

## تشخیص ژن‌های مسئول لوکمیا حاد بر مبنای داده‌های بیان ژنی

نمامعلی آزادی<sup>۱\*</sup> (M.Sc)، کرامت الله نوری جلیانی<sup>۱</sup> (Ph.D)، مروت طاهری کلاتی<sup>۲</sup> (M.Sc)

۱- دانشگاه علوم پزشکی تهران، دانشکده بهداشت، گروه اپیدمیولوژی و آمار زیستی و انستیتو تحقیقات بهداشتی

۲- دانشگاه علوم پزشکی تهران، دانشکده پزشکی، گروه میکروبی‌شناسی

### چکیده

سابقه و هدف: در سال ۱۹۹۹ در یک مقاله، گالوپ و همکارانش کوشیدند تا برخلاف روش‌های بسیار پرهزینه متداول و رایج، با ارائه روشی نوین بر مبنای داده‌های بیان ژنی این سرطان را تشخیص دهند. هدف مشخص بود: ارائه روشی جدید که در عین کم‌هزینه بودن، کاراتر نیز باشد. آن‌ها با مطالعه ۶۸۱۷ ژن مربوط به ۳۸ فرد مبتلا به لوکمیا با بهره‌گیری از روش‌های آماری خاصی در نهایت ۵۰ ژن را انتخاب و برای تشخیص افراد بیمار استفاده نمودند. ما در این مقاله بر آنیم برخلاف روش‌های پیچیده ارائه شده توسط گالوپ جهت تشخیص ژن‌های مسئول و مارکرهای ژنی لوکمیا حاد، به یک روش آماری ساده‌تر دست یابیم.

مواد و روش‌ها: داده‌های مورد استفاده در این تحقیق همان داده‌های گالوپ می‌باشند که با تکنیک پیشرفته DNA میکروآرای که اخیراً نیز وارد کشور شده است، به دست آمده‌اند. ما با انجام آزمون‌های آماری  $t$ -test، Wilcoxon و روش‌های مبتنی بر آزمون‌های جای‌گشت جهت تشخیص ژن‌های مسئول بیماری تلاش نمودیم.

یافته‌ها: ما با استفاده از روش وست فال و یانگ بر مبنای آماره  $t$  در سطح معنی‌داری  $\alpha=0/015$ ، ۴۸ ژن را به عنوان ژن‌های مسئول جهت تشخیص افراد شناسایی نمودیم که از این میان ۲۸ ژن با آن‌چه گالوپ و همکارانش به دست آورده‌اند مشترک است؛ یعنی یک هم‌خوانی ۵۸٪ را نشان می‌دهد. ۲۰ ژن باقی‌مانده (۴۲٪) با ژن‌های حاصله از مطالعه گالوپ تفاوت دارند.

نتیجه‌گیری: نتایج حاصله جهت تشخیص افراد، بر مبنای این ۴۸ ژن همان نتایجی را خواهد داد که ۵۰ ژن گالوپ به دنبال دارد.

واژه‌های کلیدی: DNA میکروآرای، بیان ژن، انتخاب ژن، لوکمیا، روش وست فال و یانگ، ریزآرایه

### مقدمه

سالیان متمادی است که بشر می‌کوشد به راز بیماری‌ها پی برده و علت وجودی آن‌ها را دریابد. بیماری‌های جسمی و روانی از جمله مسائل مهم انسانی هستند، که توجه بسیاری از متخصصین را در هر عصر و زمانی به خود معطوف نموده‌اند. از کوشش برای ارتباط دادن بسیاری از بیماری‌ها به نیروهای ماوراء الطبیعه در سالیان دور تا کشف پنی‌سیلین و در

سال‌های اخیر ظهور تکنولوژی مولکولی، جملگی تلاش‌هایی است که بشر در این راستا انجام داده است. امکان بهره‌گیری از علم و فن‌آوری سایر علوم دانشمندان عرصه بیولوژی را بر آن داشت که تحقیقات خود را متوجه کوچک‌ترین واحد زنده یک ارگانیسم، یعنی سلول نمایند. از این‌رو Sechena در سال ۱۹۹۶ روشی را ارائه نمود [۱۰] که بر مبنای آن محققین قادر شدند به طور هم‌زمان ژن‌های موجود در دو بافت سالم و

\* نویسنده مسئول. تلفن: ۰۲۱-۸۸۰۵۴۳۵۷-۸۵۸، نمابر: ۰۲۱-۸۸۰۵۴۳۵۵، E-mail: Nouri4@yahoo.com

درست آنها، علم تحلیل این داده‌ها خود جزء جدا نشدنی و عمده این مطالعات گردیده است. به دلیل فقدان داده‌ای در این زمینه در کشور، ما در این مقاله به ناچار از داده‌های لوکیما که توسط گالوپ و همکارانش در سال ۱۹۹۹ از مطالعه بر روی ۳۸ فرد مبتلا به لوکیما حاد به دست آمده است استفاده کرده‌ایم [۶]. ما معتقدیم که روش تحلیل داده‌های حاضر را می‌توان در موارد مشابه برای سایر داده‌های ژنی حاصل از DNA میکروآرای - که انتظار می‌رود در سال‌های پیش روی، توسط محققین کشور تولید شوند - تعمیم‌پذیر است. هدف ما این است که با استفاده از روشی ساده‌تر DEها را - که به احتمال زیاد حاوی مارکرهای ژنی سرطان لوکیمای حاد هستند - تشخیص و معرفی نماییم. از آنجا که هدف ما معرفی تکنولوژی DNA میکروآرای نمی‌باشد، جزئیات بیش‌تر در مورد تکنولوژی و نحوه کارکرد آن را می‌توانید در مراجع ۱، ۵ و ۸ ببینید.

**بیان ژنی و میکروآرای.** هر ژن، خود قطعه‌ای از DNA است که بنا بر اصل مرکزی در صورتی بیان می‌شود که ابتدا در هسته قطعه DNA متناظر با آن ژن به mRNA رونویسی شده، سپس در سیتوپلاسم این mRNA به پروتئین ترجمه گردد. پروتئین‌ها واحدهای عملکردی سلولند که از طریق ترجمه mRNAهای اختصاصی که از روی ژن‌های مربوطه رونویسی شده‌اند ساخته می‌شوند. میزان mRNAی رونویسی شده از یک ژن، با مقدار پروتئین تولید شده از آن ژن در سیتوپلاسم سلول رابطه مستقیم دارد. بنابراین با اندازه‌گیری مقدار mRNA رونویسی شده از یک ژن می‌توان به‌طور غیرمستقیم بیان آن ژن و میزان پروتئین مربوطه را محاسبه کرد. بر این اساس تکنولوژی DNA میکروآرای که در سال ۱۹۹۶ برای اولین بار توسط Schena [۱۰] معرفی گردید و بعدها توسط Golub [۶] و Alizadeh [۲] برجسته شد، فرصت گران‌بهایی را برای دانشمندان فراهم آورده تا بتوانند با اندازه‌گیری میزان mRNAی تولیدی در هر سلول میزان بیان اختصاصی ژن‌ها را سنجیده و در صورت لزوم با یک‌دیگر مقایسه کنند. در میکروآرای بعد از کشت سلول

سرطانی را مطالعه نمایند. این تکنولوژی که به DNA میکروآرای شهرت یافت، برخلاف روش‌های قبلی، که تنها مطالعه یک ژن را میسر می‌نمود، در یک تحول شگرف دانشمندان را قادر ساخت که در یک لحظه بیان هزاران ژن را تحت مطالعه داشته باشند. تا آنجا که بسیاری ظهور این تکنولوژی را انقلاب و تحولی بزرگ در تشخیص و درمان بیماری‌ها به‌ویژه سرطان تلقی می‌نمایند [۷].

نتایج به‌دست آمده از DNA میکروآرای، منجر به ایجاد دیدگاه‌های تازه‌ای در مورد نحوه شکل‌گیری، پیشرفت و پاسخ به درمان بیماری‌ها به‌ویژه بیماران سرطانی گردیده است. در این تکنیک میزان بیان هزاران ژن به‌طور هم‌زمان، جهت مطالعه رفتار و عملکرد سلول مورد بررسی قرار می‌گیرد و در نهایت با استفاده از روش‌های مناسب آماری، داده‌های حاصله تجزیه و تحلیل می‌شوند [۴، ۹]. در سال‌های اخیر خصوصیات ویژه و منحصر به فرد این داده‌ها منجر به تشکیل شاخه جدیدی در علم آمار جهت تحلیل این داده‌ها گردیده است. در حال حاضر با توجه به این که سکانس کامل ژنوم انسانی در دسترس می‌باشد، بررسی کامل نسخه‌برداری در سلول‌های نرمال و سرطانی امکان‌پذیر شده است و هم‌راه با تکامل هم‌زمان ابزارهای ضروری انفورماتیک و آنالیز داده‌ها جهت تبدیل و تفسیر آنها، نحوه نگرش به سرطان دچار تحولی شگرف گردیده است.

نظر به اهمیت تحقیقات مولکولی، شناخت آخرین دست‌آوردهای دانشمندان در زمینه ژنتیک، گسترش روزافزون تکنولوژی جدید به عرصه مطالعات مولکولی، اهمیت مطالعات مبتنی بر بیان ژن در جهت‌گیری تحقیقات جدید و هم‌چنین ورود تکنولوژی DNA میکروآرای در ماه‌های اخیر به کشورمان، که از آن به عنوان «ریز آرایه» نام برده می‌شود، ما را بر آن داشت تا مقاله حاضر را طراحی نماییم. این مقاله به عنوان اولین کار صورت گرفته در کشور در زمینه نحوه تحلیل داده‌های حاصله، جهت معرفی این تکنولوژی و نحوه تحلیل داده‌های حاصله می‌تواند گامی بسیار مهم تلقی گردد. از طرفی با توجه به انبوه داده‌های حاصله و ضرورت تحلیل

این بدان معناست که چنانچه نتیجه این روش به یک تشخیص منجر شود، بدون شک نتیجه آن درست خواهد بود. آن‌ها دلیل انتخاب ۵۰ ژن جهت پیش‌بینی را تا حدودی اختیاری دانستند و بیان داشتند که این ژن‌ها با یکی از طبقات ALL یا AML هم‌بستگی بالایی دارند. آن‌ها دلیل دیگر جهت اتخاذ ۵۰ ژن را کافی بودن این حجم نمونه جهت به‌دست آوردن نتایج مطمئن از آزمون‌های آماری در مقابل اریبی و هم‌چنین به اندازه کافی کوچک بودن این تعداد ژن جهت کاربردهای کلینیکی را برشمردند. هم‌چنین آن‌ها نتیجه‌گیری می‌کنند که پیش‌بینی نتایج بر اساس ۱۰ تا ۲۰۰ ژن نتایج یکسانی را به دنبال خواهد داشت. گالوپ و همکارانش در پایان با استفاده از این ۵۰ ژن نتیجه گرفتند که در حقیقت طبقه ALL خود مرکب از دو طبقه ALL B-cel و ALL C-cel می‌باشد.

داده‌های لوکمیا که از مغز استخوان (BM) یا از نمونه خون (PB) ۳۸ فرد مبتلا به لوکمیا حاد (۱۱ نمونه AML و ۲۷ نمونه ALL) به‌دست آمده است در آدرس اینترنتی <http://www.genome.wi.mit.edu/MPR> قابل دسترسی است.

## مواد و روش‌ها

در غالب مطالعات صورت گرفته بر مبنای اندازه‌گیری بیان ژنی به‌وسیله DNA میکروآرای [۴، ۳، ۲، ۱، ۹]، کشف طبقات (Class discovery)، پیش‌بینی طبقات (Class prediction) و شناخت مارکرهای ژنی (گروه اول) (Gene identification) به عنوان سه هدف اصلی این تحقیقات مدنظر بوده است. کشف طبقات به شناخت طبقات یا زیرطبقاتی مربوط می‌شود که تا کنون ناشناخته مانده‌اند، درحالی‌که پیش‌بینی طبقات مربوط به اختصاص دادن بیماران به طبقات، موجود و معلوم است. شناخت مارکرهای ژنی و تشخیص ژن‌های DE به انتخاب زیرمجموعه‌ای از ژن‌ها مربوط می‌گردد که در نمونه‌های مورد مطالعه از بیان به نسبت بالاتر (یا پایین‌تر) برخوردارند. جهت شناخت این ژن‌ها از

سرطانی و نرمال طی فرآیندی خاص mRNA از مابقی سلول جدا و به cDNA تبدیل می‌گردد. سپس محلول حاضر که مخلوطی از cDNAهای بافت نرمال و سرطانی است را بر روی یک چیب که حاوی هزاران چاله (Spot) می‌باشد و هر چاله خود در برگیرنده یک ژن خاص به تعداد بسیار زیاد است سرازیر می‌گردد. در این جا cDNAها با قطعات مکمل خود پیوند می‌یابند. از آن‌جا که cDNAهای سرطانی و نرمال قبلاً با رنگ‌های متفاوتی نشانه‌گذاری شده‌اند - معمولاً سرطانی با قرمز و نرمال با رنگ سبز - با استفاده از اسکنر به راحتی می‌توان تشخیص داد که در هر چاله، سهم cDNA نرمال و سرطانی چسبیده شده به هر ژن به چه میزان می‌باشد. به عبارت دیگر چنانچه در چاله اول تعداد cDNAهای بافت سرطانی زیاد باشند غلظت رنگ قرمز در این نقطه زیاد خواهد بود و معلوم می‌شود که ژن موجود در این چاله در افراد سرطانی بیان بیش‌تری داشته است.

**داده‌های لوسمی.** لوسمی حاد به طور عمده به دو طبقه لوسمی میلوپستی حاد (AML) و لوسمی لیمفوبلاستیک حاد (ALL) تقسیم می‌شود. این دو طبقه در سطح میکروسکوپی کاملاً یک‌سان بوده و بیماران مبتلا به این دو طی سال‌ها عملکرد یکسانی دارند. در سال ۱۹۹۹ در یک مقاله Golub و همکارانش [۵] کوشیدند تا با ارائه روش‌های نوینی برخلاف روش‌های متداول و بسیار پرهزینه رایج، بر مبنای داده‌های بیان ژنی، این سرطان را تشخیص دهند. آن‌ها دلیل انتخاب لوکمیا جهت این مطالعه را توانایی بررسی دقیق نتایج حاصله با نتایج روش‌های موجود بیان نمودند. آن‌ها عقیده داشتند که در حال حاضر لوکمیا توسط روش‌های موجود به خوبی قابل تشخیص است بنابراین می‌توان میزان صحت نتایج روش‌های پیشنهادی را با دقت ارزیابی نمود. گالوپ و همکارانش در نهایت ۵۰ ژن را به عنوان ژن‌های DE انتخاب و برای تشخیص افراد استفاده نمودند. این روش با دقت ۱۰۰٪ افراد را تشخیص داد. به‌طوری‌که توانست از ۳۸ نفر بیمار مورد مطالعه به‌درستی ۳۶ نفر را به یکی از طبقات ALL یا AML اختصاص دهد و برای ۲ نفر باقی‌مانده ابراز عقیده‌ای ننمود.

P-Value متناظر با هر ژن، ژن‌ها را رتبه‌بندی و زیرمجموعه‌ای از آن‌ها که رتبه‌ای بالا داشته باشند را برای آنالیز بیش‌تر انتخاب خواهیم نمود.

## نتایج

جهت تشخیص ژن‌های DE از میان ۳۰۵۱ ژن موجود، ابتدا با استفاده از t-test و Wilcoxon-test در سطح ۰/۰۵ آزمون انجام شد؛ که اولی ۱۰۴۹ و دومی ۱۰۵۴ ژن را معنی‌دار تشخیص داد. در سطح  $\alpha=0/001$ ، آزمون t به تعداد ۳۱۸ ژن و آزمون ویلکاکسون ۶۱۴ ژن را به عنوان DE متمایز نموده است.

جدول ۱. نتایج حاصله از اجرای آزمون t، آزمون Wilcoxon و روش تعدیل مبتنی بر هر یک از این دو آزمون جهت تشخیص ژن‌های DE در سطوح مختلف معنی‌داری ( $\alpha$ )

نوع آزمون	روش آزمون	سطح معنی‌داری ( $\alpha$ )				
		۰/۰۰۱	۰/۰۱۰	۰/۰۱۵	۰/۰۵	۰/۱
آزمون t	آزمون معمول t	۳۱۸	۶۴۲	۸۴۵	۱۰۴۹	۱۳۲۰
	روش تعدیل	۷	۳۸	۴۸	۹۳	۱۱۹
آزمون W	آزمون معمول W	۶۱۴	۶۵۳	۸۴۸	۱۰۵۴	۱۲۸۵
	روش تعدیل	۱۶	۴۸	۵۴	۱۰۱	۱۲۵

در این جدول هم‌چنین نتایج حاصل از روش تعدیل نیز آورده شده است. تفاوت روش تعدیل با هر یک از دو آزمون t و ویلکاکسون کاملاً مشهود است. به عنوان مثال این روش در سطح معنی‌داری  $\alpha=0/015$  برای آماره t، ۴۸ ژن و برای آزمون ویلکاکسون ۱۰۱ ژن را به عنوان ژن‌های مسئول می‌شناسد. ما این ۴۸ ژن را به عنوان DE پیشنهاد می‌کنیم که به منظور تشخیص افراد، می‌توان میزان بیان آن‌ها را اندازه‌گیری نمود و به تشخیص آنان مبادرت ورزید. از این تعداد، ۲۸ ژن با آن‌چه گالوپ و همکارانش معرفی نموده‌اند مشترک می‌باشد که در جدول ۲ نام آن‌ها آورده شده است و ۲۰ ژن مابقی، با آن‌چه گالوپ و همکارانش پیشنهاد داده‌اند متفاوتند که در ستون دوم جدول ۲ آورده شده‌اند.

میان هزاران ژن مورد مطالعه باید دو فاکتور عمده را در نظر داشت: ابتدا انتخاب روش (یا آماره‌ای) است که بتوان ژن‌ها را بر مبنای آن رتبه‌بندی نمود و دوم انتخاب یک مقدار بحرانی برای این آماره است به طوری که مقادیر بیش‌تر از آن را بتوان معنی‌دار تلقی کرد. از این دو، جنبه اول یعنی انتخاب یک آماره مناسب اهمیت بسیار بیش‌تری دارد [۳]. ضرورت اولیه در رتبه‌بندی ژن‌ها از این حقیقت ناشی می‌شود که در یک مطالعه ما عملاً قادر به پی‌گیری تعداد معدودی از ژن‌ها جهت انجام آنالیزهای بیش‌تر هستیم. از این‌رو، در بسیاری از مطالعات میکروارایی، هدف، مشخص نمودن تعداد معدودی (حدود ۱۰۰ ژن) از ژن‌هاست که با احتمال زیاد مارکرهای ژنی بیماری را در خود دارند. بنابراین اتخاذ یک روش مناسب جهت انتخاب این ژن‌ها (۱۰۰ ژن) که نسبت به سایر ژن‌ها از شرایط احراز بیش‌تری برخوردار باشند اهمیت به‌سزایی دارد.

داده‌های لوسمی که در این‌جا مورد بررسی قرار می‌گیرند از مطالعه ۶۸۱۷ ژن مربوط به ۳۸ فرد مبتلا به لوکمیا حاد که ۱۱ نفر لوکمیا حاد لیمفوبلاستیکی (ALL) و ۲۷ نفر لوکمیا میلوپتی حاد داشته‌اند استخراج شده است. بعد از پردازش اولیه با روش‌های نرمال‌سازی تعداد ۳۰۵۱ ژن در مطالعه باقی ماندند. در نتیجه، داده‌های حاصله به صورت یک ماتریس  $3051 \times 38$  خواهند بود. بدین منظور جهت تشخیص ژن‌های مسئول می‌توان از آزمون‌هایی که جهت مقایسه دو گروه در اختیار است مانند t-test و Wilcoxon استفاده نمود. این دو آزمون از آزمون‌های معمول جهت مقایسه میانگین یا میانه دو گروه می‌باشند. در این‌جا هدف ما آن است که بدانیم آیا تک‌تک ژن‌ها در این دو گروه به طور یکسان بیان شده‌اند یا خیر؟ به عنوان مثال آیا میانگین ژن CD33- یکی از ۳۰۵۱ ژن مورد مطالعه- در ۲۷ نفر مبتلا به ALL (گروه اول) با میانگین آن در ۱۱ نفر مبتلا به AML (گروه دوم) یکسان است. طبیعتاً لازم است که ۳۰۵۱ آزمون صورت گیرد، یعنی به ازای هر ژن یک آزمون. ما این دو آماره آزمون را برای هر یک از ژن‌ها در دو گروه محاسبه خواهیم نمود سپس بر مبنای بزرگی این آماره و با توجه به

جدول ۲. نام ژن‌هایی که در سطح معنی‌داری ۰/۰۱۵ با استفاده از روش وست فال و یانگ بر مبنای آماره t، معنی‌دار تشخیص داده شده‌اند. این ژن‌ها را می‌توان جهت اهداف کلینیکی به منظور تشخیص افراد به کار برد.

ژن‌های مشترک	ژن‌های غیرمشترک
1 INDUCED MYELOID LEUKEMIA CELL DIFFERENTIATION PROTEIN MCL1	ADM Adrenomedullin
2 Inducible protein mRNA	CYSTATIN A
3 LYN V-yes-1 Yamaguchi sarcoma viral related oncogene homolog	Macmarcks
4 CD33 CD33 antigen (differentiation antigen)"	SPTAN1 Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)
5 CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)"	IEF SSP 9502 mRNA
6 IL7R Interleukin 7 receptor	RB1 Retinoblastoma 1 (including osteosarcoma)
7 FAH Fumarylacetoacetate"	IGB Immunoglobulin-associated beta (B29)
8 CTSD Cathepsin D (lysosomal aspartyl protease)	GB DEF = Recombination activating protein (RAG2) gene, last exon
9 ACADM Acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	GLUTATHIONE S-TRANSFERASE, MICROSOMAL
10 CCND3 Cyclin D3	Putative enterocyte differentiation promoting factor mRNA, partial cds
11 HKR-T1	Lysophospholipase homolog (HU-K5) mRNA
12 Cytoplasmic dynein light chain 1 (hdcl1) mRNA	GTBP DNA G/T mismatch-binding protein
13 Leukotriene C4 synthase (LTC4S) gene	PLECKSTRIN
14 GB DEF = Homeodomain protein HoxA9 mRNA	VIL2 Villin 2 (ezrin)
15 Catalase (EC 1.11.1.6) 5'flank and exon 1 mapping to chromosome 11	ME491 gene extracted from H.sapiens gene for Me491/CD63 antigen
16 IRF2 Interferon regulatory factor 2	T-COMPLEX PROTEIN 1, GAMMA SUBUNIT
17 RETINOBLASTOMA BINDING PROTEIN P48	APLP2 Amyloid beta (A4) precursor-like protein 2
18 LYZ Lysozyme	TCRA T cell receptor alpha-chain
19 TCF3 Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)	INTERLEUKIN-8 PRECURSOR
20 Transcriptional activator hSNF2b	X-LINKED HELICASE II
21 Interleukin 8 (IL8) gene	
22 MYL1 Myosin light chain (alkali)	
23 DHPS Deoxyhypusine synthase	
24 Zyxin	
25 LEPR Leptin receptor	
26 TOP2B Topoisomerase (DNA) II beta (180kD)	
27 MEF2A gene	
28 C-myb gene extracted from Human (c-myb) gene	

را نخواهند داشت. از طرفی محدودیت منابع اقتضا می‌کند که عملاً تعداد معدودی ژن که از نظر مارکر بودن بیش‌ترین شانس را داشته باشند، جهت تحلیل‌های بیش‌تر و اهداف کلینیکی معرفی شوند. در نتیجه برای نایل شدن به این هدف ما مجبور به تعدیل این دو آزمون هستیم به طوری که بتوان مقدار  $\alpha$  را ثابت فرض نمود. جهت رفع این مشکل و دست‌یابی به نتایجی معتبر بایستی روشی را به کار برد که بتوان خطای نوع اول را در همان سطح ( $\alpha$ ) مثلاً ۰/۰۵، فرض نمود. بدین منظور چندین روش مانند بونفرونی، شفه، وست فال و یانگ و غیره وجود دارند. ما به علت حجم زیاد داده‌ها،

## بحث و نتیجه‌گیری

جدول ۱ تفاوت قابل ملاحظه‌ای را بین روش تعدیل با هر یک از دو آزمون t و ویلکاکسون نشان می‌دهد. آشکارا پیداست که تعداد زیادی از این ژن‌ها، ژن DE نبوده بلکه به غلط این دو آزمون آن‌ها را متمایز نموده‌اند. علت اصلی این انحراف را می‌توان در این واقعیت دید که جهت تشخیص DEها از میان ۳۰۵۱ ژن مورد مطالعه بر مبنای این دو آزمون، انجام ۳۰۵۱ آزمون فرضیه نیز لازم است. بی‌آمد انجام این انبوه آزمون، افزایش سریع  $\alpha$  (خطای نوع I) خواهد بود. در نتیجه آزمون‌های t-test و Wilcoxon-test کارایی لازم

آنکه با بررسی میزان بیان این ژن‌ها در افراد مظنون می‌توان به تشخیص افراد قبل از بروز علائم بالینی و رسیدن بیماری به مراحل نهایی تکامل خود، پرداخت. هم‌چنین این ژن‌ها را می‌توان به عنوان بهترین کاندیدا برای مارکر ژنی لوکمی در نظر گرفت.

### تشکر و قدردانی

منابع مالی این پژوهش از محل اجرای طرح شماره ۱۲۹۹ به تاریخ ثبت ۱۳۸۲/۷/۱۹ دانشکده بهداشت و انستیتو تحقیقات بهداشتی دانشگاه علوم پزشکی تهران تامین گردیده است. که بدین وسیله تقدیر و تشکر می‌شود.

### منابع

- [۱] آزادی نامعلی. روش‌های آماری در تحلیل داده‌های بیان ژنی به‌دست آمده از تکنولوژی DNA میکروآری. پایان‌نامه کارشناسی ارشد، تهران: دانشکده بهداشت، دانشگاه علوم پزشکی تهران، ۱۳۸۲.
- [2] 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.
- [3] 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: Technical Report # 576, Department of Statistics, U. C. Berkeley.
- [4] 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>.
- [5] Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM. Expression profiling using cDNA microarrays. *Nature Genetic*, 1999; 21:10-4.
- [6] 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.
- [7] Habek M. DNA microarray technology to revolutionise cancer treatment. *Lancet Oncol*, 2001; 2(1):5.
- [8] Nguyen DV, Arpat AB, Wang N, Carroll RJ. DNA microarray experiments: biological and technological aspects. *Biometrics*, 2002; 58(4):701-17.
- [9] Satagopan JM, Panageas KS. Tutorial in biostatistics, a statistical perspective on gene expression data analysis. *Stat Med*, 2003; 22:481-99.
- [10] 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.

ساختار آن‌ها و کارایی بهتر روش وست فال و یانگ [۳] نسبت به سایر آزمون‌ها در داده‌های میکروآری از این روش استفاده کرده‌ایم. این روش توانسته است با مبنا قرار دادن هر یک از آماره‌های  $t$ ، مشکل زیاد شدن خطای نوع اول ( $\alpha$ ) و در نتیجه نامعتبر شدن نتایج را حل کند. لازم است توضیح دهیم که این روش تعدیل را می‌توان نه تنها برای آماره‌های  $t$  و ویلکاکسون، بلکه برای هر آماره دل‌خواه دیگری نیز به‌کار برد. روش وست فال و یانگ (یا روش تعدیل) بر مبنای آماره  $t$  یا ویلکاکسون با دقت بیش‌تر  $p$ -value‌های تعدیل شده‌ای را تولید می‌کند که منجر به تشخیص تعداد کم‌تری ژن DE می‌شود. بر این اساس و با مبنا قرار دادن آماره  $t$  این روش منجر به تولید ۴۸ ژن در سطح معنی‌داری ۰/۰۱۵ شده است. این نتایج به میزان ۵۸٪ با نتایج گالوپ هم‌خوانی دارد که این می‌تواند بسیار حائز اهمیت باشد. زیرا ما توانستیم با انجام روشی به مراتب ساده‌تر که توسط نرم‌افزارهای آماری موجود مانند  $S$ -plus یا  $R$  قابل انجام است به این نتایج برسیم. این در حالی است که گالوپ و همکارانش جهت تحلیل داده‌ها از نرم‌افزاری خاص استفاده نموده‌اند که برای ما قابل دسترسی نبود. نکته مهم‌تر آن‌که با این تعداد ژن پیشنهادی، می‌تواند دقیقاً به همان نتایج گالوپ در تشخیص افراد دست یافت، زیرا بنا بر نتایج گالوپ و همکارانش صحت تشخیص ۱۰ ژن نتایجی را به همراه خواهد داشت که با ۲۰۰ ژن حاصل می‌گردد. بنابراین چنانچه ما سطح معنی‌داری را ۰/۰۵ در نظر بگیریم (یعنی ۹۳ ژن) کلیه ژن‌های تشخیص داده شده به‌وسیله گالوپ را پوشش خواهد داد. شناخت این ژن‌ها جهت درمان و حتی پیش‌گیری از آن می‌تواند بسیار مهم و حائز اهمیت باشد. با اطلاع از نحوه بیان این ژن‌ها در افراد مبتلا، پزشکان قادر خواهند بود که با تجویز داروها و روش‌های درمانی مناسب، میزان بیان آن‌ها را کنترل نمایند. نتیجه مهم‌تر