رضا طالبی، دانشجوی دکتری ژنتیک و اصلاح نژاد دام دانشگاه بوعلی سینا همدان
احمد احمدی، استادیار گروه علوم دامی دانشگاه بوعلی سینا همدان

بیان ژن و کاربرد آن در علوم زیستی

امروزه ابزارهای آزمایشگاهی پیشرفته در علم ژنتیک مولکولی و نرم‌افزارهای کامپیوتری وابسته جهت آنالیز داده‌های حاصل، سبب گسترش در علم بیوانفورماتیک شده که به طور فزاینده در سایر شاخه‌های علوم زیستی نفوذ کرده است. بیوانفورماتیک در زمینه‌های مختلف به همراه روش‌های آزمایشگاهی سبب پیشرفت در موضوعات علوم زیستی خصوصاً بیان ژن شده است. از آنجایی‌که در بیان ژن از محصول به ماده اولیه می‌رسیم این تکنولوژی امکان طراحی الگوریتم یا مدل‌سازی را بر اساس وقایع اتفاق افتاده طی بیان ژن در اختیار قرار می‌دهد. روش‌های مختلفی جهت بررسی بیان ژن از قبیل لکه‌گذاری نورترن (northern blotting)، هیبریداسیون درجا (in situ hybridization)، رونوشت بردار معکوس (reverse transcription)، شناسایی توالی رونویسی شده (expressed sequence tag)، ریزآرایه (microarray)، آنالیز سریالی بیان ژن (serial analysis of gene expression) و RNA-Seq موجود است که هرکدام فواید و معایب منحصر به فردی دارند. این مطالعه به تعدادی از کاربردهای فراوان توالی‌یابی ترانسکریپتوم و استراتژی‌های تحلیلی مرتبط بر آن اشاره دارد.

در دهه ۱۹۵۰ نقش اساسی RNAدر ترجمه در قالب نظریه اساسی مولکولی، توسط جیمز واتسون مطرح شد. این نظریه از آنجا نشأت گرفت که در سلول‌های یوکاریوتی، DNA درون فضای هسته محصور است ولی پروتئین‌ها در فضای سیتوپلاسمی و در حضور RNAهای فراوان ساخته می‌شوند. با در نظر گرفتن نقش DNA و حضور آن در هسته و نیز کد شدن این اطلاعات در قالب پروتئین در سیتوپلاسم، حضور مولکول های واسطه‌ای به نام RNA مطرح شد (۵). داده‌های بیان ژن اطلاعات ارزشمندی در مورد شبکه‌های بیولوژیک، حالات سلولی و فهم عملکرد ژن‌ها ارائه می‌دهد. یک هدف از تحلیل داده بیان ژن تعیین چگونگی تأثیر بیان هر ژن منفرد بر روی بیان ژن‌های دیگر در همان شبکه ژنتیکی است و هدف دیگر مشخص کردن این نکته است، که چگونه ژن‌ها در سلول‌های سالم و بیمار بیان می‌شوند. در علم پزشکی پروفایل بیان ژن در مدیریت و کنترل بیماری‌های عفونی و مقایسه بیان ژن در سلول‌های سالم و سرطانی کاربرد عملی دارد (۲۱). از جمله روش‌های مورد استفاده در تحلیل داده‌های بیان ژن شامل نرمال سازی، خوشه‌بندی، طبقه‌بندی و غیره است (۲۸). واحدی و همکاران (۲۰۰۸)، با خوشه‌بندی داده‌های بیان ژن در بیماران سرطانی آن‌ها را در ۲ گروه جداگانه با دقت بالا جای دادند که بیشترین نزدیکی با اطلاعات واقعی حاصل شد. بیان ژن و ارتباط آن با چند شکلی‌های ژنی می‌تواند در پاسخ افراد به داروها نیز مطرح باشد. به عنوان مثال ژن MDR۱ که به طور طبیعی در بسیاری از بافت‌های بدن انسان از قبیل مغز، کلیه، کبد، ماهیچه و روده بیان شده و نقش حفاظت این بافت‌ها را در برابر مواد سمی دارد (۲۰)، سبب مقاومت دارویی در سلول-های سرطانی نیز می‌شود. تاکنون بیش از ۵۰ چند شکلی تک نوکلئوتیدی(SNP) در آن گزارش شده که در حدود ۲۰ مورد از آن‌ها جهش‌های خاموش هستند (۲۷)، اگرچه در تحقیقات سامانیان و محجوبی (۲۰۱۲)، ارتباطی بین بیان ژن MDR۱ و چندشکلی‌های موجود یافت نشد.

امروزه تکنولوژی‌های توالی‌یابی با توان بالا به‌طور رایج در بیولوژی استفاده می‌شوند. این تکنولوژی‌ها میلیون‌ها توالی از قطعات شکسته شده‌ی ژنوم با اندازه های ۹۵ جفت باز (Read) جهت اتصال به آداپتور را تولید می‌کنند که در پروژه‌های ژنوم، اپی‌ژنوم و ترانس‌کریپتوم (کل RNA کد شونده و غیر کد شونده) استفاده می‌شوند (۱۶). توسعه‌ی تکنولوژی‌های توالی‌یابی با توان زیاد، نقش در کاهش هزینه و به دنبال آن افزایش سریع در بازده توالی‌یابی داشته، اگرچه خطاهای توالی‌یابی برخی خوانش‌ها با طول کوتاه هنوز قابل توجه است. این چالش‌ها نیاز فوری به الگوریتم-های بیوانفورماتیکی مؤثر با سیستم تأثیرگذار دارد که مقادیر بالایی داده توالی‌یابی ترانسکریپتوم تولید نماید و مطالعات مرتبط با تنوع را نیز انجام دهد. استراتژی‌های تحلیلی مرتبط بر توالی‌یابی ترانسکریپتوم شامل نقشه یابی خوانش‌ها با طول کوتاه، شناسایی محل اتصال اسپلایسینگ اگزون _ اگزون، کمیت بیان ژن یا ایزوفرم‌های آن، تجزیه و تحلیل بیانی متفاوت و نوسازی ترانس‌کریپتوم است(۳)، که در تحقیق حاضر به آن‌ها پرداخته شده است.

بررسی بیان ژن

بیوانفورماتیک و فن‌آوری‌های جدید، روش‌های بیولوژیکی سنتی را دچار تحول کرده است. تلفیق ابزارهای محاسباتی و دستگاه‌های پیچیده مهندسی، زمینه را برای کشف حیطه‌های ناشناخته خاصی همچون ژنتیک بیش از پیش فراهم کرده است (۲۸). روش‌های ابتدایی در آنالیز بیان ژن از قبیل لکه‌گذاری نورترن، هیبریداسیون درجا، رونوشت بردار معکوس و شناسایی توالی رونویسی شده EST، بازده پایین و محدودیت کارایی در تعداد زیاد ژن دارند (۱۷). از جمله فن‌آوری‌های نوین حاصل از

تلفیق بیوانفورماتیک و روش‌های بیولوژیکی در حوزه استفاده از داده‌های بیان ژن روش ریزآرایه،SAGE (آنالیز سریالی بیان ژن) و RNA Seq) (Sequencing RNA است.

◈ ریز آرایه

در دهه گذشته، با توانایی به مطالعه در اطلاعات ژنتیکی ژنوم در دامنه گسترده، ریزآرایه‌ها روشی با توان عملیاتی بالا به منظور تجزیه و تحلیل‌های بیان ژن به حساب آمدند (۱۷). ظهور فن‌آوری نوین ریز آرایه را می توان در سال ۱۹۹۵ دانست. با استفاده از این تکنیک می‌توان به طور همزمان بیان هزاران ژن را در حداقل زمان ممکن انجام داد و در سال‌های اخیر موجب تولید حجم انبوهی از داده‌های بیان ژنی شده است. این روش هر توالی ژنی شناخته شده به عنوان یک کاوشگر روی یک آرایه شیشه‌ای یا نایلونی ثبت می‌شود. mRNA استخراج شده از بافت یا نمونه خون بارنگ‌های فلورسنت علامت‌گذاری می‌شود و کاوشگرها با RNA مکمل بر روی یک آرایه هیبرید می‌شوند. دو نوع آرایه با بیشترین کاربرد عبارتند از:

◈ آرایه بر پایه DNAمکمل (Complementary DNA Spotted)

◈ آرایه بر پایه الیگونوکلئوتید (Oligo nucleotide array)

آنالیز ریز آرایه محدودیت‌های خاصی دارد که از آن جمله می‌توان عدم توانایی در شناسایی رونوشت‌های جدید، دامنه‌ی محدود، تکرارپذیری دشوار و عدم انجام مقایسه بین آزمایش‌های مختلف به علت خطاهای تصادفی و معمول محققان و آزمایشگاه‌ها را نام برد (۱۹). برای آنالیز داده‌های ریز آرایه می‌توان از شبکه‌هایی موسوم به شبکه بیانی ژن GCNs استفاده کرد. این شبکه برای انواع وسیعی از مسائل بیولوژی مانند نقش اثرات متقابل بین پروتئین‌ها، کشف جایگاه اتصال فاکتور رونویسی و مدل‌سازی اثرات متقابل ژنتیکی به کار برده می‌شود. چندین ابزار برای تصویرسازی و آنالیز شبکه‌های بیولوژی از جمله VisAnt، cytoscape و tYNA استفاده می‌شود (۳۱). در "شکل ۱" مراحل تکنیک ریز آرایه برای شناسایی و بیان ژن در سلول‌های سالم و بیمار ارائه شده است.

◈ آنالیز سریالی بیان ژن

جهت آنالیز بیان ژن در مقیاس گسترده می‌توان از تکنیک قدرتمندی به نام آنالیز سریالی بیان ژن (SAGE) استفاده نمود. این تکنیک امکان بررسی وسیع رونوشت‌های RNA را بدون داشتن اطلاعات اولیه از ترانس‌کریپتوم آن فراهم می‌کند. SAGE کتابخانه بزرگی از توالی‌های الیگونوکلئوتیدی کوتاه (tag) با منشأ mRNA فراهم می‌آورد که از بافت یا سلول خاص استخراج شده است. توالی هر tag برای جستجوی آن در پایگاه اطلاعاتی کافی است و فراوانی هر tag به طور مستقیم نشان دهنده فراوانی رونوشت مربوطه است. از کاربردهای SAGE می‌توان بررسی تغییرات ترانس‌کریپتوم در مسیرهای مربوط به سرطان، بررسی اثرات دارو بر بافت های مورد بررسی، دیابت و پارکینسون، در مطالعات گیاهی، در تعیین تفاوت بیان ژن در نمونه‌هایی که تحت شرایط پاتولوژیکی و فیزیولوژیکی مختلف قرارگرفته‌اند، شفاف ساختن مسیرهای مختلف سلولی، کمک به شناسایی مسیرهای ایجاد بیماری، شناسایی ژن‌های جدید و با فراوانی کم، تکمیل سایر مطالعات بیان ژن، شناسایی اهداف پایین دست آنکوژن‌ها و ژن های مهارکننده‌ی تومور را نام برد (۱۰). در "شکل ۲" مراحل تکنیک SAGE آورده شده است.

تفاوت روش SAGE و ریز آرایه در آنالیز بیان ژن (۱ ، ۲۲) عبارتند از:

- ✓ روش ریز آرایه، نیازمند آگاهی اولیه از توالی ژن مورد نظر است. در حالی که SAGE نیازی به آگاهی اولیه‌ی ژنوم ارگانیسم ندارد.
- ✓ در ریز آرایه نواحی مختلف رونوشت و ترکیب بازها جهت هیبریداسیون تعیین کننده است ولی در تکنیک SAGE ناحیه ی ۳ اگزون، هدف ردیابی است و حضور آنزیم محدود کننده در این جایگاه جهت رها شدن برچسب از الگو تعیین کننده است.
- ✓ نحوه شناسایی در ریز آرایه بر پایه سیگنال‌های فلورسانس و شدت آن‌ها است ولی در SAGE بر پایه تعیین کمی رونوشت است.
- ✓ اطلاعات ریز آرایه نسبت به SAGE پیچیده تر است.
- ✓ SAGE نسبت به تکنیک ریز آرایه قابلیت تعیین دقیق فراوانی mRNA و نشان دادن تفاوت‌های اندک در سطح بیان ژن بین نمونه‌ها را دارد.
- ✓ معمولاً تکنیک SAGE جهت فراهم نمودن اطلاعات اولیه از ترانس‌کریپتوم استفاده و سپس ریز آرایه‌ها این اطلاعات را گسترش و در تعداد نمونه‌های بیشتری به کار می‌گیرد.
- ✓ اطلاعات SAGE مشابه روش RNA-Seq دیجیتالی است در نتیجه در روش SAGE امکان مقایسه بین کتابخانه‌ها به سهولت انجام می‌گیرد ولی در روش ریزآرایه همان طور که در بالا اشاره شد امکان مقایسه بین آزمایش‌های مختلف آن وجود ندارد.

◈ تکنیک RNA-Seq

RNA-Seq یک تکنیک ابتکاری است که مطالعات ترانس‌کریپتوم را بر اساس تکنولوژی‌های نسل جدید توالی‌یابی انجام می‌دهد و بر بسیاری از مشکلات ریز آرایه غلبه می‌کند (۱۷). توالی‌یابی نمودن داده می‌تواند باعث آنالیز

مجدد اگزون-هایی که جدیداً کشف شده‌اند شود، در حالی که نمونه‌ها بر روی یک ریزآرایه با کاوشگرهای به‌روز، مجدداً بایستی آنالیز شوند (۹). RNA-Seq به‌طور گسترده وابسته به ابزارهای بیوانفورماتیکی توسعه یافته جهت پشتیبانی در مراحل متفاوت پردازش داده‌ها است (۱۷). تکنولوژی RNA-seq می تواند در تشخیص صفات اپی‌ژنتیک کمک کننده باشد. اپی‌ژنتیک، مطالعه تغییرات‌برگشت‌پذیر در بیان ژن‌ها است که از طریق تقسیمات میتوز، میوز و یا هر دو، قابل وراثت بوده و تغییر در توالی DNA را شامل نمی‌شود (۱٤). گام نهایی توالی‌یابی RNA، به عنوان RNA-seq شناخته می‌شود و از محدودیت‌های تکنولوژی‌های قبلی از قبیل وابستگی اولیه به دانش در مورد ارگانیسم‌ها که در روش ریز آرایه و PCR نیاز است، عاری است (۱۵).RNA-Seq یک روش قدرتمند بر پایه‌ی توالی‌یابی است که محققان را قادر می‌سازد تا توالی‌یابی و کمیت رونوشت‌های RNA را در یک محصول کامل بیان ژن (ترانسکریپتوم) تعیین کنند. این تکنیک برای بررسی ژنوم تازه توالی یابی شده و شناسایی ژن های جدید یا ایزوفرم های آن در موجوداتی که تفسیرات ژنی هنوز ناکامل مانده بسیار ارزشمند است. ضمناً به ما امکان می‌دهد تا فعالیت‌های ژنی را در بافت‌های مختلف ارگانیسم‌ها، مراحل و یا تحت شرایط گوناگون بررسی کنیم. RNA-Seq در مقایسه با ریز آرایه‌ها در یک لحظه تقریباً تصویر تمامی رونوشت‌های بیان شده را ضبط کرده در حالیکه ریز آرایه تنها تکیه بر اطلاعات پیشین دارد. از این رو قادر به شناسایی متغیرهای اسپلایسینگ، ژن‌ها و رونوشت‌های جدید نیست (۱۳ و ۳۰). تکنیک RNA-Seq احتیاجی به کاوشگر و آغازگر ندارد و داده‌های کاملاً نااریب و توانایی تجزیه و تحلیل آن، کشف رونوشت را که در روش‌های سنتی بر اساس ریز آرایه امکان پذیر نیست فراهم می‌کند. به‌علاوه RNA-Seq در تجزیه و تحلیل بیان ژن می‌تواند رونوشت‌های جدید، ایزوفرم‌های جدید، مکان‌هایی که اسپلایسینگ به طور متناوب در آن‌ها رخ داده (۷ و ۲۳)، RNAهای غیر کد کننده طویل (lncRNAs) (۸)، بیان در آلل خاص، رونوشت‌های کمیاب و متغیرهای تک‌نوکلئوتیدی در اگزون‌های بیان شده در یک تک‌آزمایش را شناسایی کند (۲).RNA-Seq برخلاف ریزآرایه‌ها کارایی در اندازه‌گیری شدت کاوشگر پیوسته، کمی نمودن بیان ژن‌ها و رونوشت‌ها (۲۵)، یافتن الحاق‌های ژنی (۱۸)، توالی‌یابی دیجیتال جهت ردیف کردن خوانش‌ها را نیز دارد. طبیعت دیجیتال این پروسه با ایجاد یک دامنه‌ی نامحدود پویا، محققان را قادر می‌سازد تا بیان کمی RNAها را با تفکیک پذیری بالا مورد بررسی قرار دهد و برای دستیابی به تغییرات ظریف بیان ژن که به فرآیندهای بیولوژیکی وابسته‌اند اهمیت دارد (۲۹). کاوشگرهای استاندارد ریز آرایه فقط بیش از ۲۰٪ از میانگین یک ژن را پوشش می‌دهند و به طور بیولوژیکی فقط به یک بخش از داده مربوطه دستیابی دارند، در حالی که RNA-Seq یک پروفایل از رونوشت کامل است (۹). علاوه بر این، RNA-Seq اختلال پایین، حساسیت بالا و در نمونه هایی با میزان کم RNA کارایی قابل توجه دارد. این تکنیک توانایی درک بهتر پیچیدگی‌ها و دید کلی بی سابقه از ترانسکریپتوم در گونه‌های مختلف را برای ما فراهم می‌کند. تکنولوژیRNA-Seq چندین مزیت بیشتر نسبت به ریز آرایه دارد که در جدول ۱، این‌دو روش باهم مقایسه شده‌اند (۹).

اگرچه تکنیک RNA-Seq فواید شگرفی به همراه داشته است ولی چالش‌های زیادی بین تکنولوژی‌های توالی‌یابی و تجزیه وتحلیل داده‌های بیوانفورماتیک وجود دارد. با این همه کتابخانه ژنومی حاصل از RNA-Seq دارای اریب بوده و کتابخانه‌های رشته‌ای خاص که اهمیت زیادی در جهت‌یابی رونوشت‌ها دارند، هنوز به آسانی ایجاد نشده‌اند. به‌علاوه این تکنیک مقادیر زیادی داده تولید می‌کند و به طور کلی طول خوانش‌ها کوتاه و دارای خطاهای توالی‌یابی هستند. به طور کلی این قبیل چالش‌ها مرتبط با روش‌ها والگوریتم‌های تأثیرگذار در تحلیل داده‌های RNA-Seq است. اهمیت توالی‌های ژنوم مرجع برای مطالعات RNA-Seq بسیار زیاد است زیرا آن‌ها الگوهای لازم برای نقشه یابی خوانش‌ها و تفسیرهای مرتبط بر الگوریتم‌های لازم برای آنالیز بهینه‌ی نتایج را فراهم می‌کنند. در "شکل ۳" مراحل تکنیک RNA-Seq به طور نمونه توضیح داده شده است (۲۹).

کاربردهای داده RNA-seq

نوسازی ترانسکریپتوم

ترانسکریپتوم، مجموعه‌ای از RNA تولید شده (کد کننده و غیر کد کننده) در یک یا جمعیتی از سلول است که روش کاربردی و محسوس در تهیه آن، تکنیک RNA-Seq است. در حال حاضر، عمدتاً ۲ دسته استراتژی به منظور ساخت ترانسکریپتوم موجود است "جدول ٦".

- **روش genome-guided**، در ابتدا نقشه تمامی ریدهای توالی‌یابی شده‌ی ترانسکریپتوم تهیه و سپس ریدهای ردیف شده بر اساس اطلاعات نقشه یابی آن‌ها به داخل رونوشت‌ها یا قطعات، سرهم می‌شوند.
- **روشgenome-independent**، بدون نیاز به ژنوم مرجع، ریدها را مستقیم به داخل رونوشت‌ها سرهم می‌کنند (۸ و ۲۲). برخی برنامه‌ها از قبیل Cufflinks و Scripture با اجرای محاسبات یکسان از ریدهای مستقیماً اتصال یافته به منظور نوسازی ترانس‌کریپتوم طی روش genome-guided استفاده می‌کنند. نرم‌افزارهای Cufflinks و Scripture به ترتیب بر اساس حداکثر دقت و حساسیت عمل می‌کنند. روش genome-guided که در ارگانیسم‌های بررسی شده در

دسترس است نیازمند ژنوم مرجع نسبتاً کامل و با کیفیت بالا اســـت. نرم‌افزارهای Velvet Trinity ، Trans-ABySS ، و Oases (منتشر نشده) بر اســـاس روش genome-independent عمـــل می‌کننـــد. نرم‌افـــزار Velvet دارای کارایی ســـرهم کردن مجدد یا دنووا ژنوم و ترانس‌کریپتوم است. این نرم‌افزار توسط یک سری الگوریتم‌ها نقش تجزیه نمودار Bruijn و در نهایت ســـرهم کردن ریدها به داخل کانتیگ‌ها یا اسکافولدها را بر عهده دارد. به طور کلی روش‌های genome-guided و genome-independent بـــه ترتیب نیازمنـــد و فاقد نیاز به ژنوم مرجع هســـتند. در جدول ٦، ابزارهای لازم جهت نوســـازی ترانس‌کریپتوم نام برده شده است.

◈ نقشه یابی خوانش‌های کوتاه

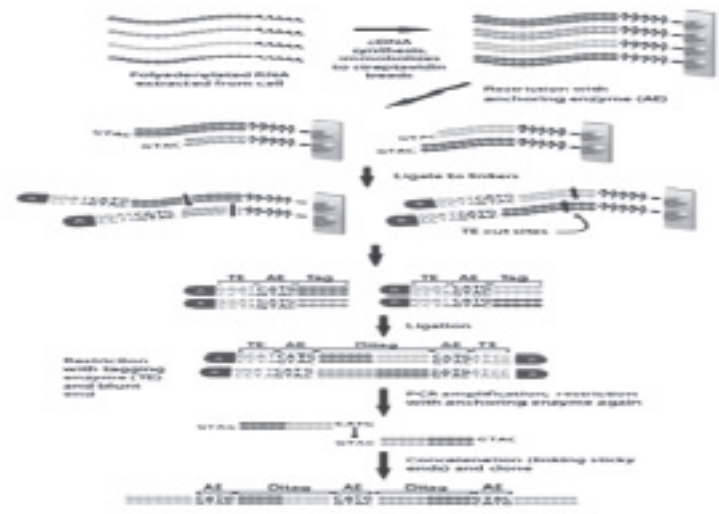
گام پایـــه و بســـیار مهـــم در آنالیز داده‌های RNA-seq ردیف‌یابی خوانش‌ها اســـت. پیچیدگی توالی‌های ژنـــوم تأثیرات مســـتقیم بر دقت نقشـــه یابی خوانش‌های کوتـــاه دارد. ژنـــوم پروکاریوتـــی به علت کوچک بودن، پیچیدگی کمتری نســـبت به ژنوم یوکاریوتی دارد. ژنوم پستانداران بسیار بزرگ و محتوی توالی‌های تکراری و همسان اســـت که این مشابهت‌ها چالش‌های بزرگی برای نقشـــه یابی خوانش‌های کوتاه هستند. تفاوت زیاد در طول اگزون‌ها و اینترون‌های یوکاریوت‌ها ســـبب دشواری‌های زیادی در اجرای الگوریتم‌های نقشه یابی در ژنوم آنها می‌شـــود. کوتاه یا بلندی اینترون‌های آن‌ها زمان محاسبات را افزایش می‌دهد. علاوه بر این به دلیـــل تعداد خوانش‌های زیاد با طول ۳۵-۴۰۰ جفت بـــاز، خطاهای توالی‌یابی بر ســـختی‌ها و ابهامات ردیف‌یابـــی می‌افزاید (۳). بر طبق نقشـــه یابی توالی‌های خوانش شـــده‌ی کوتاه، ســـرعت و دقت مهم ترین عامل تأثیرگذار بر داده‌های RNA-Seq و تکمیل نتایج تحلیلی آن به حســـاب می‌آید. برنامه‌های طراحی نقشه‌ی داده RNA-Seq با طول خوانش‌های کوتاه به ۲ دســـته متصل و غیر متصل تقســـیم بندی می‌شـــوند. نوع متصل معمولاً بـــرای ردیف کردن ژنوم‌های مرجع به خاطر وجود فواصل بزرگ ناشـــی

"شکل ۱" مراحل تکنیک ریز آرایه به منظور مقایسه الگوی بیان ژن در سلول‌های سالم و بیمار. ابتدا DNA مورد نظر (mRNA استخراج شده) که همان cDNAهای تکثیر شده توسط PCR است بر روی لام مخصوص ریزآرایه دورگ گیری می‌شود، سپس توسط مواد فلورسنتی نشاندار می‌شوند. پس از نشاندار شدن مولکول‌های cDNA عملیات شستشو انجام گرفته تا اتصالات غیراختصاصی جدا و تشعشع فلورسنتی مربوط به اتصالات اختصاصی توسط اسکنر ثبت و میزان روشنایی که بیانگر میزان بیان است توسط نرم افزار مربوطه محاسبه و آنالیز آماری انجام می‌شود (۲۲).

از اینترون‌ها کاربرد دارد. نوع غیر متصل مناســـب برای ردیف کردن خوانش‌ها در مقابل مجموعه داده‌های رونویسی است تا بیان ژن یا ایزوفرم‌های آن را کمی نماید. این برنامه‌ها اخیراً از ۲ روش کلاسیک به طور گسترده استفاده می‌کنند که شامل؛ الگوریتم‌های Hash و Burrows-Wheeler Transform (BWT) اســـت (٦). در جدول ۲، برنامه‌های طراحی نقشـــه‌ی داده RNA-Seq با طول خوانش‌های کوتاه نوشـــته شده اســـت. الگوریتم‌های Hash-based عبارتند از؛ Maq، ZOOM، RMAP، SeqMap و SOAP کـــه بر اســـاس کاربرد حافظه به ۲ دســـته متفاوت تقسیم‌بندی شده اســـت. یک نوع از این کاربردهای حافظـــه بر اندازه و طول ریدها و انواع دیگر بر اندازه ژنوم و طول دانه (Seed) وابسته هستند. الگوریتم‌های Burrows-Wheeler Transform (BWT) عبارتند از؛ Bowtie، SOAP۲ و BWA که ســـبب کاهش معنی‌دار در حافظه مورد دلخواه و افزایش معنی‌دار در ســـرعت نقشـــه یابی می‌شوند (۱۲). تمامی استراتژی‌های بر پایه‌ی BWT و Hash قادر به پردازش ریدهای کوتاه هستند ولی به خاطر روش‌های متفاوت در ردیف‌یابی خوانش‌های کوتاه دارای عملکرد متفاوت در کاربرد حافظه، زمان یا ســـرعت صرف شـــده، حمایت طول ریدها، تعداد خوانش‌های نقشه برداری شده و صحت ردیف‌یابی خوانش‌ها هستند (۳).

◈ محل اتصال اسپلایسینگ اگزون-اگزون

اسپلایســـینگ متناوب یک پدیده رایج در روند رونویســـی ژنی اســـت و اهمیت به ســـزایی در ســـاخت RNAهای ژنوم (کد کننـــده پروتئینی و غیر پروتئینـــی) دارد کـــه کارکرد طبیعی اندام‌ها را نیز تضمیـــن می‌کند (۸). در حال حاضر، تعداد مدل‌های موجود که به خوبی اتصال اسپلایسینگ اگزون-اگزون را تفسیر کنند بسیار کم بوده و از طرفی بخش عمده ژنوم گونه‌ها هنوز توالی‌یابی نشده است. تراپنل و همکاران (۲۰۱۰)، با استفاده از آنالیز داده‌های RNA-Seq



"شکل ۲" مراحل تکنیک SAGE . در این تکنیک mRNA استخراج شده به cDNA دو رشته‌ای تبدیل و با آغازگرهای متصل به استرپتاویدین تثبیت می‌شود. هضم با آنزیم محدودالاثر اول (AE) امکان اتصال به ۲ نوع لینکر از پیش طراحی شده به نام A و B را فراهم می‌آورد سپس هضم با آنزیم محدودالاثر دوم (TE) و مخلوط نمودن ۲ واکنش با همدیگر توسط آنزیم لیگاز انجام می‌گیرد، مولکول‌های حاصله در ۲ انتها دارای لینکر متفاوت (A و B) و در وسط دارای ۲ قطعه توالی یا شناسه (Tag) مربوط به ژن‌های رونویسی شده می‌باشند که آن‌ها را ۲شناسه یا DiTag می‌نامند، در مرحله بعد تکثیر توسط PCR، هضم آنزیمی و واکنش اتصال سوم منجر به تولید پلیمرهای شناسه (Concatamers) می‌شود که همسانه سازی و سپس توالی‌یابی می‌شود. در نهایت آنالیز توالی یا شناسه‌ها توسط نرم‌افزارهای مربوطه بیانگر میزان بیان آن‌ها در سلول مورد مطالعه است (۲۶).

حاصل از لاین سلولی مایوبلاست موش هزاران رونوشت تفسیر نشده یافتند. مطالعات گاتمن و همکاران (۲۰۱۰)، بیش از هزار RNAهای غیر کدکننده بین ژنی بزرگ را از توالی یابی ترانس‌کریپتوم سلول‌های بنیادی رویان موش شناسایی کردند. علاوه بر این، شناسایی محل اتصال اسپلایسینگ بین اگزون‌ها برای درک ایزوفرم‌های ژنی و کمی نمودن بیان ژن و ایزوفرم‌های آن بسیار حیاتی است (۸). جهت شناسایی محل اتصال اسپلایسینگ بین اگزون‌ها، در ابتدا بایستی یکسری نرم‌افزارها نقشه یابی قسمت‌های اتصال یافته‌ی خوانش‌ها را انجام داده و سپس خوانش‌های موجود در محل اتصالات اسپلایسینگ به قطعات کوچک‌تر جدا شده تا نقشه اگزون‌های متفاوت به وسیله crossing-checking با اینترون‌ها امکان‌پذیر شود. در جدول ۳، چندین نرم‌افزار که برای شناسایی محل اتصال اسپلایسینگ توسعه یافته‌اند نشان داده شده است. نرم افزار TopHat در ابتدا با استفاده از الگوریتم bowtie ریدهای RNA-Seq را به ژنوم‌ها ردیف و سپس محل اتصالات اسپلایسینگ بین اگزون‌ها را بر طبق نتایج نقشه یابی پیش‌بینی می‌کند. به خاطر اینکه اکثر اینترون‌ها دارای الگوی GT-AG هستند، نرم افزار TopHat فقط ردیف‌یابی خوانش‌های کوتاه‌تر از ۷۵ جفت باز را در بین تمام اینترون‌های دارای الگوهای GT-AG گزارش می‌کند این برنامه همچنین انترون‌های GC-AG و AT-AC با طول خوانش‌های بلندتر را جستجو خواهد نمود (۳). نرم افزارSpliceMap بدون توجه به تفسیر ساختارهای ژنی موجود، محل اتصالات اسپلایسینگ جدید را با دقت بالا شناسایی می‌کند. MapSplice نرم افزار کارآمد دیگری است که بدون وابستگی به خصوصیات محل اسپلایسینگ و طول انترون‌ها می‌تواند محل اتصالات اسپلایسینگ را با ویژگی و حساسیت بالایی شناسایی کند. SOAPsplice نرم افزار توسعه یافته‌ی قدرتمندی است که قادر به شناسایی محل اتصالات اسپلایسینگ جدید بدون استفاده از اطلاعات قبلی است. این برنامه احتمالاً در پیش‌بینی مجدد (دنوو) از محل اتصالات اسپلایسینگ و مطالعه‌ی اسپلایسینگ متناوب کاربرد دارد. از آنجایی که تمامی این استراتژی‌ها نیازمند تهیه‌ی نقشه اولیه خوانش‌های RNA-Seq با ژنوم مرجع است برای موجوداتی کاربرد دارند که توالی‌های ژنوم مرجع آن‌ها موجود باشد (۳۰).

کمیت‌پذیری بیان ژن و ایزوفرم آن

قبل از تکنولوژی‌های RNA-Seq، ریزآرایه‌ها تکنولوژی غالب برای بررسی پروفایل‌های بیان ژن به شمار می‌آمدند اگرچه آن‌ها در کمی نمودن بیان ژن فقط محدود به سطح ژن بودند. در مقابل، RNA-Seq می‌تواند بیان ژن را در هردو سطوح ژن و ایزوفرم‌های آن ارزیابی کند. بسیاری از ژن‌های چند اگزونی قادرند ایزوفرم‌های چندگانه با نقش‌های مختلف در خلال بیان ژن بسازند. مطالعه در سطوح ایزوفرم‌های ژن‌ها به منظور درک بهتر پیچیدگی‌های ترانس‌کریپتوم ضروری است (٤).

RNA-Seq قابلیت شناسایی ژن‌های تفسیر نشده و ایزوفرم‌های آن را برای هر گونه دارند در حالی که ریزآرایه‌ها تنها قادر به شناسایی ژن‌های شناخته شده بر اساس اطلاعات پیشین‌اند (۸ و ۲۵). در جدول ٤، نرم‌افزارهای آنالیز بیان ژن بر اساس داده RNA-Seq گزارش شده است. نرم‌افزار Cufflinks در ابتدا ردیف‌ها را به داخل یک مجموعه رونوشت‌های مقرون به صرفه سر هم نموده و سپس بر اساس تعداد خوانش‌ها در تهیه نقشه ترانس‌کریپتوم، فراوانی نسبی رونوشت‌ها را محاسبه می‌کند. این برنامه همچنین می‌تواند ژن‌های جدید

و ایزوفرم‌هـــای آن را بر طبق نتایج نقشـــه یابـــی خوانش‌ها در ژنوم مرجـــع پیش‌بینی کنـــد. نرم‌افـــزار Scripture می‌توانـــد از ابتدا ترانس‌کریپتوم را نوسازی کرده و کمی کردن بیان رونوشـــت را انجام دهد. نرم‌افزار MISO (ترکیب ایزوفرم‌ها) (۱۱) بر اساس یـــک چارچـــوب احتمالـــی از خوانش‌های تخصیـــص یافته به ایزوفـــرم جهت ارزیابی تعداد بی-شـــمار ایزوفرم‌ها استفاده می‌کند. نرم‌افـــزار ALEXA-Seq ابزاری جهت آنالیز بیان متناوب و قادر به کمی کردن بیان ایزوفرم-هاست. از این رومی‌توان با انتخاب نرم‌افزارهای مرتبط آنالیز مورد نیاز بر اساس اهـــداف تحقیق را انجـــام داد. صحت کمی نمـــودن بیان ژن یا ایزوفرم‌هـــای آن تا حد زیادی به وسیله نتایج نقشه یابی خوانش‌های RNA-Seq تخمین زده شده است. ژنوم مرجـــع معمـــولاً دارای توالی‌های همســـان و تکراری زیادی اســـت که در تقسیم‌بندی خوانش‌هـــا ابهامات زیادی ایجـــاد می‌کنند. علاوه بـــر این، تخصیص بنـــدی خوانش‌ها در موقعیت صحیح اتصالات اسپلایســـینگ بر روی ژنوم مرجع بســـیار دشوار است. با توجه بـــه این جنبه‌هـــا، دقیق‌ترین راه کمی نمودن بیان ژن یا ایزوفرم‌های آن، تهیه نقشه مســـتقیم خوانش‌های RNA-Seq از روی توالی-های محصول ترجمه (ترانسکریپتوم) است ولی بایستی به پیچیدگی ترانسکریپتوم اشـــاره نمود که ســـاختن پایگاه داده دقیق و کامـــل از آن حتی در گونه‌هایی که مطالعات خوبی روی آن‌ها انجام شده (انسان و موش) کاری بس دشوار است.

◈ آنالیز بیانی متفاوت

ژن‌هـــای یوکاریوتی تحت شـــرایط مختلـــف و تأمین نیاز موجـــودات، چندین ایزوفرم متمایز را نشان خواهند داد. با تعیین تغییـــرات بیان ژن‌ها یا ایزوفرم‌های آن بین ۲ نمونه یـــا گام متفاوت، انجام آنالیزهای بیانی متفاوت جهت شناسایی ژن‌ها یا ایزوفرم‌های بیان شده امکان پذیر خواهد شد. از داده‌های RNA-Seq می‌تـــوان جهـــت آنالیز بیانی متفاوت اســـتفاده نمود، زیرا نســـبت به ریز آرایه‌هـــا دارای فوایـــد بیشـــتر، ارزان‌تر و

جدول ۱- مقایسه تکنولوژی RNA-Seq با ریزآرایه (۲۲)

کاربرد	RNA-Seq	ریزآرایه
تکثیرپذیری با توان اجرای بالا	بله	بله
قابلیت مقایسه دامنه پویایی با فراوانیهای رونوشت داخل سلول	بله	خیر
توانایی تعیین مکانهای اسپلایسینگ متناوب و ایزوفرمهای جدید	بله	خیر
آنالیز *De novo* (مجدداً) نمونهها بدون ژنوم منبع	بله	خیر
قابلیت آنالیز مجدد داده	بله	خیر

جدول ۲- ابزارهای نقشه یابی ریدهای کوتاه

نام الگوریتم	سایت اینترنتی جهت دسترسی	استراتژی مورد استفاده
Bowtie	http://bowtie.cbcb.umd edu	BWT-based
BWA	/http://bio-bwa.sourcef orge.net	BWT-based
Soap۲	http://soap.genomics.org.cn/soapaligner.html	BWT-based
Maq	/http://maq.sourcef orge.net	Hash-based
RMAP	/http://rulai.csh۱.edu/rmap	Hash-based
SeqMap	/http://biogibbs Stanford.edu/Bjiangh/SeqMap	Hash-based
SHRiMP	/http://compbio.cs.toronto.edu/shrimp	Hash-based
SSAHA۲	/http://www.sanger.ac.uk/resources/software/ssaha۲	Hash-based
SOAP	/http://soap.genomics.org.cn/soap۱	Hash-based
ZOOM	/http://www.bioinfor.com	Hash-based

جدول ۳- لیست نرم افزاهای مورد نیاز جهت یافتن اتصالات اسپلایسینگ

نام نرمافزار	سایت اینترنتی جهت دسترسی به نرمافزار مربوطه
HMMSplicer	http://derisilab.ucsf.edu/index.php
MapSplice	http://www.netlab.uky.edu/p/bioinfo/MapSplice
SOAPsplice	http://soap.genomics.org.cn/soapsplice.html
SpliceMap	/http://www.stanford.edu/group/wonglab/SpliceMap
SplitSeek	/http://solidsoftwaretools.com/gf/project/splitseek
Supersplat	/http://supersplat.cgrb.oregonstate.edu
TopHat	/http://tophat.cbcb.umd.edu

بررســـی بیان ژن در سطوح ژنی و ایزوفرم‌های آن است در حالی که ریز آرایه‌ها ایزوفرم‌های ژن را بررسی نمی‌کنند. سطح بیان ژن یا رونوشت‌ها در RNA-Seq به تعداد خوانش‌های ســـازمان یافته ولی در ریز آرایه‌ها به انعکاس نور فلورسنتی پس از هیبریداسیون وابسته است. اگر بین شمارش خوانش‌های ژن یا رونوشت آن در ۲ وضعیت آزمایشی متفاوت تفاوت آماری معنی‌دار وجود داشته باشد، می‌توان آنالیز بیانی متفاوت را در داده‌های RNA-Seq تعیین کرد، اگرچه در این حالت داده‌های RNA-Seq اریب‌هایی از قبیل عمق توالی‌یابی، محاسبه پراکندگی در نمونه‌های مختلف و طول ژن یا رونوشـــت دارند. ضمناً محاســـبه پراکندگی در میان نمونه‌ها دارای تفاوت‌هایی اســـت. علاوه بر این، نســـبت ریدکانت‌ها برای

نام نرمافزار	سایت اینترنتی جهت دسترسی به نرمافزار مربوطه
ALEXA-Seq	http://www.alexaplatform.org/alexa_seq/index.htm
Cufflinks	/http://cufflinks.cbcb.umd.edu
IsoInfer	http://www.cs.ucr.edu/~jianxing/IsoInfer.html
MISO	/http://genes mit.edu/burgelab/miso
MMSEQ	http://bgx org.uk/software/mmseq.html
rSeq	/http://www.personal. Umich.edu/~jianghui/rseq
Scripture	/http://www.broadinstitue.org/software/scripture

جدول ۵ - ابزارهای در دسترس برای آنالیز بیانی متفاوت

نام نرمافزار	سایت اینترنتی جهت دسترسی به نرمافزار مربوطه
baySeq	http://www.bioconductor.org/packages/۲.۸/bioc/html/baySeq.html
Cuffdiff	/http://cufflinks.cbcb.umd.edu
DEGseq	/http://bioinfo.au.tsinghua.edu.cn/software/degseq
DESeq	/http://www-huber.embl.de/usrs/anders/DESeq
EdgeR	http://www.bioconductor.org/packages/release/bioc/html/edgeR.html
GPSeq	http://www-rcf.usc.edu/~liangche/software.html
Myrna	http://bowtie-bio.sourceforge.net/myrna/index.shtml
NOISeq	http://bioinfo.cipf.es/noiseq/doku.php?id=start
ASC	http://www.stat.brown.edu/Zwu/research.aspx
GENE-Counter	http://changlab.cgrb.oregonstate.edu/?q=node/view/۵۲۷

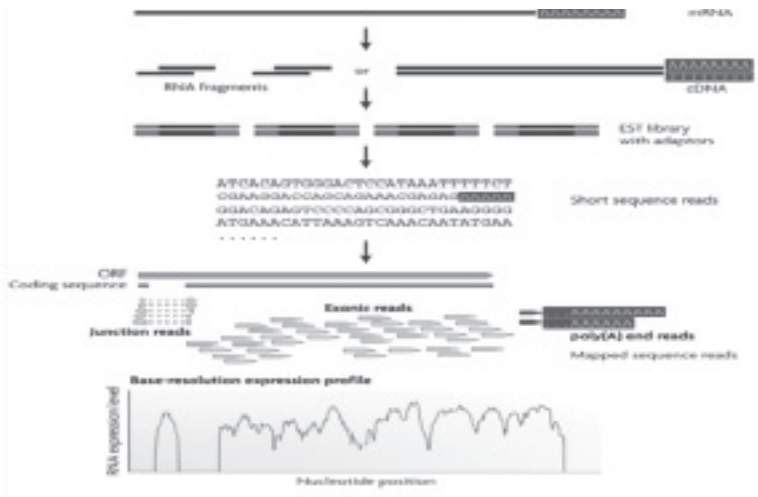
جدول ٦- ابزارهای لازم جهت نوسازی ترانسکریپتوم

نام نرمافزار	سایت اینترنتی جهت دسترسی به نرمافزار مربوطه	دستهبندی
Cufflinks	/http://cufflinks.cbcb.umd.edu	Genome-guide
Scripture	http://www.broadinstitute.org/software/scripture/?q=home	Genome-guide
Velvet	/http://www.ebi.ac.uk/~zerbino/velvet	Genome-independent
Trans-ABySS	http://www.bcgsc.ca/platform/bioinfo/software/trans-abyss	Genome-independent
Trinity	/http://trinityrnaseq.sourceforge.net	Genome-independent
Oases	http://www.ebi.ac.uk/~zerbino/oases	Genome-independent
Rnnotator	Need contact Virginia De la Puente at vtdelapuente@lbl.gov	Genome-independent

رونوشت‌های وابسته برابر با طول رونوشت ضرب در سطوح بیانی RNA مرتبط آن است. برای تعیین ژن‌ها یا ایزوفرم‌های بیان شده متمایز بایستی اریب‌های RNA-Seq به درستی مورد بررسی قرار گیرد (۳ و ۱۵). در جدول۵، یکسری استراتژی‌ها تحت نرم‌افزارهای مختلف موجود است که قادرند با استفاده از داده‌های RNA-Seq ژن‌ها یا ایزوفرم‌های بیان شده را شناسایی کنند. این روش‌ها به ۲ دسته مدل‌های پارامتریک و غیر پارامتریک تقسیم‌بندی می‌شوند. روش‌های پارامتریک بر اساس توزیع‌های احتمالی شناخته شده (دوجمله‌ای، پواسون و دوجمله‌ای منفی) و غیر پارامتریک فاقد فرضیات توزیع داده هستند. تارازونا و همکاران (۲۰۱۱)، مدل NOISeq (روش غیر پارامتریک قدرتمند) را برای برطرف کردن اریب داده‌های RNA-Seq از تغییرات عمق توالی‌یابی پیشنهاد کردند (۲٦)، نتایج آزمایشات آنها نشان داد که این مدل نسبت به روش‌های پارامتریک موجود DESeq ، baySeq و edger که برای توزیع دوجمله‌ای منفی استفاده می‌شوند، در مقابل تغییرات عمق توالی‌یابی انعطاف‌پذیرتر است در ضمن وی اثبات کرد که روش‌های پارامتریک موجود نسبت به NOISeq به عمق توالی‌یابی بالا وابسته هستند (۳).

نتیجه‌گیری

در تحقیق حاضر شماری از تکنیک‌های بررسی بیان ژن، مقایسه آن‌ها با همدیگر، مزایا و معایب آن‌ها و در ضمن، کاربردهای متنوع داده‌های تکنیک RNA-Seq که توسط نرم‌افزارهای مربوطه قابل انجام است شرح داده شد. انتخاب نرم‌افزار مناسب و پارامترهای بهینه برای آن‌ها از اهمیت بسزایی برخوردار بوده و نتایج را مستقیماً تحت تأثیر قرار می‌دهد. از این رو انتخاب بهترین روش آنالیز بر پایه طبیعت داده‌ها یکی از بخش‌های مهم ارزیابی حجم عظیم داده‌های حاصل از روش‌های پربازده (high throughput) در مطالعه بیان ژن است. در ضمن می‌توان با طراحی الگوریتم‌های مختلف توسط نرم‌افزارهای یاد شده از داده‌های بیان ژنی GEO موجود در NCBI یک سیستم طبقه بندی را برای بیماری‌هایی که سرو کارشان با بیان ژن و میزان پیشرفت بر اساس تخریب بافت‌های خاص است از قبیل سرطان خون، پارکینسون، MS و ... ایجاد کرد تا بتوان میزان پیشرفت بیماری به روش‌های بیوانفورماتیکی مشخص و تمهیدات لازم جهت درمان یا مهار پیشرفت بیماری انجام گیرد. تکنیک RNA-Seqبا آنالیز ترانسکریپتوم بافت و جمعیت سلولی، جهش‌های جدید و رونوشت‌های سرطانی را شناسایی و تومورها را بر اساس الگوهای بیان ژنی یا توالی پاتوژن‌های میکروبی جهت تشخیص بیماری طبقه‌بندی می‌کند. انتخاب روش مناسب جهت به دست آوردن داده‌های



"شکل ۳" مراحل تکنیک RNA-Seq به طور نمونه. ابتدا RNAهای بلند به کتابخانه cDNA یا قطعات RNA یا DNA تبدیل می‌شوند. آداپتورهای توالی‌یابی (به رنگ آبی متصل به کتابخانه EST) متعاقباً به قطعات و توالی‌های کوتاه cDNA که با استفاده از تکنولوژی توالی‌یابی با توان بالا (NGS) به دست آمده‌اند، اضافه می‌شوند. ریدهای توالی‌یابی شده حاصله پس از ردیف‌یابی سپس به همراه ژنوم مرجع یا ترانس‌کریپتوم در ۳ دسته (exonic reads، junction reads و poly (A) end-reads) رده‌بندی می‌شوند. این ۳ دسته منظور یک پروفایل بیانی بر اساس تفکیک پذیری را می‌سازند (۶۷).

بیان ژن به سهولت و امکانات موجود وابسته است، ولی تکنیک‌های پیشرفته و پربازده باعث تسهیلات بیشتری در آنالیز داده‌ها می‌شوند.

منابع

[1]. Aldaz CM. Serial analysis of gene expression (SAGE) in cancer research. In: Ladanyi M, Gerald WL. Editors. Expression profiling of human tumors, diagnostic and research application. New Jersey: Human press 2003; p.60–47.

[2].Chepelev I, Wei G, Tang Q, et al. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. Nucleic Acids Researches 37 ;2009: e106.

[3]. Chen G, Wang C, Shi TL. Overview of available methods for diverse RNA-Seq data analyses. Science China Life Science 1128–1121 :54 ;2011.

[4]. Chen G, Yin K, Shi L, et al. Comparative analysis of human protein- coding and noncoding RNAs between brain and 10 mixed cell lines by RNA-Seq. PLoS ONE 6 ;2011: e28318.

برای دیدن ادامه منابع به وب سایت ماهنامه مراجعه کنید.

Neek Azma Co.(Ltd)
شرکت مهندسی رهاورد نیک آزما
کنترل کیفی، آزمون و سنجش کالیبراسیون
تجهیزات پزشکی، بیمارستانی، آزمایشگاهی و رادیولوژی
نیک آزما
Neek Azma
✓ مجوز رسمی فعالیت کنترل کیفی از اداره کل تجهیزات پزشکی وزارت بهداشت ایران
✓ پروانه اشتغال کنترل کیفی تجهیزات پرتونگاری از سازمان انرژی اتمی ایران
تهران، خ جمهوری، خ کمالی، نبش جوانشیر، پلاک ۳
تلفن: ۸-۶۶۸۴۱۰۲۷ فکس: ۶۶۸۴۱۰۲۹