جهت تهیه فایل WORD این مقاله به سایت DaneshResan.com مراجعه نمایید و عنوان مقاله را جستجو کنید
بیش از ۲ میلیون مقاله فارسی در این سایت موجود میباشد

کومش- جلد ١٤، شماره ٤ (پیاپی ٤٨)، تابستان ۱۳۹۲

تحلیل خانواده‌های ژنی در ریزآرایه لنفوبلاستیک حاد

سهیلا خداکریم [۱](Ph.D)، حمید علوی‌مجد [۲*](Ph.D)، فرید زایری [۳](Ph.D)، مصطفی رضایی‌طاویرانی [۳](Ph.D)، سید محمد طباطبایی [۴](M.Sc)، نسرین دهقان نیری [۳](M.Sc)

۱- دانشگاه علوم پزشکی شهید بهشتی، دانشکده بهداشت، گروه اپیدمیولوژی

۲- دانشگاه علوم پزشکی شهید بهشتی، دانشکده پیراپزشکی، گروه آمار زیستی

۳- دانشگاه علوم پزشکی شهید بهشتی، دانشکده پیراپزشکی، مرکز تحقیقات پروتئومیکس

۴- دانشگاه علوم پزشکی شهید بهشتی، دانشکده پیراپزشکی، گروه انفورماتیک پزشکی

چکیده

سابقه و هدف: تحلیل خانواده ژنی داده‌های ریزآرایه، مسیرهای بیولوژیکی یا خانواده‌های ژنی که بین فنوتیپ‌های مورد نظر بیان متفاوتی دارند، را شناسایی می‌کند. بر خلاف تحلیل مبتنی بر ژن‌های منفرد، تحلیل خانواده‌های ژنی دانش بیولوژیکی موجود روی ژن‌ها را برای بررسی تفاوت بیان ژنی بین فنوتیپ‌ها به کار می‌گیرد. در این مطالعه به مقایسه دو روش تحلیل خانواده‌های ژنی پرداخته می‌شود.

مواد و روش‌ها: از دو روش طبقه‌ای و فراموضعی، برای شناسایی خانواده‌های ژنی با بیان متفاوت به عنوان عامل افتراق بین افراد سالم و بیمار (فنوتیپ) مبتلا به لنفوبلاستیک حاد که دارای کروموزوم BCR/ABL هستند، استفاده شد. در این مطالعه، داده‌های ریزآرایه‌ی یک کارآزمایی بالینی لوسمی لنفوبلاستیک حاد مورد استفاده قرار گرفته و برای تحلیل داده‌ها از نرم‌افزار R استفاده گردید.

یافته‌ها: از ۲۰۰ خانواده ژنی بررسی شده، روش فراموضعی ۱۱۴ خانواده را که بین جمعیت سالم و بیمار بیان متفاوتی داشتند، شناسایی کرد. در حالی‌که روش طبقه‌ای تنها موفق به شناسایی ۳۰ خانواده ژنی با بیان متفاوت گردید.

نتیجه‌گیری: در هر یک از دو مجموعه خانواده‌های ژنی شناسایی شده به‌وسیله هر روش، تعدادی خانواده‌ی ژنی موثر در بیماری لوسمی لنفوبلاستیک حاد دیده می‌شود. بنابراین، اگر چه معرفی روش تحلیلی با توان تشخیصی بالاتر بین افراد سالم و بیمار مبتلا به لنفوبلاستیک حاد، نیازمند به مطالعه‌ای دقیق‌تر است، اما بررسی خانواده‌های ژنی مشترک بین دو روش منجر گردید، آخرین یافته‌های بیولوژیست‌ها تنها با استفاده از این روش‌های آماری به دست آید.

واژه‌های کلیدی: خانواده ژنی، ریزآرایه، لوسمی لنفوبلاستیک حاد، فراموضعی، طبقه‌ای

مقدمه

فناوری ریزآرایه (Microarray) ابزاری پر توان است که بیان هم‌زمان هزاران ژن را فراهم آورده و امکان یافتن مکانیسم علیتی برای رفتار سلول‌ها و بیماری‌ها را ممکن می‌سازد [۲،۱]. یکی از کاربردهای متداول این فناوری، اندازه‌گیری بیان ژن در نمونه‌های بیولوژیکی تحت شرایط یا

* نویسنده مسئول، تلفن: ۰۲۱-۲۲۷۳۴۰۴۳ نمابر: ۰۲۱ - ۲۲۷۲۱۱۵۰ Email: Alavimajd@gmail.com

تاریخ دریافت: ۱۳۹۱/۵/۱۰ تاریخ پذیرش: ۱۳۹۱/۱۱/۲۴

جهت تهیه فایل WORD این مقاله به سایت DaneshResan.com مراجعه نمایید و عنوان مقاله را جستجو کنید
بیش از ۲ میلیون مقاله فارسی در این سایت موجود میباشد

فنوتیپ‌های متفاوت است. این فرآیند به شناسایی ژن‌های با بیان متفاوت، بین بافت نرمال و غیر نرمال (فنوتیپ) می‌انجامد و به عنوان یک ابزار تشخیصی مهم در علوم پزشکی تلقی می‌گردد [۲].

نتایج حاصل از فناوری ریزآرایه، لیست طویلی از ژن‌ها با بیان متفاوت بین فنوتیپ‌ها است. این لیست پایان مطالعه ریزآرایه نخواهد بود بلکه تنها نقطه شروعی برای بیولوژیست‌ها،جهت یافتن الگوی بیان ژنیبین فنوتیپ‌ها است [۳].

روش‌های آماری متنوعی، برای مدل‌بندی و تفسیر نتایج حاصل از فناوری ریزآرایه پیشنهاد شده است. در دهه گذشته، ابزار و روش‌های آماری که در سطح وسیعی برای تحلیل این داده‌ها به کار گرفته شده، تحلیل ژن منفرد (Individualgene analysis) است. در این تحلیل، معنی‌داری تفاوت بیان یک ژن و فنوتیپ مورد نظر مورد ارزیابی قرار می‌گرفت. معمولاً نتایج حاصل از تحلیل ژن منفرد لیستی از ژن‌هایی بود که بیان آن‌ها نسبت به یک نقطه برش (Cutpoint)، بین دو گروه فنوتیپ تفاوت معنی‌داری داشت. تفسیر بیولوژی نتایج حاصل از تحلیل ژن منفرد به علت حجم زیاد، بسیار دشوار بوده و اغلب استخراجیک الگوی منطقی جهت تفسیر رابطه بین لیست ژنیوفنوتیپ مورد نظر غیر ممکن ساخته است [۵,۴].

تفسیر دشوار نتایج حاصل از تحلیل داده‌ها در سطح ژن به علت بعد زیاد داده‌ها (تعداد ژن‌ها) و نادیده گرفتن روند زیستی سازگانی بین ژن‌ها، باعث جهش از تحلیل در سطح ژن به خانواده ژنی (GeneSet) گردید. تحلیل خانواده ژنی (Gene Set Analysis) از ترکیب دانش بیولوژی (خانواده‌های ژنی) و تحلیل‌های آماری برای یافتن اختلاف بیان ژن‌ها بین فنوتیپ‌ها استفاده می‌کند. خانواده ژنی به گروهی از ژن‌ها اطلاق می‌شود که از نظر بیولوژیکی، عمل‌کرد سلولی یا مکان کروموزومی به هم شبیه هستند. اطلاعات مربوط به این خانواده‌ها، در پایگاه‌های داده‌ای نظیر Gene Ontology (GO)، Kyoto Encyclopedia of Gene and Genomes (KEGG)، Genemap و Resnet در دسترس است [۶,۵].

انطباق و هماهنگی روش تحلیل خانواده ژنی با دانش بیولوژی، در مقایسه با روش تحلیل ژن منفرد بیش‌تر است. زیرا در روش تحلیل خانواده ژنی، عمل‌کرد گروهی ژن‌ها در نظر گرفته می‌شود. از دیدگاه بیولوژی، تفسیر نتایج حاصل از تحلیل خانواده ژنی ساده‌تر است، زیرا قرار گرفتن ژن‌ها در هر خانواده ژنی منطبق با مکان کروموزومی یا نقشی است که درون سلول ایفا می‌کنند و به این ترتیب درک شبکه پیچیده فرآیندهای سلولی نیز ساده‌تر می‌گردد [۷].

مدل‌های آماری متنوعی، در تحلیل خانواده ژنی به کار گرفته شده‌اند. اما استفاده از روش‌های گوناگون به نتایج متفاوتی می‌انجامد که سردرگمی‌هایی برای پژوهش‌گران به همراه داشته است. هنوز به سوالاتی نظیر این که کدام روش از توان آماری بیش‌تری برخوردار است؟ پاسخ داده نشده است. در یک نگاه کلی می‌توان روش‌های تحلیل خانواده ژنی را به دو طبقه تقسیم کرد:

- روی‌کردهایی که از لیست ژنی مبتنی بر تحلیل ژن منفرد، استفاده می‌کنند.
- روی‌کردهایی که از لیست ژنی مبتنی بر تحلیل ژن منفرد، استفاده نمی‌کنند.

در روی‌کرد اول با دو فرضیه روبه‌رو خواهیم شد: فرضیه نخست استقلال بین خانواده ژنی و فنوتیپ مورد نظر با آزمون‌هایی نظیر کای-اسکور و آزمون دقیق فیشر بررسی می‌شود [۸,۵]. در این حالت با دو متغیر برنولی (متعلق بودن یا نبودن ژن به خانواده ژنی مورد نظر، وجود یا عدم وجود تفاوت بیان ژنی معنی‌دار بین گروه سالم یا بیمار) روبه‌رو خواهیم بود. در فرضیه دوم، فرض صفر توالی تصادفی ژن‌های یک خانواده در طول لیست ژنی است. این لیست بر اساس میزان تفاوت در بیان ژن‌های حاصل از نتایج تحلیل مبتنی بر ژن منفرد می‌باشد. برای تحلیل این فرضیه از مجموعه روش‌های تحلیل غنی‌سازی خانواده ژنی (GeneSet Enrichment Analysis) استفاده می‌کنیم. این

جهت تهیه فایل WORD این مقاله به سایت DaneshResan.com مراجعه نمایید و عنوان مقاله را جستجو کنید
بیش از ۲ میلیون مقاله فارسی در این سایت موجود میباشد

روش اولین بار در سال ۲۰۰۳ توسط موسا (Mootha) و همکارانش [۹] ارایه شد که در مطالعه‌های بعدی بسط یافت [۱۰-۱۳].

در روی‌کرد دوم داده‌های خام یا مقادیر بیان ژن با استفاده از تکنیک‌های تک‌متغیره و چندمتغیره در جهت شناسایی خانواده‌های ژنی با بیان متفاوت، تحلیل می‌شوند. در بعضی از این روش‌ها هم‌بستگی بین ژن‌ها نیز مورد توجه قرار می‌گیرد. روش فراموضعی از مهم‌ترین تکنیک‌های تک‌متغیره ارایه شده، از برازش مدل‌های خطی تعمیم‌یافته روی مقادیر بیان ژن استفاده کرد [۱۴]. در سال ۲۰۰۵ منسمن و مستر (Mansmann و Mesterand) در سال ۲۰۰۸ هومل (Hummel) مدل تحلیل کوواریانس را پیشنهاد کردند. این روش مانند روش فراموضعی بود با این تفاوت که نقش بیان ژن و فنوتیپ جابه‌جا شده بود [۱۵،۱۶]. از روی‌کردهای چندمتغیره نیز می‌توان به آماره‌هایی نظیر T2 هتلینگ و آماره N، اشاره کرد [۱۷-۱۹].

در این مطالعه، به مقایسه دو روش طبقه‌ای و فراموضعی در داده‌های لوسمی لنفوبلاستیک حاد (AcuteLymphoblastic Leukemia) پرداختیم، این مجموعه شامل ۱۲۶۲۵ ژن و ۱۲۸ مشاهده (نمونه) است. در این مطالعه به بررسی تفاوت الگوی بیان بین افراد مبتلا به لوسمی لنفوبلاستیک حاد و سالم بر اساس خانواده‌های ژنی که از روی مکان کروموزومی آن‌ها در سایت KEGG تعریف شده‌اند، پرداخته شد.

مواد و روش‌ها

در ابتدا به تشریح دو روش تحلیل خانواده‌های ژنی طبقه‌ای و فراموضعی پرداخته می‌شود، با این فرض که فنوتیپ مورد نظر دو حالتی است. برای معنی‌داری آماری نیز هم‌واره از جای‌گشت (Permutation) از طریق نمونه‌گیری بر حسب واحدها (Subject Sampling) استفاده می‌شود [۲۰].

روش طبقه‌ای. روش طبقه‌ای به رغم سادگی محاسبات، یکی از بسط‌های پر توان تحلیل غنی‌سازی خانواده ژنی است. این روش برای هر ژن و برای هر خانواده ژنی P-value محاسبه می‌کند. آماره آزمون این روش، مجموع آماره‌های آزمون تی (statistics) ژن‌های یک خانواده ژنی تقسیم بر ریشه دوم تعداد ژن‌های هر خانواده است. بنابراین فرض صفر در این روش تصادفی بودن ژن‌ها در طول لیست ژنی نیست، بلکه برابری میانگین بیان ژن‌ها در دو گروه فنوتیپ است. این روش با بسته نرم‌افزاری Category در زیر فضای بیوکانداکتور (Bioconductor) در محیط نرم‌افزاری R قابل اجرا است [۲۱،۲۲].

روش فراموضعی. مبنای این روش برای بررسی ارتباط الگوی بیان ژن در یک مجموعه ژنی با فنوتیپ مورد نظر تحلیل مبتنی بر ژن منفرد نیست، بلکه از مقادیر بیان ژن به‌طور مستقیم استفاده می‌کند. این روش مدل‌های رگرسیون لوجستیک با اثرات تصادفی روی داده‌های بیان ژنی را به کار می‌گیرد، تا فنوتیپ را از روی مقادیر بیان ژن پیش‌بینی کند. اگر m تعداد ژن‌های یک خانواده ژنی و n تعداد نمونه‌ها باشد:

$$E(Y_i|r_i)=h^{-1}(\alpha+\sum_{j=1}^{m}x_{ij}\beta_j) \quad j=1,2,\ldots,m$$

H تابع ربط، α و β نیز ضرایب رگرسیونی مدل هستند. اگر βها در یک خانواده ژنی همگی صفر باشند، معادل است با آن که ژن‌ها در این خانواده رابطه‌ای با فنوتیپ مورد نظر ندارند. هم‌واره در چنین بررسی‌هایی m در مقایسه با n بسیار بزرگ‌تر خواهد بود و بنابراین درجه آزادی بسیار کوچک (df=n-m) خواهد شد. برای برطرف کردن این مشکل فرض می‌شود βها هم توزیع، با میانگین صفر و واریانس τ^2 هستند. حال فرض صفر (بیان ژن‌ها در یک خانواده ژنی بین دو فنوتیپ تفاوتی ندارند) از $H_0:\beta_1=\beta_2=\ldots=\beta_m$، به صورت مختصر $H_0:\tau^2=0$ در می‌آید. اگر $r_i=\sum_{j=1}^{m}x_{ij}\beta_j$، $E(r)=0$ و $Cov(r)=\tau^2X'X$ آن‌گاه برای آزمودن فرض صفر از آماره Q لی کسی و هوولینگ (LeCessie and Van Houwelingen) و هم‌چنین هووینگ-

دیوسترمات (Houwing-Duistermaat) استفاده می‌شود [۲۴،۲۳].

$$Q = \frac{(Y-\mu)R(Y-\mu)}{\mu 2}$$

که در آن $R = (1/m)X'X$ ، $\mu = h^{-1}(\alpha)$ و $\mu 2$ گشتاور دوم مرکزی Y تحت فرض صفر است. اگر حجم نمونه بزرگ باشد، آماره Q از توزیع نرمال مجانبی تبعیت می‌کند. چنان‌چه با نمونه‌های با حجم کوچک روبه‌رو شویم، آن‌گاه توزیع کای-اسکور تقریب بهتری خواهد بود. این روش با بسته نرم‌افزاری globaltest [۲۱] در زیرفضای بیوکانداکتور در محیط نرم‌افزاری R قابل اجرا است.

کاربرد

روش‌های طبقه‌ای و فراموضعی روی داده‌های لوسمی لنفوبلاستیک حاد مطالعه‌ی ریزآرایه چریتی (Chiaretti) و هم‌کارانش اجرا شد، این داده‌ها در وب‌سایت بیوکانداکتور در اختیار عموم قرار گرفته است [۲۵]. مجموعه داده‌ها شامل ۱۲۶۲۵ ژن، روی ۱۲۸ نمونه بود که نمونه‌ها به دو گروه برابر تقسیم شدند، افراد مبتلا به لوسمی لنفوبلاستیک حاد و افراد سالمی که هیچ وضعیت غیر نرمالی در سلول‌های آن‌ها مشاهده نشده است.

نتایج

ابتدا بیماران مبتلا به لوسمی لنفوبلاستیک حاد که منشا بیماری آن‌ها سلول‌های B است و عامل ABL/BRC (محصول جهش کروموزوم ۹:۲۲) را دارند، فیلتر شدند. پس از انجام پیش‌پردازش (Preprocessing) با بسته نرم‌افزاری genefiltering، ۱۸۵۷ ژن روی ۳۷ نمونه‌ی مبتلا به لوسمی لنفوبلاستیک حاد و ۴۲ نمونه‌ی سالم در مجموعه داده‌ها باقی ماند. برای خانواده‌های ژنی نیز از اطلاعات مربوط به KEGG استفاده شد و این ۱۸۵۷ ژن در ۲۰۰ خانواده ژنی قرار گرفتند [۲۶،۲۱].

با استفاده از روش‌های طبقه‌ای و فراموضعی به بررسی بیان ۲۰۰ خانواده ژنی بین افراد مبتلا به لوسمی لنفوبلاستیک حاد و سالم، پرداخته شد. بررسی نتایج نشان داد که روش طبقه‌ای، ۳۰ خانواده ژنی با بیان متفاوت شناسایی کرد. در حالی‌که روش فراموضعی موفق به شناسایی ۱۱۴ خانواده ژنی با بیان متفاوت بین فنوتیپ‌های لوسمی لنفوبلاستیک حاد و سالم گردید. جدول ۱ و۲ به ترتیب ۱۰خانواده ژنی که کم‌ترین P-value را در روش‌های طبقه‌ای و فراموضعی داشته‌اند، نشان می‌دهد.

جدول۱ .خانواده‌های ژنی شناسایی شده با روش طبقه‌ای

ردیف	خانواده ژنی	تعداد ژن ها	آماره آزمون	P-value
۱	Ribosome	۳۶	-۹/۳۷۱۳۱	۳/۵۹E-۲۱
۲	DNA replication	۲۸	-۶/۳۲۸۸۴	۱/۲۴E-۱۰
۳	Spliceosome	۸۵	-۴/۸۱۳۴۶	۷/۴۲E-۷
۴	Mismatch repair	۱۹	-۴/۷۵۹۲۲	۹/۷۲E-۷
۵	Homologous recombination	۱۵	-۳/۷۲۷۹۸	۹/۶۵E-۵
۶	Non-homologous end-joining	۷	-۳/۴۸۴۳۷	۰/۰۰۰۲۴۷
۷	Nucleotide excision repair	۳۰	-۳/۴۲۴۱۹	۰/۰۰۰۳۰۸
۸	Purine metabolism	۶۷	-۳/۳۱۱۳۱	۰/۰۰۰۴۶۴
۹	RNA polymerase	۱۲	-۳/۱۳۴۹۵	۰/۰۰۰۸۵۹
۱۰	Parkinson's disease	۷۱	-۳/۰۱۶۷۶	۰/۰۰۱۲۷۷

جدول۲. خانواده‌های ژنی شناسایی شده با روش Globaltest

ردیف	خانواده ژنی	تعداد ژن ها	آماره آزمون	P-value
۱	Adherens junction	۴۰	۱/۴۵E-۶	۷/۴۴E-۹
۲	Pathogenic Escherichia coli infection	۳۴	۷/۴۸E-۵	۳/۸۶E-۷
۳	Axon guidance	۵۴	۹/۹۰E-۵	۵/۱۳E-۷
۴	Fc gamma R-mediated phagocytosis	۴۵	۱/۱۳E-۴	۵/۹E-۷
۵	Tight junction	۵۵	۳/۸۱E-۴	۱/۹۹E-/۶
۶	Basal cell carcinoma	۱۳	۶/۲۳E-۴	۳/۲۸E-۶
۷	Complement and coagulation cascades	۱۹	۸/۳۱E-۴	۴/۳۹E-۶
۸	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	۲۴	۱/۰۴E-۳	۵/۵۴E-۶
۹	Leukocyte transendothelial migration	۹	۱/۲۹E-۳	۶/۹۲E-۶
۱۰	Melanogenesis	۵۹	۱/۹۸E-۳	۱/۷E-۵

جدول ۳. خانواده‌های ژنی مشترک بین روش طبقه‌ای و فراموضعی دخیل در لوسمی لنفوبلاستیک حاد

ردیف	خانواده ژنی
۱	Non-small cell lung cancer (27)
۲	Acute myeloid leukemia (28-30)
۳	Chronic myeloid (31,32)leukemia
۴	Melanoma(33)
۵	Prostate cancer (34)
۶	Glioma(35,36)
۷	Endometrial cancer Renal cell carcinoma(37)
۸	Insulin signaling pathway(38)
۱۰	Neurotrophin signaling pathway (39)
۱۱	Fc epsilon RI signaling pathway (40)
۱۲	B cell receptor signaling pathway (41,42)
۱۳	T cell receptor signaling pathway (43)
۱۴	VEGF signaling pathway (44-46)
۱۵	ErbB signaling pathway(47,48)

برای شناسایی خانواده ژنی موثر بر بیماری لوسمی لنفوبلاستیک حاد و کنکاش بیش‌تر روی نتایج به‌دست آمده، خانواده‌های ژنی با بیان ژنی متفاوت مشترک بین دو روش فوق مشخص گردید. نتایج نشان داد ۱۴ خانواده ژنی مشترک با بیان ژنی متفاوت بین دو روش وجود دارد. لیست این ۱۴ خانواده ژنی در جدول ۳ آمده است. هم‌چنین، ارتباط بین این ۱۴ خانواده ژنی و بیماری لوسمی لنفوبلاستیک حاد در مقاله‌های چاپ شده در Pubmed بررسی شد. نتایج نشان داد همه این ۱۴ خانواده ژنی، بر اساس مقالات چاپ شده در Pubmed نیز به عنوان خانواده ژنی موثر در لوسمی لنفوبلاستیک حاد معرفی شناخته شده بودند.

جدول ۴ – ژن‌های موثر در لوسمی لنفوبلاستیک حاد و مشترکبین ۱۴ خانواده ژنی جدول ۳

ردیف	ژن	ردیف	ژن
۱	AKT1(49)	۱۱	PIK3CG(58)
۲	AKT2(49)	۱۲	PIK3R1(59)
۳	BRAF(50)	۱۳	PIK3R2(59)
۴	GRB2(51)	۱۴	PIK3R3(60)
۵	HRAS(52)	۱۵	MAPK1(61)
۶	KRAS(53)	۱۶	MAPK3(62)
۷	NRAS(54)	۱۷	MAP2K1(63)
۸	PIK3CA(55)	۱۸	MAP2K2(64)
۹	PIK3CB(56)	۱۹	RAF1(65)
۱۰	PIK3CD(57)	۲۰	SOS1(66)

در بررسی خانواده‌های مشترک با بیان متفاوت بین دو روش، علاوه بر دست‌یابی به خانواده‌های ژنی مهم در بیماری لوسمی لنفوبلاستیک حاد به شناسایی موثرترین ژن‌های دخیل در این بیماری نیز پرداخته شد. به این ترتیب که ژن‌های ۱۴

خانواده‌ی ژنی بررسی شد و ژن‌هایی که بیش‌ترین تکرار را در این خانواده‌ها داشتند، استخراج گردید. در این بررسی، ۲۳ ژن در همه ۱۴ خانواده ژنی مذکور شناسایی شد. مقاله‌های چاپ شده در Pubmed ارتباط بین ۲۰ ژن شناسایی شده را با بیماری لوسمی لنفوبلاستیک حاد تایید کردند (جدول ۴) و تنها مقاله‌ای برای سه ژن AKT3، PIK3R5 و SOS2 یافت نشد.

بحث و نتیجه‌گیری

روش‌های طبقه‌ای و فراموضعی این امکان را برای محقق فراهم می‌آورند که از دانش بیولوژی موجود برای تحلیل داده‌های بیان ژنی استفاده کنند. نتایج نشان داد که تعداد خانواده‌های ژنی شناسایی شده با بیان متفاوت بین افراد مبتلا به لوسمی لنفوبلاستیک حاد و افراد سالم، در روش فراموضعی تقریباً ۴ برابر روش طبقه‌ای است. روش فراموضعی در مقایسه با روش طبقه‌ای، مبتنی بر تحلیل ژن منفرد نبوده و توجه بیش‌تری به عمل‌کرد سازگانی ژن‌ها در هر خانواده ژنی دارد. بنابراین روش فراموضعی با در نظر گرفتن هم‌بستگی بین ژن‌ها در هر خانواده ژنی، هم‌خوانی بیش‌تری با عمل‌کرد ژن‌ها در فعالیت‌های زیستی دارد. اما در هر دو روش هم‌بستگی بین خانواده‌های ژنی نادیده گرفته شده و به خانواده‌های ژنی را مستقل در نظر می‌گیرند.

در لیست ۱۱۴ عضوی خانواده‌های ژنی معنی‌دار با روش فراموضعی، خانواده‌هایی نظیر tight junction [۶۷]، basal cell carcinoma [۶۸] و Fc gamma R-mediated [۶۹] هستند که بر اساس مقاله‌های چاپ شده در Pubmed عامل بیماری لوسمی لنفوبلاستیک حاد شناخته شده‌اند.

اگر چه از ۲۰۰ خانواده ژنی، روش طبقه‌ای تنها بیان متفاوت ۳۰ خانواده را گزارش می‌کند. اما در این مجموعه به ارتباط خانواده‌هایی نظیر DNA replication [۷۰]، repair mismatch [۷۱]، non-homologousend-joining [۷۲]، purine metabolism [۷۳] با بیماری لوسمی لنفوبلاستیک حاد، در مطالعات پیشین اشاره شده است.

در مطالعه دینو (Dinu) و هم‌کارانش [۷۴] نشان داده شد که نرخ مثبت واقعی و منفی واقعی روش فراموضعی در داده‌های واقعی بیش از روش طبقه‌ای بود. تسای (Tsai) و هم‌کارانش [۷۵] نشان دادند که در هم‌بستگی بالا خطای نوع I مشاهده شده در هر دو روش بسیار به هم نزدیک است. در حالی‌که در هم‌بستگی پایین، خطای نوع I روش طبقه‌ای زیاد است.

وجود ساختار هم‌بستگی پیچیده و پارازیت‌های موجود در داده‌های ریزآرایه، توافق بین داده‌های بیان ژن و داده‌های شبیه‌سازی شده را کاهش می‌دهد [۷۶،۷۴]. بنابراین در این مطالعه برای مقایسه دو روش شبیه‌سازی انجام نگرفت. در این مطالعه برای ارزیابی زیستی روش‌های تحلیل خانواده ژنی فراموضعی و طبقه‌ای، به بررسی خانواده‌های ژنی مشترک معنی‌دار بین دو روش و بیماری لوسمی لنفوبلاستیک حاد پرداخته شد. هم‌چنین جستجوی ارتباط بین ژن‌های پرتکرار در ۱۴ خانواده ژنی مشترک بین دو روش و بیماری لوسمی لنفوبلاستیک حاد بر اساس پایگاه Pubmed انجام گرفت. زیرا هر روش فرضیه‌های خود را داشته و بدیهی است بررسی نتایج روش‌های مختلف می‌تواند منجر به ارایه فرضیه‌های جدید گردد.

در بررسی ارتباط بین ژن‌های پرتکرار در خانواده‌های ژنی معنی‌دار و بیماری لوسمی لنفوبلاستیک حاد، ژن AKT1 شناسایی شد. در سال ۲۰۱۰، این ژن به عنوان یکی از ژن‌های موثر در بیماری لوسمی لنفوبلاستیک حاد معرفی شده است. این مهم با طی مراحل آزمایشگاهی و بالینی به دست آمد، اما در این مطالعه تنها با استفاده از روش‌های محاسباتی کشف این مهم میسر گردید [۴۹]. بنابراین ممکن است با طراحی مطالعه‌های دقیق‌تر بتوان ارتباطی بین ژن‌های AKT3، PIK3R5، SOS2 و لوسمی لنفوبلاستیک حاد یافت.

تشکر و قدردانی

این مقاله حاصل طرح تحقیقاتی مشترک بین گروه آمار زیستی و مرکز تحقیقات پروتئومیکس دانشکده پیراپزشکی

دانشگاه علوم پزشکی شهید بهشتی است که بدینوسیله از هیئت رئیسه محترم دانشکده پیراپزشکی برای تامین بودجه طرح پژوهشی صمیمانه سپاسگزاری می شود.

منابع

[1] Luo W, Hankenson KD, Woolf PJ. Learning transcriptional regulatory networks from high throughput gene expression data using continuous three-way mutual information. BMC Bioinformatics 2008; 9: 467.

[2] Hou T-J. Comparison of multiple comparison methodsfor identifying differential gene expression in simulated and real papillary thyroid cancer microarray data. Journal [serial on the Internet]. 2009 Date: Available from: http://digitalcommons.library.tmc.edu/ dissertations/AAI1465785

[3] Al-Shahrour F, Diaz-Uriarte R, Dopazo J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. Bioinformatics 2004; 20: 578-580.

[4] Rivals I, Personnaz L, Taing L, Potier MC. Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics 2007; 23: 401-407.

[5] Khatri P, Draghici S. Ontological analysis of gene expression data: current tools, limitations, and open problems. Bioinformatics 2005; 21: 3587-3595.

[6] McLachlan GJ, Do K-A, Ambroise C. Analyzing microarray gene expression data. New Jersey: John Wiley & Sons; 2004.

[7] Dopazo J. Functional interpretation of microarray experiments. OMICS 2006; 10: 398-410.

[8] Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37: 1-13.

[9] Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genesinvolved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003; 34: 267-273.

[10] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005; 102: 15545-15550.

[11] Kim SY, Volsky DJ. PAGE: parametric analysis of gene set enrichment. BMC Bioinformatics 2005; 6: 144.

[12] Nettleton D, Recknor J, Reecy JM. Identification of differentially expressed gene categories in microarray studies using nonparametric multivariate analysis. Bioinformatics 2008 ; 24: 192-201.

[13] Ackermann M, Strimmer K. A general modular framework for gene set enrichment analysis. BMC Bioinformatics 2009; 10: 47.

[14] Goeman JJ, van de Geer SA, de Kort F, van Houwelingen HC. A global test for groups of genes: testing association with a clinical outcome. Bioinformatics 2004; 20: 93-99.

[15] Mansmann U, Meister R. Testing differential gene expression in functional groups. Goeman's global test versus an ANCOVA approach. Methods Inf Med 2005; 44: 449-453.

[16] Hummel M, Meister R, Mansmann U. GlobalANCOVA: exploration and assessment of gene group effects. Bioinformatics 2008; 24: 78-85.

[17] Kong SW, Pu WT, Park PJ. A multivariate approach for integrating genome-wide expression data and biological knowledge. Bioinformatics 2006; 22: 2373-2380.

[18] Lu Y, Liu PY, Xiao P, Deng HW. Hotelling's T2 multivariate profiling fordetecting differential expression in microarrays. Bioinformatics 2005; 21: 3105-3113.

[19] Xiong H. Non-linear tests for identifying differentially expressed genes or genetic networks. Bioinformatics 2006; 22: 919-923.

[20] Goeman JJ, Buhlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. Bioinformatics 2007; 23: 980-987.

[21] The BioConductorProject. Available from: www.bioconductor.org.

[22] The R Project for Statistical Computing. Available from: http://www.r-project.org/

[23] le Cessie S, van Houwelingen HC. Testing the fit of a regression model via score tests in random effects models. Biometrics 1995; 51: 600-614.

[24] Houwing-Duistermaat JJ, Derkx BH, Rosendaal FR, Van Houwelingen HC. Testing familial aggregation. Biometrics 1995; 51: 1292-1301.

[25] Chiaretti S, Li X, Gentleman R, Vitale A, Vignett M, Mandelli F, Ritz J, Foa R. Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. Blood 2004; 103: 2771-2778.

[26] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000; 28: 27-30.

[27] Chung HJ, Park CJ, Jang S, Chi HS, Seo EJ, Seo JJ. A case of lineage switch from acute lymphoblastic leukemia to acute myeloid leukemia. Korean J Lab Med 2007; 27: 102-105.

[28] Jongeneel CV, Quackenbush EJ, Ronco P, Verroust P, Carrel S, Letarte M. Common acute lymphoblastic leukemia antigen expressed on leukemia and melanoma cell lines has neutral endopeptidase activity. J Clin Invest 1989; 83: 713-717.

[29] Nishida H, Yoshimizu N, Ueno H, Fujita A, Kato T, Park JW, et al. Ph-positive acute lymphoblastic leukemia after long-term remission of Ph-positive acute myeloid leukemia. Leuk Res 2007; 31: 417-418 .

[30] Lounici A, Cony-Makhoul P, Dubus P, Lacombe F, Merlio JP, Reiffers J. Lineage switch from acute myeloid leukemia to acute lymphoblastic leukemia: report of an adult case and review of the literature. Am J Hematol 2000; 65: 319-321.

[31] Lima M, Coutinho J, Bernardo L, dos Anjos Teixeira M, Casais C, Canelhas A, et al. Philadelphia-positive T-cell acute lymphoblastic leukemia with polymyositis, migratory polyarthritis and hypercalcemia following a chronic myeloid leukemia. Ann Hematol 2002; 81: 174-177.

[32] Santos FP, Rodrigues M, Mac-Donald Bley do Nascimento C, Kerbauy FR, Ribeiro AA, Mauro Kutner J, Hamerschlak N. Philadelphia-negative acute lymphoblastic leukemia in a chronic myeloid leukemia patient receiving dasatinib. Cytotherapy 2010; 12: 113-115.

[33] Coleman A, Fountain JW, Nobori T, Olopade OI, Robertson G, Housman DE, Lugo TG. Distinct deletions of chromosome 9p associated with melanoma versus glioma, lung cancer, and leukemia. Cancer Res 1994; 54: 344-348.

[34] Gandhok N, Sartor O. Unexpected response of hormone-refractory prostate cancer to treatment with an antileukemic chemotherapy regimen. Urology 2004; 64: 807-809.

[35] Liang T, Tan T, Xiao Y, Yi H, Li C, Peng F, et al Methylation and expression of glioma pathogenesis-related protein 1 gene in acute myeloid leukemia. Zhong Nan Da Xue Xue Bao Yi Xue Ban 2009; 34: 388-394.

[36] Anoop P, Hargrave D, Zacharoulis S, Saran F. Diagnostic and therapeutic challenges owing to concurrent pontine glioma and acute lymphoblastic leukemia. J Pediatr Hematol Oncol 2008; 30: 454-457.

[37] Schafernak KT, Yang XJ, Hsueh W, Leestma JL, Stagl J, Goldman S. Pediatric renal cell carcinoma as second malignancy: reports of two cases and a review of the literature. Can J Urol 2007; 14: 3739-3744.

[38] Chow EJ, Pihoker C, Friedman DL, Lee SJ, McCune JS, Wharton C, et al. Glucocorticoids and insulin resistance in children with acute lymphoblastic leukemia. Pediatr Blood Cancer 2013; 60: 621-626.

[39] Meyer J, Rhein M, Schiedlmeier B, Kustikova O, Rudolph C, Kamino K, et al. Remarkable leukemogenic potency and quality of a constitutively active neurotrophin receptor, deltaTrkA. Leukemia 2007; 21: 2171-2180 .

[40] Bruserud Ø, Gjertsen BT, Ulvestad E. Expression of Fc (epsilon)-receptors by human acute myelogenous leukemia (AML) blasts: studies of high- and low- (CD23) affinity receptor expression and the effects of IgE-mediated receptor ligation on functional AML blast characteristics. Leuk Res 2002; 26: 515-521.

[41] Bicocca VT, Chang BH, Masouleh BK, Muschen M, Loriaux MM, Druker BJ, Tyner JW. Crosstalk between ROR1 and the Pre-B Cell Receptor Promotes Survival of t(1;19) Acute Lymphoblastic Leukemia. Cancer Cell 2012; 22: 656-667.

[42] Nahar R, Müschen M. Pre-B cell receptor signaling in acute lymphoblastic leukemia. Cell Cycle 2009; 8: 3874-3877.

[43] Luo CJ, Jiang H, Chen J, Dai JL, Shen XY, Xue HL, et al. Correlation between blood T-cell receptor rearrangement excision circles levels and severe infection in children with acute lymphoblastic leukemia. Zhongguo Dang Dai Er Ke Za Zhi 2011; 13: 466-470.

[44] Demacq C, Vasconcellos VB, Izidoro-Toledo TC, da Silva Silveira V, Canalle R, Queiroz RG, et al. Vascular endothelial growth factor (VEGF) and endothelial nitric oxide synthase (NOS3) polymorphisms are associated with high relapse risk in childhood acute lymphoblastic leukemia (ALL). Clin Chim Acta 2010; 411: 1335-1340.

[45] Hacıhanefioglu A, Gonullu E, Mehtap O, Keski H, Yavuz M, Ercin C. Effect of heat shock protein-90 (HSP90) and vascular endothelial growth factor (VEGF) on survival inacute lymphoblastic leukemia: an immunohistochemical study. Med Oncol 2011; 28: 846-851.

[46] Diffner E, Gauffin F, Anagnostaki L, Nordgren A, Gustafsson B, Sander B, et al. Expression of VEGF and VEGF receptors in childhood precursor B-cell acute lymphoblastic leukemia evaluated by immunohistochemistry. J Pediatr Hematol Oncol 2009; 31: 696-701.

[47] Kunami N, Yotsumoto F, Ishitsuka K, Fukami T, Odawara T, Manabe S, et al. Antitumor effects of CRM197, a specific inhibitor of HB-EGF, in T-cell acute lymphoblastic leukemia. Anticancer Res 2011; 31: 2483-2488.

[48] Johnston JB, Navaratnam S, Pitz MW, Maniate JM, Wiechec E, Baust H, et al. Targeting the EGFR pathway for cancer therapy. Curr Med Chem 2006; 13: 3483-3492.

[49] Evangelisti C, Ricci F, Tazzari P, Chiarini F, Battistelli M, Falcieri E, et al. Preclinical testing of the Akt inhibitor triciribine in T-cell acute lymphoblastic leukemia. J Cell Physiol 2011; 226: 822-831.

[50] Hou P, Liu D, Xing M. The T1790A BRAF mutation (L597Q) in childhood acute lymphoblastic leukemia is a functional oncogene. Leukemia 2007; 21: 2216-2218.

[51] Juric D, Lacayo NJ, Ramsey MC, Racevskis J, Wiernik PH, Rowe JM, et al. Differential gene expression patterns and interaction networks in BCR-ABL-positive and -negative adult acute lymphoblastic leukemias. J Clin Oncol 2007; 25: 1341-1349.

[52] Ahuja HG, Foti A, Bar-Eli M, Cline MJ. The pattern of mutational involvement of RAS genes in human hematologic malignancies determined by DNA amplification and direct sequencing. Blood 1990; 75: 1684-1690.

[53] Mansour MR. Oncogenic kras and notch-1 cooperate in T-cell acute lymphoblastic leukemia/lymphoma. Expert Rev Hematol 2009; 2: 133-136.

[54] Paulsson K, Horvat A, Strömbeck B, Nilsson F, Heldrup J, Behrendtz M, et al. Mutations of FLT3, NRAS, KRAS, and PTPN11 are frequent and possibly mutually exclusive in high hyperdiploid childhood acute lymphoblastic leukemia. Genes Chromosomes Cancer 2008; 47: 26-33.

[55] KalenderAtak Z, De Keersmaecker K, Gianfelici V, Geerdens E, Vandepoel R, Pauwels D, et al. High accuracy mutation detection in leukemia on a selected panel of cancer genes. PLoS One 2012; 7: e38463.

[56] Chiarini F, Falà F, Tazzari PL, Ricci F, Astolfi A, Pession A, et al. Dual inhibition of class IA phosphatidylinositol 3-kinase and mammalian target of rapamycin as a new therapeutic option for T-cell acute lymphoblastic leukemia. Cancer Res 2009; 69: 3520-3528.

[57] Tasian SK, Doral MY, Borowitz MJ, Wood BL, Chen IM, Harvey RC, et al. Aberrant STAT5 and PI3K/mTOR pathway signaling occurs in human CRLF2-rearranged B-precursor acute lymphoblastic leukemia. Blood 2012; 120: 833-842.

[58] Subramaniam PS, Whye DW, Efimenko E, Chen J, Tosello V, De Keersmaecker K, et al. Targeting nonclassical oncogenes for therapy in T-ALL. Cancer Cell 2012; 21: 459-472.

[59] Kharas MG, Janes MR, Scarfone VM, Lilly MB, Knight ZA, Shokat KM, et al. Ablation of PI3K blocks BCR-ABL leukemogenesis in mice, and a dual PI3K/mTOR inhibitor prevents expansion of human BCR-ABL+ leukemia cells. J Clin Invest 2008; 118: 3038-3050.

[60] Colamonici OR, Rosolen A, Cole D, Kirsch I, Felix C, Poplack DG, at al. Stimulation of the beta-subunit of the IL-2 receptor induces MHC-unrestricted cytotoxicity in T acute lymphoblastic leukemia cells and normal thymocytes. J Immunol 1988; 141: 1202-1205.

[61] Masiero M, Minuzzo S, Pusceddu I, Moserle L, Persano L, Agnusdei V, et al. Notch3-mediated regulation of MKP-1 levels promotes survival of T acute lymphoblastic leukemia cells. Leukemia 2011; 25: 588-598.

[62] D'Angelo V, Crisci S, Casale F, Addeo R, Giuliano M, Pota E, et al. High Erk-1 activation and Gadd45a expression as prognostic markers in high risk pediatric haemolymphoproliferative diseases. J Exp Clin Cancer Res 2009; 28: 39.

[63] Makita Y, Narumi Y, Yoshida M, Niihori T, Kure S, Fujieda K, et al. Leukemia in Cardio-facio-cutaneous (CFC) syndrome: a patient with a germline mutation in BRAF proto-oncogene. J Pediatr Hematol Oncol 2007; 29: 287-290.

[64] Rauen KA. Cardiofaciocutaneous Syndrome. Seattle (WA): University of Washington, Seattle; 1993-2007.

[65] Bruchova H, Kalinova M, Brdicka R. Array-based analysis of gene expression in childhood acute lymphoblastic leukemia. Leuk Res 2004; 28: 1-7.

[66] Swanson KD, Winter JM, Reis M, Bentires-Alj M, Greulich H, Grewal R, et al. SOS1 mutations are rare in human malignancies: implications for Noonan Syndrome patients. Genes Chromosomes Cancer 2008;47(3):253-9.

[67] Mochida GH, Ganesh VS, Felie JM, Gleason D, Hill RS, Clapham KR, et al. A homozygous mutation in the tight-junction protein JAM3 causes hemorrhagic destruction of the brain, subependymal calcification, and congenital cataracts. Am J Hum Genet 2010; 87: 882-889.

[68] El Demellawy D, Sur M, Ross C, DeNardi F, Alowami S. Composite multifocal basal cell carcinoma and precursor B acute lymphoblastic leukemia: Case report. Diagn Pathol 2007; 2: 32-39.

[69] Stefanescu RN, Olferiev M, Liu Y, Pricop L. Inhibitory Fc gamma receptors: from gene to disease. J Clin Immunol 2004; 24: 315-326.

[70] Tantawy AA, El-Bostany EA, Adly AA, Abou El Asrar M, El-Ghouroury EA, Abdulghaffar EE. Methylene tetrahydrofolatereductase gene polymorphism in Egyptian children with acute lymphoblastic leukemia. Blood Coagul Fibrinolysis 2010; 21: 28-34.

[71] Matheson EC, Hall AG. Expression of DNA mismatch repair proteins in acute lymphoblastic leukaemia and normal bone marrow. Adv Exp Med Biol 1999; 457: 579-583.

[72] van Laarhoven JP, Spierenburg GT, Bakkeren JA, Schouten TJ, De Bruyn CH, Geerts SJ, et al. Purine metabolism in childhood acute lymphoblastic leukemia: biochemical markers for diagnosis and chemotherapy. Leuk Res 1983; 7: 407-420.

[73] Kim SJ, Park TS, Lee ST, Song J, Suh B, Kim SH, et al. Therapy-related myelodysplastic syndrome/acute myeloid leukemia after treatment with temozolomide in a patient with glioblastomamultiforme. Ann Clin Lab Sci 2009; 39: 392-398.

[74] Dinu I, Liu Q, Potter JD, Adewale AJ, Jhangri GS, Mueller T, et al. A biological evaluation of six gene set analysis methods for identification of differentially expressed pathways in microarray data. Cancer Inform 2008; 6: 357-368.

[75] TsaiCA, Chen JJ. Multivariate analysis of variance test for gene set analysis. Bioinformatics 2009; 25: 897-903.

[76] Khodakarim S. Study and development of gene set analysis. Shahid Beheshti Univ Med Sci Dep Biostatistics 2012. (Persian).

Gene-set analysis on acute lymphoblastic leukemia microarray

Soheila Khodakarim (Ph.D)[1], Hamid AlaviMajd (Ph.D)[*2], Farid Zayeri (Ph.D)[3], Mostafa Rezaei-Tavirani (Ph.D)[3], Seyyd Mohammad Tabatabaee (M.Sc)[4], Nasrin Dehghan-Nayeri (M.Sc)[3]

1 - Dept. of Epidemiology, Faculty of Public Health, Shahid Beheshti University of Medical Sciences, Tehran, Iran

2 - Dept. of Biostatistics, Faculty 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- Dept. of Medical Informatics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran

(Received: 31 Jul 2012; Accepted: 12 Feb 2013)

Introduction: Gene-set analysis of microarray data determines biological pathways or gene setswith differential expression in a phenotype of interest. In contrast to the analysis of individual genes, gene-set analysis utilizes existing biological knowledge of genes in assessing differential expression. This paper compared the biological performance of two gene-set analysis methods.

Materials and Methods: To determinegene sets, which are differentially expressed between acute lymphoblastic leukemia (ALL) with BCR/ABL and those with no observed cytogenetic abnormalities, the real microarray data come from a clinical trial in acute lymphoblastic leukaemia were used in this study. For this reason, we used two GSA methods; GSEA-category and Global test and then the data were analyzedusing by R software.

Results: Globaltest identified 114 out of 200 gene sets introduced in KEGG with differentially expressed on comparing the group with BCR/ABL to those with no observed cytogenetic abnormalities. While Category could identify just 30 gene sets of this set.

Conclusion: Both of used methods include number of gene sets affecting ALL. So the more thorough study is needed to identify the more metric method. Evaluation of common gene sets among the two methods could identify the latest findings of biologists only by the used statistical methods.

Keywords: Gene set, Microarray, Acute lymphoblastic leukemia, Globaltest, Category

* Corresponding author: Fax: +98 21 22721150; Tel: +98 21 22734043

Alavimajd@gmail.com