

**ابزاری از MATLAB برای غنی سازی مسیر با استفاده از امتیاز نظم مسیر مبتنی بر توپولوژی**

**چکیده**

**پیشینه:** رسیدگی به مقدار وسیع داده‌های تظاهر ژنی تولیدشده توسط روش‌های نمای نسخه‌برداری ژنوم، یک کار چالش‌برانگیز است که نیازمند ترکیبی آگاه از روش‌های پیش‌پردازش، فیلتراسیون و تحلیل دارد اگر که قرار باشد نتایج زیست‌شناختی معناداری حاصل شود. برای مثال، طیفی از آمار سنتی و رویکردهای تحلیل مسیر محاسبه‌ای برای شناسایی فرایندهای بسیار ارائه‌شده در داده‌های ریزآرایه به‌دست‌آمده از حالت‌های مختلف بیماری استفاده‌شده است. اگرچه، اکثر این رویکردها تمایلی در بهره‌برداری از کل طیف داده تظاهر ژنی یا روابط مختلف و وابستگی‌ها ندارند. قبلاً، ما ابزار تحلیل غنی‌سازی مسیر را در MATLAB که یک امتیاز نظم مسیر (PRS) را با در نظر گرفتن علامت‌دهی توپولوژی مسیر و بیش نمایندگی و بزرگی ژن‌هایی که به‌طور متفاوتی ظاهرشده‌اند به دست می‌دهد. در اینجا، این رویکرد را برای شامل شدن مسیر متابولیک گسترش دادیم و استفاده از رابط کاربر گرافیکی (GUI) را توصیف کردیم.

**نتایج:** با استفاده از تعدادی از جایگاه‌های ریزآرایه و گونه‌ها، مصرف‌کنندگان قادرند تا امتیازات PRS را به همراه امتیاز z مطابق برای مقایسه محاسبه کنند. ارزیابی بیشتر اهمیت مسیر ممکن است برای افزایش اعتماد به مسیرهای به‌دست‌آمده انجام شود و مصرف‌کنندگان می‌توانند دایره المعارف Kyoto نمودارهای مسیر ژن و ژنوم را که برای تأکید بر ژن‌های نهفته علامت‌گذاری شده است، ببینند.

**نتیجه‌گیری:** ابزار PRS، فیلتری برای منزوی‌سازی بینش‌های زیستی از داده‌های نسخه‌برداری‌شده پیچیده ارائه می‌کند.

**پیشینه**

به‌طور فزاینده‌ای، روش‌های نمای نسخه‌برداری با بازده بالا (ریزآرایه‌ها یا به‌طور فزاینده‌ای، دنباله‌های RNA) تحقیقات علوم حیاتی مدرن را شکل می‌دهد. چنین روش‌هایی، یک دوربین مولکولی فراهم می‌کند که تصاویری از سراسر ژنوم فعالیت ژنتیکی می‌گیرد. اگرچه، داده‌های تحلیل مؤثر ریزآرایه‌ها، چالش‌هایی را مخصوصاً در رسیدگی به تعداد زیادی از ژن‌هایی که به‌طور هم‌زمان مطالعه می‌شوند ارائه می‌دهد.

تحلیل تظاهر ژن درزمینهٔ دانش برگزیده یا "تحلیل مسیر ناشی از پایه دانش" ازآنجایی‌که این مسئله باعث کاهش در فضای تحقیقاتی از هزارها ژن تا زیرمجموعه‌ای از فرایندهای زیست‌شناختی که برای تفسیر انسانی بیشتر مهار شدنی‌تر است ضروری است [1]. طبق Khatri و همکارانش [2]، رویکردهای غنی‌سازی مسیر می‌تواند به سه نسل تقسیم‌بندی شود:

1- تحلیل بیش نمایندگی (ORA): این تحلیل یک مسیر را با در نظر گرفتن نسبت ژن‌هایی که به‌صورت متفاوت ظاهرشده (DEG) و در هر مسیر مربوط به نسبت تمام ریزآرایه‌های DEG مشاهده می‌شود، ثبت می‌کند. این تحلیل برای ابزارهای تحلیل مسیر متعددی من‌جمله GenMAPP [3], GoMiner [4], [5] Onto-Express and FatiGo [6] به کار می‌رود.

2- امتیازدهی طبقه عملکردی (FCS): FCS یک امتیاز را به هر ژن بر اساس تظاهر آن، در مسیر می‌دهد که با آن امتیاز مسیر بر اساس امتیازات تمام ژن‌ها در مسیر محاسبه می‌شود. تعدادی از روش‌های FCS از طریق ابزارهای مستقل مانند GSEA [7]، SigPathway [8] و SAFE [9] یا ابزارهای وب مانند T-profiler [10]، Gazer [11] و GeneTrail [12] اجرا می‌شود.

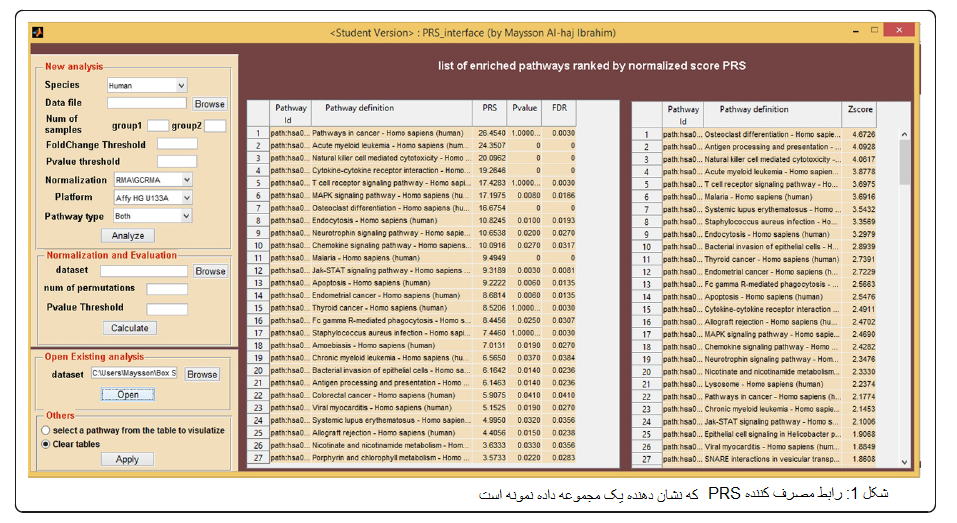
3-رویکردهای مبتنی بر توپولوژی مسیر (PT): این رویکردها از توپولوژی مسیرها با دادن اوزان برای اتصالات از پیش تعیین‌شده بین ژن‌ها که امتیازدهی مسیر را تشکیل می‌دهند، بهره می‌گیرند. برخی رویکردهای مبتنی بر توپولوژی در دهه‌های گذشته در ادبیات توصیف‌شده است. طبق Mitrea et al [13]، رویکردهای مبتنی بر PT، در روشی که اطلاعات توپولوژی مسیر را به امتیاز مسیر ترجمه می‌کند متفاوت‌اند. برخی روش‌ها تنها از داده‌های توپولوژی ژن‌هایی که به‌صورت متفاوت ظاهرشده‌اند (DEG) در امتیاز غنی‌سازی استفاده می‌کنند (مثلاً، MetaCore[14]، EnrichNet [15])، درحالی‌که دیگر روش‌ها (من‌جمله، SPIA [16] و GANPA [17]) از داده‌های تظاهر DEG به همراه داده‌های توپولوژی استفاده می‌کنند. متناوباً، برخی روش‌ها از دادهای تظاهر ناشی از تمام ژن‌های ریزآرایه استفاده می‌کنند چه بین شرایط تغییر بکنند یا نکنند، برای مثال، PathOlogist [18]، DEGraph [19] و ACST [20]. نکته مهم این‌که برخی ابزار مبتنی بر PT تنها از توصیف‌های علامت‌دهی مسیر مانند Pathway-Express [21]، NetGSA [22]، ScorePAGE [23]، TAPPA [24]، MetPA [25] و Clipper [26] استفاده می‌کنند.

پیش‌ازاین، یک روش غنی‌سازی مسیر جدید را ارائه کردیم که در آن‌هم توپولوژی مسر و هم بزرگی تظاهر ژن، ایجاد امتیاز نظم مسیر (PRS) را تغییر می‌دهد [27]. مخصوصاً، با ترکیب داده‌های تغییر کل برای آن نسخه‌هایی که از آستانه اهمیت تجاوز می‌کنند و با در نظر گرفتن پتانسیل تظاهر ژن تغییریافته در اثرگذاری بر نسخه‌برداری‌های پایین‌دست، مسیرهایی را شناسایی کردیم که به فرایند پاتوفیزیولوژیک تحت بررسی مرتبط است. رویکرد ما تعدادی از مسائل را که به‌طور بالقوه روش‌های غنی‌سازی را تضعیف می‌کند در نظر می‌گیرد. ما گام‌هایی برای کاهش تأثیر خطاها در نگاشت شناسه و کاهش خطای ایجادشده توسط مسیرهای اضافی (مانند نمونه‌های چندگانه یک ژن) برداشتیم. روش‌های توپولوژی نیز باید به‌صورت مؤثر به حلقه‌ها رسیدگی کنند، لذا ما از الگوریتم جستجویی که از نظریه گراف ناشی شده است برای حل این مشکل استفاده کردیم. هم‌چنین احساس کردیم که تقسیم‌بندی دلخواهانه فرایندها به نظم بالا یا پایین، ازآنجایی‌که تغییرات در تظاهر ژن احتمالاً در سراسر مسیرها توزیع می‌شود، ساختگی است و ازاین‌رو ارزیابی ما، یک ارزیابی کلی اثر بود.

در اینجا، اجرای رویکرد PRS خود را به‌عنوان یک ابزار مستقل که به مصرف‌کننده نهایی گزینه واردکردن داده را از جایگاه‌ها و انواع مختلف ریزآرایه می‌دهد، توصیف کردیم. این ابزار هم امتیازات z و هم PRS را به دست می‌دهد، تحلیل آماری ارائه می‌کند و اجازه مرور راه‌هایی که دارای ژن‌های نهفته‌ای هستند که بارنگ‌های مختلف علامت‌گذاری شده‌اند می‌دهد. گزارش خود را با این مطلب که مصرف‌کنندگان قادرند تا هم مسیرهای متابولیک و هم علامت‌دهی را تقویت کنند، ارتقا می‌دهیم.

**اجرا**

رویکرد PRS در MATLAB اجرا شد. مصرف‌کنندگانی که به محیط MATLAB دسترسی ندارند، می‌توانند کامپایلر اجرای MATLAB (MRC) را دانلود کنند تا نرم‌افزار توصیف‌شده در اینجا را با یک GUI کاربرپسند به‌کارگیرند. رابط PRS (شکل 1) توابع متعددی را برای مصرف‌کنندگان ارائه می‌دهد:



**پیش‌پردازش داده‌های ریزآرایه**

ما یک فیلتر را برای نرمال‌سازی داده‌ها از جایگاه‌ها مختلف مجدداً مهندسی نکردیم، در عوض، مصرف‌کنندگان ابتدا باید داده‌های نسخه‌برداری‌شده را با استفاده از یکی از ابزارهای بی‌شمار موجود پیش‌پردازش کنند. داده‌ها باید در قالب یک صفحه گسترده (Spreadsheet) Excel باشد که در آن اولین ستون باید شناسه محقق باشد و در ستون‌های بعدی باید ارزش‌های تکرار شده نرمالیزه شده تظاهر را از شرایط تست و کنترل قرار داد. اطلاعات مازاد در خصوص گونه‌ها، تعداد نمونه، تغییر کل، آستانه‌های تست t، روش نرمال‌سازی و جایگاه موردنیاز است.

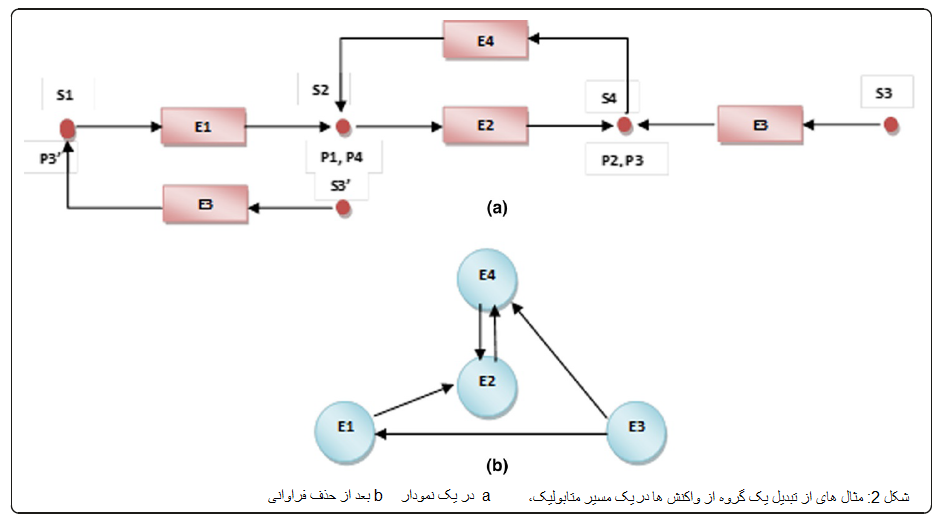
**نمایش مسیر**

الگوریتم اساسی ما پیش‌ازاین توصیه شد [27]. به‌طور خلاصه، دایره المعارف Kyoto تعاریف مسیر ژن‌ها و ژنوم ها [28] استفاده شد که در آن مسیرها در فرمت (KGML) زبان نشانه‌گذاری KEGG حفظ‌شده‌اند. مجموعاً ما 189 توصیف علامت‌دهی و متابولیک از KEGG وارد کردیم و این‌ها را در موارد MATLAB تجزیه‌وتحلیل کردیم و سپس به گراف‌های مستقیم تبدیل کردیم. پرونده‌های KGML شامل سه نوع مورد است: ورودی‌ها، روابط و واکنش‌ها. این‌ها می‌توانند به‌صورت موارد گرافیکی در نقشه مسیر مرتبط کشیده شود (پرونده اضافی 1). تنها ورودی‌ها (که متشکل از گره‌ها هستند و به‌صورت جعبه نمایش داده می‌شوند) و روابط (که به‌صورت مرز نمایش داده می‌شوند) در نمایش مسیرهای علامت‌گذاری که پروتئین‌ها (جعبه‌ها) با "روابط" پیوند داده‌شده‌اند، استفاده می‌شوند. تمام سه نوع در نمایش ساختار مسیرهای متابولیک برای گرفتن روابط لایه-آنزیم- محصول که در آن آنزیم‌ها (جعبه‌ها) با " روابط" پیوند دارند و ترکیبات (دایره‌ها) با " واکنش‌ها" پیوند دارند استفاده می‌شوند. برای تبدیل یک مسیر متابولیک به یک گراف به روشی منطقی، ما آنزیم‌ها را به‌عنوان گره‌ها در گراف نشان دادیم، درحالی‌که لایه‌ها و محصولات برای شناسایی جهت روابط (مرزها) بین گره‌ها استفاده شدند (شکل 2). درحالی‌که اذعان کردیم که ممکن نیست تا هرگونه تأثیر بر روی شار را با این منطق پیش‌بینی کرد، استدلال کردیم که هرگونه تغییر در تظاهر گره در یک مسیر متابولیک می‌تواند فیزیولوژیکی باشد مخصوصاً اگر گره‌ها متصل باشند.

نمایش مسیرها به‌صورت گراف با توجه به اینکه فراوانی را که ژن‌ها تنها یک‌بار در هر گراف مسیر نشان دادند کاهش می‌دهد، دارای مزیت است. یک الگوریتم جستجوی عمق اول (DFS) ناشی از نظریه گراف برای تضمین اینکه حلقه‌ها تنها یک‌بار شمرده می‌شوند به کار رفت.

**امتیازدهی مسیر**

روش ما اوزان را به تمام گره‌های مهم (همان DEGs) در یک مسیر اختصاص داد تا نیروی توپولوژیکی آن‌ها را نشان دهد (مخصوصاً تعداد گره‌های مهم پایین‌دست که به آن یا به‌صورت مستقیم یا از طریق دیگر گره‌های مهم اشاره شد همان‌طور که پیش‌ازاین توصیف شد [27]). یک PRS بر مبنای ارزش تغییر کل و وزن دهی تمام گره‌های مهم در مسیر و محاسبه شد و برای اندازه مسیر نرمال شد. هم‌چنین ما یک امتیاز Z را نیز محاسبه کردیم [29] (با بهبودی در اجراهای اولیه که در آن بعد از حذف ژن‌های فراوان از توصیفات مسیر، این کار انجام شد). نرم‌افزار دو فهرست از مسیرهای رتبه‌بندی شده طبق PRS و امتیاز Z را به دست می‌دهد که هم به برای تحلیل بیشتر هم به‌صورت فایل excel و هم به‌صورت.mat ذخیره شدند.

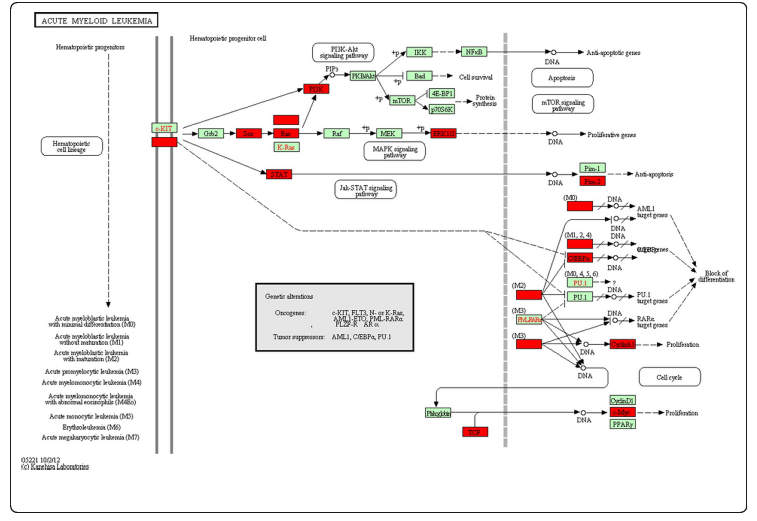


**ارزیابی اهمیت مسیر**

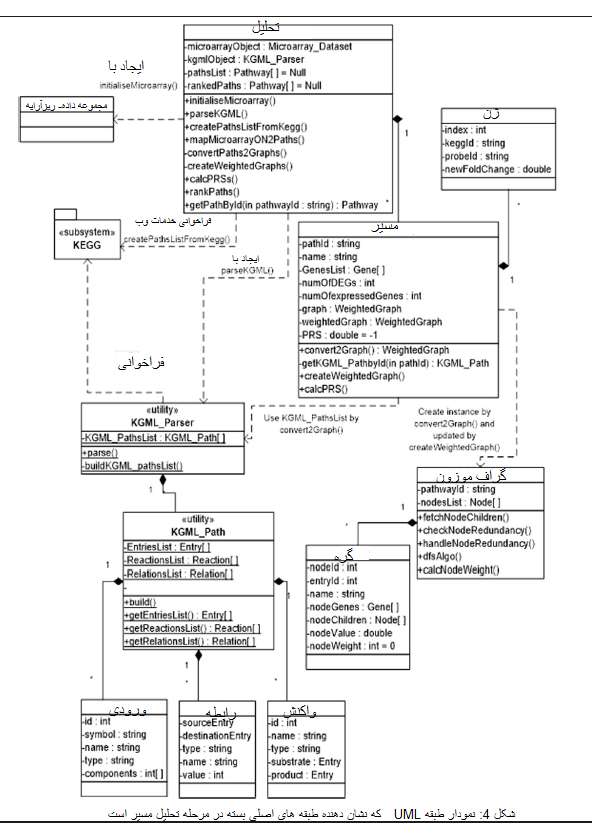
سپس ما احتمال کسب امتیازاتی حداقل به‌اندازه امتیاز PRS به‌صورت اتفاقی و با استفاده از روش جایگشت غیر پارامتریک را ایجاد کردیم. در ابتدا، مقادیر تغییر کل برای تمام ژن‌های ریزآرایه تغییر کرد. سپس این مقادیر به سمت مسیرها بازگشت و PRS دوباره محاسبه شد. این فرایند n بار تکرار شد که در آن n توسط مصرف‌کننده از طریق رابط ارائه شد (معمولاً n=1000). اهمیت آماری (مقدار p) هر امتیاز مسیر با مقایسه بین امتیاز مشاهده‌شده و امتیازات ایجادشده رندم n برآورد شد. برای کسب ارزیابی اهمیت آماری قابل‌اعتمادتر، مقادیر p برای چندین تست اصلاحی توسط روش نرخ کشف خطا (FDR) بر اساس آستانه ارائه‌شده توسط مصرف‌کننده، تنظیم شدند. این مسئله به‌صورت دقیق‌تر در گزارش اصلی ما توصیف‌شده است [27].

**تجسم مسیرهای غنی‌شده**

بعد از انجام تحلیل، نتایج به‌صورت فایل‌هایی با فرمت .mat برای بازیابی آسان ذخیره شدند. با کلیک کردن بر روی نام مسیر از فهرست مسیرهای رتبه‌بندی شده که در جدول 1 نشان داده‌شده است و انتخاب گزینه تجسم یک مسیر از رابط، یک نقشه مسیر علامت‌گذاری شده نشان داده خواهد شد. درواقع، نرم‌افزار یک خدمات وب نقشه‌کشی مسیر فرا خواهد خواند که بر روی وب‌سایت KEGG قرار دارد و تعدادی از پارامترها را من‌جمله فهرستی از ژن‌های تظاهر با تغییرات کلی و رنگ‌های مشخص را عبور می‌دهد تا بین DEG ها با ژن‌های غیر نهفته تفاوت قائل شود. شکل 3 نشان‌دهنده یک نقشه مسیر معمولی است که در آن ژن‌های مهم (بالای آستانه) به رنگ قرمز مشخص‌شده‌اند و غیر مهم‌ها (تغییر نکرده یا ظاهرنشده) به رنگ سبز.



شکل 3: یک مسیر علامت‌گذاری شده معمولی، در این مورد،KEGG" مسیر لوسمی میلوئیدی حاد " غنی‌شده در یک مجموعه داده AML (ورودی GEO، #GSE9476)؛ ژن‌های مهم قرمزرنگ و غیر مهم‌ها سبز هستند.

****

**UML برای مدل‌سازی و توصیف نرم‌افزار**

در اینجا، ما از زبان مدل‌سازی یکپارچه (UML) استفاده کردیم تا ساختار و توابع روش خود را با نمودارها توصیف، مدل‌سازی و تجسم کنیم. 14 نوع نمودار دسته‌بندی‌شده در سه‌طبقه در UML 2.0 وجود دارد [30]، اگرچه، در این مقاله ما از دوتا استفاده کردیم: نمودارهای طبقه و توالی. نمودارهای طبقه ساختارهای ثابت یا موارد اصلی در نرم‌افزار را نشان می‌دهد. شکل 4 نشان‌دهنده طبقه‌های اصلی در مرحله تحلیل مسیر است. تحلیل طبقه، طبقه اصلی است که رابطی برای اجرای تمام خدمات ارائه‌شده توسط ابزار ارائه می‌کند؛ و دارای چهار ویژگی اصلی است:

- مورد ریزآرایه: موردی از طبقه "مجموعه داده- ریزآرایه" که با فراخوانی اولیه تابع ریزآرایه () ایجاد می‌شود (فایل اضافی 2 را ببینید)؛ و داده‌های تظاهر ژنی نرمال شده را و فهرستی از تمام ژن‌ها با مقادیر تغییر کلی‌شان را دارد.

- مورد KGML: موردی از طبقه “KGML\_Parser” را با فراخوانی تابع ParseKGML(فایل اضافی 3 را ببینید) ایجاد می‌کند؛ و ساختار پایای تمام‌مسیرها را به‌عنوان فهرستی از موارد طبقه “KGML-PATH” که با فرمت KGML تعریف‌شده است، داراست. یک مورد از “KGML\_path” ساختار یک مسیر KEGG را نشان می‌دهد و متشکل از فهرست مطالب، فهرست واکنش‌ها و فهرست روابط است (فایل اضافی 1 را ببینید).

- فهرست مسیر: این فهرستی از موارد طبقه مسیر است که با فراخوانی تابع Createpathlistfromkegg() ایجاد می‌شود (فایل اضافی 4 را ببینید). این مورد درنهایت فهرستی از مسیرهای غنی‌سازی شده با ارجاع به یک مجموعه داده ریزآرایه را داراست.

- مسیرهای رتبه دهی شده: این مورد با فراخوانی تابع rankpaths() ایجاد می‌شود؛ و دارای همان فهرست مسیرهای تعریف‌شده توسط Pathslist می‌باشد، اما به‌صورت نزولی بر اساس مقادیر PRS رتبه‌بندی شده‌اند.

نمودارهای توالی برای نشان دادن توابع ابزار PRS طبق انواع مختلف روابط درونی بین موارد استفاده‌شده است. به‌عنوان‌مثال، شکل 5 نمایش‌دهنده توابع PRS اصلی با گام‌های زیر است:

1. تبدیل مسیرها به گراف توسط تابع convertpath2graph که نیاز به استفاده از مورد KGML دارد و دارای فهرستی از مطالب، روابط و واکنش‌های تمام‌مسیرهاست.

2. با استفاده از اطلاعات ذخیره‌شده در مورد KGML و pathlist برای هر گراف (شکل 4 را ببینید)، فهرستی از گره‌ها (که هر گره نشان‌دهنده یک یا چند ژن از مسیر اصلی است) و فهرستی از کودکان برای هر گره ایجاد می‌شود.

3. حذف ژن‌های فراوان که ممکن است بارها در همان مسیر نشان داده‌شده باشد. دو تابع برای رسیدگی به فراوانی گره طراحی‌شده است: checknoderedundancy() و handlenoderedundancy().

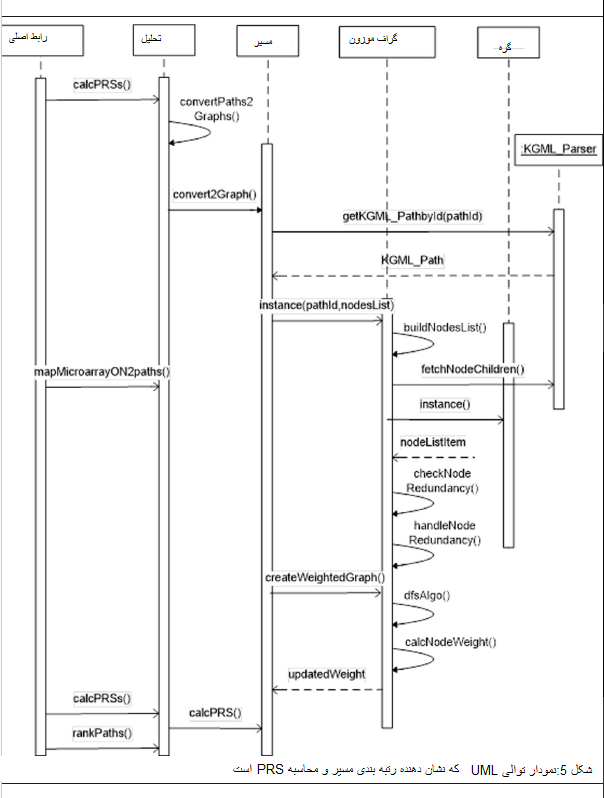
4. بعد از ایجاد یک گراف برای هر مسیر، گراف‌ها با فراخوانی تابع createwightedgraphs() وزن دهی می‌شود که از الگوریتم DFS برای عبور دادن گره‌های هر گراف و اختصاص یک وزن برای هر گره مهم با در نظر گرفتن حلقه‌های گراف استفاده می‌کند.

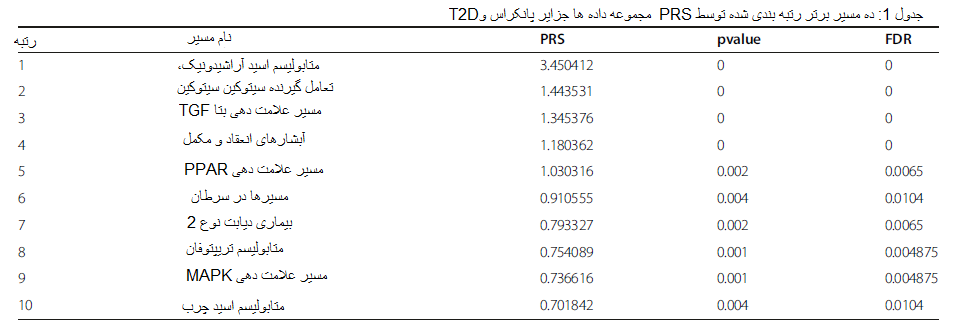
5. امتیاز نظم مسیر (PRS) به هر گراف وزن دهی شده با استفاده از اوزان گره‌های مهم در گراف و دیگر پارامترها اختصاص می‌یابد.

ما تمام این طبقه‌ها، توابع و الگوریتم DFS را با استفاده از MATLAB R2010a اجرا کردیم.

**نتایج و بحث**

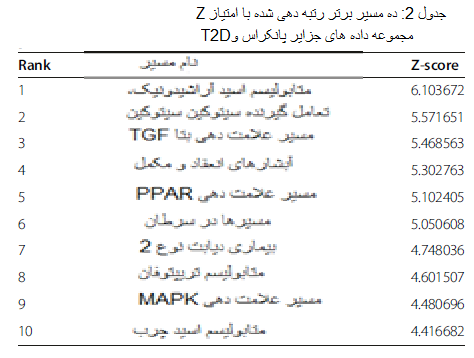
ارزیابی موردی روش‌های تحلیل غنی‌سازی جدید بسته به توانایی آن‌ها در تشخیص فرایندهای بیولوژیکی که در حالت‌های بیماری اختلال ایجاد می‌کنند، دشوار است. ما و دیگران برای این ارزیابی بامطالعه عملکرد در طی طیفی از مجموعه داده‌های ناشی از شرایط مشخص تلاش کردیم ([27] و رفرنس‌های اینجا). با گسترش دادن الگوریتم خود برای شامل شدن مسیرهای بیوشیمی، تحلیل دیگر بر روی یک مجموعه داده انجام دادیم که اختلال متابولیک رایج را توصیف می‌کند که نوع 2 بیماری دیابت است (T2DM). داده‌ها در ابتدا توسط Taneera و همکارانش [31] که سطوح تظاهر ژن را در RNA جداشده از جزایر پانکراس انسانی که از 9 اهداکننده جسد دیابت نوع 2(T2D) با نمونه‌های RNA جزایر پانکراس ناشی از 54 اهداکننده جسد غیر دیابتی مقایسه کردند، ایجاد شد. این‌ها در آرایه‌های ST ژن انسانی Affymetrix 1.0 ترکیب‌شده بودند و مقادیر تظاهر حاصل توسط تحلیل چند آرایه‌ای قوی (RMA) قبل از آپلود شدن در کل تظاهر ژن نرمال شد (www.ncbi.nlm.nih.gov/geo; accession#GDS4337). ما یک فایل ورودی شامل شناسه‌های محققین Affymetrix و داده‌های تظاهر ژن نرمال شده برای هر 63 نمونه ایجاد کردیم. دیگر پارامترها نیازمند اعداد نمونه در هر گروه (9 در گروه 1، 54 در گروه 2) و مقادیر آستانه ارزش P و تغییر کل بری فیلتر کردن ژن‌های مهم بودند (در این مورد تغییر کل ≥1.3 و مقدار p<0.05). آستانه‌های تغییر کل اتفاقی هستند و مقدار انتخاب‌شده در این مثال تعداد کافی از ژن‌های نهفته به دست داد تا نگاشت مسیر ممکن شود (در این مثال، آستانه 1.5 تنها 88 DEG به دست می‌داد). مصرف‌کننده می‌تواند انتخاب کند که برای علامت‌دهی یا مسیر متابولیک یا هردو غنی‌سازی کند. تست آماری اضافه‌تر را می‌توان در صورت نیاز با روش جایگشت انجام داد در این مثال ما از تعداد 1000 جایگشت انجام دادیم و آستانه مقدار p = 0.05 است. جدول 1 و 2 به ترتیب نشان‌دهنده ده مسیر برتر رتبه‌بندی شده طبق PRS و امتیاز Z است که در آن تنها مسیرهای مهم (FDR<0.05) انتخاب شدند. تعدادی از فرایندهای مربوط به T2DM با هر دو روش انجام شدند، مخصوصاً مسیرها متابولیک مانند "متابولیسم اسید آراشیدونیک"[32]. "متابولیسم اسید چرب" [33,34]، هم‌چنین فرایندهای علامت‌دهی مانند "مسیر علامت‌دهی”PPAR [35,36]. هر دو روش" مسیرهای سرطان " را شناسایی می‌کنند که ازآنجایی‌که این توصیف، تعدادی از فرایندهای موجود در دیابت من‌جمله آپوپتز و چرخه سلولی به همراه علامت‌دهی TGF- بتا را دربرمی گیرد، جای تعجبی ندارد [37]." تعاریف و آبشارهای انعقادی " با هر دو روش امتیاز بالایی داشتند که می‌توانست یک مثبت خطا باشد یا تغییراتی را در عروق جزایر دیابتی نشان دهد. جدای از این انتظار، گفته می‌شود که تمام دیگر مسیرهای امتیاز بالای PRS در حالت‌های بیماری نهفته می‌شوند. در مقابل، توضیح تعدادی از مسیرهای شناسایی‌شده با امتیازدهی Z دشوارتر است و لذا ممکن است مثبت‌های خطایی باشند ("شبکه ایمنی روده "،"مولکول‌های چسبندگی سلول "،"رد آلوگرافت"،" عفونت استافیلوکوکوس اورئوس "). درنهایت، روش PRS امتیاز بهتری به دو مسیر ضروری برای T2DM،" علامت‌گذاری MAPK" [38] و " دیابت نوع 2" [39] در مقایسه با امتیازدهی z ارائه کردند. درواقع، توصیف دوم به‌وضوح تأثیر بر علامت‌دهی انسولین و adipocytokine را نشان می‌دهد که برای پاتوفیزیولوژی دیابتی‌ها در مرکزیت است.





**نتیجه‌گیری**

توسعه سریع فناوری‌های پراستفاده ژنومی و نهشت خروجی آن‌ها در پایگاه‌های داده دسترسی آزاد، مقادیر زیادی از داده‌ها بیولوژیکی تولید کرده است. استخراج و تفسیر این داده‌ها، نوآوری درزمینهٔ بیولوژی محاسبه‌ای را ترغیب کرده است که منجر به ظهور ابزار پیچیده برای تولید نتایج قابل‌اعتماد، معنادار و قابل‌بررسی شده است. این از این حیث اهمیت دارد که این نوع از آزمایش‌ها گران هستند و ابزارهای جدید احتمالاً باید به تحلیل‌های قبل از این ارزش بیفزاید.



در این مقاله، ما دو حوزه را مدنظر قراردادیم: اول، گسترش الگوریتم غنی‌سازی PRS [27] برای در برگرفتن مسیرهای متابولیک و علامت‌دهی؛ و دوم، برای ارائه توصیفی دقیق از یک GUI که تحلیل آرایه را با PRS و امتیازدهی Z تسهیل می‌کند. ابزار بهبودیافته، به تعدادی از چالش‌ها مخصوصاً در نگاشت شناسه، فراوانی توصیف مسیر و ارزیابی اهمیت آماری، رسیدگی می‌کند. برخلاف امتیازدهی z، الگوریتم PRS توپولوژی یک مسیر (رابطه بین ژن‌ها) و بزرگی تغییرات تظاهر ژن را در نظر می‌گیرد تا مسیرهای نهفته را شناسایی کند. برای این دلایل، ما استدلال می‌کنیم که غنی‌سازی PRS، بینش‌های بیولوژیکی‌تری در مقایسه با روش استاندارد فوق هندسی به دست می‌دهد. ازآنجایی‌که گام‌های پیش‌پردازش اضافی اتخاذشده برای کاهش فراوانی در توصیفات KEGG به‌راحتی در دیگر روش‌ها بدون مهندسی مجدد قابل‌ملاحظه اجرا نمی‌شود، مقایسه عملکرد با دیگر روش‌های PT ممکن نیست. البته رفتار مسیرهای علامت‌گذاری و متابولیک، مشخص است. اگرچه، ازآنجایی‌که رویکرد ما به‌جای پیش‌بینی تأثیری بر تابع مسیر، ارزیابی تغییرات نسخه‌برداری در یک مسیر بود، ما احساس کردیم که معقول است تا تأثیر بر مسیرهای علامت‌دهی و بیوشیمیایی را با استفاده از یک روش منفرد ارزیابی کنیم. بدین طریق، ما قادر بودیم تا مسیرهای بیوشیمیایی که در بیماری متابولیک اختلال ایجاد می‌کند، شناسایی کنیم. یک اصل اساسی این نوع از تحلیل این است که دانشمندان زیست پزشکی در بررسی‌های متعاقب اهداف نشان داده‌شده توسط مطالعات نمای نسخه‌برداری هدایت می‌شوند. متأسفانه، هیچ تست آماری غیرمبهمی وجود ندارد که به محقان اجازه دهد تا مطمئن باشند که هر مسیر مشخص‌شده‌ای ارزش مطالعه بیشتر را دارد (و هزینه گزاف). کاربرد رویکردهای مبتنی بر جایگشت اغلب برای تعیین احتمال یک امتیاز غنی‌سازی که به‌صورت اتفاقی به‌دست‌آمده است به کار می‌رود و برای تنظیم مقادیر p با FDR می‌توان اعتمادبه‌نفس محققان را افزایش داد که یک نتیجه معنادار است.

به‌طور خلاصه، ما استدلال می‌کنیم که فراهم کردن گزینه انتخاب ابزار تحلیل برای محققان که با توجیهات مشخصی نسبت به آن آگاه شده‌اند، امکان ترکیب شدن یا متضاد شدن مدارک را برای تسهیل تصمیم‌گیری آگاهانه‌تر به وجود می‌آورد.

**در دسترس بودن و الزامات**

**نام پروژه:** نرم‌افزار PRS

**صفحه اصلی پروژه:** [**http://www.buckingham.ac.uk/research/clore-laboratory-diabetes-obesity-and-metabolic-research/staff/maysson-al-haj-ibrahim/prs-tool/**](http://www.buckingham.ac.uk/research/clore-laboratory-diabetes-obesity-and-metabolic-research/staff/maysson-al-haj-ibrahim/prs-tool/)**.**

**سیستم عملیاتی:** بسترهای نرم‌افزاری مستقل

**زبان برنامه‌نویسی:** MATLAB

**دیگر الزامات:** MATLAB 2010a یا بالاتر. اگر MATLAB بر روی کامپیوتر شخصی شما نصب نشده است، شما ابتدا نیاز به نصب محیط MCR و سپس اجرای ابزار PRS دارید.

**محدودیت‌های استفاده:** ندارد.

**فایل‌های اضافی**

فایل اضافی 1: مواردی که تشکیل‌دهنده مسیرهای KEGG هستند و در یک فایل KGML نشان داده‌شده‌اند.

فایل اضافی 2: نمودار توالی UML که نشان‌دهنده اجرای تابع ریزآرایه “initialize” است.

فایل اضافی 3: نمودار توالی UML تابع parse KGML.

فایل اضافی 4: نمودار توالی UML که نشان‌دهنده اجرای جزئیات فرایند ایجاد فهرستی از مسیرهای KEGG و نگاشت داده‌های ریزآرایه درون آن‌هاست.



