

معرفی فن آوری نوین توالی‌یابی آر.ان.آ

وحید شریعتی، فریناز به فرجام*، امیر موسوی

استادیار پژوهشگاه ملی و مهندسی ژنتیک و زیست فناوری، تهران، ایران
*دانشجوی دکتری پژوهشگاه ملی و مهندسی ژنتیک و زیست فناوری، تهران، ایران
دانشیار پژوهشگاه ملی و مهندسی ژنتیک و زیست فناوری، تهران، ایران

farinaz.cell88@gmail.com

چکیده

امروزه فناوری‌های توالی‌یابی با نتایج دقیق و راندمان بالا در علوم زیست‌شناسی بسیار مورد استفاده قرار می‌گیرند. یکی از روش‌های نسل بعدی توالی‌یابی، توالی‌یابی آر.ان.آ است که برای بررسی ترانسکریپتوم از آن استفاده می‌شود. این روش با به کارگیری فناوری‌های توالی‌یابی عمیق دی.ان.آ مکمل، سنجش فراوانی آر.ان.آ در سطح ژنومی را برای محققین میسر می‌سازد و دیدگاه ما را در ارتباط با پیچیدگی ژنوم یوکاریوت‌ها تغییر می‌دهد. پیشرفت‌های اخیر در روش‌های توالی‌یابی آر.ان.آ، خصوصیات کامل‌تری را از رونوشت‌های آر.ان.آ در مقایسه با سایر روش‌ها نشان می‌دهد. این پیشرفت‌ها شامل بهبود نقشه‌برداری جایگاه‌های شروع رونویسی، سنجش بیان اختصاصی یک رشته، بررسی اتصالات ژنی و تشخیص پیرایش‌های فرعی و ترکیب ایزوفرم‌ها می‌باشند. پیشرفت‌های مداوم در این فناوری، امکان توالی‌یابی مستقیم آر.ان.آ و دست‌یابی به توالی‌های مربوط به مولکول‌های آر.ان.آ با فراوانی پایین را نیز فراهم نموده است. در این مقاله مروری، به معرفی روش توالی‌یابی آر.ان.آ و کاربردهای آن پرداخته شده و به برخی پیشرفت‌های آن که به بسیاری از محدودیت‌های سایر روش‌ها غلبه می‌کند، اشاره خواهد شد.

کلمات کلیدی: توالی‌یابی آر.ان.آ، توالی‌یابی، ترانسکریپتوم، دی.ان.آ مکمل، رونوشت و نقشه‌برداری

مقدمه

سلول‌ها و بافت‌ها، همچنین برای درک مراحل پیشرفت بیماری ضروری می‌باشد (۳۹).

اصطلاح توالی‌یابی دی.ان.آ، به روش‌های تعیین توالی برای مشخص کردن ترتیب بازهای نوکلئوتیدی در یک مولکول دی.ان.آ اشاره می‌کند. ورود این روش به عرصه زیست‌شناسی، باعث تسریع کشف و

ترانس‌کریپتوم، سری کامل رونوشت‌های یک سلول است که میزان این رونوشت‌ها برای مراحل خاص رشد و نمو و شرایط فیزیولوژیکی مهم می‌باشد. به گونه‌ای که درک مفهوم آن برای تفسیر عناصر عملکردی ژنوم و مشخص کردن ساختار مولکولی

بررسی‌های بیولوژیکی شده است. به‌طوری که سرعت بالای توالی‌یابی دی.ان.آ با روش‌های جدید، ابزاری برای تعیین توالی کل ژنوم بسیاری از حیوانات، گیاهان، ژنوم‌های میکروبی و ژنوم انسانی در پروژه ژنوم انسان شده است (۹). پانزده سال بعد از کشف مولکول مارپیچ دی.ان.آ در سال ۱۹۵۳، اولین آزمایش‌های تعیین توالی آن انجام شد. اولین گزارش‌ها در سال ۱۹۷۵ توسط سنگر و کولسن با معرفی روش جمع و تفریق (Plus-minus) بر روی ژل اکریل امید معرفی شد که این روش ۳۰ سال بعد نیز مورد استفاده قرار گرفت. کشف تری‌فسفات‌های دی‌داکسی نوکلئوتید در سال‌های بعد، عاملی برای جایگزینی این روش با روش توالی‌یابی سنگر در سال ۱۹۷۷ شد (۹ و ۲۹).

پیشرفت در روش‌های ترانسکریپتوم ادامه پیدا کرد تا این که انواع مبتنی بر دورگه‌سازی (Hybridization-based approaches) بدست آمدند. براون و همکاران در سال ۱۹۹۵، سطح بیان ژن‌های شناخته شده را با استفاده از ریزآرایه دی.ان.آ مکمل لکه‌گذاری شده (Spotted cDNA microarray) مشخص نمودند. یکی دیگر از این روش‌ها، tiling array ارائه شده توسط شرکت Affymetrix بود که در سال ۲۰۰۲ با بررسی الگوی بیان کل ژنوم به حضور ژن‌های جدید و واریانت‌های پیرایش پی بردند. از ابتدای سال ۲۰۰۵ چندین متد جدید در آزمایشگاه جرج کرچ و 454 Life Sciences معرفی شد که تحولی در عرصه توالی‌یابی بود. همه این‌ها در دوره‌ای بعد از سنگر قرار دارند و به عنوان فن‌آوری‌های نسل بعدی توالی‌یابی (NGS) نامیده می‌شوند (۲۹ و ۳۹).

فن‌آوری‌های متنوعی برای تفسیر و کمیت سنجی ترانسکریپتوم وجود دارد. اقدامات بر اساس دورگه‌سازی که با انکوباسیون دی.ان.آ مکمل نشاندار شده با فلورسانس با استفاده از ریزآرایه‌ها انجام می‌شود، دارای یکسری محدودیت‌هایی هستند. اول این که می‌بایستی از قبل اطلاعاتی در مورد ژنوم موجود داشته باشیم. از طرفی در این روش‌ها سطح زمینه به دلیل دورگه‌سازی‌های متقاطع بالا می‌باشد و همین عامل محدود شدن رنج پویای یافته‌ها می‌شود. برخلاف روش‌های ریزآرایه، اقدامات بر اساس توالی، به‌طور مستقیم توالی دی.ان.آ مکمل را مشخص می‌کنند. در ابتدا توالی‌یابی سنگر دی.ان.آ مکمل و یا کتابخانه توالی‌های کوتاه بیان یافته استفاده شد که نسبتاً کارایی پایینی داشتند و گران بوده و به‌طور کلی کمی نبودند. روش‌های بر اساس برچسب (Tag-based approaches) مانند:

(SAGE: Serial analysis of gene expression) و (CAGE: Cap analysis of gene expression) (MPSS: Massively parallel signature sequencing) برای غلبه بر این محدودیت‌ها پیشرفت کردند. اما اساس این روش‌ها نیز فن‌آوری گران قیمت توالی‌یابی سنگر می‌باشد. با این قطعات کوتاه، برچسب‌ها نمی‌توانند به صورت اختصاصی، ژنوم مرجع را نقشه برداری کنند. علاوه بر این، تنها بخشی از رونوشت آنالیز می‌شود و ایزوفرم‌ها از همدیگر تمیز نمی‌شوند (۳۴ و ۳۹).

اخیراً و در سال ۲۰۰۸ برای غلبه بر این محدودیت‌ها، روشی دیگر از نسل بعدی توالی‌یابی برای تفسیر ترانسکریپتوم حاصل شد. این روش که به عنوان نیرویی قوی برای تعیین ویژگی ترانسکریپتوم و

"شریعتی و همکاران، معرفی فن آوری نوین توالی‌یابی آر.ان.آ"

آشکارسازی پیچیدگی آن و سنجش دقیق سطح رونوشت‌ها و ایزوفرم‌های آن در مقایسه با سایر روش‌ها به کار می‌رود، توالی‌یابی دی.ان.آ مکمل موازی (Massively parallel cDNA sequencing) و یا توالی‌یابی آر.ان.آ نامیده می‌شود (۲۱ و ۳۹).

فن آوری توالی‌یابی آر.ان.آ و مزایای آن

فن آوری توالی‌یابی آر.ان.آ از روش‌های توالی‌یابی مهم و قابل‌توجهی که امروزه با سیستم‌های Illumina، Roche 454 Life Science و Applied Biosystems پیشرفت کرده است استفاده می‌کند. قبل از بیان مراحل کار فن آوری توالی‌یابی آر.ان.آ، اشاره کوتاهی به اساس کار این سیستم‌ها داریم:

توالی‌یابی Roche 454

یکی از روش‌های توالی‌یابی، از نسل بعدی توالی‌یابی که به صورت تجاری در دسترس است، در سال ۲۰۰۵ توسط مارگولیز و همکاران و شرکت 454 Life Sciences توسعه پیدا کرد که بر اساس تکنیک pyrosequencing بود. این سیستم، توالی‌یابی کل ژنوم خرد شده بدون نیاز به کلون کردن در باکتری /شرشیاکلی و یا هر میزبان دیگر را امکان‌پذیر می‌کند (۳۰).

در این روش ابتدا دی.ان.آ به صورت تصادفی به قطعات کوتاهی تبدیل می‌شود و به توالی‌های پیوند دهنده (Adapters) متصل می‌شود. در مرحله بعدی به مولکول‌ها اجازه اتصال به قطعات موجود بر روی گلوله‌های ریز (Beads) را می‌دهند. در روش توالی‌یابی 454، تکثیر قطعات حاصله با واکنش زنجیره‌ای پلی‌مراز امولسیون (Emulsion PCR) صورت می‌گیرد که در آن قطعات در یک محلول

دارای روغن حل می‌شوند و هر کدام از آن‌ها حکم یک واکنش را دارند. بیدهای موجود در ابعاد ماکرو و یا نانو هستند و به روش نانوفن آوری ساخته شده، دارای توالی‌هایی مکمل قطعات ما هستند. با گرم و سرد کردن‌های متوالی، واکنش تکثیر درون قطره‌های مربوطه انجام می‌گیرد. سپس مجموعه‌های بسیار وسیعی از این گلوله‌های ریز را بر روی PicoTiterPlate قرار می‌دهند به طوری که هر گلوله در داخل یک حفره قرار می‌گیرد. در مرحله بعدی گلوله‌های ریز آنزیمی اضافه می‌شوند. با انجام واکنش تکثیر و افزوده شدن هر نوکلئوتید، تصویر برداری با دوربین‌های خاصی صورت می‌گیرد. آنالیز سری تصاویر حاصله توالی قطعات را مشخص می‌کند (۹ و ۲۷).

توالی‌یابی Illumina (Solexa)

یکی دیگر از فن‌آوری‌های نسل بعدی توالی‌یابی، در سال ۲۰۰۶ توالی‌یابی Illumina بود. تفاوت این روش با 454 در اساس کار تکثیر قطعات می‌باشد. در این فن‌آوری، واکنش زنجیره‌ای پلی‌مراز به صورت پل (Bridge PCR) انجام می‌گیرد. تکثیر بر روی صفحات جامدی صورت می‌گیرد که به صورت درجا و تصادفی، مکمل‌هایی برای آداپتور قطعات چسبیده شده است. قطعات پس از اتصال به توالی‌های مکمل خودشان بر روی صفحه می‌توانند با حرکت از جانب دیگر نیز به توالی دیگری بر روی صفحه متصل شوند و یک پل تشکیل دهند. گرم و سرد کردن‌های مکرر، باعث تکثیر قطعات می‌شود. در مرحله بعدی با دنا توره کردن دو رشته‌ای‌ها، تکثیر قطعات با افزودن نوکلئوتیدهای دارای نشان فلورسانس که انتهای ۳'

کوتاهی در حد ۳۰-۲۵ نوکلئوتید، می تواند ۳-۲ گیگا باز توالی را در هر دوره توالی یابی کند (۹ و ۲۷).

توالی یابی Helicos

روش Helicos پلی فرم دیگری از نسل بعدی توالی یابی است که دارای اساسی مثل سه پلت فرم اصلی فوق می باشد، اما متکی به تکثیر کلونال نمی باشد. این روش، نسبتاً گران قیمت است و از این رو در مقایسه با سه روش فوق کمتر مورد استفاده قرار می گیرد (۳۵).

از زمان معرفی سیستم Helicos، تلاش های بسیاری برای ایجاد فن آوری های توالی یابی کارآمد متکی بر تک مولکول ها صورت می گیرد که دیگر نیازی به مرحله تکثیر نباشد. مراحل کلی کار در روش Helicos به این صورت است که پس از شکافتن یک مولکول دی.ان.آ تکی به قطعاتی کوتاه که همراه دنباله های پلی A هستند، این قطعات به سطح فلوسل هایی متصل می شوند. اتصال قطعات به سطح فلوسل ها از طریق پلی T های متصل شده به سطح فلوسل ها انجام می گیرد. زمانی که پلی مرز و باز نشاندار بر روی سطح جاری می شوند، شرکت نوکلئوتید در ساختار در حال تشکیل در مکان هایی که حالت مکمل بین نوکلئوتید اضافه شده و اولین موقعیت در قطعه متصل به سطح وجود دارد، انجام می گیرد. در مرحله بعدی بازهایی که وارد ساختار نشده اند، شستشو می شوند. سپس یک دوربین کل سطح قطعه را برای شناسایی بازهای جدیدی که اضافه شده اند، اسکن می کند. در مرحله بعدی، رنگ نشاندار که مهار کننده اتصال می باشد، جدا شده و چرخه دیگری از اضافه شدن نوکلئوتیدی جدید به

آن ها بلوکه شده است انجام می گیرد و در هر مرحله اضافه شدن آن ها، بلوک برداشته شده و نور ساطع شده تصویربرداری می شود و با سنجش شدت نورهای حاصله، ردیف توالی ها مشخص می شود (۹ و ۲۷).

توالی یابی ABI's SOLiD

اساس این روش شیمی اتصال-دو رگه سازی می باشد که در سال ۲۰۰۸ معرفی شد. در این روش اتصال پرایمر به نشانگرها و ساطع شدن نور حاصله توالی مربوطه را مشخص می کند. این واکنش دارای مراحل تکثیر نمی باشد. واکنش با استفاده از یک آنزیم لیگاز و اتصال قطعات و در نهایت ساطع شدن نور حاصله انجام می گیرد. قطعات دی.ان.آ برای این روش بر روی سطوح گلوله های ریز مغناطیسی با اندازه یک میکرونی تکثیر می شوند و امکان ایجاد سیگنال کافی در طول مراحل توالی یابی را فراهم می کند و سپس بر روی اسلایدهای فلوسل قرار می گیرند. توالی یابی بر اساس لیگاز، با اتصال یک پرایمر به توالی های مشترک بر روی قطعات تکثیر شده شروع می شود. سپس دی.ان.آ لیگاز، یک ۸ تایی نشاندار شده با فلورسانس خاصی را فراهم می کند که در آن بازهای چهارم و پنجم با گروه فلورسانس متصل شده، کد می شوند. هر مرحله اتصالی همراه با آشکارسازی فلورسانس همراه است. بعد از آن در یک مرحله بازسازی، بازها از ۸ تایی متصل که دارای گروه فلورسانس است، جدا می شوند و به صورت همزمان پرایمر برای دور بعدی واکنش اتصالی آماده می شود. در نهایت با کنار هم گذاشتن تصاویر حاصله، می توان به ترتیب توالی ها پی برد. این روش با دارا بودن توالی

"شریعتی و همکاران، معرفی فن آوری نوین توالی‌یابی آر.ان.آ"

همراه پلی مرآز آغاز می‌شود و با تکرار این چرخه‌ها، کل توالی مشخص می‌شود (۱۴).

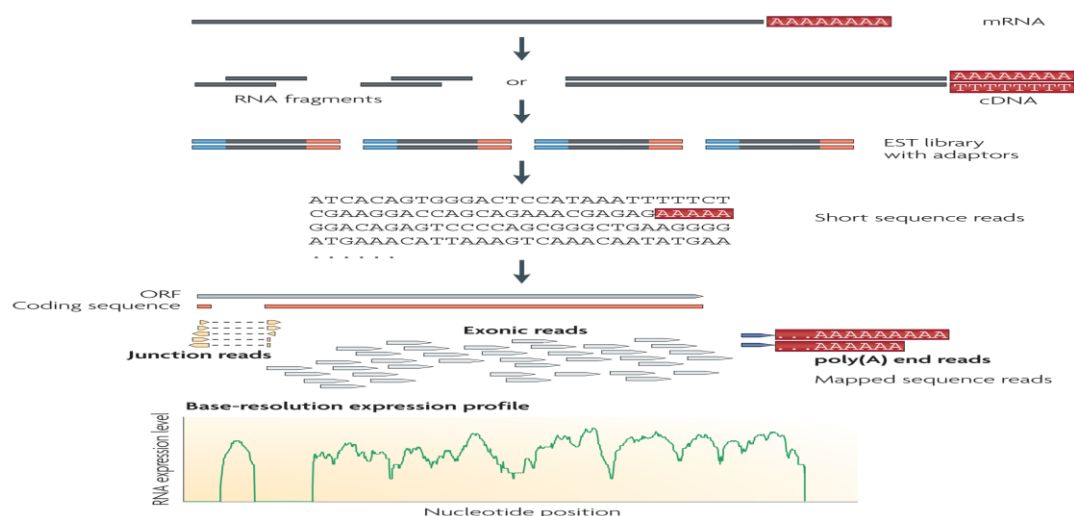
مراحل کار فن آوری توالی‌یابی آر.ان.آ

به طور کلی مراحل کار فن آوری توالی‌یابی آر.ان.آ به این صورت است که در مرحله اول آر.ان.آ استخراج شده با دو روش قطعه قطعه شدن آر.ان.آ و یا قطعه قطعه شدن دی.ان.آ مکمل، به کتابخانه‌ای از دی.ان.آ مکمل تبدیل می‌شود (شکل ۱). در روش اول، مولکول‌های آر.ان.آ ابتدا با ریزسازی و یا هیدرولیز به قطعات کوتاهی تبدیل می‌شوند و بعد به صورت دی.ان.آ مکمل در می‌آیند. اما در روش دوم ابتدا مولکول‌های آر.ان.آ به دی.ان.آ مکمل تبدیل می‌شوند و در مرحله بعدی با تیمار آنزیم تجزیه کننده دی.ان.آ و یا ریزسازی با صوت، به کتابخانه‌ای از دی.ان.آ مکمل تبدیل می‌شوند.

کتابخانه توالی‌های کوتاه بیان یافته با اتصال آداپتور به یک یا هر دو انتهای این دی.ان.آ مکمل‌ها بدست می‌آید. مرحله بعدی تعیین توالی قطعات حاصله توسط یکی از پلت‌فرم‌های چهار شرکت Illumina، Roche 454 Helicos BioSciences و Life Technology می‌باشد که هر مولکول با و یا بدون تکثیر با روشی با راندمان بالا توالی‌یابی می‌شود و قطعاتی کوتاه از یک انتها (توالی‌یابی یک طرفه: Single-end sequencing) و یا از هر دو انتها (توالی‌یابی دو طرفه: Paired-end sequencing) حاصل می‌شود. این قطعات بسته به روش توالی‌یابی که استفاده می‌شود، ۴۰۰-۳۰ جفت باز هستند.

مرحله آخر، هم‌ترازی، قطعات کوتاه توالی‌یابی شده بر روی ژنوم مرجع و یا رونوشت‌های آن می‌باشد و یا این که قطعات، بدون توالی ژنومی مرجع به صورت از نو (Denovo) کنار همدیگر قرار می‌گیرند. لازم به ذکر است که این توالی‌های کوتاه به سه دسته توالی‌های اگزونی (Exonic reads)، توالی‌های اتصال (Junction reads) و یا توالی‌های انتهای پلی‌آدنینی (Poly(A) end-reads) تقسیم می‌شوند. این سه توالی، برای ایجاد یک پروفایل بیانی و یا یک نقشه رونویسی در مقیاس ژنومی به کار می‌روند که شامل ساختار رونویسی و یا سطح بیان هر ژن می‌باشد (۳۹).

اگر چه این فن آوری هنوز در مسیر پیشرفت قرار دارد، اما در مقایسه با سایر روش‌ها، دارای مزایای زیادی می‌باشد. اول این که این فن آوری محدود به شناسایی رونوشت‌های توالی‌های ژنومی موجود نمی‌باشد و همین ویژگی آن برای ارگانیزم‌هایی که به صورت غیرمدل هستند و هنوز توالی ژنومی آن‌ها شناسایی نشده است، بسیار اهمیت دارد. ثانیاً با این روش می‌توان مرزهای رونویسی و جایگاه‌های شروع رونویسی (Transcription start sites (TSS)) را با دقت یک نوکلئوتید شناسایی کرد. علاوه بر این توالی‌یابی آر.ان.آ، در شناسایی تنوع توالی‌ها مثل پلی‌مورفیسم در مناطق رونویسی شده نقش تعیین کننده دارد. با این روش شناسایی حوادث پیرایش متناوب و رونوشت‌های ترکیب ژنی (Gene fusion) که در بیماری‌هایی مثل سرطان بسیار مهم می‌باشند، بهبود یافته است (۲۰ و ۳۹).



شکل ۱- مراحل یک آزمایش معمول توالی یابی آر.ان.آ تبدیل آر.ان.آ به دی.ان.آ مکمل و افزودن آداپتور، ایجاد توالی کوتاه بیان یافته با سیستم های توالی یابی، هم ترازی آن ها روی ژنوم یا رونوشت و در نهایت ایجاد یک پروفایل بیانی برای بررسی کمی بیان ژن

چندین پیشرفت در فن آوری توالی یابی آر.ان.آ

۱- ترسیم جایگاه های شروع رونویسی

مشخص کردن جایگاه های شروع رونویسی با دقت یک نوکلئوتید برای مشخص کردن کامل محصولات آر.ان.آ و شناسایی مناطق پروموتوری تنظیم کننده بیان هر رونوشت ضروری می باشد. یکی از اقدامات موجود برای این منظور که از روش سنگر توسعه پیدا کرده روش CAGE است (۳۲). این روش دی.ان.آ مکمل های کلون شده حاصل از آر.ان.آ را توالی یابی می کند که دارای انتهای ۵' دست نخورده ای هستند (برای مثال دارای کلاهک در انتهای ۵' خود می باشند). این روش گرچه مهم و سودمند می باشد، ولی به میزان زیادی آر.ان.آ نیاز دارد و توالی های کوتاهی در حد تقریباً ۲۰ نوکلئوتید ایجاد می کند.

برطرف کردن این محدودیت ها با سازگاری این روش با توالی یابی آر.ان.آ اتفاق افتاد و منجر به شناسایی پیچیدگی های غیرقابل پیش بینی توزیع جایگاه های شروع رونویسی در طول ژنوم و حوالی پروموتور شد.

از جمله فن آوری های ترکیب این دو روش می توان deepCAGE (۳۷)، PEAT: Paired end analysis of transcription start sites (آنالیز از دو انتهای جایگاه های شروع رونویسی) (۱۷)، nanoCAGE (CAGE کم مقدار) و CAGEscan (توالی جفت برای ترکیب انتهای ۵' CAGE با توالی های پایین دست) (۲۵) را نام برد.

ترسیم جایگاه شروع رونویسی با میزان کمتر از ۱۰ نانوگرم با NonoCAGE امکان پذیر می شود که این عمل را با استفاده از استراتژی های تکنیری متنوعی انجام می دهد. دو روش PEAT و nanoCAGE با توالی یابی از هر دو انتها، پیوستگی جایگاه های شروع رونویسی با مناطق پایین دست را برآورد می کنند. علاوه بر این، در تسهیل ارتباط جایگاه های شروع رونویسی با رونوشت های خاص نیز نقش دارند (۲۱).

۲- توالی یابی آر.ان.آ یک رشته خاص (Strand specific RNA seq)

مطالعات ترانسکرپتوم در گونه های مختلف حضور

"شریعتی و همکاران، معرفی فن آوری نوین توالی یابی آر.ان.آ"

فراگیری از حوادث رونویسی رشته بی معنی (Antisense) را نشان می دهد. امروزه مشخص شده است که رونوشت های بی معنی، دارای عملکرد هستند و نقش های متنوعی در جایگاه های فیزیولوژیکی و بیماری دارند (۵).

در طول سنتز دی.ان.آ مکمل رشته اول، به خاطر عملکرد دی.ان.آ پلی مرز وابسته به دی.ان.آ، رشته دوم نیز به صورت محصول مصنوعی تولید می شود. برای سنتز رشته بی معنی و توالی یابی آن، استراتژی هایی توسعه پیدا کرده است که اطلاعات اختصاصی یک رشته را ایجاد می کنند. اولی با اتصال آداپتورهایی به انتهای آر.ان.آها و یا مولکول های دی.ان.آ مکمل تک رشته ای در جهت از قبل تعیین شده می باشد. جهت آداپتورهای استفاده شده به عنوان نقاط مرجع برای به دست آوردن اطلاعات رشته مربوط مورد استفاده قرار می گیرد. دومی توالی یابی مستقیم محصولات دی.ان.آ مکمل تک رشته ای ایجاد شده است. نهایتاً سومی نشاندار کردن شیمیایی محصولات سنتز شده دی.ان.آ مکمل رشته دوم و یا آر.ان.آ به صورت انتخابی می باشد (۸ و ۲۴). این استراتژی ها در بررسی های ترانسکریپتوم شامل ترسیم جایگاه های ترجمه آر.ان.آها (برای مثال محل اتصال پلی زوم ها) و شناسایی آر.ان.آهای مرتبط با پروموتورهای خاص بسیار مفید می باشند (۲۱).

۳- مشخص کردن الگوهای پیرایش فرعی

اهمیت الگوهای پیرایش فرعی در رشد و نمو مشخص می باشد. از طرفی ۶۰-۱۵ درصد جهش های عامل بیماری، به دلیل پیرایش فرعی است. از این رو آشنایی با مجموعه کامل حوادث پیرایش و درک این

که چگونه این الگوهای متغیر در رشد و نمو، تمایز سلولی و بیماری انسانی نقش دارند، بسیار حیاتی است (۱۸ و ۳۸). پیرایش در انسان ها، به دلیل وجود بیش از ۹۵ درصد ژن های چند اگزونی که پیرایش فرعی در آن ها اتفاق می افتد و وجود تقریبی ۱۱۰۰۰۰ جایگاه پیرایش جدید در هر بافت، حائز اهمیت است (۲).

پیشرفت در فن آوری های توالی یابی کنونی، تعیین توالی قطعات بلندتر، ترسیم و مکان یابی توالی های کوتاه بر روی اگزون های با پیرایش فرعی را امکان پذیر می سازد. در این فن آوری ها، توالی ها به چندین قطعه تقسیم می شوند و هر قطعه به صورت مستقل بر روی ژنوم هم تراز می شود. علاوه بر این روش هایی که با استفاده از توالی های دو طرفه انجام می گیرند، ما را قادر می سازند که بتوانیم اطلاعات حاصل از دو نقطه در یک رونوشت را با حدس فاصله بین توالی ها تعیین توالی کنیم. در نتیجه امکان جستجوی الگوهای پیرایش بدون نیاز به دانش قبلی در مورد شرح رونوشت ایجاد می شود (۱ و ۳۶، ۲۱).

۴- روش های هدفدار با استفاده از توالی یابی آر.ان.آ

علی رغم توانایی های در حال پیشرفت نسل بعدی، توالی یابی در ارتباط با ویژگی های عملکردی و کاهش هزینه ها برای هر داده حاصله، دست یابی به توالی یابی قطعات برای کاربردهای بالینی دارای هزینه گزافی است. از جمله این کاربردها می توان مشخص کردن رونوشت های دارای فراوانی پایین (برای مثال آلل های رونوشت هایی که دارای بیان افتراقی هستند) را نام برد. در این موارد می بایستی زیر گروه رونوشت مورد نظر غنی سازی شود تا با رسیدن به فراوانی بالایی از

آن، بتوان توالی‌یابی را انجام داد (۱۵).

استراتژی‌های غنی‌سازی هدف (Target-enrichment strategies) برای توالی‌یابی مجدد دی.ان.آ ژنومی پیشرفت کرده‌اند. بسیاری از این فن‌آوری‌ها برای به دام‌اندازی آگزون‌های ژنوم مورد استفاده قرار می‌گیرند که بسیاری از جهش‌های عامل بیماری با احتمال بالا در ترانسکریپتوم کد کننده پروتئین رخ می‌دهند. تغییرات در استراتژی‌های غنی‌سازی دی.ان.آ ژنومی، منجر به توسعه روش توالی‌یابی آر.ان.آ هدفدار برای توالی خاصی شده است. فن‌آوری توالی‌یابی آر.ان.آ هدفدار برای شناسایی رونوشت‌های ترکیبی (۱۱)، بیان اختصاصی آلل، جهش‌ها (۴۰) و حوادث ویرایش آر.ان.آ (RNA-editing) (۱۲) در یک زیر گروه رونوشت‌ها کاربرد دارد (۲۱).

۵- توالی‌یابی آر.ان.آهای کوچک

تأثیر فن‌آوری توالی‌یابی آر.ان.آ در ارتباط با کشف آر.ان.آهای کوچک و مشخص کردن خصوصیات آن‌ها بسیار قابل توجه است (۶). در بسیاری از مطالعات، برای کشف این آر.ان.آها از روش pyrosequencing استفاده شد (۲۶ و ۲۸). اما امروزه با استفاده از سایر روش‌های با کارایی بالای نسل بعدی توالی‌یابی و با بررسی کل ژنوم، گونه‌های زیادی از آر.ان.آهای کوچک شناسایی شده است (۳۳).

استراتژی‌های آماده‌سازی نمونه برای آر.ان.آهای بلندتر (بیش از ۲۰۰ نوکلئوتید)، برای آر.ان.آهای کوچک مناسب نمی‌باشد. در این روش‌ها با رونویسی معکوس، دی.ان.آ مکمل حاصل می‌شود و توالی‌یابی می‌شود. در مورد آر.ان.آهای کوچک، دی.ان.آهای مکمل حاصله با این روش‌ها بسیار کوتاه بوده و

کارایی لازم برای توالی‌یابی و هم‌ترازی را ندارند. به همین دلیل می‌بایستی استراتژی‌های آماده‌سازی نمونه دچار تغییر شوند (۲۱). از این‌رو، یکی از محدودیت‌های روش‌های بر مبنای توالی‌یابی آر.ان.آ که امروزه مورد استفاده قرار می‌گیرد، عدم توانایی آن در بیان دقیق مقدار کمی و فراوانی این آر.ان.آها می‌باشد که با احتمال بالا به دلیل خطاهای انجام شده در طول مراحل آماده‌سازی است (۱۰).

۶- توالی‌یابی مستقیم آر.ان.آ

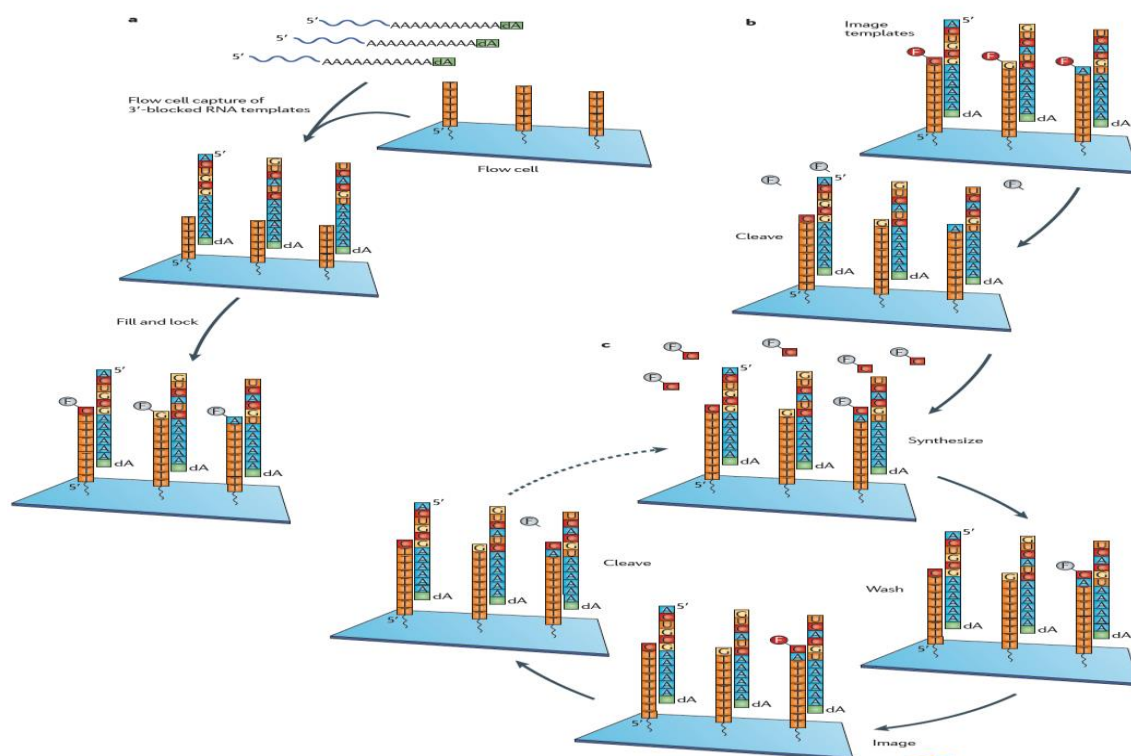
سنتز دی.ان.آ مکمل و سایر مراحل که برای آماده‌سازی نمونه‌های آر.ان.آ به منظور توالی‌یابی انجام می‌گیرد، برخی کاربردهای توالی‌یابی آر.ان.آ را محدود می‌کند. برای مثال در مورد توالی‌یابی آر.ان.آ که برای یک رشته خاص صورت می‌گیرد، همان‌طور که گفته شد، دی.ان.آ مکمل رشته دوم به صورت محصول مصنوعی نادرست ایجاد می‌شود. کتابخانه‌های اختصاصی یک رشته، برای جلوگیری از این مشکل ساخته می‌شوند، اما به دلیل استفاده از اتصالات آر.ان.آ - آر.ان.آ، کار پر زحمتی می‌باشند. محدودیت دیگر در ارتباط با تغییر الگو (Template switching) است که در زمان سنتز دی.ان.آ مکمل ممکن است دی.ان.آ مکمل در حال سنتز و ناقص از رشته آر.ان.آ الگوی خودش جدا شود و به یک رشته الگوی دیگر متصل شود که دارای توالی مشابه با الگوی اصلی می‌باشد و دی.ان.آهای مکمل کایمرا را به وجود آورد (۳ و ۱۳). تغییر الگو باعث بروز مشکلاتی در ارتباط با تشخیص مرزهای آگزون-ایترون و رونوشت‌های کایمرای واقعی می‌شود.

این محدودیت‌ها با به وجود آمدن فن‌آوری توالی‌یابی

"شریعتی و همکاران، معرفی فن آوری نوین توالی یابی آر.ان.آ"

توالی یابی آر.ان.آ می شود. توالی یابی مستقیم آر.ان.آ به مقدار آر.ان.آ کمی در حد چند فمتومول نیاز دارد که این مزیت در مورد نمونه هایی که به صورت بایگانی هستند و برای نگهداری طولانی مدت در فرمالین و پارافین تثبیت شده اند و ژنوم استخراجی از آن ها به صورت سالم نیست و همچنین برای گونه های آر.ان.آ کوتاه بسیار حائز اهمیت می باشد که به راحتی تبدیل به دی.ان.آ مکمل نمی شوند. علاوه بر این آماده سازی نمونه در این روش نسبتا ساده است (۲۱).

و آنالیز مستقیم آر.ان.آ برطرف شد که اخیرا و برای اولین بار با توالی یابی تک مولکولی Helicos توسعه پیدا کرد (۲۱ و ۲۳) (شکل ۲). این روش مبتنی بر دورگه سازی چند فمتومول از الگوهای آر.ان.آ که انتهای ۳' آن ها پلی آدنیل شده است با سطوح توالی یابی دارای پلی تیمین در کانال هایی می باشد. این روش می تواند آر.ان.آ های دارای پلی آدین را از آر.ان.آ کل یا عصاره سلولی جدا کرده و توالی یابی کند (۲۲). این فن آوری مزایایی نسبت به بقیه روش ها نشان می دهد که باعث پیشرفت و توسعه استفاده از روش



شکل ۲- توالی یابی مستقیم آر.ان.آ با روش Helicos (a): اتصال آر.ان.آ پلی آدنیل به توالی های پلی تیمیدین روی سطح و انجام مکانیسم 'Fill-and-lock': مرحله fill با تیمیدین طبیعی و پلی مراز و مرحله lock با نوکلئوتیدهای فلورسانس (b): عکس برداری برای مشخص کردن موقعیت دقیق الگو و برش شیمیایی اتصال رنگ و نوکلئوتید و آماده کردن الگو برای ورود نوکلئوتید بعدی (c): انکوباسیون سطح با یک نوکلئوتید نشاندار و فلورسانس و ادامه مراحل برش و شستشو

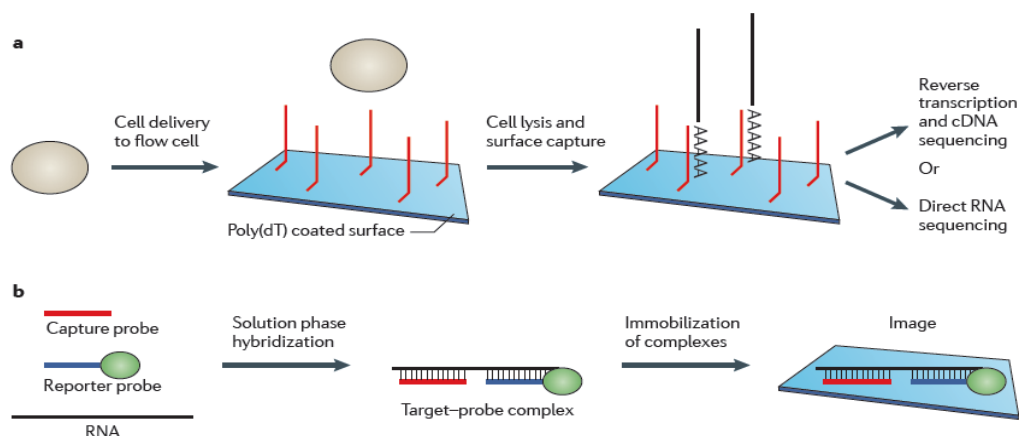
۷- توالی‌یابی نمونه‌های آر.ان.آ با فراوانی پایین

نمونه‌های بیولوژیکی (مثل بافت یا مایعات بدن) دارای ترکیب پیچیده‌ای از انواع سلول‌ها هستند که می‌بایستی برای بررسی، با روش‌های خاصی منتخب و مورد مطالعه قرار گیرند. در این میان، آنالیز نمونه‌های آر.ان.آ با فراوانی پایین، نیاز به چندین مرحله تکثیر دارد تا ماده ژنتیکی لازم برای توالی‌یابی حاصل شود که ممکن است صحت این توالی‌ها حفظ نشود (۴ و ۱۹). برای برطرف کردن این محدودیت‌ها، توالی‌یابی بر اساس توالی‌یابی آر.ان.آ مطرح شد که روشی جدید است.

در این فن آوری دو روش وجود دارد. روش اول بر مبنای استفاده از پرایمرهای دارای پلی‌تیمین بر روی سطوح رونویسی است که ام.آر.ان.آهای پلی‌آدنینی را از عصاره سلولی انتخاب کرده و با سنتز تک رشته‌ای دی.ان.آ مکمل بر روی سطح توالی‌یابی را انجام می‌دهد (شکل ۳a). روش‌های بر مبنای دورگه‌سازی نیز برای کار با آر.ان.آهای با فراوانی پایین مورد

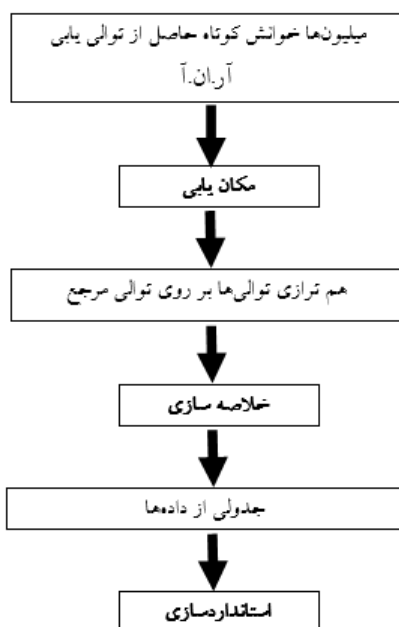
استفاده قرار می‌گیرند که مبتنی بر ایجاد نشانگرهای اختصاصی هدف است (شکل ۳b). در این روش مخلوط نشانگرها با نمونه‌های آر.ان.آ موجود در محلول دو رگه می‌شوند. از دو نشانگر مورد استفاده در این روش، یکی نشانگر تسخیری (Capture probe) است که شامل یک توالی اختصاصی هدف برای اتصال به آن و یکسری تغییراتی که باعث تثبیت هدف بر روی سطح می‌شود. نشانگر دوم گزارشگر (Reporter probe) است که دارای توالی برای اتصال در جایگاه دیگری از هدف و یک ماده فلورسانس می‌باشد. بعد از این مرحله، تصویربرداری برای تشخیص و شمارش رونوشت‌ها صورت می‌گیرد (۷ و ۲۱). با این سیستم می‌توان ۱۶۳۸۴ رونوشت را به صورت خود به خودی شناسایی کرد. این روش تقریباً ۱۰۰ نانوگرم آر.ان.آ یا ۵۰۰۰-۲۰۰۰ سلول نیاز دارد اما بهینه‌سازی دو رگه‌سازی با نشانگر و مراحل تثبیت در سطح ممکن است باعث کاهش میزان آر.ان.آ ورودی هم بشود (۲۱).

"شریعتی و همکاران، معرفی فن آوری نوین توالی‌یابی آر.ان.آ"



شکل ۳- فن آوری‌هایی برای بررسی بیان ژنی در تک سلولی‌ها یا دارای فراوانی پایین (a): انتقال سلول به سطح پوشیده از پلی‌تیمین و لیز سلولی، اتصال رونوشت‌های دارای پلی‌آدنین به سطح و توالی‌یابی آن‌ها (b): استفاده از دو پروب: پروب تسخیری با یک ناحیه مکمل برای اتصال به هدف و یک ناحیه برای اتصال به سطح و تثبیت آن، پروب گزارشگر دارای توالی مکمل ناحیه‌ای دیگر در هدف و یک بارکد فلورسانس برای شناسایی

وجود می‌آید. استانداردسازی، امکان مقایسه دقیق سطوح بیانی بین نمونه‌ها را امکان‌پذیر می‌کند (۱۰ و ۳۷).



شکل ۴- مروری بر مراحل آنالیز داده‌های حاصل از توالی‌یابی آر.ان.آ

آنالیز داده‌های حاصل از توالی‌یابی آر.ان.آ

بررسی‌های توالی‌یابی آر.ان.آ نیز مثل هر فن آوری دیگری در نهایت بانک داده‌ای وسیع و پیچیده‌ای نتیجه می‌دهد که تجزیه و تحلیل داده‌ها که همان توالی‌های کوتاه حاصله می‌باشند، برای رسیدن به نتایج دقیق، به اصول و روش‌های خاصی نیازمند است که با نرم افزارهای پیشرفته‌ای امکان‌پذیر می‌شود (۲۰ و ۳۱). در طرح کلی توالی‌یابی آر.ان.آ، ابتدا مکان‌یابی توالی‌های حاصله بر روی ژنوم یا رونوشت صورت می‌گیرد (Mapping). در مرحله دوم، توالی‌های نقشه‌یابی شده برای هر نمونه، بسته به هدف مورد آزمایش در سطح ژنی، اگزونی و یا سطح رونویسی خلاصه می‌شوند (Summarizing). در نهایت اطلاعات خلاصه شده در ارتباط با تست‌های آماری استانداردسازی می‌شوند (Normalizing) و لیستی طبقه‌بندی شده از ژن‌ها، با در نظر گرفتن احتمال آماری وقوع تصادفی آن‌ها (P-values) به

نتیجه‌گیری

با درک کاملی از همه رونوشت‌ها با این روش می‌توان به مسیرهای بیولوژیکی که در شرایط فیزیولوژیکی متفاوت فعال هستند، پی برد. پیشرفت‌های فن‌آوری توالی‌یابی آر.ان.آ، توانایی ما را برای استفاده از سنجش آر.ان.آهای موجود در تشخیص‌های بالینی بالا می‌برد که از جمله آن‌ها می‌توان به آر.ان.آهای جنینی یا سلول‌های توموری در حال گردش اشاره نمود. به‌طور کلی، این فن‌آوری‌های جدید، برآورد سریع‌تر شیوع بیماری و یا جایگاه‌های دچار جهش را امکان‌پذیر می‌کنند.

فن‌آوری توالی‌یابی آر.ان.آ و پیشرفت‌های اخیر آن، ابزاری قدرتمند برای محققین فراهم می‌کند که بتوانند به فهرست کاملی از رونوشت‌های مشتق شده از ژنوم همه موجودات از ارگانیسم‌های ساده تک سلولی تا سلول‌های پیچیده پستانداران، همچنین در بافت‌های سالم و بیمار دست یابند. علاوه بر این توانایی ما را برای کار کردن با آر.ان.آهای کم مقدار استخراج شده از بافت‌های تثبیت شده در پارافین یا فرمالین و یا گونه‌های آر.ان.آ موجود در تک سلولی‌ها، بالا می‌برد.

References

فهرست منابع

- 1- **Ameur A, Wetterbom, A., Feuk, L. and Gyllensten, U. (2010).** Global and unbiased detection of splice junctions from RNA-seq data. *Genome Biol.* 11: R34.
- 2- **Carninci P. (2009)** Is sequencing enlightenment ending the dark age of the transcriptome? *Nature Methods.* 6:711–3.
- 3- **Cocquet J, Chong, A., Zhang, G. and Veitia, R. A. (2006).** Reverse transcriptase template switching and false alternative transcripts. *Genomics.* 88:127–31.
- 4- **Dafforn A., et al. (2004)** Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis. *Biotechniques.* 37:854–7
- 5- **Faghihi M.A. and Wahlestedt C. (2009).** Regulatory roles of natural antisense transcripts. *Nature Rev Mol Cell Biol.* 10:637–43.
- 6- **Ghildiyal M. and Zamore P. D. (2009).** Small silencing RNAs: an expanding universe. *Nature Rev Genet.* 10:94–108.
- 7- **Geiss G.K., et al. (2008).** Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotech.* 26:317–25.
- 8- **He Y, Vogelstein, B., Velculescu, V. E., Papadopoulos, N. and Kinzler, K. W. (2008).** The antisense transcriptomes of human cells. *Science.* 322:1855–7.
- 9- **Hutchison CA. (2007).** DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res.* 35(18):6227–37.
- 10- **Kapranov P., et al. (2010).** New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism. *Nature.* 466:642–6.
- 11- **Levin J.Z., et al. (2009).** Targeted next-generation sequencing of a cancer transcriptome enhances detection of sequence variants and novel fusion transcripts. *Genome Biol.* 10: R115.
- 12- **Li J.B., et al. (2009).** Genome-wide identification of human RNA editing sites by parallel DNA capturing and sequencing. *Science* 324:1210–3.
- 13- **Mader R.M., et al. (2001).** Reverse transcriptase template switching during reverse transcriptase-

- polymerase chain reaction: artificial generation of deletions in ribonucleotide reductase mRNA. mRNA J Lab Clin Med. 137:422–8.
- 14- **Masoudi-Nejad A, Narimani Z and Hosseinkhan N. (2013).** Emergence of Next-Generation Sequencing. Next Generation Sequencing and Sequence Assembly: Springer. p. 11-39.
- 15- **Metzker M.L. (2010).** Sequencing technologies — the next generation. Nature Rev Genet. 11:31–46.
- 16- **Mortazavi A. Williams B., McCue K., Schaeffer L. and Wold B. (2008).** Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 5:621-8.
- 17- **Ni T., et al. (2010).** A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nature Methods. 7:521–7.
- 18- **Nilsen T.W., Graveley B.R. (2010).** Expansion of the eukaryotic proteome by alternative splicing. Nature. 463:457–63.
- 19- **Nygaard V. and Hovig E. (2006).** Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. Nucleic Acids Res. 34:996–1014.
- 20- **Oshlack A., Robinson M.D. and Young M.D. (2010).** From RNA-seq reads to differential expression results. Genome Biology. 11:220.
- 21- **Ozsolak F. and Milos P.M. (2011).** RNA sequencing: advances, challenges and opportunities. Nat Rev Genet. 11:87-98.
- 22- **Ozsolak F., et al. (2009).** Direct RNA sequencing. Nature. 461:814–8
- 23- **Ozsolak F., et al. (2010).** Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell. 143:1018–29.
- 24- **Parkhomchuk D., et al. (2009).** Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res. 37: e123.
- 25- **Plessy C., et al. (2010).** Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. Nature Methods. 7: 528–34.
- 26- **Rajagopalan R, Vaucheret, H., Trejo, J. and Bartel, D. P. (2006).** A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev. 20:3407–25.
- 27- **Reinert and Huson D.H. (2007).** bioinformatics Support for Genome-Sequencing Projects. Wiley-VCH Verlag. 25; 1-56.
- 28- **Ruby J.G., et al. (2006).** Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. Cell. 127:1193–207.
- 29- **Schuster S.C. (2008).** Next-generation sequencing transforms today's biology. Nature methods. 5:16-18.
- 30- **Shariati J.V., Tavakol E., Pietrella M., Giovanni G., Pagnotta M.A. and Porceddu E. (2014).** Comparative Transcriptome Analysis Using 454 Pyrosequencing in Two Bread Wheat Genotypes Under Non-Stressed and Water Stressed Conditions. BMC Genomics (In Press).
- 31- **Shendure J. and Ji H. (2008).** Next-generation DNA sequencing. Nat Biotechnol. 26:1135-45.
- 32- **Shiraki T., et al. (2003).** Cap analysis gene expression for high- throughput analysis of transcriptional starting point and identification of promoter usage. Proc Natl Acad Sci USA. 100:15776–81.
- 33- **Taft R.J., et al. (2009).** Tiny RNAs associated with transcription start sites in animals. Nature Genet. 41:572–8.

- 34- **Tavakol E. and Shariati J.V. (2014).** An introduction to next generation sequencing. *Mutation*. p. 11-23.
- 35- **Taylor KH. Shi H. and Caldwell CW. (2010).** Next generation sequencing: advances in characterizing the methylome. *Genes*. 1:143-65.
- 36- **Trapnell C. Pachter L. and Salzberg S. (2009).** L. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics*. 25:1105-11.
- 37- **Valen E., et al. (2009).** Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res*. 19: 255-65
- 38- **Wang G.S. and Cooper T. A. (2007).** Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Rev Genet*. 8:749-61.
- 39- **Wang Z., Gerstein M. and Snyder M. (2009).** RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 10(1):57-63.
- 40- **Zhang K., et al. (2009).** Digital RNA allelotyping reveals tissue- specific and allele-specific gene expression in human. *Nature Methods*. 6:613-8.

An introduction to RNA sequencing

Vahid Shariati, Farinaz Behfarjam*, Amir Mousavi

Assistant Professor of National Institute of Genetics and Biotechnology, Tehran, Iran

*Ph.D. Student of National Institute of Genetics and Biotechnology, Tehran, Iran

Associate Professor of National Institute of Genetics and Biotechnology, Tehran, Iran

farinaz.cell88@gmail.com

Abstract

High-throughput sequencing technologies are now in common use in biology. RNA-seq is one of the next generation sequencing that applies for transcriptome profiling. This method with deep sequencing of complementary DNA allows researchers to measure RNA abundance on a genome-wide scale and alter our view of the extent and complexity of eukaryotic transcriptomes. Recently, several developments in RNA-seq methods have provided an even more complete characterization of RNA transcripts than other methods. These developments include improvements in transcription start site mapping, strand-specific measurements, gene fusion detection and detection of alternative splicing events and isoforms. Ongoing developments of this technology get possible of direct RNA sequencing and profiling low-quantity RNA samples. In this review article will be introduced RNA-seq technology and implied some developments that overcome limitation of other methods.

Keywords: RNA-seq, sequencing, transcriptome, cDNA, transcript and mapping