابه حذف می گردد، شبکه به طور بدی تحت تاثیر قرار نمی گیرد زیرا یک کپی ژنی از آن وجود دارد، اما زمانی که هر دو آنها برداشته می شود کل شبکه دچار فروپاشی می شود. کپی ژنها به طور گسترده ای در فرآیند های بیولوژیکی واقعی جهت اطمینان از پایداری و استحکام موجودات زنده حضور دارند. در حال حاضر،برای تشخیص این اثرات ترکیبی با بکار گیری روش های IT پیشرفته جدید امکاناتی وجود دارد هر چند که محاسبات آنها ممکن است زمان بر باشد.

مشکل دیگری که وجود دارد این است که کیفیت شبکه به طور منفی تحت تاثیر نتایج است، به ویژه زمانی که لبه ها در شبکه به یک طرف تمایل دارند (دچار بایاس هستند). این اتفاق به خصوص در مطالعات انسانی به وقوع می پیوندد. برای مثال وقتی که مقالات مربوط به اینترکشن پروتئین-پروتئین مورد بررسی قرار می گیرد دیده می شود که ژنهای hot یا ژنهای مورد نظر خیلی بیشتر از ژنهای cold یا بخشهای که احتمالا می توانند hob باشند مورد مطالعه قرارگرفته اند، زیرا بیشتر اینترکشن های ژن های hot را کشف کرده اند درحالی که برای ژن های cold بیشتر اینترکشن ها ناشناخته باقی مانده است.

استنتاج اطلاعات از شبکه های شناخته شده

درک عملکردهای بیولوژیک بر پایه ی استاندارد (مدولارینی)در شبکه

وجود ساختارهای مدولار (خوشه های محکم متصل زیر شبکه ها) در شبکه های بیولوژیکی مختلف مورد توجه قرار گرفته است. در شبکه های بیولوژیکی ،این مدولها اغلب بیانگر فرآیندهای عملکردی بیولوژیکی ویژه ای هستند. مدولها می توانند به وسیله ی الگوریتم های متنوعی مانند آنچه در مدل انرژی لوگاریتمی Lin(http://www. informatik.tu-cottbus.de/_an/GD/linlog.html) الگوریتم MCODE (http:// baderlab.org/Software/MCODE) و الگوریتم خوشه بندی Marqcov(http:// www.micans.org/mcl/) شناسایی شوند. سپس با امتحان مدول ها در هستی شناسی ژنی (GO) و مسیرهای KEGG و دیگر حاشیه نویسی های عملکرد ی می توان به کشف عملکرد های بیولوژیکی آنها پی برود(۱۷-۱۸) .

استنتاج روابط عملکردی و ژنها ی عملکردی جدید از طریق شبکه ها

در چند سال گذشته ،بیشتر مطالعات انجام شده بر روی شناسایی روابط عملکردی بین ژنها تمرکز دارند. این مطالعات از همکاری مطالعات انسانی و مطالعات پیش بینی عملکرد ژن بدست آمده اند. هدف این روش ها شناسایی ژنهای مرتبط با بیماریها ی

ناشناخته از روی لیست کاندیدهای ژنی مشتق شده از مطالعات مربوطه است .معمولا این روش ها شامل نه تنها PPIهاست بلکه شامل بسیاری از انواع اطلاعات دیگر است که می تواند در قالب انواع مختلف لبه ها خلاصه شود . ایده اولیه این است که ژنهایی با عملکردهای مشابه معمولا در شبکه های PPI بسیار به هم متصل می شوند.بنابراین به منظور شناسایی ژنهای مربوط به بیماری جدید از لیست کاندید ژنی،تنها نیاز به یافتن ژنهای شناخته شده با فنوتیپ های مشابه در شبکه های PPI است.

در دهه اخیر، مطالعات متعددی از آنالیز داده های OMIM وجود دارد که از PPI و شرح شباهت بین ژنها و فنوتیپ ها که برگرفته از مطالعات انسانی است استفاده شده است(۱۹-۲۰).با پیشرفت تکنولوژی های جدید هر روز بیشتر و بیشتر مطالعات مربوطه بر روی جمعیت های بزرگ و فنوتیپ های ویژه ای با سطوح پوشش و دقت و تفکیک پذیری بالایی به اتمام رسیده است.این مطالعات گسترده ژنومی(GWAS) فرصتهایی برای کاربرد همه روش ها فراهم نموده است. همانطور که با ادغام انواع مختلفی از شبکه ها می توان یک شبکه وزن دار کلی با وزن های مختلف روی لبه های مختلف دید،مابه طور اساسی به معرفی یک روش با کاربردهای وسیع و یک اجرای کامپیوتری خوب که بر پایه الگوریتم پیاده روی تصادفی (random walk) است (۲۱)می پردازیم.

پیاده روی تصادفی روی گراف جابجایی یک پیاده رونده تکراری از گره فعلی خود به همه همسایگان خود است که از گره های منبع با استفاده از تمام لبه های وزن دار شده شروع می شود.هر گره منبع می تواند یک وزن متفاوت داشته باشد و به طور اساسی مجموع همه می تواند تا ۱نرمال سازی شود، بنابراین این مقدار همچنین می تواند به عنوان احتمال انتقال اطلاعات در کل شبکه در نظر گرفته شود. در اینجا،در مقایسه با پیاده روی تصادفی سنتی،یک فرآیند راه اندازی مجدد اضافه می شود که در هر مرحله سیگنال در گره s با احتمالr راه اندازی مجدد دارد.این بیانگر این است که در هر مرحله از گذار،تنها (1-r) از اطلاعات کل به طورپیوسته با r از راه اندازی مجدد کل انتقال یافته است.

هدف از این روش اضافه شدن یک ورودی مداوم است و زمانی که وضعیت پایدار بدست آمد،همه گره های دیگر باید یک نسبت ثابت از اطلاعات را به خروجی که جمع آن r است دارا باشند.

از نظر فرمول، پیاده روی تصادفی با راه اندازی مجدد به صورت زیر تعریف می شود.

$$Pt+1 = (1-r)*W*Pt + r*P0$$

وقتی w یک ماتریسی است که تنها برپایه توپولوژی شبکه بنا نهاده شده است در واقع ستون ماتریس مجاوت نرمال سازی شده است که مقدار غیر صفر بیانگر وزن یک لبه در شبکه است.pt برداری است که هر عنصر احتمال اطلاعات روی یک گره در مرحله t را دار است. در این برنامه،بردار احتمال اولیه P0 به عنوان احتمال های وزن دارای است که هر احتمال بیانگر نفوذ یک ژن منبع مورد نظر در بیماری است که جمع همه این احتمالات برابر۱ است.

وقتی تفاوت بین pt وpt+1 کمتر از یک آستانه قرار دادی است حالت پایدار PN بدست آمده است و نتیجه مورد نظر بدست آمده است ژنهای کاندید بیماری نیز بعدا براساس مقدار PN رتبه بندی می شوند.

اجرای الگوریتم پیاروی تصادفی نسبت به الگوریتم های قبلی بهتر انجام می شود. همچنین کار با این الگوریتم به آسانی صورت می گیرد. یکی از مزایای بارز این روش این است که PNافزودنی است که باعث می شود این الگوریتم بسیار مناسب باشد. یک مثال ساده از آن را ببینند. حالت پایدارPN از تنها گره منبع A وB به صورت(B) PN یا(A) PN است.وقتی می خواهیم به اثر ترکیبی AوB توجه کنیم،می توانیم احتمالات وزن دار از دو رگه منبع را به صورت a و(1-a)و حالت پایدارPN را برای هردوAوB به عنوان گره های منبع به صورت PN(AB)= a* PN(A) +PN(B)* (1_a) محاسبه شود این فرمول می تواند به یک مجموعه s از ژنهای منبع چند گانه گسترش یابد.بنابراین اساسا برای یک شبکه خاص،ما مجبور نیستیم PN برای هر مجموعه ژن منبع را دوباره محاسبه نماییم. در عوض ما می توانیم هر ژن

منبع را به طور جداگانه محاسبه نماییم و نتایج وزن داده شده را جمع کنیم .در این الگوریتم ، تفاوت r بیانگر میل متفاوت است.r بالا بیانگر اثر بیشتر ژنهای ورودی و کمتر گذار در شبکه است در حالی که r پایین منجر به مراحل گذار بیشتر می شود . به طور تجربی نتیجه پایدار در طی ۳۰تا۵۰ مرحله با توجه به r مختلف و آستانه استفاده شده می تواند بدست آید و الگوریتم نیز زمان زیادی را مصرف نمی کند. بنابراین این امکان وجود دارد که PN هر ژن در شبکه را محاسبه نمود.

همانطور که قبلا گفته شد همه این الگوریتم ها به طور منفی تحت تاثیر کیفیت شبکه ها و ژن های hot آنها قرار می گیرند.اگر یک شبکه بایاس دار استفاده شود به احتمال خیلی زیاد در ژنهای hot گیرمی افتیم.

تنظیم کننده های رونویسی شناخته شده از داده بیانی در شبکه های رونویسی

فاکتورهای رونویسی یک نقش تنظیمی ضروری در فرآیند های بیولوژیکی مختلف ایفا می کنند.با این حال ،بعید است که از داده های بیانی به علت بیان پایین و اغلب پراکنده تشخیص داده شوند.جهت پر کردن این شکاف Reverter و همکارانش یک الگوریتم فاکتور موثر تنظیمی (RIF) جهت شناسایی فاکتورهای رونویسی ضروری از داده های بیان ژن به وسیله شبکه بیانی ادغامی پیشنهاد دادند.

آنالیزRIF یک نمره به هر فاکتور رونویسی با توجه به هر دو ارتباط بین فاکتور رونویسی و ژنهای بیان شده متفاوت و سطح بیان متفاوت ژنهای بیان شده می دهد. به طور خاص برای یک مدول عملکردی داده شده ، پتانسیل رگولاتوری شان به میانگین همبستگی بیانی مطلق در همه ژنها در مدول نمره دهی می شود(۲۲-۲۳).

استخراج تنظیم کننده های اتصال مسیر و افکتور ها براساس جریان های شبکه

اخیرا از تکنیک های بسیار پیشرفته جهت آشکارسازی پتانسیل اجزاء شبکه های بیولوژیک به طور گسترده ای استفاده می شود. تا کنون این تکنیک های بسیار پیشرفته دو کلاس را پوشش می دهند(۱) صفحه نمایش های ژنتیکی شامل بیان بالا،حذف یاصفحه نمایش کتابخانه RNA تداخلی و(۲)پروفایل mRNA با استفاده از میکرو اری یا تکنولوژی سکونس RNA . با مقایسه نتایج این دو روش،YEGER-LOTEM و همکارانش یافتند که صفحه نمایش های ژنتیکی تمایل به شناسایی تنظیم کننده هایی دارند که برای پاسخ سلولی ضروری هستند در حالی که ژنهای بیان شده متفاوت شناسایی شده به وسیله پروفایل mRNA احتمالا افکتورهای پایین دست هستند که تغییرات به طور غیرمستقیم انعکاسی از تغییرات ژنتیکی در شبکه های تنظیمی است(۲۴) این در مورد بیماری ها نیز صدق می کند ؛ با استفاده از مطالعات دیابت نوع II و فشارخون ،ما فهمیدیم که ژنهای عامل بیماری که دارای احتمال بالایی ابتلا به دیابت نوع II و فنوتیپ های فشارخون هستند وقتی که دستکاری می شود تبدیل به HUB در شبکه های اینتراکتومی می شود و در مسیرهای سیگنالینگ پررنگ می شوند در حالی که ژنهای بیان شده با تفاوت معنی دار شناسایی شده به وسیله میکرواری اغلب در مسیرهای متابولیکی حضور پررنگی می یابند ارتباط بین این دو مجموعه به طور قابل توجهی محکم است(۲۵).

برای پرکردن فاصله بین داده های غربالگری ژنتیکی و داده های بیان mRNA از شبکه های ملکولی شناخته شده،YEGER-LOTEM و همکارانش یک رویکرد یک پارچه ای که به شبکه پاسخی مشهور است ارتقاء داده اند.به طور خلاصه شبکه پاسخی یک الگوریتم بهینه سازی جریان است که به تعریف دوباره یک زیر شبکه ضروری می پردازد که ارتباط HITهای ژنتیکی (مبدأ) و ژنهای بیان شده متفاوت (هدف) از یک شبکه وزن دار کلی را نشان می دهد که هر گره یا لبه با یک وزن که بیانگر اهمیت بیولوژیکی آنها یا درجه اطمینان آنهاست مشخص می شود. هزینه یک لبه به وسیله مقدار لوگاریتم وزن آن تعریف می شود . بنابراین هدف شبکه پاسخ را می توان با حل کردن یک مسئله بهینه سازی برنامه ریزی خطی که باعث به حداقل رساندن هزینه کلی شبکه وقتی توزیع حداکثر جریان از مبدأ به هدف است محقق ساخت. با توجه به راه حل ، آن لبه ها با جریان مثبت

به عنوان زیر شبکه ضروری پیش بینی شده تعریف می شوند.

بحث و نتیجه گیری

روشهای اصلی و برنامه های کاربردی در آنالیز شبکه جهت تفسیر فنوتیپ های پیچیده معرفی شده است.اگر چه این روش ها دارای مزایای زیادی است بیولوژی شبکه هنوز با بسیاری از چالش ها روبرو است . بیشتر روش ها پر کیفیت datasetها که تعیین کننده مثبت کاذب و پوشش محدود شده هستند عمل می کنند. بیشتر لبه ها در نقشه های شبکه هنوز فاقد ویژگی ها و جهت های دقیق هستند تغییرات پس از ترجمه به سختی در یک مقیاس بزرگ مانیتور می شوند. خصوصیت انواع سلول و بافت به سختی مورد توجه قرار می گیرد. با این حال با پیشرفت تکنولوژی های جدید مانند دستگاه های بسیار پیشرفته و تکنیک های اندازه گیری دینامیک تک سلولی و با افزایش دقت و پوشش تکنولوژی های پیشرفته و شتاب تولید داده ، نیاز به یکپارچه سازی داده و مدلسازی در سطح شبکه افزایش یافته است . بیشتر روش ها در هم ادغام شده تا ابزاری مهم را برای آنالیز شبکه ایجاد نمایند .تسلط بر این روش ها لازم است اما برای درک بیولوژی کافی نیست.مهمترین چیز انجام سوالی درست است تا ابزارهای آنالیز شبکه مناسب انتخاب شود و جهت ارزیابی نتیجه آنالیز نیز آزمایشات تجربی صورت می گیرد. بعد از همه اینها، می توان گفت بیولوژی شبکه،بیولوژی است و هدف اصلی برای بیولوژی شبکه و بیولوژی ملکولی مشابه است و آن درک بهتر فرآیند های بیو لوژیکی پایه ای و مکانیسم بیمارهای انسانی هستند.

سپاسگزاری

این مطالعه حاصل طرح تحقیقاتی با کد -1-1391 150-9146 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-Barabasi AL, Oltvai ZN. Network biology: Understanding the cells functional organization. Nat Rev Genet 2004; 5:101-13.

2-Bergman A, Siegal ML. Evolutionary capacitance as a general feature of complex gene networks. Nature 2003;424:552-49.

3-Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol 1969; 22: 437-67.

4-Li F, Long T, Lu Y, Ouyang Q, Tang C. The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci USA 2004; 101:4781-6.

5-Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon, et al. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. Nat Genet 2006; 38:228-33.

6-Huang S, Eichler G, Bar-Yam Y, Ingber DE. Cell fates as highdimensional attractor states of a complex gene regulatory network. Phys Rev Lett 2005;94:

7-Han JD. Understanding biological functions through molecular networks. Cell Res 2008; 18:224-37.

8-Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41-2.

9-Tew KL, Li XL, Tan SH. Functional centrality: detecting lethality of proteins in protein interaction networks. Genome Inform 2007;19:166-77.

10-Albert R, Jeong H, Arabasi AL. Error and attack tolerance of complex networks. Nature 2000;406: 378-82.

11-He X, ZhangJ. Why do hubs tend to be essential in protein networks? PLoS Genet 2006;2: e88.

12-Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, et al. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature 2004;430:88–93.

13-Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science 2002;298:824-7.

14-Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I. et al.M. Superfamilies of evolved and designed networks. Science 2004; 303: 1538-42.
15-Wuchty S, Oltvai ZN, Barabasi AL. Evolutionary conservation of motif constituents in the yeast protein interaction network. Nat Genet 2003; 35: 176-9.
16-Yu H, Gerstein M. Genomic analysis of the hierarchical structure of regulatory networks. Proc Natl Acad Sci USA 2006;103:14724-31.
17-Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003; 4:2-6.
18-Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95: 14863-8.
19-Lage K, Karlberg EO, Storling ZM, Olason PI, Pedersen AG, Rigina O, et al. A human phenome–interactome network of protein complexes implicated in genetic disorders. Nat Biotechnol 2007;25:309-16.
20-Wu X, Jiang R, Zhang MQ, Li S. Network-based global inference of human disease genes. Mol Syst Biol 2008; 4:189-92.
21-Kohler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. Am J Hum Genet 2008;82: 949-58.
22-Reverter A, Hudson NJ, Nagaraj SH, Perez-Enciso M, Dalrymple BP. Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. Bioinformatics 2010;26:896–904.
23-Hudson NJ, Reverter A, Wang Y, Greenwood PL, Dalrymple BP. Inferring the transcriptional landscape of bovine skeletal muscle by integrating co-expression networks. PLoS ONE 2009; 4: e7249.
24-Yeger-Lotem E, Riva L, Su LJ, Gitler AD, Cashikar AG, King OD, et al. Bridging highthroughput genetic and transcriptional data reveals cellular responses to alphasynuclein toxicity. Nat Genet 2009; 41:316-23.
25-Yu H, Huang J, Qiao N, Green CD, Han JD. Evaluating diabetes and hypertension disease causality using mouse phenotypes. Syst Biol 2010; 4:97-101.

Network Analysis Methods for Interpreting Complex Phenotypes in Biological Networks

Zali H[1], Rezaei Tavirani M[2], Heidarbigi Kh[3], Shahriarinoor M[4]*

(Received: 4 Jan. 2013 Accepted: 1 Mar.2013)

Abstract

Gene network analysis is an important part of systems in biological studies. Compared with traditional genotype/phenotype studies that focused on establishing the relationships between single genes and interested traits, network analysis give us a global view of how all the genes work together properly, which in turn leads to the correct biological functions. Network analysis also helps to derive useful information from the network and also helps the discovery of biological processes from a network. In this study, the main methods and applications in network analysis to interpret complex phenotypes basically explain three aspects. The first aspect is to identify the importance of each node in the network which determine more important or crucial genes, or less important or dispensable one. Another aspect is to identify which genes are more functionally related through the whole network view by measuring the direct gene connections and also by considering the connections through the whole network. Identifying the paths or flows through the networks with known input and output genes is the last aspect discussed in network analysis. Although these methods have many advantages, network biology still faces many challenges so more methods have emerged, which provide important tools for network analysis. Mastering these methods is necessary, but far from sufficient for understanding biology. More important things to do are to ask the right questions, to choose proper network analysis tools, and to validate analysis results by solid experimentation. Finally can express that the fundamental goal is the same for network biology and molecular biology – to better understand biological processes and the mechanisms of human diseases.

Keywords: Node, Edge, Gene expression networks, Network analysis

1.Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
2. Proteomics Research Center, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran
3.Dept of Histology, Faculty of Medicine, Ilam University of Medical Sciences, Ilam, Iran
4. Dept of Microbiology Science and Research Branch, Islamic Azad University, Gilan, Iran
**(corresponding author)*