Archive of SID

ISSN: 2476-6917; Modares Journal of Biotechnology. 2019;10(4):617-625

# Gene Expression Analysis Using RNA-Seq

ARTICLE INFO

*Article Type*
Original Research

*Authors*
Sharifi Alishah M.[1] *MSc,*
Darvishzadeh R.*[2] *PhD,*
Ahmadabadi M.[3] *PhD,*
Piri Kashtiban Y.[4] *PhD,*
Hasanpur K.[5] *PhD*

**How to cite this article**
Sharifi Alishah M, Darvishzadeh R, Ahmadabadi M, Piri Kashtiban Y, Hasanpur K. Gene Expression Analysis Using RNA-Seq. Modares Journal of Biotechnology. 2019;10(4):-617-625.

[1]Plant Breeding & Biotechnology Department, Agriculture Faculty, Urmia University, Urmia, Iran
[2]"Institute of Biotechnology" and "Plant Breeding & Biotechnology Department, Agriculture Faculty", Urmia University, Urmia, Iran
[3]Agricultural Biotechnology Department, Agriculture Faculty, Azarbaijan Shahid Madani University, Tabriz, Iran
[4]Agricultural Biotechnology Department, National Center of Genetic Engineering, Tehran, Iran
[5]Department of Animal Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

***Correspondence***
*Address*: Plant Breeding & Biotechnology Department, Agriculture Faculty, Urmia University, 11 Kilometer of Sero Road, Daneshgah Boulevard, Urmia, Iran. Postal Code: 5756151818
*Phone*: +98 (44) 31942785
*Fax*: +98 (44) 32779558
r.darvishzadeh@urmia.ac.ir

***Article History***
Received: June 6, 2018
Accepted: March 18, 2019
ePublished: December 21, 2019

ABSTRACT

Revealing DNA sequences is vital for all branches of biological sciences. Next-Generation Sequencing (NGS) is a different approach in this area so that it has created a great evolution in biology science and covers various aspects of genome, transcriptome, epigenome and metagenome-level studies. NGS is considered as a high-performance method for genomic and transcriptomic information analysis in comparison with traditional methods due to providing good genomic coverage, determining each single pairs of bases and eliminating the first generation sequencing disadvantages (Sanger sequencing). Use of NGS has begun since 2005 and 2006, after the commercialization of various apparatus companies such as ABI/SOLiD Illumina, Science Roch/454Life, and Solexa to study the transcriptome of the model and non-model organisms. Recently, RNA sequencing is used widely to identify genes associated with growth and development processes and their expression patterns in response to a variety of biological and non-biological stresses, in various organs and growth stages in different organisms. It helps scientists to determine the amounts of gene expression, differentiation of different isoforms of genes, detection of gene fusions and characterization of small RNA as well as alternative splicing events, duplicate elements, exon of genes, new transcripts, UTRs, SNPs, and somatic mutations. The RNA-seq method typically consists of providing suitable biological samples, isolation of total RNA, enrichment of non-ribosomal RNAs, conversion of RNA to cDNA, construction of a fragment library, selecting size and adding linkers and sequencing on high-throughput sequencing platform, alignment, and assembly of the reads and downstream analysis.

**Keywords** Gene Expression; High-Throughput Nucleotide Sequencing; RNA-Seq; Transcripe tome

CITATION LINKS

[1] Next-generation sequencing technologies and applications for human genetic ... [2] A tale of three next generation sequencing platforms: Comparison of Ion Torrent, ... [3] Next-generation ... [4] De novo assembly and characterization of transcriptome using Illumina paired-end sequencing and identification ... [5] Applications of next-generation sequencing technologies in functional ... [6] De novo transcriptome sequencing of desert herbaceous Achnatherum splendens ... [7] Analysis of saffron stigma (Crocus sativus L.) transcriptome using SOAPdenovo and Trinity assembly ... [8] Advances in omics and bioinformatics tools for systems analyses of plant ... [9] Expression profiling of genes involved in signaling process in ... [10] RNA-seq for gene identification and transcript profiling in relation to ... [11] Identification of salt-stress-induced genes from the ... [12] Transcriptome analysis of salt stress responsiveness in the ... [13] A decade's perspective on DNA sequencing ... [14] Informatics for RNA sequencing: A web resource for analysis on the ... [15] RNA-seq: A glance at technologies and ... [16] RNA-seq assembly-are we there ... [17] A comparison of methods for differential expression analysis of ... [18] Accuracy of allele frequency estimation using pooled ... [19] Mapping and quantifying mammalian transcriptomes ... [20] Novel methods for rRNA removal and directional, ... [21] Next-generation sequencing technologies for environmental ... [22] The emergence of nanopores in next-generation ... [23] A rapid, directional RNA-seq library preparation workflow for ... [24] Informed and automated k-Mer size selection for genome ... [25] StatsDB: Platform-agnostic storage and understanding of next generation ... [26] Advancing RNA-seq ... [27] Tools for mapping high-throughput sequencing ... [28] Transcript assembly and quantification by RNA-Seq reveals unannotated ... [29] A scaling normalization method for differential expression ... [30] Evaluation of statistical methods for normalization ... [31] Bootstrap-based differential gene expression analysis for RNA-Seq data with and ...

*Copyright© 2019, TMU Press. This open-access article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License which permits Share (copy and redistribute the material in any medium or format) and Adapt (remix, transform, and build upon the material) under the Attribution-NonCommercial terms.*

www.SID.ir

Archive of SID



# تجزیه بیان ژن به روش توالی‌یابی RNA

**معصومه شریفی‌علیشاه MSc**
گروه اصلاح و بیوتکنولوژی گیاهی، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران
**رضا درویش‌زاده* PhD**
"پژوهشکده زیست فناوری" و "گروه اصلاح و بیوتکنولوژی گیاهی، دانشکده کشاورزی"، دانشگاه ارومیه، ارومیه، ایران
**محمد احمدآبادی PhD**
گروه بیوتکنولوژی کشاورزی، دانشکده کشاورزی، دانشگاه شهید مدنی آذربایجان، تبریز، ایران
**یاسر پیری‌کشتیبان PhD**
گروه بیوتکنولوژی کشاورزی، مرکز ملی مهندسی ژنتیک، تهران، ایران
**کریم حسن‌پور PhD**
گروه علوم دامی، دانشکده کشاورزی، دانشگاه تبریز، تبریز، ایران

**چکیده**
رمزگشایی توالی DNA برای همه شاخه‌های علوم زیستی، حیاتی است. توالی‌یابی نسل جدید (NGS)، رویکردی متفاوت در توالی‌یابی است که تحولی عظیم در علم زیست‌شناسی ایجاد کرده و جنبه‌های مختلفی از مطالعات در سطح ژنوم، ترنسکریپتوم، اپی‌ژنوم و متاژنوم را پوشش می‌دهد. در مقایسه با روش‌های سنتی، نسل جدید توالی‌یابی، با فراهم‌کردن پوشش بالای ژنومی، تفکیک‌پذیری در حد تک‌تک جفت‌بازها و رفع معایب نسل اول توالی‌یابی (توالی‌یابی سانگر)، روشی با کارآیی بالا برای آنالیز اطلاعات ژنومی و ترنسکریپتومی است. استفاده از نسل جدید توالی‌یابی برای بررسی ترنسکریپتوم موجودات زنده مدل و غیرمدل از سال‌های ۲۰۰۵ و ۲۰۰۶ پس از تجاری‌شدن دستگاه‌های شرکت‌های مختلف مثل ABI/SOLiD Illumina، Roch/454 Life Science و Solexa آغاز شده است. در سال‌های اخیر تکنیک توالی‌یابی RNA برای شناسایی ژن‌های مرتبط با فرآیندهای رشدونموی و بررسی الگوهای بیان آنها در پاسخ به انواع تنش‌های زیستی و غیرزیستی، در اندام‌ها و مراحل متعدد رشدی در موجودات مختلف و همچنین میزان بیان ژن‌ها، تفکیک ایزوفرم‌های مختلف از یکدیگر، شناسایی امتزاج ژن‌ها، یافتن چندشکلی‌های تک‌نوکلئوتیدی (SNP)، عناصر تکراری، وقایع اتصال جایگزین، پیداکردن اسیدهای ریبونوکلئیک غیررمزگر، پیداکردن اگزون‌های ژنی و رونوشت‌های جدید، مناطق غیرترجمه‌شونده (UTR)، جهش‌های پیکری و غیره، به‌طور گسترده استفاده می‌شود. روش توالی‌یابی RNA شامل شناسایی نمونه‌های زیستی مناسب، استخراج RNAکل، غنی‌سازی RNAهای غیرریبوزومی، تبدیل RNA به cDNA و ساخت کتابخانه‌های توالی‌یابی، انتخاب قطعات براساس اندازه، اضافه‌کردن لینکرها، استفاده از توالی‌یاب‌ها یا پلت‌فرم‌های با توان عملیاتی بالا به‌منظور تولید صدها میلیون خوانش کوتاه، استفاده از نرم‌افزارهای خاصی به‌منظور کنترل کیفیت و تطابق خوانش‌ها با ژنوم مرجع یا ترنسکریپتوم، نقشه‌بردای و آنالیزهای پایین‌دستی است.

**کلیدواژه‌ها:** بیان ژن، ترنسکریپتوم، NGS، توالی‌یابی RNA

تاریخ دریافت: ۱۳۹۷/۳/۱۶
تاریخ پذیرش: ۱۳۹۷/۱۲/۲۷
*نویسنده مسئول: r.darvishzadeh@urmia.ac.ir

## مقدمه

در دهه ۱۹۷۰، *سانجر* و همکاران و همچنین *ماکسام* و *گیلبرت*، روش‌هایی برای توالی‌یابی DNA توسط تکنیک‌های خاتمه زنجیره و تکه‌تکه‌شدن ارایه دادند. در این بین به سبب این که روش سانجر، نیاز به دست‌ورزی و استفاده کمتر از مواد شیمیایی داشت، به‌عنوان روش تعیین توالی اصلی انتخاب شد و این امر علم زیست‌شناسی را متحول نمود[1].

اما وجود معایبی در روش‌های توالی‌یابی نسل اول، از جمله محدودیت توان عملیاتی، مقیاس‌پذیری، سرعت، تفکیک‌پذیری و هزینه‌ها سبب کندشدن دسترسی محققین به اطلاعات اساسی مورد نیاز می‌شد. بنابراین برای غلبه بر این موانع همچنین نیاز روبه‌رشد به توالی‌یابی ژنوم فردی به‌صورت سریع، کم‌هزینه و دقیق، به تغییر روش سنتی تعیین توالی سانجر با استفاده از تکنولوژی جدید یا توالی‌یابی نسل جدید (NGS) نیاز بود. این تکنولوژی رویکردی متفاوت در توالی‌یابی است که تاکنون منجر به تحولی عظیم در علوم ژنومیک شده است. این تکنیک همچنین به‌عنوان روش سانجر خودکار یا توالی‌یابی انبوه موازی (MPS) در سال ۲۰۰۴ معرفی شد[2, 3].

دستگاه‌های نسل جدید توالی‌یابی در اواخر دهه ۱۹۹۰ توسط شرکت‌های مختلف نظیر Illumina/Solexa، Roch/454 Life Science و ABI/SOLiD ایجاد شده است که عمدتاً از توالی‌یابی مبتنی بر الکتروفورز مویرگی روش سانجر متفاوت هستند، برای توالی‌یابی ژنوم انسان و موجودات مدل استفاده می‌شود و طی سال‌های ۲۰۰۶-۲۰۰۵ تجاری شدند. شرکت Roch/454 Life Science اولین موسسه پیشگام در تولید و تجاری‌سازی تکنولوژی NGS است. این شرکت، در اواخر دهه ۱۹۹۰ برای اولین‌بار از تکنولوژی موسوم به پایروسکوئنسینگ برای تعیین توالی DNA در مقیاس وسیع استفاده کرد. *رونقی* و همکاران در سال ۱۹۹۸ از پایه‌گذاران پایروسکوئنسینگ بودند. شرکت Solexa در سال ۲۰۰۶ دستگاه Genome analyzer را به بازار ارایه کرد که در سال ۲۰۰۷ توسط شرکت Illumina خریداری شد و از آن به بعد با نام Illumina/Solexa شناخته می‌شود. شرکت Sequencing by Oligonucleotide Ligation and Detection (SOLiD) در سال ۲۰۰۶ تکنولوژی توالی‌یابی از طریق اتصال را به بازار معرفی کرد و در همین سال نیز توسط شرکت Applied Biosystems (ABI) خریداری شد. در حال حاضر جدیدترین سیستم توالی‌یابی نسل سوم، که اولین‌بار توسط شرکت Helicos Bioscience معرفی شد، قادر به توالی‌یابی یک مولکول منفرد است. شرکت Ion Torrent در اواخر سال ۲۰۱۰، دستگاهی موسوم به ماشین ژنوم شخصی یون تورنت (Ion PGM) را روانه بازار کرد که نوع دیگری از دستگاه‌های نسل سوم توالی‌یابی است[3, 4]. هر چند که هر کدام از این ابزارها دارای ویژگی‌های خاص خود هستند، با این حال همه دستگاه‌های توالی‌یابی انبوه موازی، دارای برخی ویژگی‌های قابل توجه نیز هستند. اول این که مراحل اولیه آماده‌سازی کاهش و ساده شده است. دوم، افزایش قطعات کتابخانه برای تمام پلت‌فرم‌ها ضروری است. سوم، واکنش‌های توالی‌یابی به‌طور خودکار شناسایی و انجام می‌شوند. در دهه گذشته، میزان توالی خروجی در هر اجرا به‌طور


www.SID.ir

Archive of SID

چشمگیری افزایش یافته است و هزینه پایه توالی‌یابی DNA در حدود ۱۰۰،۰۰۰ برابر کاهش و دقت عملیات تا حد زیادی بهبود یافته است. دستگاه‌های توالی‌یابی نسل دوم فعلی می‌توانند حدود یک‌میلیارد باز را در یک هفته بخوانند.

ساده و ارزان‌شدن فرآیند توالی‌یابی باعث افزایش پژوهش‌های زیست‌پزشکی شده است. برای ایجاد انواع کاتالوگ‌های جامع ژنتیکی، تکنیک‌های توالی‌یابی نسل جدید برای تولید داده‌های حاصل از توالی‌یابی در ۱۰۰۰ پروژه ژنومی به کار گرفته شده است. این برنامه قصد دارد بیش از ۲۰۰۰ نفر را برای یافتن چندشکلی‌های تک‌نوکلئوتیدی (SNP)، درج‌ها/حذف‌ها و همچنین تنوعات ساختاری با فراوانی بیش از ۱٪ در سراسر ژنوم و فراوانی بیش از ۰/۱% در مناطق کدکننده پروتئین‌ها، مورد مطالعه قرار دهد[3, 5].

در سال‌های اخیر، آنالیز ترنسکریپتوم به‌عنوان یکی از مراحل بنیادی در مطالعات زیست‌شناسی مطرح است[5]. ترنسکریپتوم، به سری کامل از رونوشت‌ها یا توالی‌های رونویسی‌شده در یک سلول در یک مرحله خاص از رشدونمو یا تحت یک شرایط فیزیولوژیک خاص گفته می‌شود[4]. آنالیز ترنسکریپتوم برای تفسیر عملکرد عناصر ژنومی و آشکارسازی اجزای اصلی مولکولی سلول‌ها و بافت‌ها ضروری است. بر این اساس ارزیابی ترنسکریپتوم، یک رویکرد مناسب برای به‌دست‌آوردن اطلاعاتی در زمینه سطوح عملکردی ژنومی موجودات غیرمدل است[6].

روش‌های مختلفی برای آنالیز ترنسکریپتوم وجود دارد، که از آن جمله می‌توان نورترن‌بلات، رونوشت‌برداری معکوس واکنش زنجیره‌ای پلیمراز، ریزآرایه‌ها (میکرواری) و توالی‌یابی به روش سنتی را نام برد[5, 7]. هر یک از روش‌های مذکور مشکلات و معایب مربوط به خود را دارند، از آن جمله می‌توان به استفاده از مواد رادیواکتیو در نورترن‌بلات، کارآیی پایین در روش رونوشت‌برداری معکوس واکنش زنجیره‌ای پلیمراز، عدم شناسایی رونوشت‌های جدید توسط ریزآرایه‌ها و هزینه، زمان و پیچیدگی بالای توالی‌یابی سنتی اشاره کرد. اخیراً با توسعه تکنولوژی‌های نسل جدید توالی‌یابی (NGS)، امکان مطالعه جامع ترنسکریپتوم‌ها، حتی در گونه‌های غیرمدل به کمک روش توالی‌یابی RNA فراهم شده است. فناوری توالی‌یابی نسل جدید (NGS) در سال‌های اخیر توسعه چشمگیری یافته است، به‌طوری که فناوری توالی‌یابی Illumina/Solexa ابزار قدرتمندی را برای توالی‌یابی RNA، سرهم‌بندی ترنسکریپتوم و پروفایلینگ بیان ژن فراهم کرده است[8]. هدف از این مطالعه توصیف روش توالی‌یابی RNA، تکنولوژی‌ها، روش‌ها، ابزارها و روش‌های انجام‌شده روی پروفایل ترنسکریپتوم‌ها است.

شناسایی اجزای شبکه سیگنالی در پاسخ به تنش شوری با استفاده از روش توالی‌یابی RNA در گیاه آجیلوپس (*Aegilops tauschii*) انجام شده است. نتایج بیان ژن نشان داد که از بین ۴۵۰۶ رونوشت دارای تغییر بیان معنی‌دار، فرآیندهای سیگنالی با ۶۰۳ عدد BIN (۱۱/۵۹%) بیشترین شمار را به خود اختصاص داده است. همچنین مشخص شد که از بین اجزای شبکه سیگنالی، گیرنده (رسپتور) کینازها بیشترین شمار (BIN ۲۲۲ عدد) از ژن‌های تغییر بیان یافته را شامل می‌شوند که درواقع بیان‌کننده اهمیت آنها است. نتایج این تحقیق، زمینه لازم برای انجام بررسی‌های انتقال ژن‌های مقاومت به شوری به گیاه گندم را فراهم کرده است[9]. همچنین آنالیز ترنسکریپتوم چمن (*Cynodon dactylon*) به روش توالی‌یابی RNA نمونه دیگری از بررسی مکانیزم مقاومت به تنش شوری است که به‌منظور شناسایی ژن و پروفایل رونوشت مربوط به رشد ریشه‌ها انجام شد. به‌وسیله توالی‌یابی RNA ژن‌های کدکننده فاکتورهای رونویسی، که در تنظیم سنتز لیگنین دخیل هستند، هومئوستازی ROS، که به‌وسیله پراکسیدازها کد می‌شوند و تنظیم سیگنالی فیتوهورمون‌ها که باعث ازبین‌رفتن دیواره سلول‌ها، در نتیجه تنش شوری می‌شوند، شناسایی شد. در مجموع، این اطلاعات نشان می‌دهد که تنظیم رشد ریشه‌ها، تحت تنش شوری شامل تغییر در بسیاری از جنبه‌های مختلف متابولیزم سلولی، سیگنالینگ، و جابه‌جایی مواد است[10]. شناسایی ژن‌های القاشونده تحت تنش شوری در گیاه ریبوموریا تری‌جینا (*Reaumuria trigyna*) از طریق داده‌های حاصل از توالی‌یابی RNA، با استفاده از DDRT-PCR انجام شده است، که شناسایی ژن‌های دخیل در تنش شوری و ایجاد درک عمیقی از مکانیزم مقاومت به شوری در گیاه *R. trigyna* را فراهم کرده است[11]. آنالیز ترنسکریپتوم برنج وحشی دونگسیانگ [12] (*Oryza rufipogon Griff*) و همچنین توالی‌یابی ترانسکریپتوم جوانه‌های علف بیابانی (*Achnatherum splendens*) [6]، نیز نمونه‌های دیگری از شناسایی ژن‌های دخیل در مقاومت به تنش شوری، به کمک روش توالی‌یابی RNA است.

**توالی‌یابی RNA**

توالی‌یابی RNA که به توالی‌یابی شاتگانی کل ترانسکریپتوم نیز معروف است، فناوری است که با بهره‌گیری از توانایی‌های توالی‌یابی نسل جدید (NGS) برای به‌دست‌آوردن تصویری کلی از حضور و مقدار RNA در یک بازه زمانی خاص استفاده می‌کند. این فناوری قابلیت توالی‌یابی همزمان میلیون‌ها خوانش را دارد و توصیف کمی و کیفی دقیق‌تری را از بیان ژن نسبت به فناوری ریزآرایه ارایه می‌دهد. علاوه بر این، تجزیه و تحلیل سریع، مقرون‌به‌صرفه و جامع جمعیت‌های پیچیده اسید نوکلئیکی برای گیاهان مدل یا گیاهان غیرمدل که ژنومشان هرگز توالی‌یابی نشده است را فراهم می‌کند[13].

به کمک تکنیک توالی‌یابی RNA اکثر ژن‌های بیان‌شونده در مرحله‌ای خاص شناسایی شده، ایزوفرم‌های مختلف از هم تفکیک و میزان بیان رونوشت‌ها اندازه‌گیری می‌شود. همچنین می‌توان به تعیین نمایه ترنسکریپتوم و سطوح بیان ژن، امتزاج ژن‌ها، چندشکلی تک‌نوکلئوتیدی یا SNP، عناصر تکراری، وقایع اتصال جایگزین، پیداکردن اسیدهای ریبونوکلئیک غیررمزگر، بهبود تفسیر ژنی، پیداکردن اگزون‌های ژنی، ژن‌ها و ترنسکریپت‌های جدید، پیداکردن ترنسکریپت‌های در جهت مخالف، نواحی غیرترجمه‌شونده یا UTR، کدون‌های آغازگر جایگزین رونویسی، قاب‌های بازکننده

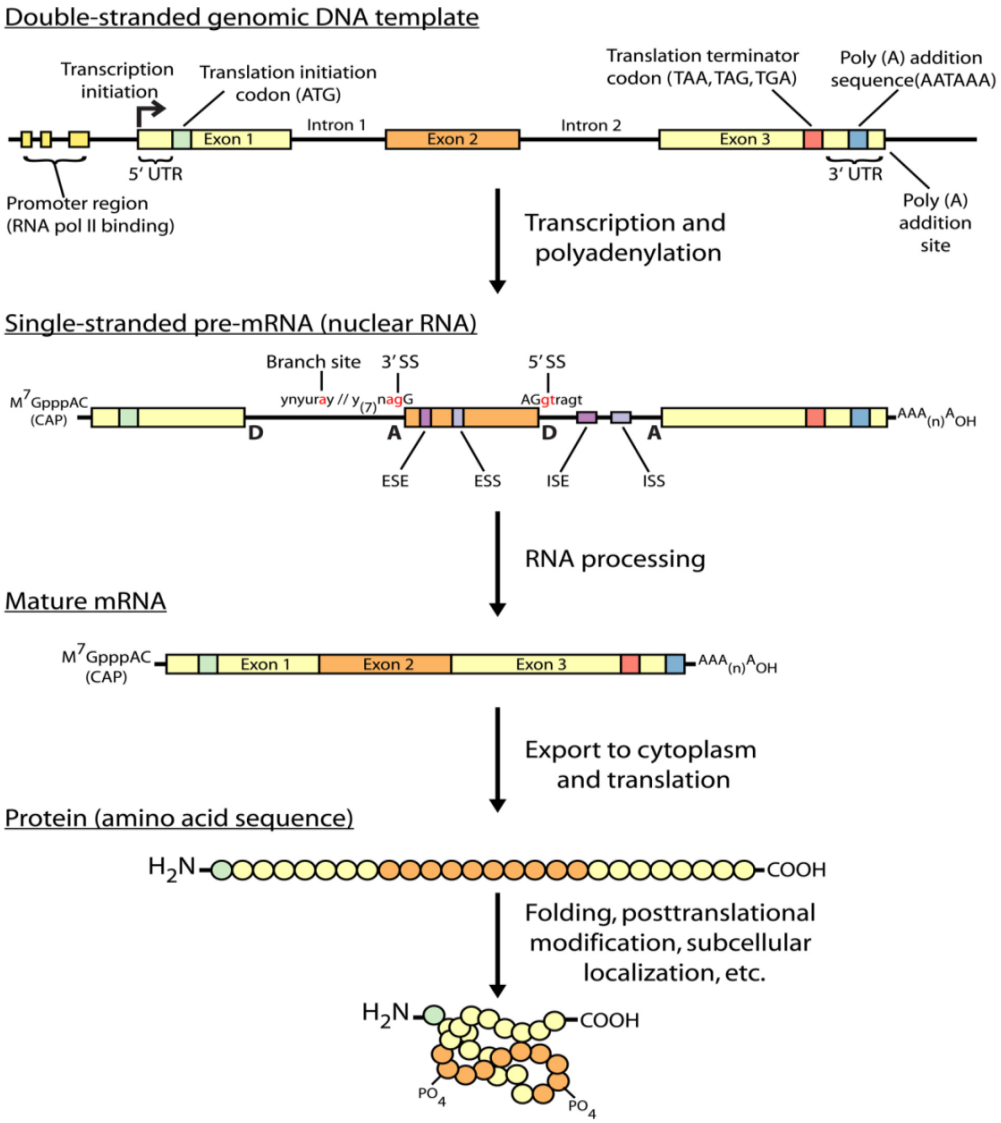

www.SID.ir

Archive of SID



آغازگر رونویسی، کاتالوگ‌های ژن‌های بیان‌شده اشاره نمود. توالی‌یابی RNA یک روش پویا و جدید برای آنالیز تفاوت بیان ژن است که بسیار دقیق‌تر از روش‌های دیگر است. شناسایی ژن‌های با بیان بسیار کم مرتبط با آزمایش را میسر می‌کند. با هزینه معقول و رو به کاهش اطلاعات زیادی را تولید می‌کند. علاوه بر میزان بیان ژن‌ها، توالی ژن‌ها، جهش‌های پیکری و سایر تنوعات موجود در توالی‌ها را فراهم کرده و برای بررسی بیماری‌های ژنتیکی از جمله سرطان بسیار مناسب است. شکل ۱، شماتیکی از اصل مرکزی دوگما است[14].

سه مرحله اساسی در توالی‌یابی RNA شامل بخش *wet-lab*، بخش *equipo* و درنهایت بخش این‌سیلیکو (*in silico*) است. مرحله *wet-lab* به‌عنوان بخش آزمایشی و شامل طراحی آزمایش با توجه به اهداف مورد نظر، استخراج RNA، خالص‌سازی و غنی‌سازی mRNAها، سنتز cDNA و ساخت کتابخانه‌های توالی‌یابی، قطعه‌قطعه‌کردن، افزودن آداپتور و تکثیر قطعات است. بخش *equipo in* که به‌عنوان بخش توالی‌یابی است، براساس پروتکل‌ها، روش‌ها و با استفاده از کیت‌های ارایه‌شده توسط هر تکنولوژی پلت‌فرم، صورت می‌گیرد. این بخش شامل انتخاب پلت‌فرم‌های مناسب، طراحی طریقه ران‌کردن نمونه‌ها و توالی‌یابی است. تهیه کتابخانه cDNA که آخرین مرحله بخش *wet-lab* است، می‌تواند جزء این بخش نیز محسوب شود. بخش این‌سیلیکو، به‌عنوان بخش بیوانفورماتیکی، مرحله چگونگی تجزیه و تحلیل داده‌های حاصل از توالی‌یابی براساس اهداف پروژه، الزامات، ابزارها و برنامه‌های موجود به‌منظور دست‌یابی به نتایج قابل اطمینان‌تر است. این بخش شامل کنترل کیفیت، پاکسازی آلودگی‌ها و تطابق خوانش‌ها با ژنوم مرجع، نقشه‌برداری، شناسایی ژن‌ها و آنالیزهای دیگری از قبیل شناسایی تفاوت بیان ژن‌ها، فیلوژنتیک، تنوع ژنتیکی، SNPها و غیره است[15-17] (شکل ۲).

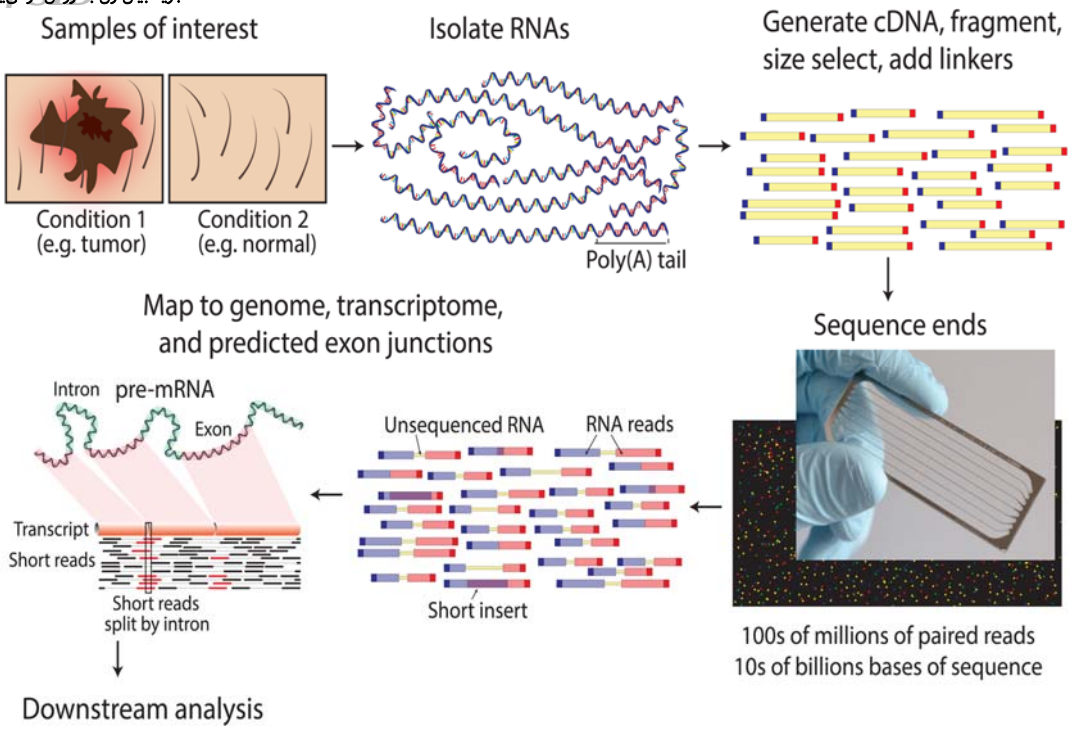



**شکل ۱)** شماتیکی از اصل مرکزی دوگما که نشان‌دهنده جریان اطلاعات ژنومی از DNA دورشته‌ای به پروتئین‌های تغییریافته پس از ترجمه است که هر مرحله با ویژگی‌های مولکولی ویژه، مورد بررسی قرار گرفته است. اختصارات شامل D: جایگاه اتصال‌دهنده، A: جایگاه اتصال‌پذیرنده، UTR: مناطق غیرترجمه‌شونده، SS: مناطق اسپلایسینگپ، ESE: مناطق تقویت‌کننده اسپلایسینگ اگزونی، ESS: مناطق خاموش‌کننده اسپلایسینگ اگزونی، ISE: مناطق تقویت‌کننده اسپلایسینگ اینترونی، ISS: مناطق خاموش‌کننده اسپلایسینگ اینترونی



www.SID.ir

Archive of SID




**شکل ۲)** روش توالی‌یابی RNA شامل شناسایی نمونه‌های زیستی مناسب (و تکرار)، استخراج RNA کل، غنی‌سازی RNAهای غیرریبوزومی، تبدیل RNA به cDNA و ساخت کتابخانه‌های توالی‌یابی، انتخاب قطعات براساس اندازه، اضافه‌کردن لینکرها، استفاده از توالی‌یاب‌ها یا پلت‌فرم‌های با توان عملیاتی بالا به‌منظور تولید صدها میلیون خوانش کوتاه، الاینمنت خوانش‌ها با ژنوم مرجع یا ترنسکریپتوم و آنالیزهای پایین‌دستی برای تخمین بیان، شناسایی ایزوفرم‌های رونوشت و غیره است.

عوامل مختلفی می‌تواند هر مرحله از پروژه توالی‌یابی RNA را تحت تاثیر قرار دهد. در بخش *wet-lab*، کیفیت، مقدار و خلوص RNA، تخریب RNA و حذف RNAهای ریبوزومی می‌تواند بر نتایج توالی‌یابی RNA بسیار تاثیرگذار باشد، زیرا تمام تجزیه و تحلیل‌ها براساس RNA ورودی انجام می‌شود. همچنین تفاوت در طول خوانش‌ها، دقت خواندن و زمان توالی‌یابی در بخش *equipo in* و نیز استفاده از نرم‌افزارها و سخت‌افزارهای خاص و کارآیی متفاوت آنها در بخش این‌سیلیکو بر نتایج حاصل از توالی‌یابی RNA موثر است[15].

### ایزولاسیون RNA

به‌منظور بررسی تفاوت بیان ژن‌ها تحت شرایط نرمال و غیرنرمال (تنش)، با استفاده از روش توالی‌یابی RNA، ابتدا نمونه‌های زیستی مناسب (و تکرارها)، تعریف و استخراج RNA از آنها صورت می‌گیرد. اساساً برای بررسی ترنسکریپتوم با استفاده از تکنیک توالی‌یابی RNA، mRNA مورد نیاز است، در حالی که بسته به اهداف پروژه، انواع دیگر RNA مانند RNA کل، RNAهای کوچک، میکروRNAها و RNAهای غیرکدکننده نیز به‌عنوان مواد مورد نیاز تکنیک توالی‌یابی RNA به کار برده می‌شوند[18].

بین mRNA یوکاریوت‌ها و پروکاریوت‌ها و روش‌های بررسی RNA در آنها تفاوت‌هایی وجود دارد. در انتهای ´۳ اکثر mRNAهای یوکاریوتی، دم پلیA وجود دارد که موجب اتصال توالی‌های الیگوdT به دم پلیA می‌شود. سه روش برای ایزولاسیون RNA به کار می‌رود: روش اول با استفاده از ستون‌های دارای پروب‌های dT انجام می‌شود که با عبور RNA کل از داخل این ستون‌ها، RNAهای دارای دم پلیA، به ستون‌ها متصل می‌شوند. در روش دوم، که روشی بسیار سریع است، RNAها به‌طور مستقیم از سلول‌های لیزشده با اتصال به توالی‌های الیگوdT رسوب داده‌شده روی بیدهای مغناطیسی، با مگنت‌های قوی بیرون کشیده می‌شوند. روش سوم لیزکردن سلول‌ها و سپس تهیه کمپلکس ریبوزوم-mRNA براساس گرادیانت‌های ساکارز است[15, 19, 20].

خلوص و عدم خردشدگی RNA حین استخراج بسیار حائز اهمیت است. محاسبه نسبت جذب ۲۶۰ به ۲۸۰نانومتر برای نمونه‌های RNA، تعیین‌کننده خلوص اسیدهای نوکلئیک استخراجی است. محدوده قابل قبول نسبت ۲۶۰ به ۲۸۰نانومتر، برای نمونه‌های RNA ۱/۸-۲/۱ است و عدد مطلوب، ۲ در نظر گرفته می‌شود. عدد کمتر از ۱/۸ نشان‌دهنده وجود ناخالصی و عدد بیشتر از ۲/۳ نشان‌دهنده تخریب RNA است. عدد خردشدگی RNA با RIN از یک تا ۱۰ تعریف می‌شود. در صورتی که ارزش RIN محاسبه‌شده برای نمونه‌ها کمتر از ۷ باشد، توالی‌یابی یا سکانسینگ انجام نمی‌شود[15].

### تهیه کتابخانه

به‌منظور تهیه کتابخانه‌های توالی‌یابی، mRNA باید با واکنش رونویسی معکوس با استفاده از آغازگرهای تصادفی یا پرایمرهای الیگوdT به cDNA تبدیل شود. استفاده از روش دوم، به این دلیل که نتایج حاصل از توالی‌یابی، حاوی اطلاعات مطلوبی (فاقد RNAهای ریبوزومی) است، مناسب‌تر است. البته این روش باعث غنی‌شدن کتابخانه از نواحی ´۵ می‌شود. بهتر است RNAهای ریبوزومی در مرحله استخراج به‌وسیله ستون پلیA حذف شوند. در این صورت روش اول ترجیح داده می‌شود. به‌منظور ساختن cDNA


www.SID.ir

Archive of SID


و کتابخانه‌های مربوطه، کیت‌های خاصی برای هر پلت‌فرم با توجه به تکنولوژی و دستورالعمل‌های آن، مورد استفاده قرار می‌گیرد. بعد از تهیه کتابخانه و قطعه‌قطعه‌کردن cDNA، نگهداری اطلاعات mRNA، مهم‌ترین فاکتور برای داشتن نتایج و داده‌های تمیز و سالم است.

**توالی‌یابی و پلت‌فرم‌های مورد استفاده**

اولین و مهم‌ترین امر در تکنیک‌های تعیین توالی این است که محققین براساس ویژگی‌های طرح پژوهشی خود، مناسب‌ترین روش توالی‌یابی را از نظر کارآیی انتخاب نمایند. از زمان معرفی روش تعیین توالی بر پایه سانجر، بسیاری از تکنیک‌ها در جهت افزایش بازدهی تعیین توالی و بالطبع کاهش هزینه این پروسه، توسط شرکت‌های مختلف ارایه شده است. شرکت‌هایی از قبیل Pacific Biosciences و 454 Life Sciences، به‌علت ارایه طول خوانش‌های بیشتر، از سایر شرکت‌ها متمایز شده‌اند. شرکت‌هایی از قبیل Illumina و Ion Torrent به‌واسطه هزینه‌های پایین‌تر تعیین توالی و تکنیک‌هایی از قبیل Solid با وجود طول خوانش‌های کوچک‌تر، به‌دلیل بازدهی بیشتر، مورد توجه محققین هستند. همچنین شرکت Oxford Nanopore Technologies با ارایه ابزارهای جدیدی از قبیل نانوپورها، با هزینه بسیار اندک ولی با میزان خطای بیشتر عمل تعیین توالی را انجام می‌دهند[21, 22]. تمام پلت‌فرم‌ها، دارای یک اصل اساسی شامل ایزولاسیون RNA کل، غنی‌سازی mRNA، ساختن cDNA و اتصال cDNA به یک ماتریکس جامد تک‌قطعه‌ای، با محدودکردن رقت و سپس تکثیر این مولکول‌ها به‌وسیله یک امولسیون اختصاصی PCR (emPCR) در SOLiD، 454 Pyrosequencing، Ion Torrent، یا اتصال براساس واکنش پل (Illumina) هستند، در حالی که فناوری Pacbio یا Pacificbio نیازی به تکثیر cDNA ندارد و توالی‌یابی به‌طور مستقیم از cDNA هدف انجام می‌شود.

نسخه‌های cDNA مولکول‌های RNA همسان می‌توانند به‌طور موازی یا با اندازه‌گیری الحاق نوکلئوتیدهای فلورسنت (Illumina)، اتصال‌دهنده‌های کوتاه فلورسنت (SOLiD)، یا به‌وسیله آزادکردن محصولات جانبی حاصل از الحاق نوکلئوتیدهای نرمال (۴۵۴)، انتشارات فلوئورسانس یا با اندازه‌گیری تغییر pH (یون‌تورنت) توالی‌یابی شوند. با توجه به اطلاعات موجود در مورد استراتژی کتابخانه توالی‌یابی RNA که برای پلت‌فرم‌های مختلف در SRA اجرا می‌شوند، Illumina با بیش از ۹۴/۲% در رتبه اول و سپس به‌ترتیب ABI/SOLiD، 454 Roche، Ion Torrent، Helicos HeliScope و Pacific Bio در رتبه‌های بعدی قرار دارند[15]. به‌علت هزینه پایین‌تر و مقدار بیشتر توالی خوانده‌شده توسط Illumina، این تکنولوژی به‌ویژه در سال‌های اخیر مورد توجه محققان مختلف قرار گرفته است[23].

**بیوانفورماتیک و ابزارهای آنالیز توالی‌یابی RNA**

بیوانفورماتیک یک ابزار قوی برای تجزیه و تحلیل داده‌های حاصل از مطالعات زیستی است. آنالیز داده‌های توالی‌یابی RNA به‌منظور تصحیح، ترمیم و پاک‌سازی خوانش‌های پرایمر، خوانش‌های با کیفیت پایین و سایر آلودگی‌های ممکن، تطابق و جمع‌آوری خوانش‌ها، تطابق یا نقشه‌برداری با ترانسکریپتوم یا ژنوم، اندازه‌گیری اگزون‌ها یا ژن‌ها و درنهایت تجزیه و تحلیل داده‌های به‌دست‌آمده از تفاوت بیان ژن به کار برده می‌شود[15].

توالی‌های به‌دست‌آمده از تکنیک توالی‌یابی RNA خوانش خام نامیده می‌شوند که در قالب فستکیو ذخیره شده‌اند. در این قالب هر قرائت خام با یک @ شروع می‌شود. خطوط اول و سوم نام توالی را نشان می‌دهند که توسط دستگاه توالی‌یاب ایجاد شده است. خط دوم شامل توالی قرائت خام به‌دست‌آمده است و خط چهارم حروفی هستند که بیانگر کیفیت نوکلئوتیدهای قرائت‌شده در هر توالی هستند. درواقع هر حرف از خط چهارم مربوط به یکی از بازهای خط دوم است. ارزش کیفی توالی‌یابی در محدوده ۴۰-۱ قرار دارد. قرائت‌های نامطلوب موجود در داده‌های خام می‌تواند در آنالیزهای بیوانفورماتیکی بعدی اثر منفی داشته باشد[24]. کنترل کیفیت توالی‌ها، اولین مرحله از توالی‌یابی RNA است. برنامه‌ای که به‌طور رایج در سراسر دنیا برای کنترل کیفیت توالی‌های حاصل از توالی‌یابی RNA به کار می‌رود، نرم‌افزار FastQC است. این نرم‌افزار فایل توالی‌ها را کنترل می‌کند و اطلاعاتی را درباره توالی خوانش‌ها ارایه می‌دهد (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) [25].

از برنامه‌های دیگر کنترل کیفیت توالی‌ها، می‌توان به NGS QC Toolkit اشاره کرد که توالی‌های Illumina و Roche 454 را بررسی و تصحیح می‌کند (http://59.163.192.90:8080/ngsqctoolkit/). برای حذف توالی‌ها از قطعات با کیفیت پایین، برنامه‌های مختلفی مانند ConDeTri، trimmomatic BlastX Toolkit، DynamicTrim،NGS QC Toolkit و Quake استفاده می‌شود. برنامه‌هایی مانند Biostring (www.bioconductor.org)، Seqclean (http://compbio.dfci.harvard.edu/tgi/software/) و CANGS DB برای بهبود کیفیت توالی‌ها و برنامه SEECER برای تصحیح خطای توالی‌یابی و بهبود کیفیت تطابق (الاینمنت) خوانش‌ها با ژنوم و دقت مونتاژ مورد استفاده قرار گیرند. زمانی که خوانش‌های با کیفیت بالا به دست آمد، اولین مرحله آنالیز، تعیین خوانش‌های کوتاه حاصل از توالی‌یابی RNA، به‌عنوان ترنسکریپتوم برای شناسایی ساختارهای رونویسی و ترنسکریپتوم است. دو رویکرد یا استراتژی عمومی برای هم‌ردیفی و مونتاژ داده‌های توالی‌یابی RNA به‌عنوان مونتاژ *de novo* و نقشه‌برداری با استفاده از یک ژنوم مرجع وجود دارد[26]. استراتژی اول یعنی مونتاژ *de novo* برای شناسایی رونوشت‌های موجودات غیرمدل یا موجوداتی که ژنوم آنها توالی‌یابی نشده یا ناقص است، قابل استفاده است. نرم‌افزارهایی مانند SOAP denovo.Trinity، SOAP denovo- و


www.SID.ir

Archive of SID

Trans، Trans-ABySS و Velvet/Oases برنامه‌هایی هستند که به‌منظور مونتاژ *de novo* مورد استفاده قرار می‌گیرند. در استراتژی دوم که به‌منظور هم‌ردیفی‌کردن داده‌های توالی‌یابی RNA به یک ترانسکریپتوم، نیاز به یک ژنوم مرجع دارد، از نرم‌افزارهای Bowtie2، GMAP و TopHat2 استفاده می‌شود[27]. شناسایی رونوشت‌ها، مرحله بعدی است که برای این منظور از دو برنامه بسیار پرکاربرد BLAST+ و Blast2GO و همچنین برنامه KOBAS2.0 استفاده می‌شود.

بررسی تفاوت بیان ژن، یکی از اهداف اصلی توالی‌یابی RNA است که ابزارهای بیوانفورماتیکی بسیاری برای این منظور وجود دارد. تجزیه بیان ژن به روش توالی‌یابی RNA بر این اساس است که چه تعداد خوانش می‌توانند به یک ژن اختصاصی منطبق شود. برای مقایسه تفاوت بیان ژن‌ها، نیاز به نرمال‌سازی است. به‌منظور نرمال‌سازی و محاسبه تفاوت بیان ژن‌ها، روش‌های مختلفی از قبیل RPKM [19]، FPKM [28] و TMM [29] وجود دارد. آزمون آماری برای تعیین تفاوت معنی‌دار در بیان ژن‌ها، می‌تواند براساس یکی از روش‌های زیر صورت گیرد: توزیع دوتایی منفی (DESeq, Cufflinks)، روش‌های بیز مبتنی بر مدل پواسون بیش- پراکنده (EdgeR, BaySeq, BitSeq)، روش بیز تجربی (Alexa-Seq)، آلگوریتم امیدریاضی- بیشینه‌سازی (RSEM, EBSeq)، آماره های ناپارامتری و مدل‌های تجربی روی توزیع نویز تغییرات شمارشی با کنترست‌کردن تفاوت‌های تغییر در برابر (M) و تفاوت‌های بیان مطلق (NOISeq)، روش ناپارامتری بر پایه بیز تجربی (NPEBseq)، آزمون دقیق فیشر که توزیع تعداد nامین ژن را با توزیع تعداد ژن دیگر از طریق اندازه‌گیری ارتباط بین متغیرهای تعداد ژن مورد علاقه و فاکتور نرمال‌کننده مقایسه می‌کند[30, 31].

پلی‌مورفیسم تک‌نوکلئوتیدی (SNP) را نیز می‌توان با استفاده از نتایج توالی‌یابی RNA تجزیه و تحلیل کرد. برنامه‌هایی از قبیل Samtools/Bcftools، GATK و VCFtools برای این منظور به کار می‌روند. آلترناتیواسپلایسینگ فرآیندی است که پس از رونویسی روی می‌دهد و به موجب آن چندین ایزوفرم mRNA بالغ از یک mRNA نابالغ ایجاد می‌شود و می‌تواند توسط توالی‌یابی RNA با استفاده از برنامه‌های Cufflinks/Cuffdiff، DEXseq، MISO و غیره مورد بررسی قرار گیرد. آنالیزهای دیگر توالی‌یابی RNA، یافتن و تشخیص اتصالات ژنی است. همجوشی ژن به‌وسیله انتقال، حذف یا وارونگی کروموزومی به وجود می‌آید که سبب ایجاد تغییراتی در کروموزوم‌ها شده و موجب ایجاد مشکلاتی مانند سرطان می‌شود. برای یافتن اتصالات ژنی، برنامه‌هایی مانند FusionSeq، SOAPfusion، SOAPFuse و TopHat-Fusion کاربرد دارند[15].

**کاربردهای NGS**

از کاربردهای این تکنولوژی پیشرفته می‌توان به موارد زیر اشاره کرد:

۱) کاربرد در سطح ژنوم شامل توالی‌یابی *denovo*، توالی‌یابی مجدد کل ژنوم، بررسی‌های جامع تمام SNPها، تغییرات ساختاری و تغییر در تعداد نسخه ژن‌ها، شناسایی چندشکلی‌ها و جهش‌های نقطه‌ای

۲) کاربرد در سطح ترنسکریپتوم شامل بررسی بیان ژن‌ها، شناسایی جهش‌های پیکری و شناسایی پردازش‌های متناوب، تعیین پروفایل بیان small RNAها

۳) کاربرد در سطح اپی‌ژنوم شامل تعیین جایگاه‌های هدف فاکتورهای رونویسی، پروفایل ژنومی اصلاحات مربوط به هیستون‌ها، الگوهای متیلاسیون DNA

۴) کاربرد در سطح متاژنوم شامل بررسی اثرات محیط و پاتوژن‌ها بر ژنوم انسان[30]

اگرچه تکنیک توالی‌یابی RNA فواید شگرفی به‌همراه داشته است، ولی چالش‌های زیادی بین تکنولوژی‌های توالی‌یابی و تجزیه و تحلیل داده‌های بیوانفورماتیک وجود دارد. با این همه کتابخانه ژنومی حاصل از توالی‌یابی RNA دارای اریب بوده و کتابخانه‌های رشته‌ای خاصی که اهمیت زیادی در جهت‌یابی رونوشت‌ها دارند، هنوز به‌آسانی ایجاد نشده‌اند. به علاوه، این تکنیک مقادیر زیادی داده تولید می‌کند و به‌طور کلی طول خوانش‌ها کوتاه و دارای خطاهای توالی‌یابی هستند. به‌طور کلی این قبیل چالش‌ها مرتبط با روش‌ها و الگوریتم‌های تاثیرگذار در تحلیل داده‌های توالی‌یابی RNA است. اهمیت توالی‌های ژنوم مرجع برای مطالعات توالی یابی RNA بسیار زیاد است، زیرا آنها الگوهای لازم برای نقشه‌یابی خوانش‌ها و تفسیرهای مرتبط بر الگوریتم‌های لازم برای آنالیز بهینه نتایج را فراهم می‌کنند. میزان بیان برآوردشده برای هر ژن، به دلیل این که جزئی از کل است، تحت تاثیر بیان سایر ژن‌ها قرار می‌گیرد. به‌خصوص ژن‌هایی که دارای بیان بسیار زیادی در یک بافت خاص هستند بر برآورد بیان سایر ژن‌ها، که بیان جزئی‌تری دارند، تاثیر سوء می‌گذارند. از این روی، آماره‌های متعددی برای برطرف‌کردن این مشکل ابداع شده‌اند که علی‌رغم بهبود نسبی، هنوز نتوانسته‌اند بر این مشکل فائق بیایند. بنابراین تنها روشی که به نظر می‌رسد بتواند بر این مشکل فائق شود حذف ژن‌های با بیان بسیار زیاد در بافت مورد بررسی و آنالیز بقیه ژن‌ها است. آنالیز تعداد زیادی ژن در یک آزمایش، با این که مزیت اصلی RNA-seq است، هنوز نقشه‌یابی میلیون‌ها قطعه کوچک روی ژنوم مرجع چالش‌های حل‌نشده دارد، زیرا نقشه‌یابی اشتباه خوانش‌های حاصل از یک ژن روی ژن‌های دیگری که از نظر توالی مشابه آن هستند، بر مشکلات بررسی بیان ژن صحیح می‌افزاید. به علاوه، مشکلات اسمبلی از نو در گیاهانی که توالی‌های تکراری و مشابه متعددی در داخل ژن‌ها وجود دارد، بررسی بیان ژن را با مشکلات عدیده‌ای همراه می‌کند.

در این مقاله تکنیک توالی‌یابی RNA برای تجزیه بیان ژن مرور شده است. نمونه‌هایی از استفاده از این تکنیک در مطالعات تنش شوری در گیاهان آورده شده است. بعد از تجزیه بیان ژن با این تکنیک موارد تکمیلی زیر پیشنهاد می‌شود:

۱- بررسی ژن‌های با تفرق بیانی در سطح پروتئوم

۲- پیشگویی ساختار پروتئینی در ژن‌های با تفرق بیانی معنی‌دار


www.SID.ir

Archive of SID


۳- دست‌ورزی ژنتیکی ترنسکریپت‌های جدیداً کشف‌شده به‌منظور دستیابی به عملکرد دقیق ژن‌های جدید دخیل در تنش شوری
٤- دستکاری ژنتیکی ژن‌های دخیل در تنش شوری به‌منظور ایجاد ارقام مقاوم به شوری

## نتیجه‌گیری

تکنولوژی NGS تحول عظیمی در علوم ژنومیکس و ترنسکریپتومیکس ایجاد نموده و جنبه‌های بی‌حد و حصری از ژنوم، رونویسی و اپی‌ژنوم گونه‌ها را آشکار کرده است. این قابلیت، تعدادی از موانع مهم را برطرف کرده و حوزه‌هایی از علوم در مورد بیماری‌های انسان تا کشاورزی و علوم تکاملی را توسعه داده است. مزایای بسیاری برای توالی‌یابی از قبیل ارزان‌بودن، کارآیی بالا و زمان کم و عدم نیاز به ژنوم مرجع RNA، وجود دارد روش توالی‌یابی RNA به‌عنوان روشی بسیار قوی در مطالعات ترنسکریپتوم جایگزین روش‌هایی از قبیل ریزآرایه شده است. توالی‌یابی RNA یک روشی چندرشته‌ای است که از چندین علوم از جمله زیست‌شناسی، شیمی، فیزیک، ریاضیات و علوم کاربردی مانند بیوشیمی، آمار زیستی، بیوانفورماتیک و غیره تشکیل شده است. این ویژگی‌ها روش توالی‌یابی RNA را به یک روش ایده‌آل تبدیل کرده است تا پاسخگوی سئوالاتی مانند چگونه زندگی براساس ژن‌ها، رونوشت‌ها، پروتئین‌ها و محصولات زیستی صورت می‌پذیرد؛ چگونه موجودات زنده با محیط ارتباط برقرار می‌کنند؟ (چگونگی بیان ژن و عملکرد ترنسکریپتوم و تاثیر بر موجودات زنده)؛ چگونه سطوح مختلف سامانه‌های زیستی از مولکول، سلول و موجود به عملکرد اکوسیستم، بیوم و بیوسفر مرتبط هستند؟ این تکنیک توانایی درک بهتر پیچیدگی‌ها و دید کلی بی‌سابقه از ترانسکریپتوم در گونه‌های مختلف را برای ما فراهم می‌کند. از طریق کاربرد تکنیک RNA-Seq توان بالقوه محققین در زمینه شناسایی مسیرهای بیوسنتزی ترکیبات ثانویه و نیز شناسایی و افزایش یا کاهش در میزان بیان ژن‌های هدف، به‌منظور افزایش مقاومت به تنش‌ها از طریق کاربرد تکنیک‌های مهندسی ژنتیک افزایش خواهد یافت. بنابراین گام اول در بهبود مقاومت موجودات زنده به انواع تنش‌ها و افزایش میزان بیان ژن‌های درگیر در مسیرهای بیوسنتزی ترکیبات آلی، نقشه‌یابی رونوشت‌ها خواهد بود. بنابراین، با استفاده از تکنیک RNA-Seq کلیه اطلاعات مربوط به ژنوم و عملکردهای بیولوژیکی ژن‌ها، شناسایی شده و پس از انجام آزمایش‌های تکمیلی در انواع دست‌ورزی‌های ژنتیکی مورد هدف قرار خواهند گرفت.

**تشکر و قدردانی:** از دانشکده کشاورزی و پژوهشکده زیست‌فناوری دانشگاه ارومیه به‌خاطر فراهم‌نمودن امکانات لازم برای انجام پژوهش، تشکر و قدردانی می‌شود.

**تاییدیه‌اخلاقی:** بدین وسیله نگارندگان مقاله اعلام می‌دارند مقاله حاضر برداشت مستقیم از نوشته‌های چاپ‌شده یا در حال چاپ دیگر تحقیقات نیست.

**تعارض منافع:** هیچ گونه تعارض منافعی وجود ندارد.

**سهم نویسندگان:** معصومه شریفی‌علیشاه (نویسنده اول)، نگارنده مقاله/پژوهشگر اصلی (٤٠%)؛ رضا درویش‌زاده (نویسنده دوم)، روش‌شناس/پژوهشگر اصلی (۲۰%)؛ محمد احمدآبادی (نویسنده سوم)، پژوهشگر کمکی (۱۰%)؛ یاسر پیری‌کشتیبان (نویسنده چهارم)، پژوهشگر کمکی (۱۰%)؛ کریم حسن‌پور (نویسنده پنجم)، روش‌شناس/پژوهشگر کمکی (۲۰%)

**منابع مالی:** منابع مالی تحقیق حاضر توسط دانشگاه ارومیه تامین شده است.

## منابع

1- Berglund EC, Kiialainen A, Syvänen AC. Next-generation sequencing technologies and applications for human genetic history and forensics. Investig Genet. 2011;2:23.
2- Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genom. 2012;13(1):341.
3- Xiong M, Zhao Z, Arnold J, Yu F. Next-generation sequencing. J Biomed Biotechnol. 2010;2010:370710.
4- Liu T, Zhu S, Tang Q, Chen P, Yu Y, Tang S. De novo assembly and characterization of transcriptome using Illumina paired-end sequencing and identification of CesA gene in ramie (Boehmeria nivea L. Gaud). BMC Genom. 2013;14(1):125.
5- Morozova O, Marra MA. Applications of next-generation sequencing technologies in functional genomics. Genomics. 2008;92(5):255-64.
6- Liu J, Zhou Y, Luo C, Xiang Y, An L. De novo transcriptome sequencing of desert herbaceous Achnatherum splendens (Achnatherum) seedlings and identification of salt tolerance genes. Genes. 2016;7(4):12.
7- Mahmoudi P, Moieni A, Khayam Nekoei SM, Mardi M, Hosseini Salekdeh GH. Analysis of saffron stigma (Crocus sativus L.) transcriptome using SOAPdenovo and Trinity assembly software. Crop Biotechnol. 2014;6:35-46. [Persian]
8- Mochida K, Shinozaki K. Advances in omics and bioinformatics tools for systems analyses of plant functions. Plant Cell Physiol. 2011;52(12):2017-38.
9- Mansouri M, Naghavi MR, Alizadeh H, Mohammadinejad G, Mousavi SA, Hosseini Salekdeh G. Expression profiling of genes involved in signaling process in Aegilops tauschii under salinity stress. Iran J Filed Crop Sci. 2016;47(2):205-16. [Persian]
10- Hu L, Li H, Chen L, Lou Y, Amombo E, Fu J. RNA-seq for gene identification and transcript profiling in relation to root growth of bermudagrass (Cynodon dactylon) under salinity stress. BMC Genom. 2015;16(1):575.
11- Dang ZH, Qi Q, Zhang HR, Li HY, Wu SB, Wang YC. Identification of salt-stress-induced genes from the RNA-Seq data of Reaumuria trigyna using differential-display reverse transcription PCR. Int J Genom. 2014;2014:381501.
12- Zhou Y, Yang P, Cui F, Zhang F, Luo X, Xie J. Transcriptome analysis of salt stress responsiveness in the seedlings of Dongxiang wild rice (Oryza rufipogon


www.SID.ir

Archive of SID

23- Pease J, Sooknanan R. A rapid, directional RNA-seq library preparation workflow for Illumina® sequencing. Nat Methods. 2012;9:i-ii.
24- Chikhi R, Medvedev P. Informed and automated k-Mer size selection for genome assembly. Bioinformatics. 2014;30(1):31-7.
25- Ramirez-Gonzalez RH, Leggett RM, Waite D, Thanki A, Drou N, Caccamo M, et al. StatsDB: Platform-agnostic storage and understanding of next generation sequencing run metrics. F1000Res. 2013;2:248.
26- Haas BJ, Zody MC. Advancing RNA-seq analysis. Nat Biotechnol. 2010;28(5):421-3.
27- Fonseca NA, Rung J, Brazma A, Marioni JC. Tools for mapping high-throughput sequencing data. Bioinformatics. 2012;28(24):3169-77.
28- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511-5.
29- Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.
30- Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinform. 2010;11(1):94.
31- Al Seesi S, Tiagueu YT, Zelikovsky A, Măndoiu II. Bootstrap-based differential gene expression analysis for RNA-Seq data with and without replicates. BMC Genom. 2014;15(Suppl 8):S2.
Griff.). PLoS One. 2016;11(1):e0146242.
13- Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011;470(7333):198-203.
14- Griffith M, Walker JR, Spies NC, Ainscough BJ, Griffith OL. Informatics for RNA sequencing: A web resource for analysis on the cloud. PLoS Comput Biol. 2015;11(8):e1004393.
15- Jazayeri SM, Melgarejo Muñoz LM, Romero HM. RNA-seq: A glance at technologies and methodologies. Acta Biológica Colombiana. 2015;20(2):23-35.
16- Schliesky S, Gowik U, Weber AP, Bräutigam A. RNA-seq assembly-are we there yet?. Front Plant Sci. 2012;3:220.
17- Soneson C, Delorenzi M. A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinform. 2013;14(1):91.
18- Konczal M, Koteja P, Stuglik MT, Radwan J, Babik W. Accuracy of allele frequency estimation using pooled RNA-Seq. Mol Ecol Resour. 2014;14(2):381-92.
19- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621-8.
20- Sooknanan R, Pease J, Doyle K. Novel methods for rRNA removal and directional, ligation-free RNA-seq library preparation. Nat Methods. 2010;7(10):858.
21- Shokralla S, Spall JL, Gibson JF, Hajibabaei M. Next-generation sequencing technologies for environmental DNA research. Mol Ecol. 2012;21(8):1794-805.
22- Steinbock LJ, Radenovic A. The emergence of nanopores in next-generation sequencing. Nanotechnology. 2015;26(7):074003.


www.SID.ir