



Principles and Advanced Applications of Next-Generation Sequencing (NGS)

Technology in Life Sciences with a Focus on Cereal crops breeding

Omid Mohammadalizadeh¹, Reza Darvishzadeh², Valiollah Mohammadi³, Somaieh Soufimaliky⁴
& Danial Kahrizi⁵

¹ Former MSc student, Department of Agronomy and Plant Breeding, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran.

² Professor, Department of Plant Production and Genetics, Faculty of Agriculture, Urmia University, Urmia, Iran.

³ Associate Professor, Department of Agronomy and Plant Breeding, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran.

⁴ MSc Graduate, Institut des Sciences du Cerveau de Toulouse, France.

⁵ Professor, Department of Agricultural Biotechnology, Faculty of Agriculture, Tarbiat Modares University, Tehran, Iran.

✉ Corresponding author. E-mail: r.darvishzadeh@urmia.ac.ir

ABSTRACT

Introduction: Since the introduction of Next-Generation Sequencing (NGS) in the early 2000s, this technology has emerged as a transformative advancement in the life sciences, significantly propelling genomic, transcriptomic, epigenomic, and other related research fields. The core principles of NGS technology encompass library preparation, sequencing, and the analysis of the resulting data. By enabling the parallel sequencing of millions of DNA fragments with high accuracy, low cost, and rapid turnaround, NGS has effectively replaced older methods like Sanger sequencing. It has revolutionized our understanding of genetic complexities, genome structures, and genetic diversity through the swift and precise sequencing of entire genomes and target regions. Key applications of NGS in the life sciences include the identification and study of genes related to quantitative and qualitative traits, genetic diversity studies, population genetics, the diagnosis of genetic diseases, epidemiology, microbiome analysis, forensic science, phylogenetics, systems biology, genetic engineering, genome editing, and plant and animal breeding. However, the effective use of NGS data necessitates the development of robust computational infrastructure and advanced algorithms, as well as the expansion of researchers' knowledge regarding the bioinformatic applications and challenges associated with NGS data analysis and interpretation.

Materials and methods: The present article is a review paper, conducted through content analysis by searching for keywords related to Next-Generation Sequencing (NGS), types of NGS sequencing, NGS data analysis, and the applications of NGS in relevant articles found in online databases such as PubMed, Web of Science, Google Scholar, and Scopus.

Results: This study aims to provide a comprehensive guide for the efficient and optimal analysis of NGS data by thoroughly reviewing first-, second-, and third-generation sequencing methods, examining NGS data analysis pipelines, and exploring the broad applications of NGS in various fields, including cereal research. The first section reviews first-generation sequencing (Maxam-Gilbert and Sanger), second-generation sequencing (Illumina, ABI/SOLID, Roche/454 pyrosequencing, Ion Torrent), and third-generation sequencing (Heliscope, SMRT, and Oxford Nanopore). The second section introduces various NGS sequencing methods, such as Whole Genome Sequencing (WGS), Whole Exome Sequencing (WES), Bulk RNA-Seq, and others, and examines their analysis pathways. The subsequent discussion elaborates on the application of NGS in diverse areas, including the identification of structural genomic variations (SVs), the study of epigenetic changes, microbial population analysis, and agriculture (with an emphasis on cereal breeding). Finally, the advantages and challenges of NGS are discussed.

Conclusion: As a revolutionary technology in genomics, Next-Generation Sequencing has profoundly impacted life sciences research. The reduction in sequencing costs, coupled with increased accuracy and the development of new methods, has positioned NGS as a critical tool for a deeper understanding of genetics and the development of personalized therapeutic strategies. With ongoing advancements in this field and the integration of NGS with artificial intelligence, the future of NGS in enhancing the precision of genetic data analysis and improving therapeutic processes appears promising.

Keywords: Big Data Analysis / Bioinformatics / Sequencing Platforms / Next-Generation Sequencing (NGS).

Article Type: Review Article

Article history: Received: 30 Jan 2024, Revised: 02 Feb 2024, Accepted: 19 Feb 2024, Published online: 27 Mar 2024

Cite this article: Mohammadalizadeh, O., Darvishzadeh, R., Mohammadi, V., Soufimaliky, S. & Kahrizi, D. (2024). Principles and Advanced Applications of Next-Generation Sequencing (NGS) Technology in Life Sciences with a Focus on Cereal crops breeding. *Cereal Biotechnology and Biochemistry*, 3(1), 110-231. DOI: [10.22126/cbb.2024.11011.1080](https://doi.org/10.22126/cbb.2024.11011.1080)





اصول و کاربردهای فناوری توالی‌یابی نسل جدید (NGS) در علوم زیستی (با رویکرد به‌نژادی غلات)

امید محمدعلیزاده^۱، رضا درویش‌زاده^۲✉، ولی‌اله محمدی^۳، سمیه صوفی ملکی^۴ و دانیال کهریزی^۵

^۱ دانش آموخته کارشناسی ارشد، گروه زراعت و اصلاح نباتات، دانشکده علوم و مهندسی کشاورزی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.
^۲ استاد، گروه تولید و ژنتیک گیاهی، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران.
^۳ دانشیار، گروه زراعت و اصلاح نباتات، دانشکده علوم و مهندسی کشاورزی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.
^۴ دانش آموخته کارشناسی ارشد، انستیتو علوم اعصاب تولوز، فرانسه.
^۵ استاد، گروه بیوتکنولوژی کشاورزی، دانشکده کشاورزی، دانشگاه تربیت مدرس، تهران، ایران.
✉ نویسنده مسئول. رایانامه: r.darvishzadeh@urmia.ac.ir

چکیده

مقدمه: از زمان معرفی توالی‌یابی نسل جدید (NGS)، در اوایل دهه ۲۰۰۰، این فناوری به عنوان یکی از تحولات بنیادین در علوم زیستی، باعث پیشرفت چشمگیر در تحقیقات ژنومیک، ترنسکریپتوم، اپیژنوم و... شده است. اصول فناوری NGS شامل مباحث مربوط به تهیه کتابخانه، توالی‌یابی و تحلیل داده‌های حاصل از آن است. این فناوری با توالی‌یابی میلیون‌ها قطعه DNA به صورت موازی و با دقت بالا و هزینه پایین و همچنین تولید حجم زیادی از داده‌های ژنومی در زمان کوتاه، توانسته است جایگزین روش‌های قدیمی‌تری مانند توالی‌یابی سنگر شود و با توالی‌یابی سریع، دقیق و کامل ژنوم‌ها و نواحی هدف انقلابی بزرگ در درک پیچیدگی‌های ژنتیکی، ساختار ژنوم و تعیین تنوع ژنتیکی ایجاد کند. از مهم‌ترین کاربردهای NGS در علوم زیستی می‌توان به شناسایی و مطالعه ژن‌های مرتبط با صفات کمی و کیفی، مطالعات تنوع ژنتیکی و ژنتیک جمعیت، تشخیص بیماری‌های ژنتیکی، اپیدمیولوژی، میکروبیوم‌شناسی، پزشکی قانونی، فیلوژنتیک، زیست‌شناسی سامانه‌ای، مهندسی ژنتیک و ویرایش ژنوم و اصلاح نبات و دام اشاره کرد. باین حال، استفاده مؤثر از داده‌های NGS مستلزم توسعه زیرساخت‌های محاسباتی قوی و الگوریتم‌های پیشرفته و همچنین گسترش اطلاعات محققان در رابطه با کاربردها و چالش‌های بیوانفورماتیک مرتبط با داده‌های NGS، برای تحلیل و تفسیر این حجم عظیم از اطلاعات است.

مواد و روش‌ها: مقاله حاضر یک مقاله مروری می‌باشد که به شیوه تحلیل محتوا (Content analysis) با جستجوی کلید واژه‌های توالی‌یابی نسل جدید (NGS)، انواع توالی‌یابی NGS، تجزیه و تحلیل داده‌های NGS، کاربردهای توالی‌یابی NGS، در مقاله‌های مرتبط در پایگاه‌های اینترنتی Google Scholar و Scopus به دست آمده است.

یافته‌ها: این مطالعه قصد دارد با مرور تفصیلی توالی‌یابی‌های نسل اول، دوم و سوم، بررسی مسیر تجزیه و تحلیل داده‌های NGS و کاربردهای گسترده NGS در زمینه‌های مختلف از جمله تحقیقات غلات، راهنمایی تقریباً کاملی برای تجزیه و تحلیل کارآمد و بهینه داده‌های حاصل از توالی‌یابی نسل جدید ارائه نماید. به این منظور در بخش اول به مرور توالی‌یابی‌های نسل اول (ماکسام-گیلبرت و سنگر)، نسل دوم (Illumina, ABI/SOLID, Ion Torrent) و نسل سوم (Oxford Nanopore, SMRT, Heliscope) پرداخته شد. سپس در بخش دوم انواع توالی‌یابی‌های NGS مانند: Whole genome sequencing; WGS, Bulk RNA-seq, Whole exome sequencing; WES, sequencing; WGS و سایر روش‌ها معرفی و مسیر تجزیه و تحلیل آنها بررسی شده‌اند و در ادامه کاربرد توالی‌یابی نسل جدید در حوزه‌های مختلف مانند شناسایی تنوع ساختمانی ژنومی (SVs)، مطالعه تغییرات اپی‌ژنتیکی، تجزیه و تحلیل جمعیت میکروبی، کشاورزی (با تأکید بر به‌نژادی غلات) توضیح داده شد. در نهایت مزایا و چالش‌های پیشروی توالی‌یابی نسل جدید بیان گردید.

نتیجه‌گیری: توالی‌یابی نسل جدید به عنوان یک فناوری انقلابی در ژنومیک، تأثیر بسزایی در تحقیقات علوم زیستی داشته است. کاهش هزینه‌ها و افزایش دقت توالی‌یابی به همراه توسعه روش‌های جدید، باعث شده است تا NGS به ابزاری کلیدی برای درک بهتر ژنتیک و توسعه راهبردهای درمانی شخصی‌سازی شده تبدیل شود. با پیشرفت‌های مداوم در این حوزه و ترکیب این فناوری با هوش مصنوعی، آینده‌ی NGS در تحلیل دقیق‌تر داده‌های ژنتیکی و بهبود فرایندهای درمانی بسیار روشن به نظر می‌رسد.

واژه‌های کلیدی: آنالیز داده‌های حجیم؛ بیوانفورماتیک؛ پلتفرم‌های توالی‌یابی؛ توالی‌یابی نسل جدید (NGS).

نوع مقاله: مقاله مروری

نوع مقاله: دریافت: ۱۴۰۲/۱۱/۳۰ **اصلاح:** ۱۴۰۲/۱۱/۱۳ **پذیرش:** ۱۴۰۲/۱۱/۳۰ **انتشار آنلاین:** ۱۴۰۳/۰۱/۰۸

استناد: محمدعلیزاده، ا.، درویش‌زاده، ر.، محمدی، و.، صوفی ملکی، س. و کهریزی، د. (۱۴۰۳). اصول و کاربردهای فناوری توالی‌یابی نسل جدید (NGS) در علوم

زیستی (با رویکرد به‌نژادی غلات). *بیوتکنولوژی و بیوشیمی غلات*, ۲(۳), ۱۱۰-۱۳۱. DOI: [10.22126/cbb.2024.11011.1080](https://doi.org/10.22126/cbb.2024.11011.1080)



مقدمه

شخصی‌سازی شده استفاده می‌شود (Behjati & Tarpey, 2013). در تحقیقات میکروبیولوژی، این فناوری برای مطالعه میکروبیوم‌های انسانی و فهم تأثیر آن‌ها بر سلامت و بیماری به کار می‌رود (Human Microbiome Project Consortium, 2012). علاوه بر این، NGS در زیست‌شناسی تکاملی برای بررسی تغییرات ژنتیکی در طول زمان و درک روابط تکاملی بین گونه‌ها استفاده می‌شود (Metzker, 2010). می‌توان اذعان نمود که معرفی این فناوری منجر به افزایش چشمگیر سرعت مطالعات در حوزه‌های مختلف تحقیقات علمی شده است. با وجود مزایای متعدد، تحلیل داده‌های NGS با چالش‌هایی همراه است. حجم بالای داده‌ها نیازمند ابزارهای محاسباتی و بیوانفورماتیکی پیشرفته برای پردازش، تجزیه و تحلیل و تفسیر صحیح است (Nielsen *et al.*, 2011). از این رو، توسعه نرم‌افزارهای کارآمد و روش‌های جدید برای مدیریت و تحلیل داده‌های NGS امری ضروری است. آینده NGS با پیشرفت‌های نوین در فناوری‌های توالی‌یابی و تحلیل داده‌ها روشن به نظر می‌رسد. کاهش هزینه‌ها و افزایش دقت توالی‌یابی به همراه توسعه روش‌های جدید، انتظار می‌رود که NGS در تحقیقات علمی و کاربردهای بالینی به‌طور گسترده‌تری مورد استفاده قرار گیرد (Mardis, 2017). همچنین، ترکیب NGS با فناوری‌های دیگری مانند هوش مصنوعی می‌تواند به بهبود فرایند تفسیر داده‌ها و توسعه درمان‌های جدید کمک کند (Libbrecht & Noble, 2015).

توالی‌یابی نسل جدید (NGS)¹ به عنوان یک فناوری پیشرفته در زمینه ژنومیک، به‌طور قابل توجهی توانایی‌های پژوهشگران را در مطالعه و تحلیل ژنوم‌های پیچیده افزایش داده است. این فناوری که در اوایل دهه ۲۰۰۰ معرفی شد، امکان توالی‌یابی سریع و با هزینه کمتر DNA و RNA را فراهم کرده است (Mardis, 2013). فناوری NGS توانسته است به سرعت جایگزین روش‌های قدیمی‌تر مانند توالی‌یابی سانگر (Sanger sequencing) شود، زیرا قادر است میلیون‌ها قطعه DNA را به صورت موازی توالی‌یابی کند (Schuster, 2008). این قابلیت باعث شده تا NGS به یکی از ابزارهای اصلی در تحقیقات ژنتیکی و پزشکی تبدیل شود. از زمان معرفی، فناوری NGS پیشرفت‌های چشمگیری را تجربه کرده است. پلتفرم‌های مختلفی از جمله Illumina، Ion Torrent و Pacific Biosciences هر یک با ویژگی‌های خاص خود، امکان توالی‌یابی با دقت و سرعت بالا را فراهم کرده‌اند (Mardis, 2008). این پلتفرم‌ها با استفاده از روش‌های شیمیایی و فیزیکی متفاوت، به محققان امکان می‌دهند تا بر اساس نیازهای خاص پروژه‌های خود، بهترین گزینه را انتخاب کنند.

توالی‌یابی نسل جدید کاربردهای گسترده‌ای در زمینه‌های مختلف علمی دارد. بطور مثال در پزشکی، NGS برای تشخیص بیماری‌های ژنتیکی، شناسایی جهش‌های سرطان و تدوین راهبردهای درمان

¹ Next generation sequencing

مطالعات مقایسه‌ای کمک می‌کنند. روش‌های توالی‌یابی نسل جدید (NGS)⁵ با توان عملیاتی بالا⁶، درجه‌ای جدید در تحقیقات زیست پزشکی باز کرده‌اند. همچنین، وجود دستگاه‌های توالی‌یابی نسل جدید انقلاب بزرگی در ژنتیک ایجاد کرده است (Church & Gilbert, 1984). توالی‌یابی فرایندی است که از طریق آن ترتیب قرار گرفتن نوکلئوتیدها در یک رشته DNA تعیین می‌شود. از زمان کشف ساختار DNA تا به امروز، تلاش‌های زیادی برای توالی‌یابی کامل ژنوم جانداران مختلف صورت گرفته است. اولین توالی‌های DNA در اوایل دهه ۱۹۷۰ توسط پژوهشگران با استفاده از روش‌های بسیار دشوار و پرهزینه بر اساس کروماتوگرافی دو بعدی⁷ به دست آمد و به تدریج با پیشرفت روش‌های تجزیه و تحلیل خودکار مبتنی بر رنگ، انجام توالی‌یابی DNA آسان‌تر و سریع‌تر شد (Olsvik *et al.*, 1993).

یکی از نخستین نمونه‌های تعیین توالی کامل، توالی‌یابی mRNA مربوط به ژنوم باکتری *MS2* بود که توسط والتر فیرز⁸ و همکارانش در دانشگاه گنت⁹ بین سال‌های ۱۹۷۲ و ۱۹۷۶ انجام شد. پیش از پیشرفت سریع روش‌های تعیین توالی DNA، در اوایل دهه ۱۹۷۰، روش‌های بسیار پرهزینه‌تری توسط فردریک سنگر در دانشگاه کمبریج¹⁰ انگلستان و همچنین والتر گیلبرت¹¹ و

در مجموع توالی‌یابی نسل جدید به عنوان یک فناوری انقلابی در ژنومیک، توانسته است تأثیر بسزایی در تحقیقات علمی و پزشکی داشته باشد. با ادامه پیشرفت‌ها و کاهش هزینه‌ها، NGS به ابزار کلیدی برای درک بهتر ژنتیک و توسعه راهبردهای درمانی شخصی‌سازی شده تبدیل خواهد شد. در این مطالعه ابتدا توالی‌یابی‌های نسل اول، دوم و سوم به تفصیل مرور خواهند شد سپس مسیر تجزیه و تحلیل داده انواع روش‌های توالی‌یابی نسل دوم بررسی و در نهایت کاربرد توالی‌یابی نسل جدید مورد بحث قرار خواهد گرفت.

نسل اول توالی‌یابی

بر اساس پژوهش‌های آوری و همکاران (Avery *et al.*, 1944)، مشخص شد که مولکول‌های دئوکسی‌ریبونوکلئیک اسید (DNA)² یا ماده وراثتی سلول‌ها هستند. ساختار هلیکس دو رشته‌ای DNA که شامل چهار نوکلئوتید است، در سال ۱۹۵۳ توسط واتسون و کریک تعیین شد و این کشف به شکل‌گیری نظریه سنترال دوگما³ در زیست‌شناسی مولکولی منجر شد (Watson & Crick, 1953; Crick *et al.*, 1970). در بیشتر موارد، DNA ژنومی به صورت گونه‌ای و فردی تعریف می‌شود و از توالی‌های آن برای بررسی عملکرد سلول و رمزگشایی اسرار موجودات زنده استفاده می‌شود. روش‌های توالی‌یابی به زیست‌شناسان در زمینه‌های بسیاری مانند هسانه‌سازی⁴، یافتن ژن‌های پاتوژن و

⁵ Next generation sequencing

⁶ High throughput

⁷ Two-dimensional chromatography

⁸ Walter Fiers

⁹ Ghent

¹⁰ Cambridge

¹¹ Walter Gilbert

² Deoxyribonucleic acid

³ Central dogma of molecular biology

⁴ Cloning

سال پس از مقاله پیشرو سنگر و کولسون^{۱۷} درباره تعیین توالی رشته‌های مثبت-منفی^{۱۸} صورت گرفت، اما به دلیل قابلیت استفاده مستقیم از DNA خالص شده، به سرعت مورد توجه قرار گرفت.

روش اولیه سنگر نیازمند کلون کردن هر نقطه شروع برای تولید DNA تک رشته‌ای بود. با این حال، با پیشرفت روش ختم زنجیره‌سازی^{۱۹}، روش ماکسام-گیلبرت به تدریج محبوبیت خود را از دست داد. دلایل این امر شامل پیچیدگی روش، استفاده زیاد از مواد شیمیایی خطرناک، دشواری کاربرد در حجم بالای نمونه و عدم امکان استفاده در کیت^{۲۰}های استاندارد زیست‌شناسی بود. این روش مستلزم نشان‌دار کردن انتهای ۵' DNA با مواد پرتوزا (به‌طور معمول از طریق واکنش کیناز و با استفاده از DNA (Gamma-³²P ATP) است. پس از آن، قطعه DNA برای تعیین توالی خالص‌سازی می‌شود. سپس با استفاده از تیمار شیمیایی، شکست در بخش کوچکی از یک یا دو نوکلئوتید از مجموع چهار نوکلئوتید در هر یک از ۴ واکنش (گوانین^{۲۱}، آدنین^{۲۲}+گوانین، سیتوزین^{۲۳}، سیتوزین+تیمین^{۲۴}) ایجاد می‌شود. به عنوان مثال: فرمیک اسید برای حذف پورین‌ها (A+G) از ساختار DNA استفاده می‌شود. این فرایند را پورین‌زدایی^{۲۵} می‌نامند.

آلن ماکسام^{۱۲} در دانشگاه هاروارد^{۱۳} به کار گرفته شد (Olsvik et al., 1976; Jou et al., 1972;) Pettersson et al., 2009; Maxam et al., 1977) به عنوان مثال، در سال ۱۹۷۳ گیلبرت و ماکسام با استفاده از روش شناخته شده تجزیه و تحلیل نقطه‌ای^{۱۴} تعیین توالی ۲۴ جفت باز را گزارش دادند. این روش به عنوان روش برش شیمیایی^{۱۵} شناخته شد و یک روش پرمشقت و مبتنی بر استفاده از مواد پرتوزا و برش شیمیایی بود (Gilbert et al., 1973). در سال ۱۹۷۷، سنگر و همکارانش روش تعیین توالی بر مبنای ختم زنجیره^{۱۶} را پیشنهاد کردند که به دلیل آسان بودن نسبی و قابل اطمینان بودن، به روش انتخابی تعیین توالی تبدیل شد. اساس این روش بر مبنای تعیین توالی بازهای DNA بر اساس ختم سنتز DNA در حال تکثیر در نقاط مختلف DNA و جداسازی قطعات کوچک DNA است (Maxam et al., 1977; Sanger et al., 1975;) (Sanger et al., 1977).

توالی‌یابی ماکسام-گیلبرت

در فاصله سال‌های ۱۹۷۶ تا ۱۹۷۷، آلن ماکسام و والتر گیلبرت روشی نوین برای تعیین توالی DNA ابداع کردند. این روش بر پایه تغییرات شیمیایی DNA و ایجاد شکاف در بازهای خاص استوار بود. گرچه انتشار این روش دو

¹⁷ Ground-breaking

¹⁸ Plus-minus sequencing

¹⁹ Chain termination method

²⁰ Kit

²¹ Guanine (G)

²² Adenine (A)

²³ Cytosine (C)

²⁴ Thymine (T)

²⁵ Depurinated

¹² Allan Maxam

¹³ Harvard

¹⁴ Wandering spot analysis

¹⁵ Chemical cleavage

¹⁶ Chain-termination

اتصال پروتئین‌ها به DNA مورد استفاده قرار می‌گیرد (Maxam *et al.*, 1977). استنتاج رشته: قطعات حاصل از چهار واکنش مختلف روی ژل آکريل‌آمید جداسازی و الکتروفورز می‌شوند تا بر اساس اندازه تفکیک شوند. برای نمایان ساختن این بخش‌ها، ژل در معرض پرتو X قرار می‌گیرد تا فرایند اتورادیوگرافی^{۲۹} انجام شود. این کار باعث ظهور نوارهای تیره‌ای می‌شود که نشان‌دهنده موقعیت مولکول‌های پرتوزا شده هستند. با بررسی وجود یا عدم وجود برخی قطعات، می‌توان توالی رشته را استنباط کرد. به عنوان مثال، اگر نوری هم در واکنش C+T و هم در واکنش C مشاهده شود، این بدان معناست که در آن موقعیت، باز C قرار دارد. اگر نوار فقط در واکنش C+T دیده شود، نشان‌دهنده حضور باز T در آن مکان است. به‌طور مشابه، می‌توان استدلال‌هایی برای تشخیص بازهای A و G نیز ارائه داد. این روش امکان بازسازی توالی کامل رشته DNA را فراهم می‌کند (Maxam *et al.*, 1977).

برای متیله کردن گوانین‌ها و تا حدودی آدنین‌ها، از دی‌متیل سولفات استفاده می‌شود. هیدرازین برای متیله کردن پیریمیدین‌ها (سیتوزین+ تیمین) به کار می‌رود. افزودن نمک سدیم کلراید به واکنش هیدروژنی باعث می‌شود که متیله شدن تیمین در واکنش‌هایی که فقط حاوی سیتوزین هستند، مهار شود (Maxam *et al.*, 1977). در مرحله بعد، DNA که تغییر شکل یافته است، توسط پیریدین^{۲۶} گرم در محل نوکلئوتید تغییر یافته شکسته می‌شود. میزان مواد شیمیایی که موجب این تغییر می‌شوند به گونه‌ای تنظیم می‌شود که به‌طور میانگین هر مولکول DNA تنها یک تغییر را تجربه کند. این فرایند منجر به تولید مجموعه‌ای از قطعات نشان‌دار می‌شود که فاصله‌های متفاوتی بین انتهای نشان‌دار و اولین نقطه برش در هر مولکول دارند. قطعات حاصل از چهار واکنش مختلف در کنار هم قرار داده شده و برای جداسازی بر اساس اندازه، روی ژل آکريل‌آمید تغییر ماهیت یافته^{۲۷} الکتروفورز می‌شوند. برای مشاهده این قطعات، ژل و فیلم رادیولوژی کنار یکدیگر قرار می‌گیرند. فیلم در نقاط مختلف روشن و تاریک خواهد بود، که نوارهای تیره نشان‌دهنده قطعات نشان‌دار DNA هستند و از طریق آن‌ها توالی DNA استنتاج می‌شود (شکل ۱).

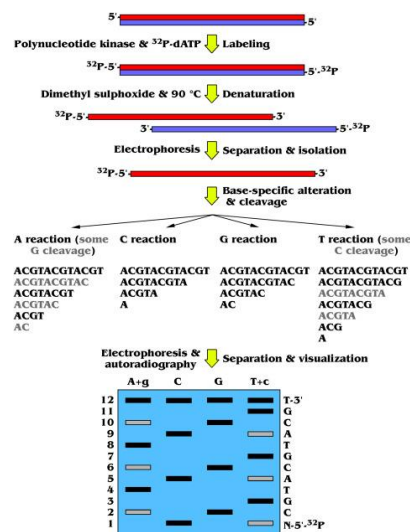
این روش که به عنوان تعیین توالی شیمیایی شناخته می‌شود، بعدها منجر به ابداع روش سنجش تداخلی متیلاسیون^{۲۸} شد که این روش برای شناسایی مکان‌های

²⁶ Piperidine

²⁷ Denaturing acrylamide

²⁸ Methylation interference assay

²⁹ Autoradiograph



شکل ۱- تعیین توالی مکسام-گیلبرت. این روش تخریب شیمیایی است که اجازه می‌دهد DNA دو رشته‌ای (dsDNA) بدون نیاز به مراحل همسانه‌سازی درون جاندار (*in vivo*) تعیین توالی شود.

Figure 1. Maxam-Gilbert sequencing. It is a chemical degradation method that allows double-stranded DNA (dsDNA) to be sequenced without the need for *in vivo* homogenization steps.

یافته (ddNTPs) برای پایان دادن به تولید شدن رشته DNA است. در این روش، ddNTPs به‌طور معمول با S35 پرتوزا نشان‌دار می‌شوند و روی ژل آکریل‌آمید الکتروفورز می‌گردند. سپس ژل از صفحه شیشه‌ای جدا شده و یک فیلم رادیولوژی روی آن قرار می‌گیرد. پس از اثرگذاری مواد پرتوزا بر فیلم، نتایج تفسیر می‌شوند (Sanger *et al.*, 1975; Sanger *et al.*, 1977). بعدها، با بهره‌گیری از ddNTP نشان‌دار شده با فلورسانس برای تعیین توالی و الکتروفورز روی ژل آکریل‌آمید در دستگاه، فرایند تعیین توالی به صورت نیمه‌خودکار درآمد. در این روش، نمونه DNA برای واکنش تعیین توالی به چهار گروه مجزا تقسیم می‌شود و هر گروه شامل تمامی چهار نوع دئوکسی‌نوکلئوتید استاندارد (dTTP, dATP,) و DNA پلی‌مراز است. سپس به هر لوله

توالی‌یابی ختم زنجیره‌سازی^{۳۰}

روش توالی‌یابی ختم زنجیره‌سازی، که به نام روش سنگر نیز شناخته می‌شود، به دلیل کارایی بالاتر، استفاده کمتر از مواد شیمیایی سمی و سطح پایین‌تر پرتوزایی در مقایسه با روش مکسام و گیلبرت، به سرعت به روش ترجیحی تبدیل شد. اساس روش سنگر^{۳۱} بر استفاده از دی‌دئوکسی‌نوکلئوزید تری فسفات (ddNTPs)^{۳۲} به عنوان عامل خاتمه‌دهنده زنجیره DNA استوار است (Sanger *et al.*, 1975; Sanger *et al.*, 1977). روش کلاسیک ختم زنجیره‌سازی نیازمند DNA تک رشته‌ای به عنوان الگو، آغازگر^{۳۳} DNA پلیمراز، دئوکسی‌نوکلئوتید تری فسفات (dNTPs) و نوکلئوتیدهای تغییر

³⁰ Chain-termination sequencing

³¹ Sanger

³² Di-deoxynucleotide

³³ Primer

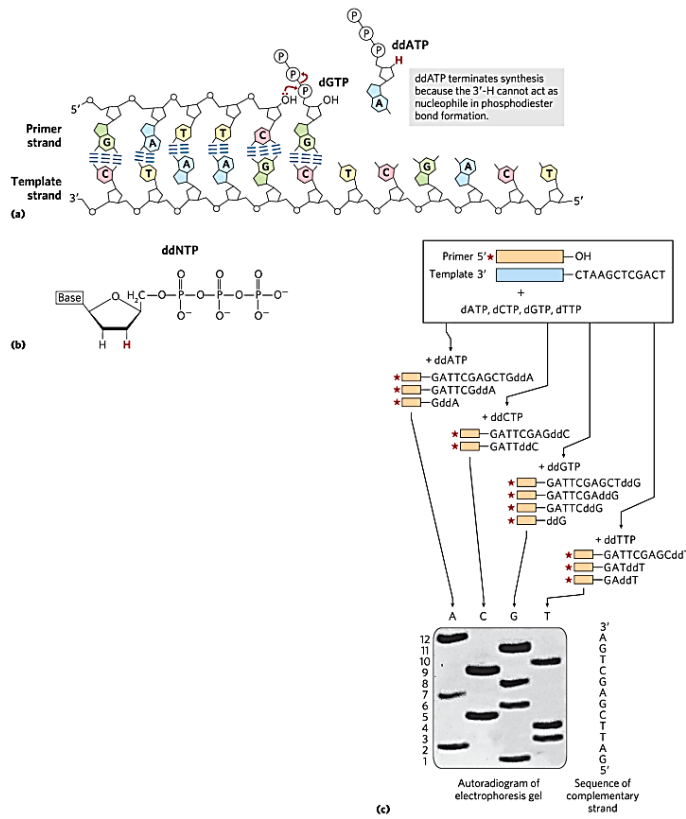
واکنش، تنها یکی از چهار دی‌نوکلئوسید نوکلئوتید (ddTTP, ddATP, ddGTP, ddCPT) که نقش خاتمه‌دهنده زنجیره را دارند، اضافه می‌شود. این نوکلئوتیدها فاقد گروه 3'OH ضروری برای تشکیل پیوند فسفودی‌استر بین دو نوکلئوتید هستند. نکته کلیدی در این روش، استفاده بسیار اندک از هر یک از ddNTPها در هر لوله است. این امر باعث می‌شود زنجیره در حال ساخت به‌طور تصادفی ddNTP را به جای dNTP استفاده کرده و در نتیجه تکثیر متوقف شود (Sanger et al., 1977; Sanger et al., 1977). بنابراین، تکثیر در نقاط مختلفی که مثلاً باز dATP (در لوله حاوی ddATP) قرار دارد، متوقف می‌شود. این فرایند منجر به تشکیل قطعات DNA با طول‌های متفاوت می‌گردد. قطعات DNA نشان‌دار و تازه ساخته شده توسط حرارت واسرشت^{۳۴} می‌شوند و بر اساس اندازه (با دقت تفکیک تک نوکلئوتیدی) روی ژل آکریل‌آمید-اوره تغییر شکل یافته، الکتروفورز می‌شوند. هر یک از چهار واکنش به‌طور جداگانه در یکی از چهار چاهک A, C, G و T الکتروفورز می‌شود. سپس باندهای DNA از طریق تصویربرداری خودکار رادیوایزوتوپی^{۳۵} قابل مشاهده شده و توالی DNA مستقیماً از روی تصویر ژل یا فیلم رادیولوژی خوانده می‌شود. حضور هر خط، قطعه یا باند تیره در یک خط عمودی نشان‌دهنده یک بخش از توالی DNA است که نتیجه ختم زنجیره پس از تلفیق یک دی‌نوکلئوسید نوکلئوتید ddATP, ddCTP, ddGTP یا

ddTTP به زنجیره است. در این روش، موقعیت نسبی باندهای مختلف در بین چهار مسیر عمودی (از پایین به بالا) برای خواندن توالی DNA استفاده می‌شود (Sanger et al., 1977; Sanger et al., 1977). به این معنی که پایین‌ترین باند مربوط به همان بازی است که در مسیر آن دیده شده است. بطور مثال اگر پایین‌ترین باند مربوط به مسیر رشته A باشد نشان می‌دهد که شروع با این باز است و اگر دومین باز نیز بالاتر از این باز باشد نشان می‌دهد که دومین باز نیز A است. اگر باز بعدی در خانه مربوط به T باشد باز بعدی T خواهد بود. بنابراین، برای سه باز اول توالی AAT خواهیم داشت. این روند خواندن به سمت بالا بین چهار ستون ادامه می‌یابد تا توالی تا حد امکان مشخص شود (شکل ۲). از طرفی دیگر، برای تعیین توالی در روش ختم زنجیره می‌توان از نوکلئوتیدهای نشان‌دار شده با فسفر پرتوزا یا از یک آغازگر نشان‌دار در انتهای 5' با رنگ فلورسانس استفاده کرد. استفاده از آغازگر دارای رنگ فلورسانس در توالی-یابی، خوانش در سامانه نوری را تسریع کرده و امکان انجام خودکار فرایند را فراهم می‌کند، که این امر این روش را به عنوان روشی مقرون به صرفه معرفی می‌کند. پیشرفت‌های بعدی توسط لروی هوود^{۳۶} و همکارانش با استفاده از ddNTPs نشان‌دار و آغازگرها، زمینه را برای اتوماسیون این روش توسط دستگاه‌ها و تعیین توالی DNA در مقیاس وسیع فراهم کرد (Sanger et al., 1977; Sanger et al., 1977).

³⁴ DNA denaturation

³⁵ Automated radioisotope imaging

³⁶ Leroy Hood



(1993).

شکل ۲- توالی‌یابی DNA با روش سنگر
Figure 2. DNA sequencing by Sanger method

جداگانه بودند و جداسازی با استفاده از ژل آکریل آمید انجام می‌شد (Ansorge *et al.*, 1993).

در اوایل دهه ۹۰ میلادی، شرکت Applied Biosystems آمریکا دستگاهی را به بازار عرضه کرد که با استفاده از چهار رنگ فلورسنتی و لیزرهای مختلف با طول موج‌های متفاوت، تعیین توالی می‌کرد. در این روش، به جای چهار چاهک برای یک نمونه، از یک چاهک

استفاده می‌شد؛ بنابراین ظرفیت تعیین توالی چهار برابر بیشتر از روش ALF بود. این دستگاه نیز به ژل آکریل

در ابتدا، این روش از مواد پرتوزا استفاده می‌کرد. بعدها، شرکت Pharmacia سوئد در دستگاه تعیین توالی نیمه اتوماتیک ³⁷ALF از از تک رنگ فلوروسین³⁸ و در دستگاه ALF Express از رنگ ³⁹CY5 بهره گرفت. هر دو دستگاه نیازمند ریختن نمونه‌ها در چهار چاهک

³⁷ ALF: Amplification of linearly-enriched fragments

³⁸ Fluorescein

³⁹ Cyanine, Cy5 is fluorescent in the far-red region (~650 excitation, 670 nm emission)

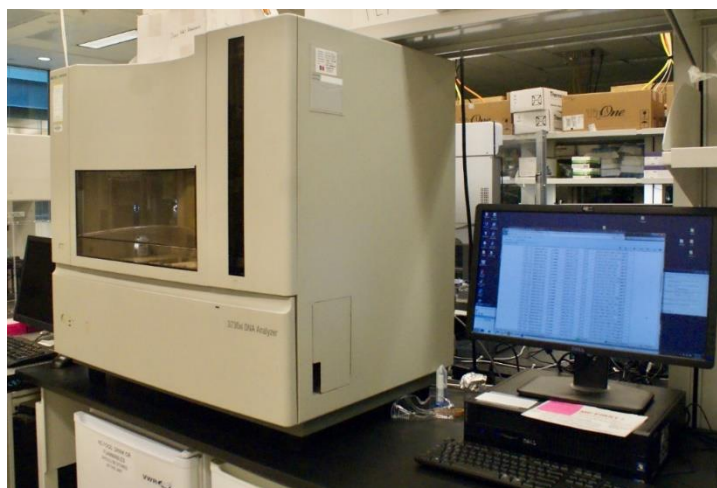
نسبت بلندتری را توالی‌یابی کند. با این شیوه سنگر توانست در ۱۹۷۷ ژنوم ویروس ϕ x174 را به‌طور کامل تعیین توالی کند (Sanger *et al.*, 1975; Sanger *et al.*, 1977).

این رویکرد در اولین ماشین خودکار توالی‌یابی بکار گرفته شد که توسط آن توالی ژنوم چندین جاندار ساده، این بار فراتر از ویروس‌ها به‌طور کامل تعیین شد. ماشین توالی-یابی Applied Biosystems capillary tube مدل ABI3730X) اولین بار مورد استفاده قرار گرفت. صحت درستی توکلوتیدهای خوانده شده با افزایش طول رشته کاهش می‌یابد؛ بنابراین برای تعیین توالی یک ژنوم به‌طور کامل ترفندهای خاصی بکار گرفته شد.

آمید وابسته بود (Ansorge *et al.*, 1993). این روش‌ها تا حد زیادی تعیین توالی DNA را ساده‌تر کردند. به عنوان مثال، در حال حاضر کیت‌هایی بر اساس ختم زنجیره به صورت تجاری در دسترس است که به صورت آماده حاوی معرف‌های مورد نیاز برای تعیین توالی هستند. با این حال، این روش کاستی‌هایی مانند غیر اختصاصی بودن محل اتصال آغازگر به DNA یا ساختار دوم DNA دارد که می‌تواند صحت نتایج به دست آمده از توالی DNA را تحت تأثیر قرار دهد (Ansorge *et al.*,). همانطور که بیان شد شیوه سنگر⁴⁰ مبتنی بر استفاده از آنزیم DNA پلی‌مراز و فرایندی مشابه فرایند طبیعی همانندسازی DNA از یک رشته الگو است. سنگر ماده خاصی بنام دی‌دئوکسی نوکلئوتید⁴¹ را شناسایی کرد که بکارگیری آن در محیط همانندسازی باعث می‌شود ادامه همانندسازی DNA پس از جفت شدن این ماده با نوکلئوتید متناظرش متوقف شود. با هدف استفاده از حسگرهای نوری برای خواندن توالی DNA در ترکیب دی‌دئوکسی نوکلئوتید متناظرش با هر نوع نوکلئوتید عنصر فلورسانس با طول موج معینی مورد استفاده قرار می‌گیرد و از این طریق نوع نوکلئوتید انتهای رشته همانندسازی شده قابل تشخیص خواهد بود. روش مربوط به سنگر (۱۹۷۷) بر خلاف شیوه ماکسام و گیلبرت که مبتنی بر انجام تعداد زیادی واکنش شیمیایی روی DNA است و از این رو به مراحل متعدد خالص‌سازی نیاز دارد؛ با طی مراحل آزمایشگاهی ساده‌تر می‌تواند رشته‌های به

⁴⁰ Sanger

⁴¹ Di-deoxynucleotide



شکل ۳- دستگاه توالی‌یاب Applied biosystems ABI3730X-DNA analyzer
Figure 3. Applied biosystems ABI3730X-DNA analyzer sequencing device

رنگ‌های فلورسنت متفاوت نشانه‌گذاری شده‌اند، آماده می‌شود. ddNTP فاقد گروه ۳'-OH هستند که برای تشکیل پیوند فسفو دی استر ضروری است، بنابراین وقتی به زنجیره در حال سنتز اضافه می‌شوند، سنتز زنجیره متوقف می‌شود (Smith *et al.*, 1986). در هر مرحله از سنتز، احتمال اضافه شدن یک ddNTP به جای dNTP وجود دارد که این فرایند به‌طور تصادفی منجر به تولید مجموعه‌ای از قطعات DNA با طول‌های مختلف می‌شود که هر کدام به یک ddNTP خاتمه یافته‌اند (Prober *et al.*, 1987). در ادامه فرایند، این قطعات DNA با استفاده از الکتروفورز موئینه‌ای (Capillary electrophoresis) جداسازی می‌شوند. در این مرحله، قطعات DNA از طریق یک لوله نازک که در آن میدان الکتریکی اعمال می‌شود، حرکت می‌کنند. قطعات کوچکتر سریع‌تر حرکت کرده و زودتر به انتهای لوله می‌رسند. در انتهای لوله، یک لیزر به رنگ‌های فلورسنت متصل به -

ختم زنجیرسازی با استفاده از رنگ Dye-termination

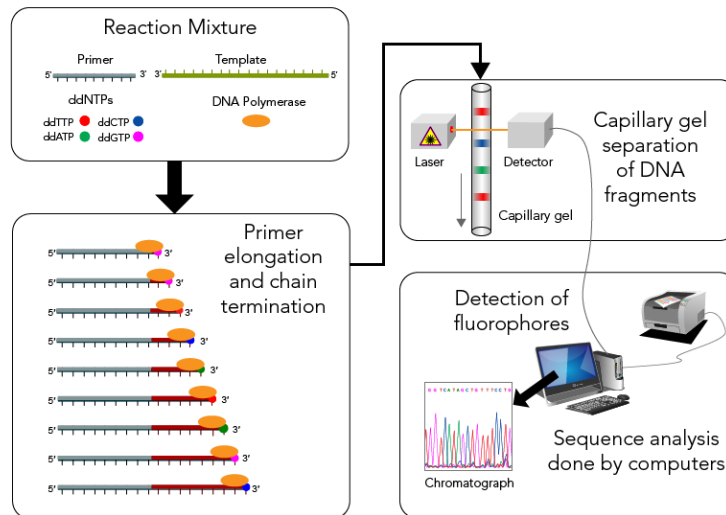
روش توالی‌یابی ختم زنجیره‌سازی با استفاده از رنگ^{۴۲} یکی از روش‌های پیشرفته برای تعیین توالی DNA است که به نام روش سنگر نیز شناخته می‌شود. این روش توسط فردریک سنگر طراحی شده و به‌طور گسترده‌ای برای شناسایی دقیق توالی بازهای نیتروژنی در مولکول‌های DNA به کار می‌رود (Sanger, Nicklen, & Coulson, 1977). این فرایند شامل چند مرحله کلیدی است که در نهایت به شناسایی دقیق توالی DNA منجر می‌شود. در ابتدا، DNA مورد نظر از نمونه استخراج و خالص‌سازی می‌شود تا به عنوان الگو برای واکنش زنجیره‌سازی مورد استفاده قرار گیرد. سپس، مخلوط واکنش شامل DNA الگو، آغازگر، DNA پلی‌مراز، دئوکسی‌نوکلئوتیدهای طبیعی (dNTPs) و دئوکسی‌نوکلئوتیدهای خاتمه‌دهنده (ddNTPs) که با

⁴² Dye-terminator sequencing

منجر به تولید محصولات با طول‌های متفاوت می‌شود که هرکدام در یک باز نیتروژنی خاص خاتمه یافته‌اند. سپس این محصولات با استفاده از الکتروفورز در ژل پلی‌اکریل‌آمید جداسازی می‌شوند و نتایج با استفاده از اتورادیوگرافی یا فلورسانس تحلیل می‌شوند (Maxam & Gilbert, 1977). این روش به دلیل زمان‌بر بودن و پیچیدگی‌اش، نیاز به استفاده از مواد پرتوزا خطرناک داشت. روش Dye-termination که به عنوان نسخه بهبودیافته روش سنگر شناخته می‌شود، با استفاده از ddNTP‌های نشان‌گذاری شده با رنگ‌های فلورسنت، این مشکلات را برطرف کرده و امکان استفاده از فناوری‌های جدیدتری مانند الکتروفورز موئینه‌ای و لیزرهای فلورسنت را فراهم می‌کند، که به‌طور قابل توجهی دقت و کارایی فرایند توالی‌یابی را افزایش داده است (Prober *et al.*, 1987). این تغییرات نه‌تنها ایمنی کار با این روش را بهبود بخشید، بلکه امکان تحلیل خودکار و سریع‌تر نتایج را نیز فراهم کرد. این روش هنوز به عنوان استاندارد طلایی برای توالی‌یابی DNA شناخته می‌شود و در بسیاری از کاربردهای بالینی و تحقیقاتی مورد استفاده قرار می‌گیرد (Shendure & Ji, 2008).

ddNTPها تابیده می‌شود. هر رنگ فلورسنت نمایانگر یکی از چهار باز نیتروژنی (آدنین، تیمین، سیتوزین، و گوانین) است و دستگاه الکتروفورز موئینه‌ای با تشخیص این رنگ‌ها، ترتیب بازهای DNA را مشخص می‌کند (Prober *et al.*, 1987; Ansorge, 2009). نتایج به دست آمده به صورت کروماتوگرام نمایش داده می‌شود که در آن هر قله نمایانگر یک باز نیتروژنی است. این داده‌ها سپس توسط نرم‌افزارهای خاص تحلیل شده و توالی نهایی DNA تعیین می‌شود (شکل ۴). یکی از مزایای اصلی این روش، دقت بالای آن در تعیین توالی DNA است، به طوری که می‌توان توالی‌های به نسبت بلند (تا حدود ۱۰۰۰ باز نیتروژنی) را با دقت بالا تعیین کرد (Metzker, 2010).

روش اصلی توالی‌یابی سنگر (Sanger sequencing) که در دهه ۱۹۷۰ توسط فردریک سنگر توسعه یافت، از روش خاتمه زنجیره بهره می‌برد، اما به جای استفاده از رنگ‌های فلورسنت، از مواد پرتوزا یا روش‌های نشان‌گذاری دیگر استفاده می‌شد. در این روش، از دنوکسی‌نوکلئوتیدهای خاتمه‌دهنده (ddNTPs) به‌طور جداگانه و در چهار واکنش مختلف برای هر یک از چهار باز نیتروژنی (A، T، C و G) استفاده می‌شود. این فرایند



شکل ۴ - ختم زنجیرسازی با استفاده از رنگ **Dye-termination**
Figure 4. Termination of chaining using Dye-termination color

(NGS)⁴³ انقلاب عظیمی در ژنتیک ایجاد نموده‌اند و قادر به توالی‌یابی ژنوم انسان با قیمت کم در مدت چند روز و با کیفیت بالا هستند. با فراهم آمدن زمینه‌های نظری و عملی توالی‌یابی در مقیاس، ژنومی پروژه توالی‌یابی ژنوم انسان به‌طور رسمی در ۱۹۹۰ با پشتیبانی مالی مؤسسه NIH⁴⁴ در ایالات متحده مطرح شد (Metzker *et al.*, 2010).

توالی‌یابی قطعات اسیدهای نوکلئیک با استفاده از

روش‌های جدید توالی‌یابی (نسل دوم)

نسل دوم توالی‌یابی DNA، که از آن به‌عنوان توالی‌یابی با توان عملیاتی بالا یا روش‌های توالی‌یابی نسل جدید⁴⁵ نیز یاد می‌کنند، انقلابی در زمینه ژنومیک ایجاد کرد و امکان توالی‌یابی سریع‌تر و ارزان‌تر DNA را فراهم نمود. زیرا این

در اکثر روش‌های تعیین توالی کامل، ژنوم مکرراً و از نقاط متفاوت به تعداد زیادی قطعات کوتاه قابل توالی‌یابی بریده می‌شود تا پس از توالی‌یابی این قطعات با رویهم قرار دادن آنها توالی کامل ژنوم بدست آید. از آنجائیکه در این روش همپوشانی قطعات با یکدیگر تصادفی روی می‌دهد لازم است هر نقطه از ژنوم توسط دست کم تعداد معینی از قطعات پوشانیده شود تا احتمال به وجود آمدن شکاف در توالی کامل ژنوم به حداقل مقدار قابل پذیرش برسد.

روش تعیین توالی سنگر بیش از ۳۰ سال استفاده شد و در پروژه ژنوم انسان نیز این روش کاربرد زیادی داشت. با این حال در سال‌های اخیر پیشرفت چشمگیری در فناوری تعیین توالی شکل گرفت. در سال ۲۰۰۵ نخستین روش تعیین توالی نسل جدید معرفی شد که خروجی بسیار بالاتر و هزینه‌های بسیار کمتری نسبت به روش سنگر داشت. وجود دستگاه‌های نسل جدید توالی‌یابی

⁴³ Next generation sequencing

⁴⁴ National Institutes of Health (NIH)

⁴⁵ Next generation sequencing

SOLID را معرفی کرد که از فناوری توالی‌یابی با لیگاسیون^{۴۷} استفاده می‌نمود (Valouev *et al.*, 2008). این روش بر اساس دورگه‌شدن^{۴۸} و لیگاسیون الیگونوکلوئوتیدهای کوتاه نشان‌دار عمل می‌کرد و دقت بالایی را ارائه می‌داد. در سال ۲۰۱۰، شرکت Ion Torrent که بعدتر توسط Life Technologies خریداری شد، پلتفرم Ion Torrent را معرفی کرد که از فناوری نیمه‌هادی برای شناسایی تغییرات pH ناشی از آزادسازی یون هیدروژن هنگام سنتز DNA استفاده می‌نمود (Rothberg *et al.*, 2011). این روش اولین فناوری توالی‌یابی بود که از تشخیص شیمیایی به جای سیگنال‌های نوری استفاده می‌کرد.

هر یک از این فناوری‌ها مزایا و کاستی‌های خاص خود را داشتند، اما همگی به‌طور قابل توجهی توان عملیاتی و سرعت توالی‌یابی را افزایش دادند و هزینه‌ها را کاهش دادند و استفاده از فناوری‌های نسل جدید توالی‌یابی با فراهم‌نمودن پوشش بالای ژنومی، امکان توالی‌یابی مجدد کل ژنوم یا کل ترنسکرپتوم^{۴۹} نمونه را به گونه‌ای کارآمد، اقتصادی و دقیق‌تر از گذشته فراهم آورد (Metzker *et al.*, 2010; Salk *et al.*, 2018). این پیشرفت‌ها منجر به گسترش کاربردهای توالی‌یابی در زمینه‌های مختلف از جمله تشخیص بیماری‌ها، مطالعات تکاملی، اکولوژی میکروبی و کشاورزی شد. باوجود این، چالش‌هایی مانند خطاهای توالی‌یابی، کاستی در طول

نسل برخلاف روش سنگر، نیاز به مراحل همسانه‌سازی ندارد؛ این ویژگی باعث کاهش چشمگیر زمان و هزینه توالی‌یابی شده است و از طرفی این نوع توالی‌یابی، می‌تواند میلیون‌ها توالی با طول متوسط ۳۰ تا ۷۰۰ جفت باز (بسته به هدف مورد نظر و نوع سامانه استفاده شده) در طی چند روز با دقت و صحت بالا تولید کند. این ویژگی‌ها باعث شده تا این فناوری جایگزین روش توالی‌یابی سنگر شود و اغلب آزمایشگاه‌ها بتوانند به‌طور مستقل از مراکز توالی‌یابی ژنومی بزرگ از آن بهره‌برداری کنند (Shendure *et al.*, 2008).

این نسل با معرفی پلتفرم Pyrosequencing 454 توسط شرکت LifeSciences 454 در سال ۲۰۰۵ آغاز شد (Margulies *et al.*, 2005). این فناوری از روش Pyrosequencing استفاده می‌کرد که در آن نوکلئوتیدها به صورت متوالی اضافه می‌شدند و سیگنال‌های نوری حاصل از ترکیب آنزیمی ثبت می‌شد. در سال ۲۰۰۶، شرکت Solexa (که بعدتر توسط Illumina خریداری شد) پلتفرم توالی‌یابی خود را معرفی کرد که بر اساس روش توالی‌یابی با سنتز^{۴۶} (SBS) عمل می‌کرد (Bentley *et al.*, 2008). این فناوری از نوکلئوتیدهای نشان‌دار فلورسنت و تصویربرداری با وضوح بالا برای شناسایی توالی استفاده می‌کرد و به سرعت به یکی از پرکاربردترین پلتفرم‌های توالی‌یابی تبدیل شد. در همان سال، شرکت Applied Biosystems (اکنون بخشی از Thermo Fisher Scientific) سامانه

⁴⁷ Sequencing by ligation (SBL)

⁴⁸ Hybridation

⁴⁹ Transcriptome

⁴⁶ Sequencing by synthesis (SBS)

توسط پال نیرن^{۵۱}، مصطفی رونقی و همکارانشان در سال ۱۹۹۰ توسعه یافت و سپس در سال ۲۰۰۵ توسط شرکت Roche/454 Life Sciences به یک پلتفرم تجاری تبدیل شد (Margulies *et al.*, 2005). این روش شامل چندین مرحله کلیدی است که در نهایت به شناسایی دقیق توالی بازهای نیتروژنی در یک نمونه DNA منجر می‌شود.

فناوری توالی‌یابی پائروسکوئسنسینگ از DNA تک‌رشته‌ای که از واسرشت‌سازی^{۵۲} محصولات PCR به دست می‌آید، استفاده می‌کند. فرایند با اتصال آغازگر به توالی مکمل شروع شده و با کمک آنزیم پلی‌مراز، ساخت رشته جدید آغاز می‌شود. در طی واکنش‌های پلیمریزه‌شدن، به ازای هر پیوند فسفودی‌استر، یک مولکول پیروفسفات (pp_i) و یک یون هیدروژن آزاد می‌شود. سپس آنزیم ATP سولفوریلاز با استفاده از پیروفسفات و آدنوزین فسفوسولفات (APS) موجود در محیط واکنش، ATP تولید می‌کند. این ATP انرژی لازم برای فعالیت آنزیم لوسیفراز را فراهم کرده و لوسیفراز با اکسید کردن لوسیفیرین، اکسی‌لوسیفیرین تولید می‌کند که این واکنش منجر به تولید نور می‌شود. سیگنال‌های نور تولید شده توسط دوربین‌های ویژه‌ای حساس به نام (CCD)^{۵۳} به صورت یک قله ثبت می‌شود. از آنجا که نوکلئوتیدهای اضافه شده در هر مرحله مشخص و یکسان هستند، می‌توان با تحلیل این قله‌ها، توالی DNA را همزمان با

خوانش و نیاز به تجهیزات پیچیده و گران‌قیمت همچنان وجود داشت و با وجود مزایای ارزشمند روش‌های NGS، تجزیه و تحلیل داده‌های حاصل از توالی‌یابی که طول کوتاهی دارند، بزرگترین چالش در این زمینه محسوب می‌شود و نیازمند رایانه‌های پیشرفته، نرم‌افزارهای خاص برای تجزیه و تحلیل داده‌ها و دانش و تخصص کافی بود (Liu *et al.*, 2012). نسل دوم توالی‌یابی DNA راه را برای پیشرفت‌های بیشتر در زمینه ژنومیک هموار کرد و زمینه را برای ظهور نسل سوم توالی‌یابی، مانند فناوری‌های Single Molecule Real-Time sequencing; SMRT و Pacific Biosciences و Nanopore sequencing را فراهم نمود. این نسل‌های جدیدتر قادر به تولید خوانش‌های طولانی‌تر و حل برخی از کاستی‌های نسل دوم هستند. جزئیات و ویژگی تعدادی از این پلتفرم‌های توالی‌یابی ژنوم و ترنسکریپتوم با توان بالا به شرح زیر است:

روش تعیین توالی پائروسکوئسنسینگ Roche/454 pyrosequencing

فناوری توالی‌یابی Roche/454 Life Sciences که یک روش توالی‌یابی DNA در زمان واقعی است و از اولین سامانه‌های توالی‌یابی نسل جدید (NGS) محسوب می‌شود، توانست تحولی عظیم در علم ژنتیک و ژنومیک ایجاد کند. این فناوری بر پایه روش پائروسکوئسنسینگ^{۵۰} عمل می‌کند که شامل استفاده از آنزیم‌ها و واکنش‌های شیمیایی مختلف است. پائروسکوئسنسینگ برای اولین بار

⁵¹ Pál Nyrén

⁵² Denaturation

⁵³ Charged-coupled device

⁵⁰ Pyrosequencing

به ذرات، در این مخلوط واکنش‌های PCR امولسیون^{۵۴} انجام می‌شود. از آنجاکه این تکنیک نیازمند DNA الگو تک رشته می‌باشد، پس از اتمام مرحله تکثیر، محصولات PCR واسرشت می‌شوند و هر مهره در این مرحله، حاوی میلیون‌ها کپی تک رشته از یک قطعه DNA منحصر به فرد است که باعث تقویت شدت سیگنال و شناسایی بهتر توسط آشکارساز^{۵۵} مرحله تعیین توالی می‌شود (Dressman *et al.*, 2003). پس از شکستن امولسیون، مهره‌های حاوی DNA تکثیر شده ابتدا در یک محلول مایع به حالت تعلیق در می‌آیند. این کار به جلوگیری از چسبیدن مهره‌ها به یکدیگر و تسهیل توزیع یکنواخت آنها کمک می‌کند. سپس محلول حاوی مهره‌ها به‌طور مستقیم بر روی صفحه پیکوتیتر ریخته می‌شود. صفحه پیکوتیتر دارای تعداد بسیار زیادی چاهک کوچک است، و طراحی آن به گونه‌ای است که مهره‌ها به‌طور تصادفی در این چاهک‌ها توزیع می‌شوند. برای اطمینان از قرارگیری مهره‌ها در چاهک‌ها، ممکن است از سانتریفیوژ یا نیروی جاذبه استفاده شود. در صورت استفاده از سانتریفیوژ، صفحه پیکوتیتر^{۵۶} به آرامی چرخانده می‌شود تا مهره‌ها به سمت چاهک‌ها حرکت کرده و در آنها جای گیرند. همچنین ممکن است از نیروی جاذبه یا لرزش خفیف استفاده شود تا مهره‌ها به ته چاهک‌ها بیفتند. پس از توزیع اولیه مهره‌ها، چاهک‌هایی که ممکن است بیش از یک مهره دریافت کرده باشند، شسته می‌شوند تا مهره‌های

سنتز آن تعیین کرد. برای افزودن نوکلئوتیدهای جدید در هر مرحله، لازم است نوکلئوتیدهای قبلی از محیط واکنش حذف شوند. این وظیفه بر عهده آنزیمی به نام آپیراز است که یک نوع نوکلئوتیداز محسوب می‌شود. روش تجاری تعیین توالی به شیوه Pyrosequencing توسط شرکت Life Sciences 454 در سال ۲۰۰۵ معرفی شد. سازوکار عمل این فناوری که به عنوان Pyrosequencing 454 نیز شناخته می‌شود شامل چندین مرحله اصلی زیر است: ابتدا DNA ژنومی به قطعاتی با طول ۳۰۰ تا ۵۰۰ جفت باز شکسته می‌شود. سپس این قطعات به صورت تک‌رشته‌ای به مهره‌های فلزی با قطر حدودی ۲۸ میکرومتری متصل می‌شوند. برای این منظور، پیش از واسرشت‌سازی، آداپتورها به دو انتهای قطعات DNA متصل می‌گردند. این آداپتورها به دو صورت عمل می‌کنند: اتصال قطعات تک‌رشته‌ای به مهره‌های فلزی کوچک پوشیده شده با استرپتوآویدین از طریق گروه بیوتین در انتهای ۵' و فراهم کردن جایگاهی برای اتصال آغازگرها به منظور تکثیر قطعه مورد نظر به دلیل تمایل بالای استرپتوآویدین به بیوتین. پس از اتصال این قطعات DNA به مهره‌های کوچک، این مهره‌ها در یک امولسیون روغن-آب قرار می‌گیرند، در این مرحله آداپتورها دومین نقش خود را ایفا می‌نمایند به‌طوری‌که آداپتورها بعنوان جایگاهی برای اتصال آغازگرها و در نتیجه تکثیر قطعه مورد نظر، عمل می‌کنند و با افزوده شدن سایر ترکیبات لازم برای تکثیر قطعات DNA متصل

⁵⁴ Emulsion PCR

⁵⁵ Detector

⁵⁶ Picotiter plate

به‌طور مداوم تکرار می‌شود تا توالی کامل DNA مشخص شود (شکل ۵). یکی از مزایای اصلی این روش، توانایی خواندن توالی‌های به نسبت طولانی (تا حدود ۷۰۰ باز) در مقایسه با سایر روش‌های NGS اولیه است (Rothberg & Leamon, 2008). با این حال، این فناوری کاستی‌هایی نیز دارد. از جمله چالش‌های اصلی می‌توان به دشواری در تعیین دقیق تعداد نوکلئوتیدها در توالی‌های هموپلیمری (توالی‌های تکراری از یک نوکلئوتید) اشاره کرد، زیرا رابطه بین شدت نور و تعداد نوکلئوتیدها در توالی‌های طولانی‌تر غیرخطی می‌شود (Huse *et al.*, 2007). علاوه بر این، هزینه به نسبت بالای هر اجرا و کاستی در تعداد نمونه‌هایی که می‌توان همزمان تجزیه و تحلیل کرد، از دیگر کاستی‌های این روش هستند. با وجود این کاستی‌ها، فناوری Roche/454 نقش مهمی در پیشرفت علم ژنومیک ایفا کرد و در بسیاری از پروژه‌های مهم توالی‌یابی، از جمله پروژه ژنوم انسان، مورد استفاده قرار گرفت (Wheeler *et al.*, 2008). این فناوری همچنین در مطالعات متاژنومیکس، بررسی تنوع میکروبی و شناسایی جهش‌های نادر در جمعیت‌های سلولی کاربرد گسترده‌ای داشت (Sogin *et al.*, 2006; Mardis, 2008). با پیشرفت سایر فناوری‌های NGS که قادر به تولید داده‌های بیشتر با هزینه کمتر بودند، استفاده از پلتفرم Roche/454 به تدریج کاهش یافت و در نهایت در سال ۲۰۱۶، تولید آن متوقف شد.

با این حال، تأثیر این فناوری بر پیشرفت علم ژنومیک و توسعه روش‌های جدید توالی‌یابی انکار ناپذیر است.

اضافی حذف شوند. این فرایند کمک می‌کند که در نهایت هر چاهک فقط یک مهره داشته باشد. پس از اینکه در هر چاهک فقط یک مهره قرار گرفت، در مرحله بعد، فرایند پیروسکوئسنینگ آغاز می‌شود. در این فرایند، نوکلئوتیدها: Deoxycytidine triphosphate; dCTP، Deoxythymidine triphosphate; dTTP، Deoxyguanosine triphosphate; dGTP و Deoxyadenosine triphosphate; dATP به صورت متوالی و مجزا و در چرخه‌های مشخص به محیط واکنش اضافه می‌شوند. هنگامی که یک نوکلئوتید مکمل با توالی الگو است، DNA پلی‌مراز آن را به رشته در حال سنتز اضافه می‌کند. این عمل منجر به آزاد شدن یک مولکول پیروفسفات (PPi) می‌شود. آنزیم ATP سولفوریلاز، پیروفسفات آزاد شده را با آدنوزین فسفوسولفات (APS) ترکیب می‌کند تا ATP تولید کند. سپس، آنزیم لوسیفراز از ATP تولید شده برای تبدیل لوسیفیرین به اکسی‌لوسیفیرین استفاده می‌کند که این واکنش با تولید نور همراه است (Ronaghi, 2001). شدت نور تولید شده متناسب با تعداد نوکلئوتیدهای اضافه شده است. این سیگنال نوری توسط یک دوربین CCD⁵⁷ ثبت می‌شود و به صورت یک قله⁵⁸ در پیروگرام⁵⁹ نمایش داده می‌شود. پس از هر چرخه، آنزیم آپیراز نوکلئوتیدهای اضافه نشده و ATP باقی‌مانده را تجزیه می‌کند تا محیط برای چرخه بعدی آماده شود (Ahmadian *et al.*, 2006). این فرایند

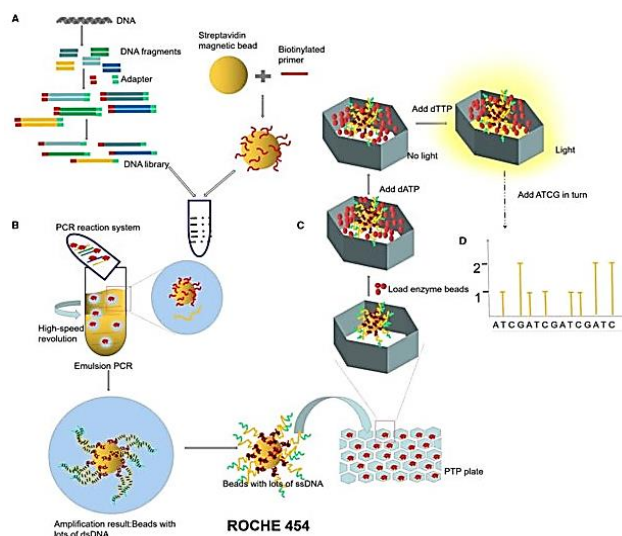
⁵⁷ Charged-coupled device

⁵⁸ Peak

⁵⁹ Pyrogram

توالی‌یابی Roche/454 Life Sciences نقطه عطفی در تاریخ علم ژنتیک و ژنومیک بود که راه را برای پیشرفت‌های بعدی در زمینه توالی‌یابی DNA هموار کرد و تأثیر عمیقی بر درک ما از ژنوم‌ها و تنوع زیستی گذاشت.

امروزه، روش‌های پیشرفته‌تری مانند توالی‌یابی Illumina، Ion Torrent و PacBio و نسل سوم توالی‌یابی مانند Oxford Nanopore، جایگزین Roche/454 شده‌اند، اما بسیاری از اصول و مفاهیم توسعه یافته در این پلتفرم همچنان در فناوری‌های جدید مورد استفاده قرار می‌گیرند (Goodwin *et al.*, 2016). در مجموع، فناوری



شکل ۵- نمایی از تکنیک توالی‌یابی پیروسیکوئسنسینگ
Figure 5. View of pyrosequencing technique

توالی‌یابی با سنتز (SBS)⁶⁰ عمل می‌کند (Metzker, 2010). طول متوسط توالی‌های خوانده شده توسط این فناوری ۳۷ تا ۱۵۰ جفت باز می‌باشد (Mardis, 2008). سازوکار عمل این فناوری شامل چندین مرحله اصلی است. در ابتدا، DNA ژنومی به قطعات کوچک (به‌طور معمول ۲۰۰-۶۰۰ جفت باز) شکسته می‌شود که این قطعات دارای انتهای صاف هستند. سپس، آداپتورهای اختصاصی به انتهای این قطعات متصل می‌شوند. این

فناوری توالی‌یابی Illumina/Solexa

فناوری توالی‌یابی Illumina/Solexa یکی از پرکاربردترین و موفق‌ترین روش‌های توالی‌یابی نسل بعدی (NGS) است که از زمان معرفی در سال ۲۰۰۶ تا به امروز تحولی عظیم در عرصه ژنومیک ایجاد کرده است (Bentley *et al.*, 2008). این فناوری که در ابتدا توسط شرکت Solexa توسعه یافت و سپس در سال ۲۰۰۷ توسط شرکت Illumina خریداری شد، بر اساس روش

⁶⁰ Sequencing by synthesis

رنگ‌های فلورسنت قابل برگشت استفاده می‌شود. هر نوکلئوتید (A, T, C و G) دارای یک رنگ فلورسنت منحصر به فرد است و همچنین یک گروه بلوکه‌کننده^۳ دارد که از اضافه شدن بیش از یک نوکلئوتید در هر چرخه جلوگیری می‌کند (Turcatti *et al.*, 2008). در هر چرخه توالی‌یابی، چهار نوع نوکلئوتید به همراه آنزیم DNA پلی‌مراز به محیط واکنش اضافه می‌شوند. پس از اتصال نوکلئوتید مکمل و شستشوی نوکلئوتیدهای اضافی، یک تصویربرداری با استفاده از لیزر و دوربین‌های حساس انجام می‌شود تا مشخص شود کدام نوکلئوتید در هر موقعیت اضافه شده است. سپس، گروه بلوکه‌کننده^۳ و رنگ فلورسنت با استفاده از یک عامل احیاکننده برداشته می‌شوند تا برای چرخه بعدی آماده شوند (Bentley *et al.*, 2008). این فرایند برای صدها چرخه تکرار می‌شود تا توالی کامل هر قطعه DNA مشخص شود (شکل ۶). برخلاف تکنولوژی ۴۵۴ فرایند اضافه شدن بازها در این سامانه به خوبی قادر به تعیین توالی DNA با دقت بالا در نواحی هموپلیمر و توالی‌های تکراری است. یکی دیگر از مزایای اصلی تکنولوژی Illumina، توانایی آن در تولید میلیاردها خوانش کوتاه با دقت بالا در یک اجرای واحد است (Goodwin *et al.*, 2016). این ویژگی امکان توالی‌یابی عمیق و با پوشش بالا را فراهم می‌کند که برای کاربردهایی مانند شناسایی واریانت‌های نادر و مطالعات ترنسکریپتومیکس بسیار مفید است. علاوه بر این، هزینه به نسبت پایین هر باز توالی‌یابی شده، این فناوری را به یکی از مقرون به صرفه‌ترین گزینه‌ها برای پروژه‌های بزرگ

آداپتورها حاوی توالی‌هایی هستند که برای اتصال به سطح فلوسل^{۶۱} و همچنین برای شروع واکنش‌های تکثیر و توالی‌یابی ضروری هستند (Fedurco *et al.*, 2006). در مرحله بعد، قطعات DNA حاصل شده واسرشت شده و به صورت تک‌رشته‌ای در می‌آیند و به توالی مکمل در سطح فلوسل دورگ می‌شوند. فلوسل یک اسلاید شیشه‌ای است که سطح آن با الیگونوکلئوتیدهایی پوشیده شده که مکمل توالی‌های آداپتور هستند. پس از اتصال، هر قطعه DNA از طریق فرایندی به نام ازدیاد پل مانند^{۶۲} که نوعی تکثیر از طریق واکنش PCR است تکثیر می‌شود (Adessi *et al.*, 2000). در این فرایند، انتهای آزاد هر رشته DNA متصل به سطح، خم شده و به یک آداپتور مکمل در نزدیکی خود شبیه یک پل متصل می‌شود. سپس، با استفاده از آنزیم DNA پلی‌مراز، یک رشته مکمل سنتز می‌شود در این پروسه، توالی‌های مکمل آداپتور بعنوان آغازگر عمل می‌کنند. با تکرار این فرایند، هر قطعه DNA اولیه به یک خوشه^{۶۳} حاوی هزاران کپی یکسان تبدیل می‌شود (Voelkerding *et al.*, 2009). پس از تشکیل خوشه‌ها، فرایند توالی‌یابی آغاز می‌شود. در مرحله بعد به منظور تعیین توالی، قطعات DNA واسرشت و خطی می‌شوند. پس از دورگه شدن آغازگر مخصوص توالی‌یابی به توالی مکمل آداپتوری در انتهای هر رشته، فلوسل برای تعیین توالی در دستگاه مربوط آماده می‌شود. در این مرحله، از نوکلئوتیدهای نشان‌دار با

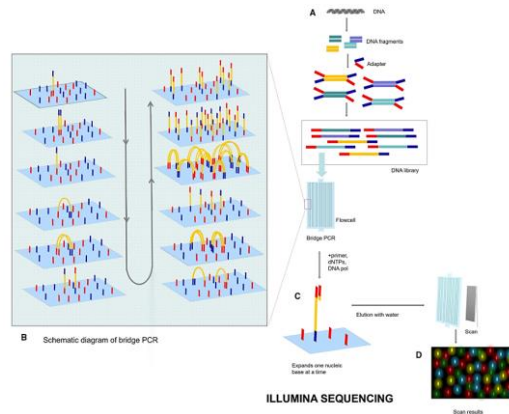
⁶¹ Flow cell

⁶² Bridge amplification (Bridge-PCR)

⁶³ Cluster

مقیاس ژنومیک تبدیل کرده است (Reuter *et al.*, 2015). با این حال، این فناوری کاستی‌هایی نیز دارد. یکی از چالش‌های اصلی، طول به نسبت کوتاه خوانش‌ها (به‌طور معمول ۳۰۰-۷۵ جفت باز) است که می‌تواند مونتاژ ژنوم، به ویژه در مناطق تکراری، را دشوار کند (van Dijk *et al.*, 2014). همچنین، خطاهای نظام-مند^{۶۴} مانند سوگیری^{۶۵} در محتوای GC و افزایش نرخ خطا در انتهای خوانش‌ها می‌تواند بر کیفیت داده‌ها تأثیر بگذارد (Nakamura *et al.*, 2011).

⁶⁴ Systematic
⁶⁵ Bias



شکل ۶ - فرایند توالی‌یابی از طریق فناوری توالی‌یابی Illumina/Solexa
Figure 6. Sequencing process through Illumina/Solexa sequencing technology

با وجود این کاستی‌ها، فناوری Illumina به دلیل دقت بالا، بازده بالا و هزینه پایین، همچنان به عنوان استاندارد طلایی در بسیاری از کاربردهای توالی‌یابی باقی‌مانده است. در سال‌های ابتدایی دهه ۲۰۱۰، شرکت ایلومینا، سامانه HiSeq را که از بازدهی بسیار بالاتر نسبت به نسل‌های پیشین برخوردار بود را معرفی نمود که از همان اصول توالی‌یابی قبلی بهره می‌برد. این دستگاه در ابتدا قادر به تولید ۲۰۰ گیگابایت داده توالی در هر اجرا بود که بعدتر به ۶۰۰ گیگابایت ارتقا یافت (Liu *et al.*, 2011). گرچه توالی‌های کوتاه تولید شده توسط فناوری ایلومینا، تحلیل‌های دقیق را خصوصاً در موجودات با اطلاعات ژنتیکی محدود با چالش مواجه می‌کند، اما قابلیت توالی‌یابی دو طرفه^{۶۶} تا حد زیادی این مشکل را مرتفع ساخته است. در نتیجه، امروزه این سامانه، گاه به تنهایی و گاه در ترکیب با سامانه ۴۵۴، برای توالی‌یابی ژنوم طیف وسیعی از موجودات مورد استفاده قرار می‌گیرد (Mardis,

2008). این فناوری در طول سال‌ها پیشرفت‌های قابل توجهی داشته است. از جمله این پیشرفت‌ها می‌توان به افزایش طول خوانش‌ها، بهبود دقت، افزایش تعداد خوانش‌ها در هر اجرا و کاهش زمان اجرا اشاره کرد (Goodwin *et al.*, 2016). پلتفرم‌های جدیدتر Illumina مانند NovaSeq و NextSeq می‌توانند تا ۶ ترابایت داده در یک اجرای واحد تولید کنند، که این امر امکان توالی‌یابی کل ژنوم انسان را در کمتر از یک روز فراهم می‌کند (Illumina, 2021). فناوری Illumina در طیف گسترده‌ای از کاربردها مورد استفاده قرار می‌گیرد. از جمله این کاربردها می‌توان به توالی‌یابی کل ژنوم (WGS)^{۶۷}، توالی‌یابی اگزوم (WES)^{۶۸}، توالی‌یابی RNA (RNA-Seq)، تعیین متیلاسیون DNA، مطالعات اپی‌ژنتیک، متاژنومیکس و تعیین ژنوتیپ اشاره کرد (Metzker, 2010; Wang *et al.*, 2009). این فناوری

⁶⁷ Whole genome sequencing

⁶⁸ Whole exome sequencing

⁶⁶ Paired-end sequencing

Thermo Fisher Biosystems که اکنون بخشی از Scientific است در سال ۲۰۰۷ خریداری و با نام فناوری توالی‌یابی ABI/SOLID معرفی شد (Mardis, 2008). این فناوری بر اساس روش توالی‌یابی با لیگاسیون عمل می‌کند که اولین بار توسط شندر و همکارانش در سال ۲۰۰۵ توصیف شد (Shendure et al., 2005). مراحل آماده‌سازی نمونه‌ی DNA به‌طور تقریبی مشابه با سامانه توالی‌یابی ۴۵۴ است و سازوکار عمل SOLID شامل چندین مرحله اصلی است. ابتدا، DNA ژنومی به قطعات کوچک (به‌طور معمول ۲۰۰-۱۵۰ جفت باز) شکسته می‌شود. سپس، آداپتورهای اختصاصی به انتهای این قطعات متصل می‌شوند. این قطعات DNA به مهره‌های مغناطیسی متصل شده و در یک امولسیون روغن-آب قرار می‌گیرند تا واکنش‌های Emulsion-PCR انجام شود (Dressman et al., 2003). در این مرحله، هر مهره حاوی میلیون‌ها کپی از یک قطعه DNA منحصر به فرد می‌شود. پس از شکستن امولسیون، مهره‌های حاوی DNA تکثیر شده به سطح یک اسلاید شیشه‌ای متصل می‌شوند (Valouev et al., 2008). در مرحله بعد، فرایند توالی‌یابی با لیگاسیون آغاز می‌شود و آغازگرها که مکمل توالی آداپتور بوده و مخلوطی از کاوشگرهای الیگونوکلوئوتیدی کوتاه (به‌طور معمول ۸ نوکلئوتید) به این سطح افزوده می‌شود. در مرحله اول آغازگر به آداپتور متصل می‌گردد. سپس مخلوطی از الیگونوکلوئوتیدهای اکتامر که انتهای آنها با رنگ‌های فلورسانس مختلف که برای ۱۶ ترکیب ممکن از

نقش مهمی در پیشرفت‌های اخیر در زمینه‌هایی مانند پزشکی شخصی، تشخیص سرطان، شناسایی بیماری‌های ژنتیکی و مطالعات تکاملی ایفا کرده است (Shendure et al., 2017). علاوه بر این، در دوران همه‌گیری COVID-19، فناوری Illumina نقش حیاتی در توالی‌یابی ژنوم ویروس SARS-CoV-2 و شناسایی واریانت‌های جدید ایفا کرد (Rambaut et al., 2020). با نگاهی به آینده، انتظار می‌رود که فناوری Illumina همچنان به پیشرفت خود ادامه دهد. پیشرفت‌های جدید مانند توالی‌یابی تک سلولی، توالی‌یابی طولانی مصنوعی^{۶۹} و بهبود الگوریتم‌های تجزیه و تحلیل داده، افق‌های جدیدی را در عرصه ژنومیک می‌گشاید (Slatko et al., 2018). در عین حال، ظهور فناوری‌های جدید مانند توالی‌یابی نانومفرد^{۷۰} و توالی‌یابی مولکول منفرد^{۷۱} در حال ایجاد رقابت و نوآوری بیشتر در این حوزه است (Logsdon et al., 2020). با این وجود، به دلیل زیرساخت گسترده، دقت بالا و هزینه پایین، انتظار می‌رود که فناوری Illumina همچنان نقش مهمی در آینده توالی‌یابی DNA ایفا کند.

فناوری توالی‌یابی ABI/SOLID

در سال ۲۰۰۶ شرکت SOLID^{۷۲} فناوری توالی‌یابی از طریق اتصال^{۷۳} که یکی از روش‌های نسل دوم توالی‌یابی است را معرفی کرد، سپس توسط شرکت Applied

⁶⁹ Synthetic long read sequencing

⁷⁰ Nanopore

⁷¹ Single-molecule sequencing, Single-molecule real-time sequencing; SMRT

⁷² Sequencing by oligonucleotide ligation and detection

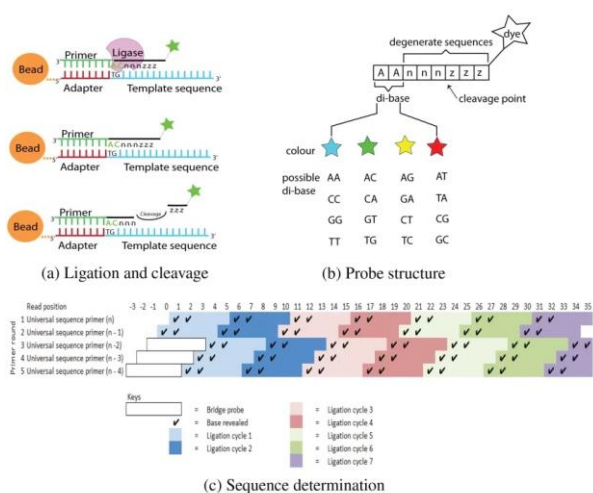
⁷³ Sequencing by ligation; SBL

در دو حالت مکمل و معکوس نیز رنگ و کد یکسان دارند. مثلاً توالی AT و حالت مکمل معکوس آن یعنی CG هر دو دارای رنگ قرمز یا کد سه هستند. (۴) دو نوکلئوتید یکسان مثلاً AA دارای کد صفر و رنگ آبی می‌باشند.

دی‌نوکلئوتید ابتدایی آنها برچسب شده‌اند، به قسمت‌های مکمل خود بر روی رشته الگو متصل می‌شوند (McKernan et al., 2009). به طبع این الیگو نوکلئوتیدها باید توالی متنوعی داشته باشند تا امکان حضور توالی مکمل بر روی رشته DNA افزایش یابد. برای این امر از Di-base probe ها استفاده می‌شود. همانطور که از اسم این کاوشگرها مشخص است، این کاوشگرها در ابتدای خود حاوی دو باز مشخص می‌باشند و مابقی بازها از توالی‌های دژنره^{۷۴} و یونیورسال^{۷۵} می‌باشند که با حرف n و Z نشان داده می‌شوند. ویژگی این کاوشگرها مربوط به دو باز اول می‌باشد و حتی برچسب رنگی که به انتهای این کاوشگرها متصل می‌گردد بسته به این دو باز تعیین می‌شود. بطوریکه یک الفبای دو حرفی و ۱۶ حالت بر اساس این دو باز در روش SOLID موجود می‌باشد (شکل ۷، قسمت B) که از هر دو حرف نوکلئوتیدی یک کد یا رنگ خاصی حاصل می‌شود. الفبای فوق چهار قانون دارد که شامل موارد زیر می‌باشد. (۱) قانون معکوس: دو نوکلئوتید مختلف در هر دو حالت ترکیبی خود، یک رنگ یا کد مشترک دارند. مثلاً دو نوکلئوتید T و C در هر دو حالت CT و معکوس آن TC، به رنگ نارنجی و کد دو شناخته می‌شوند. (۲) قانون مکمل: نوکلئوتیدهای دو بازی مکمل نیز رنگ و کد یکسان دارند. مثلاً دو باز CA و مکمل آن GT هر دو دارای رنگ سبز یا کد یک می‌باشند. (۳) قانون مکمل معکوس: با توجه به دو قانون قبل، نوکلئوتیدهای دو بازی

⁷⁴ Degenerated

⁷⁵ Universal



شکل ۷ - توالی‌یابی ABI/SOLiD
Figure 7. ABI/SOLiD sequencing

می‌شود. پس از اتصال یکی از این کاوشگرها، برچسب فلورسانس مربوط به این کاوشگر به همراه سه باز آخر بصورت آنزیماتیک جدا می‌گردد. همانطور که در شکل ۷ قسمت (B) نشان داده شده است این کاوشگرها حاوی یک Cleavage site می‌باشند که پس از اتصال به توالی مکمل، کاوشگر از این ناحیه بصورت آنزیماتیک شکسته می‌شود. بواسطه جدا شدن برچسب رنگی، نوری ساطع می‌شود که یک تصویربرداری با استفاده از لیزر و دوربین‌های حساس انجام می‌شود تا مشخص شود کدام رنگ (و در نتیجه کدام دو نوکلئوتید) در آن موقعیت قرار گرفته است (Pandey *et al.*, 2008). این مرحله آنقدر تکرار می‌شود تا رشته جدید بطور کامل سنتز شود. نکته قابل توجه این است که قبل از ورود به هر مرحله، کاوشگرهای متصل نشده بواسطه شستشو از محیط حذف

تعیین توالی بر پایه Ligation در پنج چرخه که هر چرخه نیز شامل چندین مرحله می‌باشد، بواسطه آغازگرهای مختلف از لحاظ طول انجام می‌گیرد. بدین ترتیب که در هر چرخه یک نوکلئوتید از آغازگر مربوطه کم می‌شود یعنی در دور اول از پرایمر n ، در دور دوم از آغازگر $n-1$ ، در دور سوم از آغازگر $n-2$ ، در دور چهارم از آغازگر $n-3$ و در دور پنجم از پرایمر $n-4$ استفاده می‌گردد. در چرخه توالی‌یابی اول، پس از اتصال پرایمر n به توالی آداپتور، یک مجموعه چهارتایی از این کاوشگرها به همراه آنزیم DNA لیگاز به محیط واکنش اضافه می‌شوند که برای اتصال به قطعه پرایمر با یکدیگر رقابت می‌کنند و اتصال این قطعات به یکدیگر بواسطه آنزیم لیگاز صورت می‌گیرد. در نهایت کاوشگری که به‌طور کامل مکمل با توالی هدف است، توسط DNA لیگاز به آغازگر متصل

داده‌های رنگی به توالی نوکلئوتیدی استفاده می‌شود. به عنوان مثال یک فرایند توضیح داده می‌شود: در چرخه اول ابتدا آغازگر به آداپتور متصل می‌شود و از آنجا که توالی آداپتور مشخص است و نوکلئوتید انتهایی آداپتور بدون اتصال باقی مانده است، توالی‌یابی با یک نوکلئوتید مشخص (به‌طور معمول در آداپتور) شروع می‌شود که به عنوان نقطه مرجع عمل می‌کند. در سامانه SOLiD، هر رنگ فلورسنت نماینده چهار ترکیب دی‌نوکلئوتیدی ممکن است. مثلاً: رنگ قرمز: AT, CG, GC, TA، رنگ سبز: AC, CA, GT, TG، رنگ زرد: AG, CT, GA، و رنگ آبی: AA, CC, GG, TT. بعد از اتصال اولین کاوشگر به نوکلئوتید انتهایی آداپتور و نوکلئوتید اول توالی اصلی عملیات برش کاوشگر از محل نوکلئوتید چهارم و پنج انجام شده و پرتوفلورسانس ساطع می‌شود، سپس با استفاده از آنزیم لیگاز^{۷۶} اتصال آغازگر و کاوشگر صورت می‌گیرد. بعد از این اتصال، برش و لیگاسیون^{۷۷} کاوشگر-های بعدی هم ادامه می‌یابد و در انتها پس از برش و لیگاسیون‌ها توالی بطور کامل سنتز می‌شود اما بین هر کدام از دی‌نوکلئوتیدها سه باز فاصله به وجود می‌آید و چرخه اول خاتمه می‌یابد. در چرخه دوم طول آغازگر n-1 می‌شود و کاوشگرها و توالی دی‌نوکلئوتیدی آنها نیز به موقعیت n-1 در چرخه قبلی متصل می‌شوند و همانند چرخه قبلی عملیات‌های برش و انتشار سیگنال فلورسنت و همچنین لیگاسیون انجام می‌شود این روند تا چرخه پنج و آغازگر n-4 ادامه می‌یابد و باعث می‌شود کاوشگرها

می‌گردند تا اختلالی در مراحل بعدی ایجاد نکنند. پس از اتمام چرخه اول وارد چرخه دوم می‌شود. قبل از شروع چرخه دوم، رشته حاوی آغازگر n بواسطه حرارت از رشته الگو جدا شده و از طریق سستشو از محیط حذف می‌گردد و بدین ترتیب واکنش با آغازگر n-1 در چرخه دوم آغاز می‌شود. این واکنش نیز همانند چرخه اول تا تکمیل شدن رشته در حال سنتز بواسطه کاوشگرهای مختلفی که بصورت مرحله‌ای (در هر مرحله یک مجموعه چهارتایی) به واکنش افزوده می‌شود، ادامه می‌یابد. نمایی از این تکنیک در شکل ۷ نشان داده شده است (Mardis, 2008). در فناوری ABI/SOLiD، جهت توالی‌یابی DNA سامانه کدگذاری دو رنگی (Two-color encoding system) به کار می‌رود که پیچیده‌تر از آن چیزی است که در نگاه اول به نظر می‌رسد (Shendure & Ji, 2008). اینکه چگونه دستگاه از یک رنگ فلورسنت به توالی دقیق دو نوکلئوتیدی پی می‌برد، و از تجمیع نور فلورسنت ساطع شده در چرخه‌های متوالی، توالی DNA را تشخیص می‌دهد، به روش هوشمندانه‌ای طراحی شده است. روش کار به این صورت است که، هر رنگ فلورسنت نماینده چهار ترکیب دو نوکلئوتیدی ممکن است. اما نکته کلیدی این است که هر نوکلئوتید در دو چرخه متوالی خوانده می‌شود. فرایند توالی‌یابی در پنج چرخه انجام می‌شود، هر بار با یک آغازگر که نسبت به قبلی یک نوکلئوتید جابجا شده است. دستگاه با ترکیب اطلاعات به دست آمده از تمام چرخه‌های توالی‌یابی، توالی دقیق را مشخص می‌کند و یک الگوریتم پیچیده برای تبدیل

⁷⁶ Ligase⁷⁷ Ligation

های چرخه اول ادامه پیدا می‌کند و دستگاه با ترکیب اطلاعات به دست آمده از تمام مراحل توالی‌یابی، توالی دقیق را مشخص می‌کند و یک الگوریتم پیچیده برای تبدیل داده‌های رنگی به توالی نوکلئوتیدی استفاده می‌شود. این سامانه نه‌تنها امکان تعیین توالی دقیق را فراهم می‌کند، بلکه به تشخیص و تصحیح خطاهای احتمالی نیز کمک می‌کند. صحت توالی‌های خوانده شده با استفاده از این فناوری در مقایسه با سامانه‌های دیگر در سطح بالایی است. به کمک این روش تعیین توالی میتوان بین خطای توالی‌یابی^{۷۸} و چند شکلی تک نوکلئوتیدی (SNP)^{۷۹} تمایز قائل شد. این تمایز با در نظر گرفتن این مسئله است که خطای توالی‌یابی تنها در یک واکنش تعیین توالی قابل ردیابی است اما چند شکلی در هر دو واکنش شناسایی می‌شود از این رو دقت توالی‌یابی در این سامانه در مقایسه با سامانه‌های دیگر بالاتر است (Ansoorge *et al.*, 2009). فناوری SOLiD دارای مزایا و کاستی‌هایی است. از جمله مزایای آن می‌توان به دقت بالا (به دلیل سامانه کدگذاری دو رنگی)، هزینه به نسبت پایین هر باز توالی‌یابی شده و قابلیت تولید حجم بالای داده اشاره کرد (Metzker, 2010). با این حال، این فناوری کاستی‌هایی نیز دارد. یکی از چالش‌های اصلی، طول به نسبت کوتاه خوانش‌ها (به‌طور معمول ۳۵-۷۵ جفت باز) است که می‌تواند مونتاژ ژنوم، به ویژه در مناطق تکراری، را دشوار کند (van Dijk *et al.*, 2014). علاوه بر این، فرایند توالی‌یابی به نسبت کند است و زمان اجرای

به مناطق مابین دی‌نوکلئوتیدها که در چرخه اول ناشناخته بود هم متصل شوند و عملاً در چرخه‌های متوالی به تمام طول توالی DNA کاوشگرها متصل شده باشند. حال استنتاج رشته به این صورت هست که ما میدانیم اولین نوکلئوتید از دی‌نوکلئوتید اول متصل شده در چرخه اول همان نوکلئوتید انتهایی اداپتور و مرجع هست و بر فرض مثال اطلاع داریم که این نوکلئوتید T بوده و رنگ ساطع شده از کاوشگر آن نیز قرمز بوده است؛ برای اینکه متوجه شویم نوکلئوتید دوم چه نوکلئوتیدی هست به این توجه می‌کنیم که رنگ قرمز شامل دی‌نوکلئوتیدهای AT, CG, GC, TA هست و با توجه به اینکه نوکلئوتید اول T بوده است، از بین دی‌نوکلئوتیدهای رنگ قرمز تنها دی‌نوکلئوتیدی که حرف اول آن با T شروع می‌شود TA است، بنابراین توالی نوکلئوتید دوم A خواهد بود و اینگونه توالی دو نوکلئوتید اول شناسایی می‌شود. برای شناسایی نوکلئوتید سوم بررسی می‌کنیم که در کدام چرخه به نوکلئوتید دوم، حرف اول یک دی‌نوکلئوتید متصل شده است، این چرخه n-4 خواهد بود، با توجه به اتصال کاوشگر در چرخه پنجم به نوکلئوتید دوم، این نوکلئوتید، نوکلئوتید اول دی‌نوکلئوتید آن کاوشگر خواهد بود و اگر رنگ این کاوشگر زرد باشد، با توجه به این که نوکلئوتید دوم A بوده است و توالی دی‌نوکلئوتیدهای کاوشگرهایی که رنگ زرد ساطع می‌کنند TC, GA, CT, AG می‌باشد، پس نوکلئوتید سوم G خواهد بود. این فرایند برای هر نوکلئوتید از تمامی جایگاه اتصال دی‌نوکلئوتید کاوشگر-

⁷⁸ Sequencing error⁷⁹ Single nucleotide polymorphism

نیست، اما نقش مهمی در پیشرفت علم ژنومیک و توسعه روش‌های جدید توالی‌یابی ایفا کرده است. بسیاری از اصول و مفاهیم توسعه یافته در این پلتفرم، مانند استفاده از کدگذاری رنگی برای افزایش دقت، همچنان در فناوری‌های جدید مورد استفاده قرار می‌گیرند. با نگاهی به آینده، انتظار می‌رود که پیشرفت‌های جدید در زمینه توالی‌یابی، مانند توالی‌یابی نانومنظور و توالی‌یابی مولکول منفرد، چشم‌انداز این حوزه را بیش از پیش متحول کنند (Logsdon *et al.*, 2020). با این حال، تجربیات و دانش به دست آمده از توسعه و استفاده از فناوری SOLiD، همچنان در شکل دادن به آینده توالی‌یابی DNA نقش مهمی ایفا خواهد کرد.

فناوری توالی‌یابی؛ Semiconductor Sequencing

Ion Torrent

فناوری توالی‌یابی نیمه‌هادی یا Ion Torrent، که در سال ۲۰۱۰ توسط Jonathan Rothberg و همکارانش معرفی شد، یک روش انقلابی در زمینه توالی‌یابی DNA است که بر اساس تشخیص یون‌های هیدروژن آزاد شده در طی سنتز DNA عمل می‌کند (Rothberg *et al.*, 2011). این فناوری از یک تراشه نیمه‌هادی حاوی میلیون‌ها چاهک میکروسکوپی استفاده می‌کند که هر کدام حاوی یک رشته الگوی DNA و DNA پلیمر است. در هر چرخه توالی‌یابی، یکی از چهار نوکلئوتید dATP، dCTP، dGTP یا dTTP به محلول واکنش اضافه می‌شود. اگر نوکلئوتید اضافه شده مکمل نوکلئوتید بعدی در رشته الگو باشد، DNA پلی‌مر از آن‌ها به رشته در حال سنتز اضافه

طولانی‌تری نسبت به برخی روش‌های دیگر دارد (Liu *et al.*, 2012). فناوری SOLiD در طول سال‌ها پیشرفت‌هایی داشته است. در ابتدا در هر اجرای دستگاه تا سه گیگا باز توالی خوانده می‌شد اما در دسامبر ۲۰۰۹، نسخه‌ی جدیدی از این فناوری تحت عنوان Plus Plus 3T SOLiD معرفی شد که قادر بود در هر اجراء بیش از ۶۰ گیگا باز توالی را بخواند (Ansorge *et al.*, 2009). نسل‌های جدیدتر این پلتفرم، مانند SOLiD ۵۵۰۰، قادر به تولید تا ۳۲۰ گیگاباز داده در هر اجرا هستند (Thermo Fisher Scientific, 2021). این فناوری در طیف گسترده‌ای از کاربردها مورد استفاده قرار گرفته است، از جمله توالی‌یابی کل ژنوم (WGS)، توالی‌یابی اگزوم (WES)، توالی‌یابی RNA (RNA-Seq)، تعیین متیلاسیون DNA و مطالعات اپی‌ژنتیک (Mardis, 2008; Wang *et al.*, 2009). با ظهور و گسترش فناوری‌های رقیب، به ویژه پلتفرم Illumina که قادر به تولید خوانش‌های طولانی‌تر و با سرعت بیشتر است، استفاده از SOLiD در سال‌های اخیر کاهش یافته است (Goodwin *et al.*, 2016). با وجود این، فناوری SOLiD همچنان در برخی حوزه‌ها، به ویژه در مواردی که دقت بسیار بالا مورد نیاز است، مورد استفاده قرار می‌گیرد. به عنوان مثال، در مطالعات تشخیصی واریانت‌های نادر و در برخی کاربردهای پزشکی دقیق، دقت بالای این روش می‌تواند مزیت محسوب شود (Koboldt *et al.*, 2013). در مجموع، اگرچه فناوری SOLiD دیگر پلتفرم غالب در توالی‌یابی نسل بعدی

سنتر اضافه می‌کند. این فرایند با آزاد شدن یک یون هیدروژن همراه است (Rothberg *et al.*, 2011).

واکنش شیمیایی اصلی در این فرایند به صورت زیر است:

$$(DNA)_n + dNTP \rightarrow (DNA)_{n+1} + PPi + H^+$$

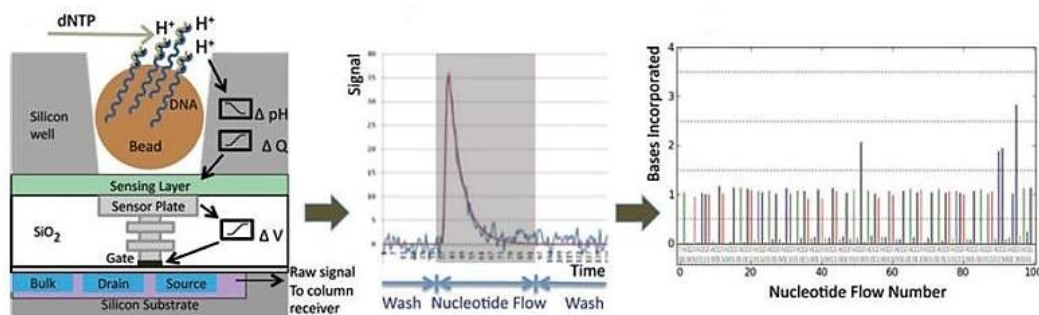
در این واکنش، $(DNA)_n$ رشته DNA در حال رشد است، dNTP نوکلئوتید اضافه شده، PPi پیروفسفات آزاد شده و H^+ یون هیدروژن آزاد شده است (Metzker, 2010). یون هیدروژن آزاد شده باعث تغییر pH در چاهک می‌شود. این تغییر pH توسط حسگر ISFET در کف چاهک تشخیص داده می‌شود. ISFET یک ترانزیستور اثر میدانی است که نسبت به تغییرات غلظت یون حساس است. تغییر pH باعث تغییر در جریان الکتریکی عبوری از ISFET می‌شود که این تغییر به عنوان سیگنال خروجی ثبت می‌شود (Toumazou *et al.*, 2013). اگر دو یا چند نوکلئوتید یکسان پشت سر هم در رشته الگو وجود داشته باشد، چندین نوکلئوتید در یک چرخه اضافه می‌شود و سیگنال قوی‌تری تولید می‌شود. به عنوان مثال، اگر دو نوکلئوتید A پشت سر هم باشند، دو یون هیدروژن آزاد می‌شود و سیگنال دو برابر قوی‌تر خواهد بود. این ویژگی به سامانه اجازه می‌دهد تا تعداد نوکلئوتیدهای یکسان پشت سر هم را تشخیص دهد (Rothberg *et al.*, 2011). پس از هر چرخه، نوکلئوتیدهای اضافه نشده شسته می‌شوند و چرخه بعدی با نوکلئوتید دیگری آغاز می‌شود. این فرایند برای صدها چرخه تکرار می‌شود تا توالی کامل قطعات DNA تعیین شود (شکل ۸). داده‌های خام حاصل از این فرایند شامل

می‌کند (Merriman *et al.*, 2012). این فرایند با آزاد شدن یک یون هیدروژن همراه است که منجر به تغییر pH در چاهک می‌شود. این تغییر pH توسط یک حسگر ISFET⁸⁰ در کف هر چاهک تشخیص داده شده و به یک سیگنال الکتریکی تبدیل می‌شود (Toumazou *et al.*, 2013). سازوکار دقیق این فرایند شامل چندین مرحله است. ابتدا، کتابخانه DNA با استفاده از روش‌های استاندارد آماده‌سازی نمونه تهیه می‌شود. این کتابخانه شامل قطعات DNA با طول حدود ۲۰۰-۴۰۰ جفت باز است که به آداپتورهای خاصی متصل شده‌اند. سپس، این قطعات به ذرات کوچک کرومی متصل می‌شوند. هر کره حاوی هزاران کپی از یک قطعه DNA منحصر به فرد است. این کره‌ها در چاهک‌های تراشه نیمه‌هادی قرار داده می‌شوند، به طوری که هر چاهک حاوی یک کره است (Quail *et al.*, 2012). در مرحله بعد، آنزیم DNA پلیمراز به هر چاهک اضافه می‌شود. این آنزیم مسئول سنتز رشته مکمل DNA است. DNA پلیمراز مورد استفاده در این فناوری به‌طور معمول یک نسخه اصلاح شده از Taq پلیمراز است که برای عملکرد بهتر در شرایط توالی‌یابی بهینه شده است (Ambardar *et al.*, 2016). سپس، چرخه‌های توالی‌یابی آغاز می‌شود. در هر چرخه، یکی از چهار نوکلئوتید به صورت محلول به تراشه اضافه می‌شود. اگر نوکلئوتید اضافه شده مکمل نوکلئوتید بعدی در رشته الگو باشد، DNA پلیمراز آن را به رشته در حال

⁸⁰ Ion-sensitive field-effect transistor

دقت در تشخیص تعداد نوکلئوتیدهای یکسان پشت سر هم (هومو پلیمرها) است. با افزایش طول هومو پلیمر، تفکیک سیگنال‌ها دشوارتر می‌شود و این می‌تواند منجر به خطا در تعیین توالی شود. علاوه بر این، طول خوانش در این روش به نسبت کوتاه است (به‌طور معمول حدود ۲۰۰-۴۰۰ جفت باز) که می‌تواند برای برخی کاربردها محدودکننده باشد (Quail *et al.*, 2012). با وجود این کاستی‌ها، فناوری Ion Torrent در بسیاری از زمینه‌های تحقیقاتی و بالینی مورد استفاده قرار گرفته است. از جمله کاربردهای آن می‌توان به توالی‌یابی آگروم، توالی‌یابی هدفمند ژن‌ها، تشخیص جهش‌های سرطانی و مطالعات متازنومیکس اشاره کرد (Ambardar *et al.*, 2016).

یک سری سیگنال‌های الکتریکی است که باید به توالی DNA ترجمه شود. این کار توسط الگوریتم‌های پیچیده نرم‌افزاری انجام می‌شود که سیگنال‌ها را تفسیر کرده و به حروف A، C، G و T ترجمه می‌کنند (Merriman *et al.*, 2012). یکی از مزایای اصلی تکنولوژی Ion Torrent، سرعت بالای آن است. از آنجاکه این روش بر اساس تشخیص الکترونیکی یون‌ها عمل می‌کند و نیازی به تصویربرداری نوری ندارد، می‌تواند بسیار سریع‌تر از روش‌های توالی‌یابی مبتنی بر فلورسانس عمل کند. علاوه بر این، این فناوری به نسبت ارزان‌تر است زیرا از تراشه‌های نیمه‌هادی استاندارد استفاده می‌کند که تولید آن‌ها ارزان‌تر است (Liu *et al.*, 2012). با این حال، این فناوری کاستی‌هایی نیز دارد. یکی از چالش‌های اصلی،



شکل ۸ - توالی‌یابی از طریق فناوری توالی‌یابی Ion Torrent؛ Semiconductor Sequencing
Figure 8. Sequencing through Semiconductor Sequencing Technology (Ion Torrent)

همچنین، الگوریتم‌های جدید برای تفسیر داده‌ها و تصحیح خطاها توسعه یافته‌اند که دقت کلی سامانه را بهبود بخشیده‌اند (Goodwin *et al.*, 2016). در مجموع، فناوری توالی‌یابی نیمه‌هادی Ion Torrent یک روش

در سال‌های اخیر، پیشرفت‌های قابل توجهی در این فناوری صورت گرفته است. به عنوان مثال، نسل جدید سامانه‌های Ion Torrent قادر به تولید خوانش‌های طولانی‌تر (تا حدود ۶۰۰ جفت باز) و با دقت بالاتر هستند.

فراهم می‌کند. این ویژگی باعث می‌شود که نوکلئوتیدهای اضافه‌شده به زنجیره در حال ساخت بلافاصله پس از اضافه شدن در زمان واقعی (Real-time) شناسایی شوند، که این امر می‌تواند منجر به دقت بسیار بالا و کاهش خطاها در هنگام همانندسازی شود. این ویژگی‌ها سبب شده‌اند که SMS به یکی از روش‌های محبوب برای مطالعه ژنوم‌های پیچیده و ناهمگن تبدیل شود، زیرا امکان تحلیل دقیق و بدون خطا را فراهم می‌کند (Schadt *et al.*, 2010).

در میان فناوری‌های اصلی این نسل، Heliscope اولین فناوری بود که در سال ۲۰۰۸ توسط شرکت Helicos Biosciences معرفی شد (Thompson & Milos, 2011). این فناوری از روش توالی‌یابی فلورسنت تک مولکولی استفاده می‌کرد و توانایی تولید میلیون‌ها خوانش کوتاه به‌طور همزمان را داشت. روش Heliscope با استفاده از یک سطح شیشه‌ای پوشیده شده با DNA تک رشته‌ای و افزودن نوکلئوتیدهای نشان‌دار فلورسنت، امکان شناسایی توالی را فراهم می‌کرد. این روش به دلیل عدم نیاز به تکثیر DNA، خطاهای مرتبط با PCR را کاهش می‌داد و امکان مطالعه مستقیم تغییرات اپی‌ژنتیکی را فراهم می‌کرد (Braslavsky *et al.*, 2003). پس از Heliscope، در سال ۲۰۱۱، شرکت Pacific Biosciences فناوری PacBio را معرفی کرد که بر اساس توالی‌یابی تک مولکولی زمان واقعی (SMRT) عمل می‌کند (Eid *et al.*, 2009). این فناوری از یک چاهک

قدرتمند و مقرون به صرفه برای توالی‌یابی DNA است که با ترکیب اصول بیوشیمیایی و الکترونیکی، امکان توالی‌یابی سریع و دقیق را فراهم می‌کند. با پیشرفت‌های مداوم در این زمینه، انتظار می‌رود که این فناوری نقش مهمی در آینده ژنومیکس و پزشکی شخصی‌سازی شده ایفا کند.

تعیین توالی مولکول منفرد (Singel molecule sequencing)

سامانه توالی‌یابی مولکول منفرد (Single molecule sequencing)، که به عنوان نسل سوم توالی‌یابی (Next-generation sequencing) نیز شناخته می‌شود، یک روش پیشرفته و کارآمد برای توالی‌یابی DNA و RNA است که به تحلیل مستقیم مولکول‌های منفرد اسید نوکلئیک بدون نیاز به تکثیر از طریق PCR می‌پردازد. این روش با حذف مراحل تکثیر، خطاهای مرتبط با PCR را کاهش می‌دهد و امکان بررسی تنوع ژنتیکی و اپی‌ژنتیکی در سطح تک مولکول را فراهم می‌سازد، و هدف آن غلبه بر کاستی‌های روش‌های سنتی توالی‌یابی است. در روش‌های نسل دوم، قبل از توالی‌یابی، برای افزایش شدت سیگنال، مرحله تکثیر از طریق PCR انجام می‌شد تا تحلیل بهتری ممکن شود، اما این فرایند می‌توانست منجر به خطاهایی شود و دقت نتایج را تحت تأثیر قرار دهد (Mardis, 2013). فناوری توالی‌یابی مولکول منفرد (SMS) امکان شناسایی توالی بازهای نیتروژنی در یک مولکول DNA تک‌رشته‌ای را به صورت مستقیم و در طول فرایند سنتز، بدون نیاز به تکثیر،

در ابتدا با چالش‌هایی در زمینه دقت روبرو بود. روش توالی‌یابی نانومفذ قابلیت تولید خوانش‌های بسیار بلند را دارد، اما در مراحل اولیه با مشکلاتی در زمینه دقت خوانش مواجه بود. با این حال، هر سه فناوری از زمان معرفی خود، پیشرفت‌های قابل توجهی در زمینه دقت، طول خوانش و سرعت توالی‌یابی داشته‌اند.

فناوری‌های نسل سوم توالی‌یابی DNA کاربردهای گسترده‌ای در زمینه‌های مختلف علوم زیستی و پزشکی پیدا کرده‌اند. این فناوری‌ها امکان مطالعه ساختارهای پیچیده ژنومی مانند تکرارهای طولانی، جایجایی‌های کروموزومی و تغییرات ساختاری بزرگ را فراهم کرده‌اند (Chaisson *et al.*, 2015). همچنین، این فناوری‌ها در زمینه مطالعات اپی‌ژنتیک، بررسی بیان ژن و توالی‌یابی ترنسکریپتوم کامل نقش مهمی ایفا می‌کنند (Stark *et al.*, 2019a). در حوزه تشخیص پزشکی، فناوری‌های نسل سوم توالی‌یابی DNA امکان شناسایی سریع و دقیق عوامل بیماری‌زا، بررسی مقاومت آنتی‌بیوتیکی و تشخیص بیماری‌های ژنتیکی پیچیده را فراهم کرده‌اند (Goldberg *et al.*, 2015). در زمینه مطالعات تکاملی و اکولوژیکی، این فناوری‌ها امکان بررسی جامع‌تر تنوع ژنتیکی در جمعیت‌های مختلف و مطالعه میکروبیوم‌های پیچیده را فراهم کرده‌اند (Goodwin *et al.*, 2016). علاوه بر این، فناوری‌های نسل سوم توالی‌یابی DNA در زمینه کشاورزی و اصلاح نباتات نیز کاربردهای گسترده‌ای پیدا کرده‌اند. این فناوری‌ها امکان شناسایی دقیق‌تر ژن‌های مرتبط با صفات مهم زراعی و بهبود فرایندهای

نانوفوتونیک که ^{۸۱} ZMWs نامیده می‌شوند استفاده می‌کند که در آن یک مولکول DNA پلی‌مرز به کف چاهک متصل شده است. با افزودن نوکلئوتیدهای نشان‌دار فلورسنت، فرایند سنتز DNA در زمان واقعی مشاهده می‌شود. روش PacBio قادر به تولید خوانش‌های بلندتر نسبت به Heliscope بود و امکان مطالعه ساختارهای پیچیده‌تر ژنومی را فراهم کرد. این فناوری همچنین امکان شناسایی تغییرات اپی‌ژنتیکی مانند متیلاسیون DNA را بدون نیاز به پردازش شیمیایی اضافی فراهم می‌کند (Flusberg *et al.*, 2010). در سال ۲۰۱۴، شرکت Oxford Nanopore Technologies فناوری Nanopore sequencing را به بازار عرضه کرد (Jain *et al.*, 2016). این فناوری از یک نانومفذ^{۸۲} پروتئینی برای عبور رشته DNA و شناسایی نوکلئوتیدها استفاده می‌کند. در این روش، یک رشته DNA از طریق یک نانومفذ پروتئینی که در یک غشای دو لایه لیپیدی قرار دارد، عبور داده می‌شود. با عبور هر نوکلئوتید از نانومفذ، تغییرات جریان یونی ایجاد شده اندازه‌گیری می‌شود و بر اساس این تغییرات، نوع نوکلئوتید شناسایی می‌شود. روش نانومفذ قادر به تولید خوانش‌های بسیار بلند است و امکان توالی‌یابی مستقیم RNA را نیز فراهم می‌کند (Garalde *et al.*, 2018). هر یک از این فناوری‌ها مزایا و چالش‌های خاص خود را دارند. روش Heliscope با وجود دقت بالا، کاستی‌هایی در طول خوانش داشت. PacBio امکان تولید خوانش‌های بلندتر را فراهم کرد، اما

⁸¹ Zero-Mode Waveguides

⁸² Nanopore

ژنومی به قطعات کوچک‌تر (حدود ۳۰-۳۵ نوکلئوتید) شکسته می‌شود و تک رشته می‌شود و با استفاده از آنزیم ترمینال ترانسفراز به انتهای ۳' این قطعات تک رشته، دزوکسی آدنوزین تری فسفات را اضافه کرده تا یک دم پلی dA به طول بیش از ۵۰ نوکلئوتید ایجاد شود. به سطح یک سلول جریان که با اولیگو dT به طول ۵۰ نوکلئوتید پوشیده شده است، متصل می‌شوند (Harris *et al.*, 2008; Shokralla *et al.*, 2012; Ansorge *et al.*, 2009; Gupta *et al.*, 2008; Thompson & Steinmann, 2010). پس از اتصال DNA به سطح فلوسل، فرایند توالی‌یابی آغاز می‌شود. در هر چرخه توالی‌یابی، یکی از چهار نوکلئوتید ختم دهنده برگشت پذیر A، C، G، یا T که با رنگ‌های فلورسنت نشان‌دار شده‌اند و گروه ترمیناتور که به تنهایی و یا بصورت متصل به واحد رنگی، به باز آلی متصل است، به محلول واکنش اضافه می‌شود و الیگو dTها بر روی فلوسل به عنوان آغازگر عمل می‌کنند و این نوکلئوتیدها توسط آنزیم DNA پلی‌مراز به رشته در حال رشد اضافه می‌شوند. در این روش به دلیل مشخص بودن نوکلئوتیدهای مشابهی که در هر چرخه اضافه می‌شود، بر خلاف روش Solexa تمامی نوکلئوتیدهای ختم دهنده برگشت‌پذیر، تنها با یک رنگ فلورسانس برچسب‌گذاری می‌شوند. آنزیم DNA پلی‌مراز مورد استفاده در این سامانه نیز، یک نسخه اصلاح شده از آنزیم DNA پلی‌مراز 29q است که برای عملکرد بهتر در شرایط توالی‌یابی بهینه شده است (Bowers *et al.*, 2009; Thompson & Steinmann, 2010). این آنزیم به دلیل دقت بالا و

به‌نژادی‌گیاهی را فراهم کرده‌اند (Michael & VanBuren, 2020).

با توجه به روند سریع پیشرفت در این حوزه، انتظار می‌رود که فناوری‌های نسل سوم توالی‌یابی DNA همچنان به تکامل خود ادامه دهند. پیش‌بینی می‌شود که در آینده، این فناوری‌ها با بهبود بیشتر در دقت، سرعت و هزینه توالی‌یابی، نقش مهم‌تری در تحقیقات پایه و کاربردی علوم زیستی و پزشکی ایفا کنند. همچنین، انتظار می‌رود که با توسعه روش‌های جدید تجزیه و تحلیل داده‌ها و الگوریتم‌های پیشرفته، امکان استخراج اطلاعات بیشتر و دقیق‌تر از داده‌های توالی‌یابی فراهم شود (Logsdon *et al.*, 2020). جزئیات و ویژگی پلتفرم‌های اصلی توالی‌یابی نسل سوم به شرح زیر است.

فناوری توالی‌یابی Heliscope

فناوری توالی‌یابی Heliscope، که توسط شرکت Helicos Biosciences در سال ۲۰۰۸ معرفی شد، اولین پلتفرم تجاری توالی‌یابی مولکول منفرد بود که انقلابی در زمینه توالی‌یابی DNA ایجاد کرد (Thompson & Milos, 2011; Thompson & Steinmann, 2010). این روش، که بر اساس روش سنتز توالی‌یابی عمل می‌کند، امکان توالی‌یابی مستقیم مولکول‌های DNA منفرد را بدون نیاز به تکثیر PCR فراهم می‌کند، و در نتیجه خطاهای مرتبط با تکثیر را حذف می‌نماید (Braslavsky *et al.*, 2003). سازوکار عمل Heliscope شامل چندین مرحله اصلی است که با آماده‌سازی نمونه آغاز می‌شود. در این مرحله، DNA

توانایی اضافه کردن نوکلئوتیدهای نشان‌دار شده انتخاب شده است (Shokralla *et al.*, 2012; Ansorge *et al.*, 2009; Gupta *et al.*, 2008; Thompson & Steinmann, 2010). همانطور که بیان شد در هر مرحله توالی‌یابی نوکلئوتیدهای ختم دهنده برگشت‌پذیر مشابهی وارد واکنش می‌شوند. پس از قرارگیری نوکلئوتیدها در زنجیره در حال ساخت، بدلیل تداخلی که بواسطه گروه رنگی و ترمیناتور ایجاد می‌شود، از عملکرد آنزیم پلی‌مراز ممانعت به عمل می‌آید. در نتیجه سنتز رشته جدید بطور موقت، متوقف می‌شود. سپس یک شستشو انجام می‌شود تا نوکلئوتیدهایی که به زنجیره‌های در حال ساخت وارد نشده‌اند از محیط حذف گردند. پس از اضافه شدن نوکلئوتید، سطح سلول جریان با استفاده از میکروسکوپ فلورسانس تصویربرداری می‌شود. نوکلئوتیدهای متصل شده به عنوان نقاط روشن در تصویر ظاهر می‌شوند. پس از تصویربرداری، گروه فلورسنت با استفاده از تابش نور لیزر شکسته می‌شود (Pushkarev *et al.*, 2009; Shokralla *et al.*, 2012; Ansorge *et al.*, 2009; Shokralla *et al.*, 2012; Ansorge *et al.*, 2009; Gupta *et al.*, 2008; Steinmann, 2010) و در نهایت جدا شدن رنگ فلورسانس به همراه گروه ترمیناتور آن دسته از نوکلئوتیدهایی که به زنجیره در حال سنتز وارد شده‌اند، نه‌تنها منجر به ساطع شدن نور و ثبت سیگنال مربوطه توسط دستگاه می‌شود بلکه منجر به از بین رفتن اثر مهاری آنها و در نتیجه ادامه پروسه سنتز نیز می‌گردد (Pushkarev *et al.*, 2009; Shokralla *et al.*, 2012; Ansorge *et al.*, 2009; Gupta *et al.*, 2008;

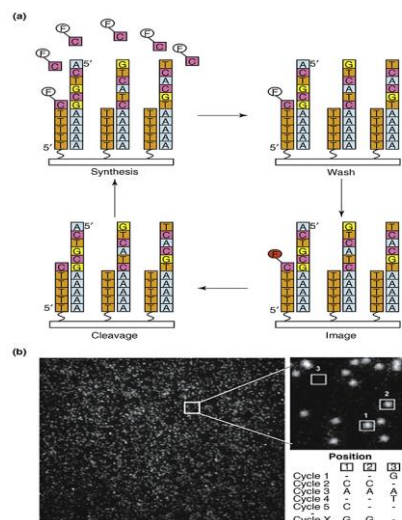
Thompson & Steinmann, 2010). این فرایند برای صدها چرخه تکرار می‌شود تا توالی کامل قطعات DNA تعیین شود (شکل ۹). یکی از ویژگی‌های منحصر به فرد Heliscope، توانایی آن در توالی‌یابی مولکول‌های RNA به‌طور مستقیم است. در این روش، RNA ابتدا به cDNA تبدیل می‌شود و سپس مراحل مشابه توالی‌یابی DNA انجام می‌شود. این قابلیت، Heliscope را به ابزاری قدرتمند برای مطالعات ترنسکرپتوم تبدیل کرده است (Ozsolak *et al.*, 2009; Thompson & Steinmann, 2010). سامانه Heliscope قادر به تولید میلیون‌ها خوانش در هر اجرا است، با طول خوانش متوسط حدود ۳۵ نوکلئوتید. اگرچه این طول خوانش به نسبت کوتاه است، اما به دلیل حجم بالای داده‌های تولید شده و دقت بالا، این سامانه برای بسیاری از کاربردها مناسب است (Thompson & Milos, 2011).

یکی از مزایای اصلی Heliscope، توانایی آن در تشخیص تغییرات اپی‌ژنتیکی مانند متیلاسیون DNA است. این سامانه می‌تواند نوکلئوتیدهای متیله شده را به‌طور مستقیم تشخیص دهد، بدون نیاز به پیش‌پردازش شیمیایی نمونه که در سایر روش‌های توالی‌یابی مورد نیاز است (Flusberg *et al.*, 2010). با این حال، Heliscope نیز با چالش‌هایی روبرو است. یکی از کاستی‌های اصلی آن، نرخ خطای به نسبت بالا در توالی‌های هوموپلیمر (توالی‌های متشکل از نوکلئوتیدهای یکسان پشت سر هم) است. این مشکل به دلیل دشواری در تشخیص تعداد دقیق نوکلئوتیدهای اضافه شده در هر چرخه رخ می‌دهد

Heliscope، توسعه روش‌های جدید برای آماده‌سازی نمونه است.

این روش‌ها امکان کار با مقادیر بسیار کم DNA یا RNA را فراهم می‌کنند، که برای کاربردهایی مانند توالی‌یابی سلول‌های منفرد بسیار مهم است (Ozsolak *et al.*, 2010). در مجموع، فناوری توالی‌یابی Heliscope یک روش قدرتمند برای توالی‌یابی مولکول منفرد است که با حذف نیاز به تکثیر PCR، امکان توالی‌یابی مستقیم و دقیق DNA و RNA را فراهم می‌کند. این فناوری با ترکیب اصول بیوشیمیایی و تصویربرداری پیشرفته، امکان توالی‌یابی با حجم بالا و دقت زیاد را فراهم می‌کند. با پیشرفت‌های مداوم در این زمینه، انتظار می‌رود که Heliscope همچنان نقش مهمی در تحقیقات ژنومیکس و ترنسکرپتومیکس ایفا کند.

(Treangen & Salzberg, 2012). با وجود این چالش‌ها، Heliscope در بسیاری از زمینه‌های تحقیقاتی مورد استفاده قرار گرفته است. از جمله کاربردهای آن می‌توان به توالی‌یابی ژنوم کامل، تجزیه و تحلیل بیان ژن، مطالعات اپی‌ژنتیک و تشخیص واریانت‌های ژنتیکی اشاره کرد (Ozsolak & Milos, 2011). در سال‌های اخیر، پیشرفت‌های قابل توجهی در این فناوری صورت گرفته است. به عنوان مثال، نسل جدید سامانه‌های Heliscope قادر به تولید خوانش‌های طولانی‌تر (تا حدود ۵۵ نوکلئوتید) و با دقت بالاتر هستند. همچنین، الگوریتم‌های جدید برای تفسیر داده‌ها و تصحیح خطاها توسعه یافته‌اند که دقت کلی سامانه را بهبود بخشیده‌اند (Pushkarev *et al.*, 2009). یکی دیگر از پیشرفت‌های مهم در تکنولوژی



شکل ۹ - فناوری توالی‌یابی Heliscope
Figure 9. Heliscope sequencing technology

قطعات بلند (به‌طور معمول ۱/۵ کیلوباز) شکسته می‌شود و به هر دو طرف آن آداپتورهای مخصوص غیر خطی و که ساختارهای سنجاق سری و حلقوی به نام SMRTbell دارند متصل می‌شود (Travers *et al.*, 2010). این ساختارهای حلقوی آداپتورها باعث می‌شود که قطعه مورد نظر پس از واسرشت‌سازی به حالت حلقوی تبدیل گردد و همچنین امکان خوانش مکرر یک قطعه DNA را فراهم می‌کنند، که به بهبود دقت توالی‌یابی کمک می‌کند. پیش از اینکه قطعات DNA به چاهک‌های ZMW منتقل شوند، آغازگرهایی به هر کدام از آداپتورهای حلقوی متصل می‌شود و در نهایت بعد از اینکه اتصال آنزیم پلیمرز صورت گرفت، نوکلئوتیدهایی که با رنگ فلورسنت برچسب‌گذاری شده‌اند همراه با دیگر عوامل مورد نیاز برای انجام واکنش پلیمریزاسیون، جهت انجام عملیات توالی‌یابی به داخل چاهک‌های ZMW منتقل می‌شوند. در طول پلیمریزاسیون، از طریق اتصال هر نوکلئوتید، گروه رنگی فلورسنت خاص آن نوکلئوتید نیز آزاد شده و نور آن ساطع می‌شود که نور فلورسنت ساطع شده بواسطه دوربین‌های CCD ثبت و پردازش می‌شود. در این حالت هر ۲ رشته DNA توالی‌یابی میشوند که این موضوع میتواند به افزایش میزان صحت توالی‌یابی کمک کند. DNA پلیمرز مورد استفاده در SMRT یک نسخه اصلاح شده از DNA پلیمرز $\phi 29$ است که برای عملکرد بهتر در شرایط توالی‌یابی بهینه شده است (Korlach *et al.*, 2010). این آنزیم به دلیل دقت بالا، سرعت زیاد و توانایی خواندن قطعات طولانی

فناوری Single Molecule Real-Time; SMRT

فناوری SMRT (Single Molecule Real-Time) یک روش پیشرفته توالی‌یابی DNA است که توسط شرکت Pacific Biosciences (PacBio) در سال ۲۰۱۱ معرفی شد (Eid *et al.*, 2009). این فناوری انقلابی در زمینه توالی‌یابی نسل سوم ایجاد کرد، زیرا امکان توالی‌یابی مولکول‌های DNA منفرد را در زمان واقعی و با طول خوانش بسیار بلند فراهم می‌کند. سازوکار اصلی SMRT بر اساس مشاهده مستقیم سنتز DNA توسط یک DNA پلیمرز منفرد است. این تکنیک از تراشه‌هایی^{۸۳} استفاده می‌کند که دارای یک فیلم فلزی بسیار باریک می‌باشند. این فیلم فلزی حاوی هزاران چاهک نانوفوتونیک است که ZMWs^{۸۴} نامیده می‌شوند. ZMWs ساختارهای نانومقیاسی هستند که حجم مشاهده را به حدود ۲۰ زپتولیترا (۱۰^{-۲۱}-لیتر) محدود می‌کنند. در این روش، از یک آنزیم DNA پلیمرز اصلاح شده استفاده می‌شود که به کف چاهک‌های ZMW متصل شده است (Levene *et al.*, 2003). محفظه داخلی این چاهک‌ها بطور پیوسته توسط دوربین‌های CCD^{۸۵} رصد می‌شود که این امر امکان مشاهده فعالیت یک مولکول منفرد و تعیین توالی همزمان با سنتز در زمان واقعی DNA را فراهم می‌کند (Korlach *et al.*, 2008). فرایند توالی‌یابی SMRT با آماده‌سازی نمونه آغاز می‌شود. در این مرحله، DNA ژنومی به

⁸³ Chip

⁸⁴ Zero-Mode Waveguides

⁸⁵ Charged-Coupled Device

غیرمتمیله شده تأخیر خواهد داشت (Flusberg *et al.*, 2010).

در روش توالی‌یابی SMRT برخلاف سایر روش‌های توالی‌یابی، نیازی نیست که پس از هر مرحله که نوکلئوتیدها به مخلوط واکنش افزوده میشوند، یک مرحله شستشو انجام شود. این مسئله، در کاهش زمان انجام این پروسه نقش مهمی ایفا میکند شکل ۱۰ نمایی از این روش توالی‌یابی را نشان داده است. یکی از مزایای اصلی SMRT، توانایی آن در تولید خوانش‌های بسیار بلند است. درحالی که طول متوسط خوانش در نسخه‌های اولیه این فناوری حدود ۱-۲ کیلوباز بود، نسخه‌های جدیدتر قادر به تولید خوانش‌هایی با طول متوسط ۱۰-۲۰ کیلوباز و حتی تا ۱۰۰ کیلوباز هستند (Rhoads & Au, 2015). این خوانش‌های بلند برای حل چالش‌های مربوط به مناطق تکراری ژنوم و سرهم‌بندی⁸⁸ *de novo* بسیار ارزشمند هستند. علاوه بر این، SMRT قادر به تشخیص تغییرات اپی‌ژنتیکی مانند متیلاسیون DNA به‌طور مستقیم و بدون نیاز به پیش‌پردازش شیمیایی نمونه است. این امر به دلیل حساسیت سامانه به تغییرات کینتیکی در فعالیت DNA پلیمراز است که توسط تغییرات اپی‌ژنتیکی ایجاد می‌شود (Clark *et al.*, 2012). با این حال، SMRT نیز با چالش‌هایی روبرو است. یکی از کاستی‌های اصلی آن، نرخ خطای به نسبت بالا در هر خوانش منفرد (حدود ۱۱-۱۵٪) است. این مشکل تا حدی با استفاده از

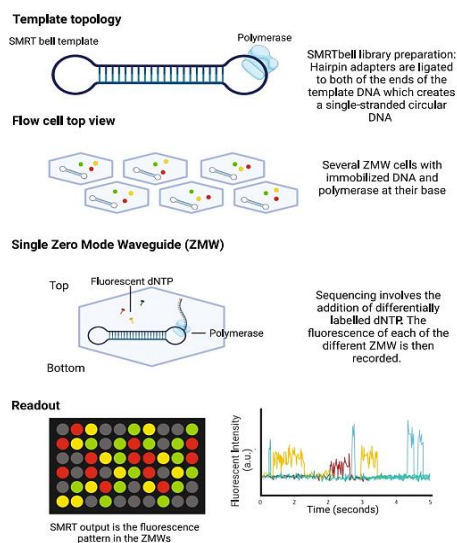
DNA انتخاب شده است. نوکلئوتیدهای مورد استفاده در این سامانه با چهار رنگ فلورسنت متفاوت (قرمز، آبی، نارنجی و سبز) نشان‌دار شده‌اند و نه به فلوروفرم باز نوکلئوتیدی بلکه به گروه فسفات γ نوکلئوتید متصل هستند. این طراحی منحصر به فرد منجر به کاهش ممانعت فضایی ناشی از رنگ فلورسانس شده و اجازه می‌دهد تا پس از اتصال نوکلئوتید به رشته در حال رشد، گروه فلورسنت به‌طور طبیعی جدا شود و پلیمراز بتواند به سنتز DNA ادامه دهد و راندمان پلیمریزه‌شدن افزایش یابد (Eid *et al.*, 2009). هنگامی که DNA پلیمراز یک نوکلئوتید را به رشته در حال رشد اضافه می‌کند، پالس فلورسانس مربوط به آن نوکلئوتید ثبت می‌شود و تعیین توالی صورت گیرد. طول مدت این پالس که به آن زمان اقامت⁸⁶ یا گفته می‌شود نشان‌دهنده زمانی است که نوکلئوتید در سایت فعال آنزیم باقی می‌ماند. فاصله بین دو پالس متوالی که به آن زمان بین پالس‌ها (IPD)⁸⁷ گفته می‌شود نشان‌دهنده سرعت اضافه شدن نوکلئوتیدها است (Chaisson *et al.*, 2015). این اطلاعات کینتیکی نه‌تنها برای تعیین توالی DNA استفاده می‌شود، بلکه می‌تواند برای شناسایی تغییرات اپی‌ژنتیکی مانند متیلاسیون DNA نیز مورد استفاده قرار گیرد. به این صورت که هنگامی که یک پلیمراز توالی‌یابی با نوکلئوتیدهایی در رشته DNA حاوی یک اصلاح (اپی‌ژنتیکی) مانند برای مثال 6-Methyl adenosine modification مواجه می‌شود، IPD نسبت به DNA

⁸⁶ Pulse width

⁸⁷ Interpulse duration (IPD)

⁸⁸ Assembly

خوانش‌های مکرر یک قطعه DNA به نام ^{89}CCS و الگوریتم‌های پیشرفته تصحیح خطا قابل حل است (Wenger *et al.*, 2019). با وجود این چالش‌ها، SMRT در بسیاری از زمینه‌های تحقیقاتی مورد استفاده قرار گرفته است. از جمله کاربردهای آن می‌توان به توالی‌یابی ژنوم کامل، مطالعات ترنسکریپتوم، تجزیه و تحلیل ساختارهای کروموزومی، مطالعات اپی‌ژنتیک و تشخیص واریانت‌های ساختاری اشاره کرد (Ardui *et al.*, 2018).



شکل ۱۰- نمایشی از روش توالی‌یابی SMRT Single Molecule Real-Time; SMRT
Figure 10. A view of Single Molecule Real-Time; SMRT sequencing method

ژنوم‌های پیچیده مهم است (Pollard *et al.*, 2018). در مجموع، فناوری توالی‌یابی SMRT یک روش قدرتمند برای توالی‌یابی مولکول منفرد است که با ترکیب خوانش‌های بلند، توانایی تشخیص تغییرات اپی‌ژنتیکی و امکان توالی‌یابی در زمان واقعی، افق‌های جدیدی را در زمینه ژنومیکس گشوده است. با پیشرفت‌های مداوم در این زمینه، انتظار می‌رود که SMRT همچنان نقش مهمی در تحقیقات ژنومیکس، از جمله در زمینه‌هایی مانند پزشکی شخصی‌سازی شده، مطالعات تکاملی و اکولوژی مولکولی ایفا کند.

فناوری توالی‌یابی Oxford Nanopore Technologies

در سال ۲۰۱۱، شرکت Oxford Nanopore Technologies روش نوینی در عرصه توالی‌یابی ژنوم به

در سال‌های اخیر، پیشرفت‌های قابل توجهی در فناوری SMRT صورت گرفته است. به عنوان مثال، معرفی سامانه Sequel II در سال ۲۰۱۹ منجر به افزایش چشمگیر در خروجی داده و بهبود کیفیت خوانش‌ها شده است. این سامانه از تراشه‌های SMRT Cell 8M استفاده می‌کند که حاوی حدود ۸ میلیون ZMW است و می‌تواند تا چهار ترابایت داده در هر اجرا تولید کند (Wenger *et al.*, 2019). یکی دیگر از پیشرفت‌های مهم، توسعه روش‌های جدید برای آماده‌سازی نمونه است. این روش‌ها امکان کار با مقادیر کمتر DNA را فراهم می‌کنند و می‌توانند قطعات DNA با طول بیش از ۱۰۰ کیلوباز را حفظ کنند. این امر به ویژه برای کاربردهایی مانند تشخیص واریانت‌های ساختاری بزرگ و سرهم‌بندی

(به‌طور معمول ۵ کیلو باز یا بیشتر) شکسته می‌شود و به آداپتورهای مخصوص متصل می‌شود. این آداپتورها شامل یک "موتور مولکولی" هستند که به‌طور معمول یک هلیکاز یا یک پروتئین اتصال‌یابنده به DNA تک رشته‌ای است. نقش این موتور مولکولی، کنترل سرعت عبور DNA از نانومنفذ (Caldwell & Spies, 2017; Japrun et al., 2014; Manrao et al., 2012). پس از آماده‌سازی، نمونه به سلول جریان MinION که حاوی نانومنفذها است، اضافه می‌شود. این نانومنفذها طوری طراحی شده‌اند که تنها DNA تک‌رشته‌ای می‌تواند از آن‌ها عبور کند. در دو طرف این کانال‌ها، ۲ الکتروود قرار داده شده‌اند که جریان یونی را درون کانال‌ها ایجاد می‌کنند. هنگامی که یک مولکول DNA یا RNA وارد نانومنفذ می‌شود، حضور آن باعث تغییر در جریان یونی می‌شود. این تغییرات جریان برای هر نوکلئوتید یا توالی کوتاه از نوکلئوتیدها (به‌طور معمول ۶-۵ نوکلئوتید) منحصر به فرد است و به عنوان "امضای الکتریکی" آن توالی عمل می‌کند و دستگاه می‌تواند توالی نوکلئوتیدی را به ترتیب عبور DNA تشخیص دهد (Laszlo et al., 2014). این سیگنال‌های الکتریکی توسط یک حسگر ASIC^{۹۳} ثبت می‌شوند و سپس توسط الگوریتم‌های پیچیده یادگیری ماشین به توالی DNA یا RNA ترجمه می‌شوند (شکل ۱۱). ورود DNA تک‌رشته‌ای به نانومنفذ نیز به وسیله ولتاژ مثبت ایجاد شده در طرف دیگر نانومنفذ انجام می‌شود، زیرا DNA

نام Nanopore Sequencing معرفی کرد. پلتفرم تجاری این فناوری پیشرفته، که با نام MinION در سال ۲۰۱۴ به بازار عرضه شد، تحولی شگرف در زمینه تجزیه و تحلیل مولکول‌های منفرد ایجاد نمود. این روش مبتکرانه، قابلیت خواندن مستقیم و طولانی توالی‌های DNA و RNA را میسر ساخت (Jain et al., 2016). سازوکار اصلی این روش بر پایه تغییرات جریان یونی که هنگام عبور مولکول‌های DNA یا RNA از نانومنفذهای پروتئینی به قطر ۱/۵ نانومتر و از جنس آلفا-همولایزین^{۹۰} که در یک غشای مصنوعی تعبیه شده‌اند ایجاد می‌شود (Deamer et al., 2016). در واقع در این سامانه، برای ایجاد نانومنفذها از آلفا-همولایزین (که یک پروتئین هپتامر^{۹۱} است و توسط باکتری استافیلوکوکوس اورئوس^{۹۲} ترشح می‌گردد و به‌طور نرمال در غشاهای بیولوژیکی منفذ ایجاد می‌کند) و یا نانومنفذهای شیمیایی، استفاده می‌شود. این نانومنفذها در غشای لیپیدی مصنوعی دو محفظه الکتروولت را از هم جدا می‌کند و با اعمال یک ولتاژ در عرض این غشا، یک جریان یونی ثابت از طریق نانومنفذها برقرار می‌شود. دستگاه MinION، که به اندازه یک درایو USB است، شامل تراشه‌ای با صدها نانومنفذ است که به‌طور همزمان کار می‌کنند و امکان توالی‌یابی در مقیاس بزرگ را فراهم می‌کنند. فرایند توالی‌یابی با آماده‌سازی نمونه آغاز می‌شود. در این مرحله، DNA ژنومی به قطعات بلند

⁹⁰ Alpha-hemolysin

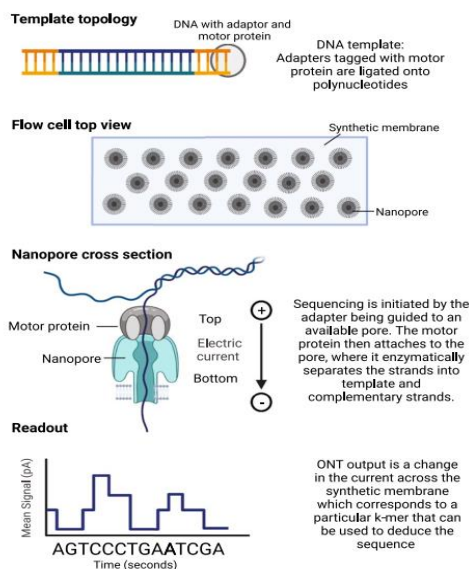
⁹¹ Heptamer

⁹² Staphylococcus aureus

⁹³ Application-Specific Integrated Circuit

هزینه بسیار پایین‌تر انجام شود (Pareek *et al.*, 2011; Steinbock *et al.*, 2015; Gupta *et al.*, 2008; Ansorge, 2009).

دارای بار منفی است و به سوی ولتاژ مثبت جذب می‌شود. فناوری توالی‌یابی نانومغذ از پلیمریزه‌شدن برای تعیین توالی‌یابی استفاده نمی‌کند و این موضوع باعث شده که توالی‌یابی موجوداتی مانند انسان طی چند ساعت و با



شکل ۱۱- توالی‌یابی از طریق روش فناوری توالی‌یابی Oxford Nanopore

Figure 11. Sequencing through Oxford Nanopore sequencing technology

است که توسط تغییرات اپی‌ژنتیکی ایجاد می‌شود (Simpson *et al.*, 2017). یکی از مزایای اصلی MinION، توانایی آن در تولید خوانش‌های بسیار بلند است. در حالی که طول متوسط خوانش در این سامانه حدود ۲۰-۱۰ کیلوباز است، خوانش‌هایی با طول بیش از ۲ مگاباز نیز گزارش شده است (Payne *et al.*, 2019). این خوانش‌های بلند برای حل چالش‌های مربوط به مناطق تکراری ژنوم، تشخیص واریانت‌های ساختاری و سرهم‌بندی *de novo* بسیار ارزشمند هستند و از طرفی این موضوع باعث شده که توالی‌یابی موجوداتی مانند

یکی از ویژگی‌های منحصر به فرد این فناوری استفاده نکردن پلیمریزه‌شدن برای تعیین توالی‌یابی و توانایی آن در توالی‌یابی مستقیم RNA است. در این روش، RNA ابتدا به cDNA تبدیل نمی‌شود، بلکه مستقیماً از نانومغذ عبور داده می‌شود. این امر امکان مطالعه مستقیم ساختار و تغییرات پس از رونویسی RNA را فراهم می‌کند (Garalde *et al.*, 2018). علاوه بر این، فناوری نانومغذ قادر به تشخیص تغییرات اپی‌ژنتیکی مانند متیلاسیون DNA به‌طور مستقیم است. این امر به دلیل حساسیت سامانه به تغییرات جزئی در ساختار مولکولی نوکلئوتیدها

هدایت می‌کند. یک توالی الیگونوکلئوتیدی خاص به انتهای آغازگر متصل می‌شود و بار مثبت در سمت دیگر نانومنفذ ایجاد می‌شود تا DNA به درون نانومنفذ جذب شود. نیروی الکتریکی باعث می‌شود که دو رشته DNA در محل این توالی ویژه به صورت مکانیکی از هم جدا شوند و در نهایت، رشته الگو از انتهای ۳' وارد نانومنفذ می‌شود. مزیت استفاده از DNA پلیمرز باکتریوفاز $\phi 29$ این است که این آنزیم هنگام اعمال ولتاژ از DNA جدا نمی‌شود و به عملکرد خود ادامه می‌دهد، که می‌تواند به کاهش سرعت ورود DNA تک‌رشته‌ای به نانومنفذ کمک کرده و شناسایی دقیق‌تر نوکلئوتیدها را برای تعیین توالی فراهم کند (Pareek et al., 2011; Steinbock et al., 2009; Ansorge, 2009; Gupta et al., 2008). با وجود این چالش‌ها، MinION در بسیاری از زمینه‌های تحقیقاتی و کاربردی مورد استفاده قرار گرفته است. از جمله کاربردهای آن می‌توان به توالی‌یابی ژنوم کامل، مطالعات ترنسکریپتوم، تجزیه و تحلیل ساختارهای کروموزومی، مطالعات اپی‌ژنتیک، تشخیص سریع بیماری‌های عفونی و مطالعات متاژنومیکس اشاره کرد (Leggett & Clark, 2017). در سال‌های اخیر، پیشرفت‌های قابل توجهی در فناوری نانومنفذ صورت گرفته است. به عنوان مثال، معرفی نسل جدید نانومنفذها (R10) منجر به بهبود دقت توالی‌یابی، به‌ویژه در مناطق هوموپلیمر، شده است (Rang et al., 2018). همچنین، توسعه روش‌های جدید برای آماده‌سازی نمونه، مانند

انسان طی چند ساعت و با هزینه بسیار پایین‌تر انجام شود (Pareek et al., 2011; Steinbock et al., 2009; Ansorge, 2009; Gupta et al., 2008). فناوری توالی‌یابی نانومنفذ با چالش‌هایی نیز مواجه است. یکی از اصلی‌ترین مشکلات این روش این است که DNA ممکن است سریع‌تر از حد مورد نیاز از نانومنفذها عبور کند، که می‌تواند به خطاهایی در تعیین توالی منجر شود. برای حل این کاستی‌ها، روش‌های جدیدتری بر اساس توالی‌یابی نانومنفذ توسعه یافته‌اند. به عنوان مثال، یک جزء اگزونوکلئاز در ورودی نانومنفذ قرار داده شده است که نوکلئوتیدهای موجود در رشته الگو را به صورت یک‌به‌یک جدا می‌کند. این فرایند اجازه می‌دهد که نوکلئوتیدها به‌طور جداگانه و متوالی وارد نانومنفذ شوند. علاوه بر این، در داخل نانومنفذ یک مولکول چرخه‌دکسترین قرار داده شده که با نوکلئوتیدهای آزاد واکنش می‌دهد و باعث می‌شود که نوکلئوتیدها زمان بیشتری در نانومنفذ بمانند، که این امر به تفکیک بهتر نوکلئوتیدها و کاهش خطا در توالی‌یابی کمک می‌کند. همچنین، می‌توان از یک تنظیم‌کننده برای کنترل سرعت ورود DNA به نانومنفذ استفاده کرد تا رشته DNA با سرعت مناسب وارد نانومنفذ شود. در این روش، از آنزیم DNA پلیمرز باکتریوفاز $\phi 29$ استفاده می‌شود. رشته DNA تک‌رشته‌ای که با یک آغازگر پیوند یافته است، توسط این آنزیم به داخل نانومنفذ کشیده می‌شود. این آنزیم به آغازگر متصل می‌شود و رشته مکمل را سنتز می‌کند، در حالی که رشته الگو را به داخل نانومنفذ

de novo de novo Genome Assembly
Chromatin Transcriptome assembly
immunoprecipitation sequencing (ChIP-seq)
Epigenomics by DNA methylation sequencing
 و *Targeted sequencing (TGS)*, *(Methyl-seq)*
Whole metagenome sequencing (WMS)
 می‌باشند. توالی‌یابی کل ژنوم (WGS) یک روش جامع برای تجزیه و تحلیل کامل توالی DNA یک موجود زنده است که امکان شناسایی انواع تغییرات ژنتیکی از جمله واریانت‌های تک نوکلئوتیدی، حذف‌ها، اضافه‌ها و تغییرات ساختاری بزرگ را فراهم می‌کند (Schwarze *et al.*, 2020). این روش در مطالعات تکاملی، تشخیص بیماری‌های ژنتیکی پیچیده و بررسی تنوع ژنتیکی جمعیت‌ها کاربرد گسترده‌ای دارد. در مقابل، توالی‌یابی کل اگزوم (WES) بر روی بخش‌های کدکننده پروتئین ژنوم (حدود ۱-۲٪ کل ژنوم) تمرکز می‌کند و برای شناسایی واریانت‌های مرتبط با بیماری در مناطق اگزونی مورد استفاده قرار می‌گیرد (Ng *et al.*, 2021). این روش به دلیل هزینه کمتر و تمرکز بر مناطق عملکردی ژنوم، برای تشخیص اختلالات ژنتیکی نادر و مطالعات ژنتیک پزشکی بسیار مفید است. روش *Transcriptomics by Bulk RNA-Seq* یک روش قدرتمند برای بررسی پروفایل بیان ژن در سطح کل ترانسکرپتوم است که امکان شناسایی و کمی‌سازی RNAها را در یک نمونه‌ی بافتی یا جمعیت سلولی فراهم می‌کند (Wang *et al.*, 2022). این روش برای مطالعه‌ی تغییرات بیان ژن در شرایط مختلف فیزیولوژیک یا پاتولوژیک، شناسایی

روش خط تولید و پیوندی‌محو^{۹۴} و روش بدون تکثیر^{۹۵}، امکان کار با مقادیر کمتر DNA و حفظ قطعات بلندتر را فراهم کرده است (Loose *et al.*, 2016). یکی از ویژگی‌های منحصر به فرد MinION، قابلیت حمل و استفاده آسان آن است. این ویژگی امکان توالی‌یابی در محل را فراهم می‌کند، که برای کاربردهایی مانند نظارت بر شیوع بیماری‌ها، مطالعات اکولوژیکی در میدان و حتی توالی‌یابی در فضا بسیار ارزشمند است (Castro-Wallace *et al.*, 2017). در مجموع، فناوری توالی‌یابی نانومفرد Oxford یک روش انقلابی در زمینه توالی‌یابی مولکول منفرد است که با ترکیب خوانش‌های بلند، توانایی تشخیص تغییرات اپی‌ژنتیکی، امکان توالی‌یابی مستقیم RNA و قابلیت حمل، افق‌های جدیدی را در زمینه ژنومیکس و کاربردهای آن گشوده است. با پیشرفت‌های مداوم در این زمینه، انتظار می‌رود که این فناوری همچنان نقش مهمی در تحقیقات ژنومیکس، از جمله در زمینه‌هایی مانند پزشکی شخصی‌سازی شده، تشخیص سریع بیماری‌ها، مطالعات تکاملی و اکولوژی مولکولی ایفا کند.

انواع توالی‌یابی‌های نسل جدید (NGS)

انواع توالی‌یابی‌های نسل جدید (NGS) شامل: Whole genome sequencing (WGS), Whole exome sequencing (WES), Transcriptomics by bulk RNA-seq, Transcriptomics by (scRNAseq), RNA-seq, Single-cell RNA-seq, small RNA sequencing

⁹⁴ Ligation-based

⁹⁵ Amplification-free

novo Transcriptome Assembly یک روش برای بازسازی ترنسکرپتوم کامل یک ارگانسیم بدون نیاز به ژنوم مرجع است که امکان شناسایی ژن‌ها و ایزوفرم‌های جدید، واریانت‌های اسپلیسینگ و مطالعات بیان ژن‌ها با عملکرد بالا را در گونه‌های غیرمدل فراهم می‌کند (Haas *et al.*, 2023). این روش برای مطالعات تکاملی، شناسایی ژن‌های مرتبط با صفات خاص و بررسی تنوع ژنتیکی در سطح ترنسکرپتوم کاربرد دارد. Mapping Protein- DNA Interactions with ChIP-Seq یک روش قدرتمند برای مطالعه برهمکنش‌های پروتئین-DNA در سطح ژنوم است که امکان شناسایی مناطق اتصال فاکتورهای رونویسی، تغییرات هیستونی و سایر پروتئین‌های متصل به DNA را فراهم می‌کند (Nakato & Shirahige, 2023). این روش برای مطالعه سازوکارهای تنظیم ژن، شناسایی عناصر تنظیمی و بررسی تغییرات اپی‌ژنتیکی کاربرد گسترده‌ای دارد. Epigenomics by DNA Methylation Sequencing یک روش برای بررسی الگوهای متیلاسیون DNA در سطح ژنوم است که نقش مهمی در تنظیم بیان ژن، خاموش‌سازی ژن و فرایندهای تکوینی دارد (Luo *et al.*, 2022). این روش برای مطالعه تغییرات اپی‌ژنتیکی در بیماری‌ها، فرایندهای پیری و تمایز سلولی کاربرد دارد. روش‌های مختلفی مانند Bisulfite Sequencing، Methylation-Specific Enzyme Sequencing و Specific Antibody Methylation Sequencing برای این منظور استفاده

ایزوفرم‌های جدید و بررسی RNAهای غیرکدکننده کاربرد دارد. Transcriptomics by Single-Cell RNA-Seq; scRNA-Seq یک پیشرفت قابل توجه در زمینه تجزیه و تحلیل ترنسکرپتوم است که امکان بررسی بیان ژن در سطح تک سلولی را فراهم می‌کند (Lafzi *et al.*, 2023). این روش برای مطالعه ناهمگونی^{۹۶} سلولی، شناسایی انواع سلولی نادر، بررسی مسیرهای تمایز سلولی و مطالعه پاسخ‌های سلولی به محرک‌های مختلف بسیار ارزشمند است. روش Small RNA Sequencing، روشی اختصاصی برای مطالعه RNAهای کوچک غیرکدکننده مانند microRNA، siRNA و piRNAها است که نقش مهمی در تنظیم بیان ژن و فرایندهای سلولی دارند (Stark *et al.*, 2022). این روش برای شناسایی Biomarkerها، مطالعه سازوکارهای تنظیمی در سطح-Post transcriptional و بررسی نقش RNAهای کوچک در بیماری‌ها کاربرد دارد. *de novo* Genome Assembly یک روش برای ساخت توالی ژنومی کامل بدون استفاده از ژنوم مرجع است که از ترکیب خوانش‌های کوتاه مانند Illumina و PacBio یا Oxford Nanopore برای بهبود کیفیت و پیوستگی Assembly استفاده می‌کند (Amarasinghe *et al.*, 2021). این روش برای مطالعه ژنوم گونه‌های جدید یا ناشناخته، بررسی تغییرات ساختاری بزرگ در ژنوم و مطالعات مقایسه‌ای ژنوم و مطالعات تکاملی و اکولوژیک کاربرد دارد. روش *de*

⁹⁶ Heterogeneity

توجه و بررسی دقیق است. در ادامه به تفصیل در مورد هر یک از این روش‌های توالی‌یابی بحث می‌شود.

توالی‌یابی کل ژنوم (WGS)

توالی‌یابی کل ژنوم یا WGS⁹⁷ یک روش پیشرفته در زمینه توالی‌یابی DNA است که امکان تعیین توالی کامل ژنوم یک ارگانیسم را فراهم می‌کند. این روش با هدف شناسایی تمامی تغییرات ژنتیکی از جمله واریانت‌های تک‌نوکلئوتیدی (SNVs)، حذف‌ها، اضافه‌ها، تغییرات ساختاری و تغییرات تعداد کپی (CNVs) در سراسر ژنوم انجام می‌شود (Schwarze *et al.*, 2020). ماهیت

WGS بر اساس توالی‌یابی تمام DNA ژنومی یک ارگانیسم استوار است که شامل مناطق کدکننده (اگزون‌ها) و غیرکدکننده (اینترون‌ها، مناطق بین ژنی و عناصر تنظیمی) می‌شود. این روش در مقایسه با سایر

روش‌های توالی‌یابی مانند Whole exome sequencing; WES

دیدگاه جامع‌تری از تغییرات ژنتیکی ارائه می‌دهد (Metzker, 2021). مراحل آزمایشگاهی و Wet-lab

روش Whole genome sequencing; WGS شامل چندین گام اساسی است. ابتدا، استخراج DNA با کیفیت بالا از نمونه مورد نظر انجام می‌شود. این مرحله به‌طور

معمول با استفاده از کیت‌های استخراج DNA شرکت‌های Qiagen، Thermo Fisher Scientific یا

Promega انجام می‌گیرد که بر اساس روش‌های

شیمیایی یا مکانیکی عمل می‌کنند (Maes *et al.*,

می‌شوند. Targeted sequencing یک روش اختصاصی برای توالی‌یابی مناطق خاصی از ژنوم است که امکان بررسی دقیق و عمیق ژن‌ها یا مناطق مورد نظر را با هزینه و زمان کمتر نسبت به WGS یا WES فراهم می‌کند (Metzker, 2021). این روش برای تشخیص جهش‌های ژنتیکی در بیماری‌های خاص، مطالعات فارماکوژنومیک و غربالگری ژنتیکی کاربرد دارد. Whole metagenome sequencing یک روش جامع برای مطالعه‌ی ترکیب و عملکرد جوامع میکروبی در محیط‌های مختلف است که امکان شناسایی و تجزیه و تحلیل ژنوم تمام میکروارگانیسم‌های موجود در یک نمونه را بدون نیاز به کشت فراهم می‌کند (Quince *et al.*, 2023). این روش برای مطالعه‌ی اکوسامانه‌های میکروبی، بررسی تعاملات میکروب-میزبان، شناسایی ژن‌های جدید با پتانسیل کاربردی و مطالعه‌ی تغییرات میکروبیوم در شرایط مختلف کاربرد دارد. در مجموع، این روش‌های توالی‌یابی نسل جدید امکان مطالعه‌ی عمیق و گسترده‌ی ژنوم، ترنسکریپتوم و اپی‌ژنوم را در مقیاس‌های مختلف از تک سلول تا جوامع پیچیده فراهم می‌کنند. این روش‌ها نقش مهمی در پیشرفت تحقیقات زیست‌شناسی مولکولی، کشاورزی، پزشکی شخصی‌سازی شده، اکولوژی مولکولی و توسعه‌ی درمان‌های جدید دارند. با این حال، چالش‌هایی مانند حجم زیاد داده‌های تولید شده، نیاز به ابزارهای بیوانفورماتیکی پیشرفته برای تجزیه و تحلیل داده‌ها و تفسیر نتایج، و مسائل اخلاقی مربوط به حفظ حریم خصوصی اطلاعات ژنتیکی نیز وجود دارد که نیازمند

⁹⁷ Whole genome sequencing

2022). طی این فرایند، هریک از قطعات DNA چندین بار توالی‌یابی می‌شود تا مشاهدات متعدد از هر ناحیه ژنومی حاصل شود. این رویکرد کمک می‌کند تا خطاهای احتمالی حین توالی‌یابی تشخیص داده شده و دقت نتایج افزایش یابد (Goodwin *et al.*, 2022). در نهایت، خوانش‌های حاصل از توالی‌یابی برای انجام مراحل تجزیه و تحلیل داده‌های کامپیوتری آماده می‌شوند. در مجموع، مراحل آزمایشگاهی و Wet-lab روش WGS شامل استخراج DNA با کیفیت بالا با استفاده از کیت‌های استخراج شرکت‌های معتبر، شکستن DNA به قطعات کوچکتر با روش‌های مکانیکی مانند فراصوت، ساخت کتابخانه‌های DNA با اتصال آداپتورهای استاندارد، تکثیر کتابخانه‌ها با PCR یا روش‌های کلونال، و انجام توالی‌یابی موازی با استفاده از پلتفرم‌های NGS پیشرفته است. این مراحل حیاتی برای تولید داده‌های با کیفیت بالا و پیوستگی مناسب هستند که در مراحل بعدی تجزیه و تحلیل داده‌ها مورد استفاده قرار می‌گیرند. پس از اتمام توالی‌یابی، مرحله تجزیه و تحلیل داده‌ها آغاز می‌شود که شامل چندین گام پیچیده است. ابتدا، کنترل کیفیت خوانش‌های خام انجام می‌شود تا خوانش‌های با کیفیت پایین یا آداپتورها حذف شوند. سپس، خوانش‌ها به ژنوم مرجع همتراز^{۹۸} می‌شوند تا موقعیت هر خوانش در ژنوم مشخص شود. پس از همترازی و حذف موارد تکراری، مرحله شناسایی واریانت‌ها انجام می‌شود که

2014). برای افزایش کیفیت و یکپارچگی DNA استخراج شده، از ابزارهایی مانند Nanodrop یا Qubit برای سنجش غلظت و خلوص DNA استفاده می‌شود. پس از استخراج، DNA به قطعات کوچکتر (به‌طور معمول ۳۰۰-۵۰۰ جفت باز) شکسته می‌شود. این مرحله با استفاده از روش‌های مکانیکی مانند فراصوت (Bioruptor, Diagenode) یا همزن مغناطیسی (Covaris) انجام می‌گیرد (Auer *et al.*, 2014). در مرحله بعد، کتابخانه‌های DNA با اتصال آداپتورها به انتهای قطعات DNA ساخته می‌شوند. این آداپتورها به‌طور معمول توالی‌های استاندارد مانند Illumina را دارند که برای تکثیر و توالی‌یابی قطعات مورد استفاده قرار می‌گیرند. ساخت کتابخانه‌های DNA به‌طور معمول با استفاده از آنزیم‌های محدودکننده مانند Nextera DNA Flex ساخت Illumina یا روش‌های لیگاسیون مانند TruSeq DNA Nano ساخت Illumina انجام می‌شود (Casbon *et al.*, 2011). سپس، کتابخانه‌های ساخته شده تکثیر شده و برای توالی‌یابی آماده می‌شوند. تکثیر کتابخانه‌ها به‌طور معمول با استفاده از PCR مانند کیت‌های (Kapa HyperPlus) یا روش‌های کلونال (مانند اسپری قطرات در فلوسیتومتری) انجام می‌گیرد (Nekrutenko & Taylor, 2012). توالی‌یابی به‌طور معمول با استفاده از پلتفرم‌های توالی‌یابی نسل بعدی (NGS) مانند Illumina NovaSeq یا BGI DNBSEQ انجام می‌شود که امکان توالی‌یابی موازی میلیون‌ها قطعه DNA را فراهم می‌کنند (Bentley *et*

⁹⁸ Align

امضای موتاسیونی را فراهم می‌کند که می‌تواند به درک بهتر سازوکارهای سرطان‌زایی و توسعه درمان‌های هدفمند کمک کند (Priestley *et al.*, 2021). در زمینه ژنتیک جمعیت، WGS برای مطالعه تنوع ژنتیکی در جمعیت‌های مختلف، بررسی الگوهای مهاجرت و تاریخ تکاملی گونه‌ها استفاده می‌شود (Mallick *et al.*, 2022). در زمینه کشاورزی و دامپروری، WGS برای بهبود نژاد گیاهان و حیوانات، شناسایی ژن‌های مرتبط با صفات مطلوب و توسعه استراتژی‌های به‌نژادی مورد استفاده قرار می‌گیرد (Purugganan *et al.*, 2021; Mathur *et al.*, 2021). در حوزه میکروبیولوژی، WGS برای شناسایی و طبقه‌بندی میکروارگانیسم‌ها، بررسی مقاومت آنتی‌بیوتیکی و ردیابی شیوع بیماری‌های عفونی کاربرد دارد (Price *et al.*, 2023; Soni *et al.*, 2023). با وجود مزایای فراوان، WGS با چالش‌هایی نیز روبرو است. حجم بالای داده‌های تولید شده نیازمند زیرساخت‌های قوی ذخیره‌سازی و پردازش داده است. تفسیر نتایج WGS نیز پیچیده است و نیاز به دانش تخصصی در زمینه‌های ژنتیک، بیوانفورماتیک و زیست‌شناسی سامانه‌های دارد. علاوه بر این، مسائل اخلاقی مربوط به حفظ حریم خصوصی و استفاده از اطلاعات ژنتیکی نیز باید مورد توجه قرار گیرند (McGuire *et al.*, 2008; Pinxten & Howard, 2014). با این حال، پیشرفت‌های مداوم در فناوری‌های توالی‌یابی و روش‌های تجزیه و تحلیل داده، همراه با کاهش هزینه‌ها، باعث گسترش استفاده از WGS در زمینه‌های مختلف علمی و

شامل شناسایی SNVs⁹⁹، indels¹⁰⁰، CNVs¹⁰¹ و تغییرات ساختاری است. این مرحله به‌طور معمول با استفاده از نرم‌افزارهایی مانند GATK، Strelka یا FreeBayes انجام می‌شود (Van der Auwera *et al.*, 2023). پس از شناسایی واریانت‌ها، واریانت‌ها کالیبره شده و کیفیت آنها با VQSR از GATK بهبود می‌یابد و همچنین واریانت‌های ساختاری با نرم‌افزارهایی مانند Delly یا Manta شناسایی می‌شوند. در مرحله بعد واریانت‌های شناسایی شده حاشیه‌نویسی (Annotation) می‌شوند تا اطلاعات بیولوژیکی و عملکردی مرتبط با هر واریانت مشخص شود. این مرحله با استفاده از پایگاه‌های داده و ابزارهایی مانند ANNOVAR، VEP یا SnpEff انجام می‌شود (McLaren *et al.*, 2022). در ادامه فیلتر کردن و اعتبارسنجی واریانت‌ها با استفاده از ابزارهایی مانند GATK Variant Filtration، VCFtools، Bcftools filter صورت می‌گیرد، تا واریانت‌های با کیفیت بالا از واریانت‌های کاذب جدا شوند و در نهایت تجزیه و تحلیل‌های عملکردی انجام می‌شود. کاربردهای WGS بسیار گسترده و متنوع است. در زمینه پزشکی، WGS برای تشخیص بیماری‌های ژنتیکی نادر، شناسایی واریانت‌های مرتبط با بیماری‌های پیچیده، و پزشکی شخصی‌سازی شده استفاده می‌شود (Royer *et al.*, 2021). در تحقیقات سرطان، WGS امکان شناسایی جهش‌های سوماتیک، تغییرات ساختاری و الگوهای

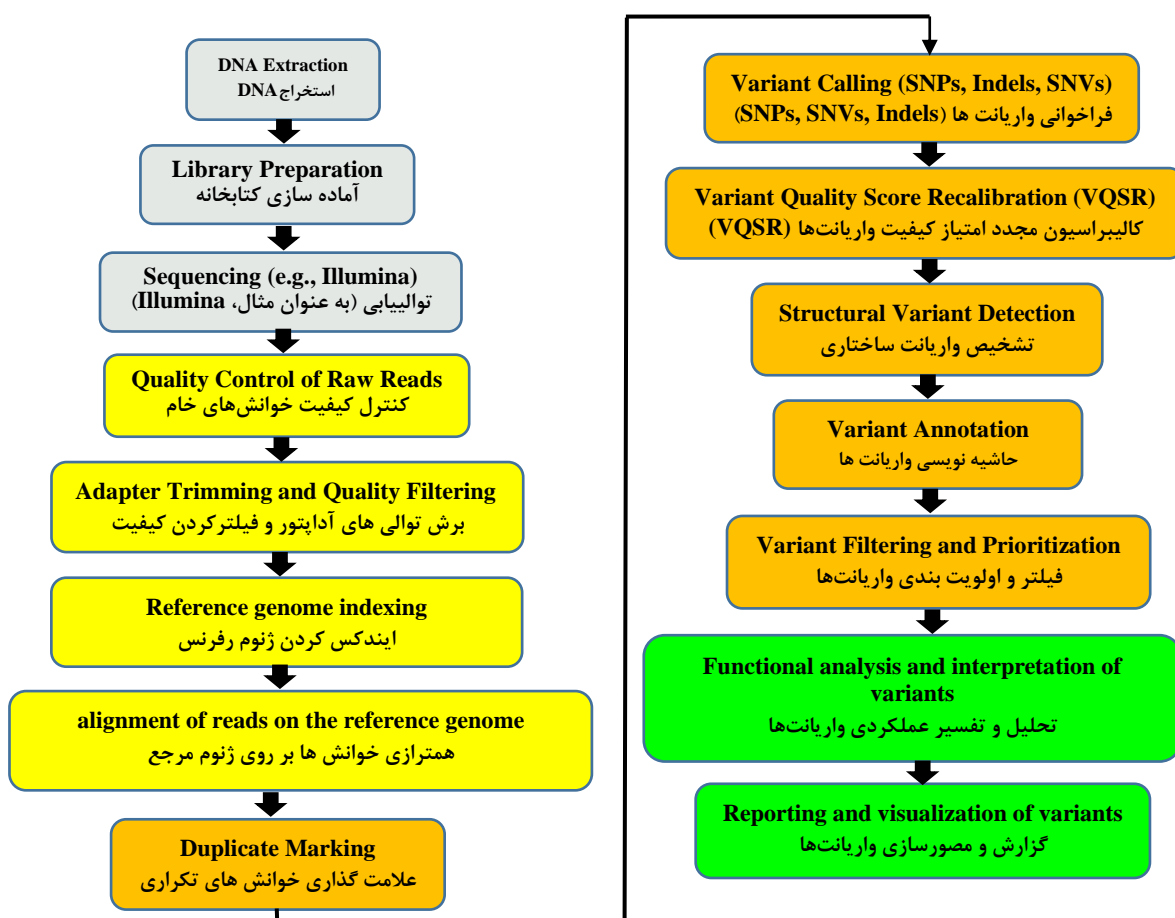
⁹⁹ Single-nucleotide variants (SNVs)

¹⁰⁰ Insertion-deletion (indels)

¹⁰¹ Copy number variation (CNV)

تکمیلی آورده شده‌اند (Van der Auwera *et al.*, 2013; McKenna *et al.*, 2010; Li & Durbin, 2009; Cingolani *et al.*, 2012; DePristo *et al.*, 2011; Robinson *et al.*, 2011; Rausch *et al.*, 2012; Fuentes-Pardo & Ruzzante, 2017; Ugur (Sezerman *et al.*, 2019).

پزشکی شده است. انتظار می‌رود که در آینده، WGS نقش مهم‌تری در تشخیص و درمان بیماری‌ها، تحقیقات زیست‌شناسی و توسعه فناوری‌های زیستی ایفا کند. در ادامه به مراحل اصلی تجزیه و تحلیل داده‌های WGS حاصل از توالی‌یابی کل ژنوم (WGS) اشاره می‌شود (شکل ۱۲). اسکرپ‌های تجزیه و تحلیل نیز در فایل



شکل ۱۲- مراحل تجزیه و تحلیل داده‌های WGS Whole genome sequencing;
Figure 12. Whole genome sequencing (WGS) data analysis steps

توالی‌یابی کل اگزوم (WES)

توالی‌یابی کل اگزوم یا WES¹⁰² یک روش قدرتمند و پر کاربرد در زمینه‌های مختلف پزشکی و ژنتیک است که قادر به تعیین توالی کل اگزوم (بخش کدکننده ژنوم) در یک آزمایش واحد است. این روش به‌طور گسترده‌ای برای شناسایی جهش‌های ژنتیکی مرتبط با بیماری‌های ژنتیکی، سرطان، و سایر اختلالات مورد استفاده قرار می‌گیرد و به محققان و پزشکان بالینی این امکان را می‌دهد تا به‌طور کارآمد و جامع به بررسی و تحلیل ژنوم فرد بپردازند (Bamshad *et al.*, 2011). این روش، Whole exome sequencing به محققان امکان می‌دهد تا با استفاده از فناوری‌های تعیین توالی نسل جدید، کل بخش کدکننده (اگزوم) ژنوم را در یک آزمایش واحد مورد بررسی قرار دهند. اگزوم تنها حدود ۱-۲ درصد از کل ژنوم را تشکیل می‌دهد اما در عین حال دارای بیشتر جهش‌های مرتبط با بیماری است (Choi *et al.*, 2009). این امر باعث می‌شود که WES در مقایسه با توالی‌یابی کل ژنوم (Whole genome sequencing) هزینه‌ی کمتری داشته و منابع محاسباتی کمتری را نیاز داشته باشد. در مورد مراحل آزمایشگاهی و Wet-lab انجام WES، این فرایند شامل چندین مرحله اصلی است که با دقت و طی شدن این مراحل می‌توان به نتایج مطلوب و قابل اطمینان دست یافت. در ابتدا، نمونه بیولوژیکی مانند خون یا بافت از بیمار یا فرد مورد مطالعه جمع‌آوری می‌شود و سپس DNA آن با استفاده از

کیت‌های استخراج DNA تجاری مانند DNeasy Blood Tissue Kit (Qiagen) & استخراج می‌گردد (Bamshad *et al.*, 2011). پس از استخراج DNA، کیفیت و غلظت آن با استفاده از دستگاه‌های اسپکتروفتومتری یا فلورومتری مانند NanoDrop (Thermo Fisher Scientific) و Qubit (Invitrogen) سنجیده می‌شود. در مرحله بعد، برای افزایش غنای اگزوم در کتابخانه DNA، از روش‌های دورگه‌شدن یا پرچم‌گذاری^{۱۰۳} استفاده می‌شود. در این مرحله، از کاوشگرهای مکمل اگزوم‌ها که به‌طور سنتتیک تولید شده‌اند، استفاده می‌شود تا بخش‌های کدکننده ژنوم (اگزوم) به‌طور انتخابی جداسازی و غنی‌سازی شوند (Bamshad *et al.*, 2011; Rabbani *et al.*, 2016). برای این منظور، از کیت‌های تجاری مانند SureSelect Human All Exon (Agilent Technologies) یا TruSeq Exome Enrichment Kit (Illumina) استفاده می‌شود. در این کیت‌ها، کاوشگرهای مکمل اگزوم‌ها به‌طور پیش‌ساخته وجود دارند و کاربران تنها باید مراحل دورگه‌شدن و پرچم‌گذاری را طبق شیوه‌نامه‌های آزمایشگاهی مربوطه انجام دهند. پس از غنی‌سازی اگزوم‌ها، کتابخانه DNA با استفاده از آنزیم‌های برش‌دهنده مانند آنزیم‌های محدودکننده یا آنزیم‌های قطع‌کننده انتهاها (End-repair enzymes) آماده‌سازی می‌شود. این مرحله شامل برش DNA، افزودن آداپتورها و بارکدها، و آماده‌سازی نهایی کتابخانه برای

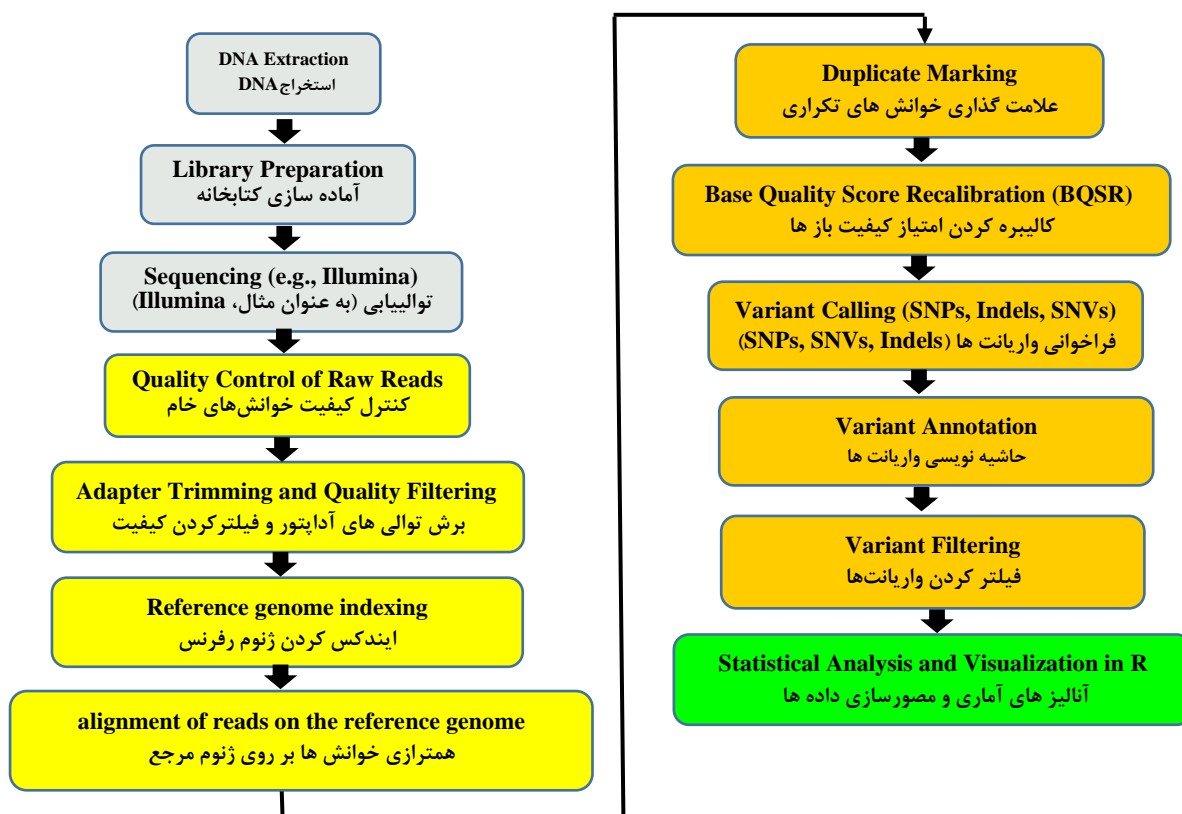
¹⁰³ Capture¹⁰² Whole exome sequencing

می‌شوند. ابزارهایی همچون BWA و Bowtie2 برای انجام این مرحله استفاده می‌شوند. پس از هم‌ترازی خوانش‌ها، تغییرات DNA در مقایسه با ژنوم مرجع مانند جایگزینی‌ها (SNP) و حذف یا اضافه شدن نوکلئوتیدها (indel) شناسایی می‌شوند. ابزارهایی مانند GATK، SAMtools و Freebayes برای این‌هدف به کار برده می‌شوند. در مرحله بعد پیش‌بینی اثرات عملکردی واریانت‌ها صورت می‌گیرد و ابزارهایی مانند ANNOVAR، VEP، SnpEff و CADD برای پیش‌بینی اثرات عملکردی هر یک از واریانت‌های شناسایی‌شده استفاده می‌شوند. در نهایت، بر اساس نتایج مراحل قبل، واریانت‌هایی که احتمال دارد با بیماری مرتبط باشند، از طریق نرم‌افزارهایی مانند BCFtools اولویت‌بندی و فیلتر می‌شوند تا بررسی‌های بیشتر بر روی آنها انجام شود (Wang *et al.*, 2010). در زمینه کاربردهای WES، این روش به‌طور گسترده‌ای در شناسایی جهش‌های مرتبط با بیماری‌های ژنتیکی نادر و متابولیک مورد استفاده قرار می‌گیرد. همچنین WES در شناسایی جهش‌های سوماتیک در سرطان‌ها، تشخیص علل ژنتیکی ناباروری، تعیین مسیرهای درگیر در بیماری‌های پیچیده، و حتی در پزشکی قانونی و تشخیص هویت نیز کاربرد دارد (Bamshad *et al.*, 2011; Rabbani *et al.*, 2014). علاوه بر این، WES می‌تواند به عنوان یک ابزار تشخیصی در کلینیک مورد استفاده قرار گیرد و به کمک آن می‌توان بیماران مبتلا به بیماری‌های ژنتیکی را شناسایی و مدیریت کرد. در

مرحله تعیین توالی است (Bamshad *et al.*, 2011). در این مرحله نیز از کیت‌های تجاری مانند KAPA HyperPrep (Roche) یا NEBNext Ultra (New England Biolabs) استفاده می‌شود. در مرحله نهایی، کتابخانه DNA آماده‌شده توسط دستگاه‌های تعیین توالی نسل جدید مانند Ion Proton، Illumina HiSeq یا Ion S5 مورد تعیین توالی قرار می‌گیرد. این دستگاه‌ها قادر به تولید میلیون‌ها خوانش از توالی‌های DNA در هر آزمایش هستند (Bamshad *et al.*, 2011). پس از مرحله تعیین توالی، داده‌های خام به‌دست‌آمده برای مراحل تجزیه و تحلیل بیوانفورماتیکی آماده می‌شوند. در مجموع، فرایند Wet-lab آزمایش WES شامل مراحل استخراج DNA، غنی‌سازی اگزوم‌ها، آماده‌سازی کتابخانه، و در نهایت تعیین توالی با استفاده از دستگاه‌های نسل جدید است که به دقت و طبق پروتکل‌های استاندارد آزمایشگاهی انجام می‌شود تا نتایج قابل اطمینان به‌دست آید. در تجزیه و تحلیل داده‌های WES، مراحل مختلفی برای شناسایی و اولویت‌بندی جهش‌های مرتبط با بیماری وجود دارد. این مراحل شامل چندین مرحله است: نخست تصفیه و فیلتر کردن خوانش‌ها که در این مرحله، ابزارهایی مانند Trimmomatic و Cutadapt برای حذف آداپتورها و حذف خوانش‌های با کیفیت پایین مورد استفاده قرار می‌گیرند. هم‌ترازی خوانش‌ها به ژنوم مرجع پس از تصفیه و فیلتر کردن خوانش‌ها اتفاق می‌افتد و خوانش‌ها به یک ژنوم مرجع نمایه^{۱۰۴} شده هم‌تراز

مراحل اصلی تجزیه و تحلیل داده‌های WES حاصل از توالی‌یابی کل اگزوم (WES) اشاره می‌شود (شکل ۱۳). اسکرپ‌های تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (DePristo *et al.*, 2011; Van der Auwera *et al.*, 2013; Li & Durbin, 2009; McKenna *et al.*, 2010; Wang *et al.*, 2010; Bao *et al.*, 2014; Hanssen *et al.*, 2024; Koboldt, 2020; Pereira *et al.*, 2020).

مجموع، Whole exome sequencing یک روش انقلابی در زمینه ژنوم‌شناسی و پزشکی شخصی‌سازی شده است که به کمک آن می‌توان به‌طور جامع و کارآمد به بررسی و تحلیل ژنوم افراد پرداخت و جهش‌های مرتبط با بیماری‌های مختلف را شناسایی کرد. این روش در حال حاضر به‌طور گسترده‌ای در محیط‌های آزمایشگاهی و کلینیکی مورد استفاده قرار می‌گیرد و دامنه کاربردهای آن همچنان در حال توسعه و گسترش است. در ادامه به



شکل ۱۳- مراحل تجزیه و تحلیل داده‌های WES Whole exome sequencing;

Figure 13. Whole exome sequencing (WES) data analysis steps

مطالعه ترنسکریپتوم از طریق Bulk RNA-Seq

ترنسکریپتومیکس، مطالعه مجموعه کامل رونوشت‌های RNA در یک سلول یا بافت، به عنوان ابزاری قدرتمند برای درک بیان ژن و عملکرد سلولی ظهور کرده است. یکی از رایج‌ترین روش‌های مورد استفاده در این زمینه، توالی‌یابی RNA انبوه (Bulk RNA-Seq) است که یک نمای جامع از نمای ترنسکریپتومی نمونه‌ی بیولوژیکی را فراهم می‌کند. Bulk RNA-Seq، شامل فرایند استخراج و توالی‌یابی پرتوان مولکول‌های RNA از یک جمعیت ناهمگن سلول‌ها است. این رویکرد به محققان امکان می‌دهد تا سطوح بیان هزاران ژن را به‌طور همزمان کمی‌سازی کنند، امکانی برای درک عمیق‌تر سازوکارهای مولکولی پایه‌ای که فرایندهای سلولی، تکوین و پاتوژنز بیماری را تنظیم می‌کنند (Wang *et al.*, 2009). ماهیت روش Bulk RNA-Seq در توانایی آن برای ثبت نمای کلی ترنسکریپتومی یک نمونه نهفته است، که بازتابی از الگوهای بیانی جمعی همه سلول‌های موجود در یک بافت یا اندام را منعکس می‌کند. این دید جامع در تضاد با سایر روش‌ها مانند توالی‌یابی RNA تک‌سلولی (scRNA-Seq) است که بر روی نمای ترنسکریپتومی سلول‌های انفرادی تمرکز دارند (Kolodziejczyk *et al.*, 2015). Bulk RNA-Seq به ویژه در مطالعه بافت‌های پیچیده ارزشمند است، زیرا تعامل بین انواع مختلف سلول‌ها و امضاهای ترنسکریپتومی منحصر به فرد آنها می‌تواند بینشی در مورد همئوستاز، بازسازی و بیماری بافت ارائه دهد. روش Bulk RNA-Seq شامل مراحل

آزمایشگاهی مختلفی است که برای استخراج، آماده‌سازی و تهیه کتابخانه‌های DNA مورد نیاز انجام می‌شوند. در ابتدا، نمونه مورد مطالعه که می‌تواند شامل یک سلول، بافت یا ارگانیسم کامل باشد، باید به درستی جمع‌آوری و نگهداری شود تا کیفیت و کمیت RNA استخراجی حفظ گردد (Bizouarn, 2014; Kanagal-Shamanna, 2016). پس از جمع‌آوری نمونه، مرحله استخراج RNA آغاز می‌شود. این مرحله شامل چندین گام است که با استفاده از کیت‌های تجاری مخصوص استخراج RNA مانند TRIzol و RNeasy انجام می‌شود. در ابتدا، نمونه‌ها در حضور بافر لیز کننده قرار گرفته و سپس با استفاده از روش‌های مناسب مانند سانتریفیوژ یا ستون‌های فیلتراسیون، RNA از سایر اجزای سلولی جداسازی می‌شود. در این مرحله، به‌طور معمول از تیمار با آنزیم DNase نیز برای حذف آلودگی‌های DNA استفاده می‌شود. در نهایت، RNA استخراجی به منظور حذف آلودگی‌های احتمالی و افزایش کیفیت، تحت مرحله خالص‌سازی قرار می‌گیرد (Becker *et al.*, 2010). پس از استخراج، کمیت و کیفیت RNA با استفاده از روش‌های مختلف از جمله اندازه‌گیری جذب در طول موج ۲۶۰ نانومتر و تجزیه و تحلیل الکتروفورز در ژل آگارز بررسی می‌شود. این ارزیابی اولیه به منظور تأیید کفایت و سلامت نمونه‌ها برای مراحل بعدی ضروری است (Bizouarn, 2014; Kanagal-Shamanna, 2016). مرحله بعدی، تبدیل RNA به cDNA است که با استفاده از آنزیم رونوشت‌برداری معکوس (Reverse

از توالی‌یابی تأثیر مستقیم دارند. پس از انجام مراحل آزمایشگاهی در روش Bulk RNA-Seq و تولید توالی‌های خام (Raw reads)، مرحله تجزیه و تحلیل داده‌ها آغاز می‌شود. این مرحله شامل چندین گام اصلی است که با استفاده از نرم‌افزارها و ابزارهای بیوانفورماتیکی متنوعی انجام می‌شود (Conesa *et al.*, 2016). در ابتدا، پیش‌پردازش توالی‌های خام انجام می‌شود. این گام به منظور حذف قسمت‌های نامناسب و ارتقای کیفیت توالی‌ها انجام می‌گیرد. در این مرحله، توالی‌های با کیفیت پایین (Based on quality scores)، توالی‌های کوتاه و آداپتورها حذف می‌شوند. همچنین در برخی موارد ممکن است نیاز به حذف توالی‌های متعلق به ریبوزوم یا سایر آلودگی‌ها باشد. برای این منظور از ابزارهایی مانند Trimmomatic و Cutadapt استفاده می‌شود (Chen *et al.*, 2018). پس از پیش‌پردازش، توالی‌های خالص شده به ژنوم مرجع هم‌تراز (Mapping) می‌شوند. این مرحله به منظور تعیین منشأ هر یک از توالی‌ها و ارتباط آن‌ها با ژن‌ها انجام می‌شود. برای این منظور از ابزارهایی مانند HISAT2، STAR و Bowtie2 استفاده می‌شود که توانایی هم‌ترازی توالی‌ها به ژنوم مرجع را با دقت و سرعت بالا دارند (Dobin *et al.*, 2013). پس از هم‌ترازی توالی‌ها، مرحله کمی‌سازی بیان ژن‌ها انجام می‌شود. در این مرحله میزان بیان هر ژن براساس تعداد توالی‌هایی که به آن هم‌تراز شده‌اند، محاسبه می‌شود. معیارهایی مانند Fragments Per Kilobase of transcript per Million mapped reads; FPKM

(Transcriptase) انجام می‌شود. در این مرحله، به‌طور معمول از آغازگرهای هگزامر تصادفی یا آغازگرهای Oligo-dT استفاده می‌شود که به ترتیب به توالی‌های تصادفی یا دنباله پلی A انتهای mRNA متصل شده و سپس cDNA سنتز می‌گردد (Kukurba & Montgomery, 2015). پس از سنتز cDNA، مرحله ساخت کتابخانه‌های DNA آغاز می‌شود. در این مرحله، cDNA به قطعات کوچک‌تر شکسته شده و در انتها، قطعات مربوط به هر نمونه با استفاده از آداپتورها و بارکدهای اختصاصی، برچسب‌گذاری می‌شوند. این مرحله که با استفاده از کیت‌های تجاری انجام می‌شود، به منظور افزایش تنوع کتابخانه و امکان تفکیک نمونه‌ها در مرحله توالی‌گیری انجام می‌گیرد (Bizouarn, 2014; Kanagal-Shamanna, 2016). در نهایت، کتابخانه‌های DNA آماده‌شده برای مرحله توالی‌گیری به دستگاه توالی‌یاب انتقال می‌یابند. این دستگاه‌ها با استفاده از فناوری‌های پیشرفته توالی‌یابی، توالی هر قطعه DNA را تعیین می‌کنند. خروجی این مرحله، توالی‌های کوتاه^{۱۰۵} است که در مراحل تحلیل داده مورد استفاده قرار می‌گیرند (Kukurba & Montgomery, 2015). در مجموع، مراحل آزمایشگاهی روش Bulk RNA-Seq شامل استخراج RNA، سنتز cDNA و ساخت کتابخانه‌های DNA است که با استفاده از پروتکل‌ها و کیت‌های استاندارد انجام می‌شود. این مراحل حساس و دقیق هستند و بر کیفیت و کمیت داده‌های نهایی حاصل

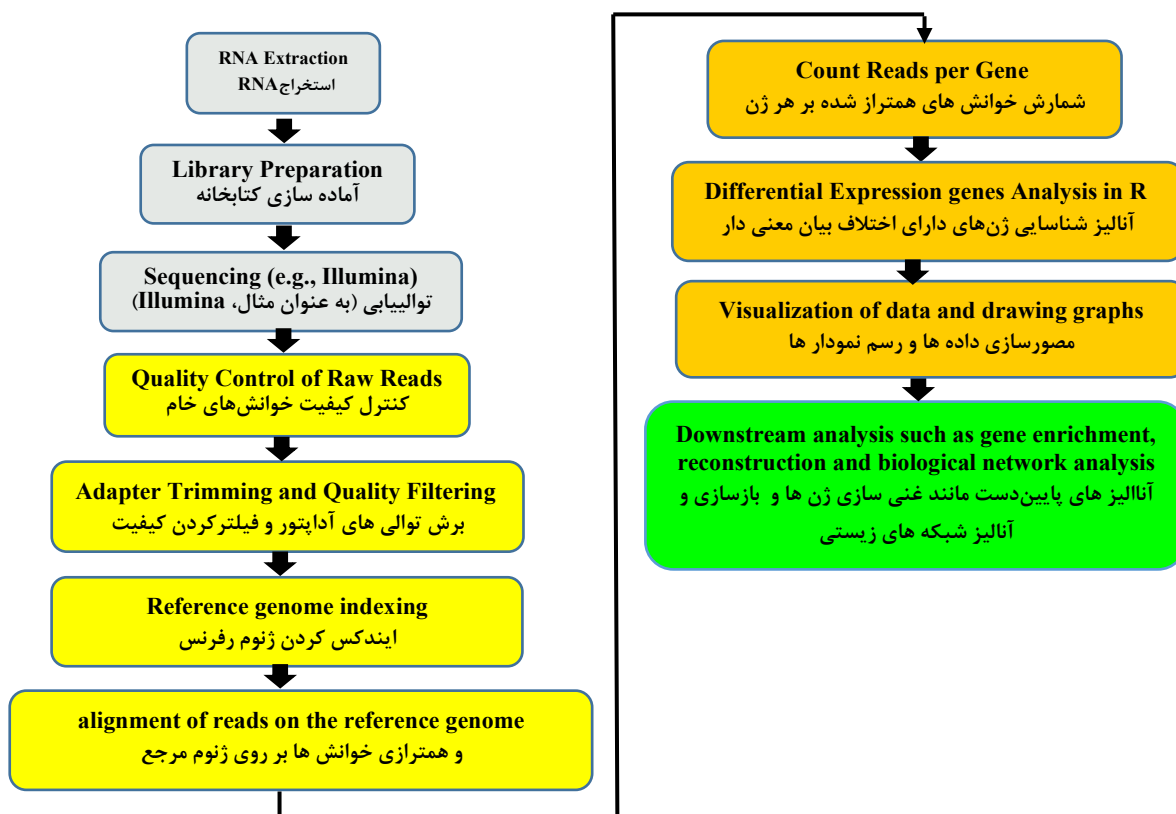
است که شامل کنترل کیفیت خوانش، تراز به ژنوم یا ترنسکریپتوم مرجع، کمی‌سازی بیان ژن و تجزیه و تحلیل بیان متفاوت است. این تجزیه و تحلیل‌ها اغلب با استفاده از نرم‌افزارهای تخصصی و خط‌مشی‌های بیوانفورماتیکی انجام می‌شوند که از الگوریتم‌ها و روش‌های آماری پیشرفته برای تفسیر دقیق حجم زیاد داده‌های توالی‌یابی استفاده می‌کنند (Mortazavi *et al.*, 2008). مزایای Bulk RNA-Seq در توانایی آن برای ارائه یک دید جامع و بی‌طرفانه از نمای ترنسکریپتومی یک نمونه بیولوژیکی نهفته است. این روش را می‌توان در طیف گسترده‌ای از کاربردهای پژوهشی به کار برد، از جمله شناسایی ژن‌های بیان متفاوت، کشف رونوشت‌های جدید، مشخصه‌سازی الگوهای جایگزینی جایگاه‌های برش، و روشن‌سازی شبکه‌های تنظیمی ژن (Ozsolak & Milos, 2011). علاوه بر این، Bulk RNA-Seq به ابزاری ارزشمند در حوزه تحقیقات ترجمانی^{۱۰۷} تبدیل شده است، جایی که برای شناسایی اهداف درمانی محتمل، توسعه نشانگرهای زیستی تشخیصی و پایش پاسخ به درمان‌های مختلف استفاده می‌شود (Stark *et al.*, 2019a). انعطاف‌پذیری و مقیاس‌پذیری این رویکرد آن را به یک گزینه محبوب برای محققان مطالعه طیف متنوعی از ارگانیسم‌ها از میکروب‌ها تا ارگانیسم‌های چندسلولی پیچیده تبدیل کرده است. با وجود گسترش فراگیر آن Bulk RNA-Seq دارای کاستی‌هایی است. ناهمگنی ذاتی نمونه‌های بیولوژیکی گاهی می‌تواند نمای ترنسکریپتومی انواع

Reads Per Kilobase of transcript per Million Transcripts Per و mapped reads; RPKM Million; TPM برای این منظور به کار می‌روند (Wagner *et al.*, 2012). در ادامه، تجزیه و تحلیل‌های آماری مختلفی برای شناسایی ژن‌های بیان متفاوت یا DEGs^{۱۰۶} بین گروه‌های مورد مطالعه انجام می‌شود. این تجزیه و تحلیل‌ها با استفاده از روش‌های آماری نظیر آزمون t، آزمون ANOVA و روش‌های مبتنی بر مدل‌های خطی عمومی انجام می‌پذیرد. ژن‌های شناسایی شده به عنوان DEG، ممکن است در فرایندهای بیولوژیکی مرتبط با شرایط مورد مطالعه نقش داشته باشند (Love *et al.*, 2014). پس از شناسایی DEG‌ها، تجزیه و تحلیل‌های بیوانفورماتیکی دیگری نظیر خوشه‌بندی ژن‌ها (Clustering)، تجزیه و تحلیل مسیرهای بیولوژیکی (Pathway analysis) و همچنین پیش‌بینی عملکرد ژن‌ها (Functional annotation) انجام می‌شود. این تجزیه و تحلیل‌ها به درک عمیق‌تر از نقش ژن‌ها در فرایندهای سلولی و سازوکارهای مولکولی کمک می‌کند (Conesa *et al.*, 2016). در مجموع، مراحل اصلی تجزیه و تحلیل داده‌های Bulk RNA-Seq شامل پیش‌پردازش توالی‌ها، هم‌ترازی به ژنوم مرجع، کمی‌سازی بیان ژن‌ها و سپس تجزیه و تحلیل‌های آماری و بیوانفورماتیکی پیشرفته‌تر است که با استفاده از ابزارهای قدرتمند این حوزه انجام می‌شود. مرحله تجزیه و تحلیل داده Bulk RNA-Seq شامل سری پیچیده‌ای از مراحل

¹⁰⁷ Translational researches¹⁰⁶ Differentially expressed genes

توالی‌یابی پرتوان و تجزیه و تحلیل محاسباتی پیشرفته، Bulk RNA-Seq به ابزاری ایده‌آل برای کشف سازوکارهای پایه‌ای که عملکرد سلولی، تکوین و بیماری را کنترل می‌کنند تبدیل شده است. در ادامه، به مراحل تجزیه و تحلیل داده‌های Bulk RNAseq حاصل از توالی‌یابی کل ترنسکرپتوم یک بافت زنده اشاره می‌شود (شکل ۱۴). اسکرپ‌های تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (Conesa *et al.*, 2016; Love *et al.*, 2014; Pertea *et al.*, 2016; Jiang

سلولی یا زیر جمعیت‌های نادر را پنهان کند که ممکن است در داده‌ها نادیده گرفته شده یا کم‌رنگ باشند. علاوه بر این، تفسیر داده‌های Bulk RNA-Seq می‌تواند پیچیده باشد و نیازمند درک عمیق از بیوانفورماتیک و روش‌های آماری برای اطمینان از نتیجه‌گیری‌های دقیق و معنی‌دار است (Haque *et al.*, 2017). در نتیجه، Transcriptomics by Bulk RNA-Seq به یک روش دگرگون‌ساز در حوزه زیست‌شناسی مولکولی تبدیل شده است که به محققان توانایی بی‌سابقه‌ای برای کشف پویایی و پیچیدگی بیان ژن می‌دهد. با بهره‌گیری از قدرت



شکل ۱۴- مراحل تجزیه و تحلیل داده‌های RNA-seq
Figure 14. RNA-seq data analysis steps

امکان ساخت کتابخانه‌های قابل توالی‌گیری فراهم شود (Ziegenhain *et al.*, 2017). بعد از استخراج و تقویت RNA، مرحله سنتز cDNA و ساخت کتابخانه‌های DNA مشابه روش Bulk RNA-Seq انجام می‌شود. در این مرحله، cDNA هر سلول به‌طور جداگانه سنتز شده و سپس با استفاده از آداپتورها و بارکدهای اختصاصی برچسب‌گذاری می‌شود. این امر به منظور افزایش تنوع کتابخانه و امکان تفکیک توالی‌های مربوط به هر سلول در مرحله توالی‌گیری صورت می‌گیرد (Ziegenhain *et al.*, 2017). در نهایت، کتابخانه‌های آماده‌شده توسط دستگاه‌های پیشرفته توالی‌یاب مورد توالی‌گیری قرار می‌گیرند. خروجی این مرحله، مجموعه‌ای از توالی‌های کوتاه است که هر کدام به یک سلول خاص تعلق دارند. این توالی‌ها در مرحله تجزیه و تحلیل داده‌ها مورد استفاده قرار می‌گیرند. در مجموع، مراحل اصلی Wet-lab روش scRNA-Seq شامل تهیه نمونه تک‌سلولی، استخراج و تقویت RNA، سنتز cDNA و ساخت کتابخانه‌های DNA، و در نهایت توالی‌گیری است که با استفاده از پروتکل‌ها و ابزارهای تخصصی انجام می‌شود. این مراحل حساس و دقیق هستند و کیفیت داده‌های نهایی را تحت تأثیر قرار می‌دهند. پس از انجام مراحل آزمایشگاهی و تهیه کتابخانه‌های DNA مربوط به هر سلول در روش scRNA-Seq و در ادامه تولید توالی‌های خام (Reads) مربوط به هر سلول از طریق دستگاه‌های پیشرفته توالی‌یاب، مرحله تجزیه و تحلیل داده‌ها آغاز می‌شود. این مرحله شامل چندین گام اصلی است که با

مطالعه ترنسکریپتوم از طریق Single-cell RNA-seq

توالی‌یابی RNA در سطح تک سلول یا-scRNA-Seq¹⁰⁸ یک روش پیشرفته در علوم زیستی است که توانایی بررسی ترنسکریپتوم (که شامل mRNA، tRNA، rRNA و سایر انواع RNA است) را در سطح تک‌سلولی فراهم می‌کند. با استفاده از این روش، امکان شناسایی و کمی‌سازی بیان ژن در سطح تک‌سلولی فراهم می‌شود که به درک عمیق‌تر از تنوع و ویژگی‌های سلول‌های مختلف در یک جمعیت کمک می‌کند (Ziegenhain *et al.*, 2017). فرایند اصلی در روش scRNA-Seq مانند روش Bulk RNA-Seq چندین مرحله است. در ابتدا، نمونه مورد مطالعه که می‌تواند شامل یک بافت، ارگان یا جمعیت سلولی باشد، باید به صورت تک‌سلولی تهیه شود. این کار با استفاده از روش‌های مکانیکی، آنزیمی یا میکروسکوپی انجام می‌شود. پس از تک‌سلولی‌سازی، سلول‌ها به‌طور جداگانه در چاهک‌های مخصوص یا کپسول‌های میکروفلوئیدی جمع‌آوری می‌شوند. سپس، مراحل استخراج RNA، سنتز cDNA و ساخت کتابخانه‌های DNA مشابه روش Bulk RNA-Seq انجام می‌شود. البته در این مرحله به دلیل مقدار بسیار پایین RNA در تک‌سلول‌ها، از روش‌های تقویت سیگنال RNA همچون اسکرایب (SCRIB-Seq) و SMART-Seq استفاده می‌شود. این روش‌ها امکان افزایش و تکثیر مولکول‌های RNA را فراهم می‌کنند تا

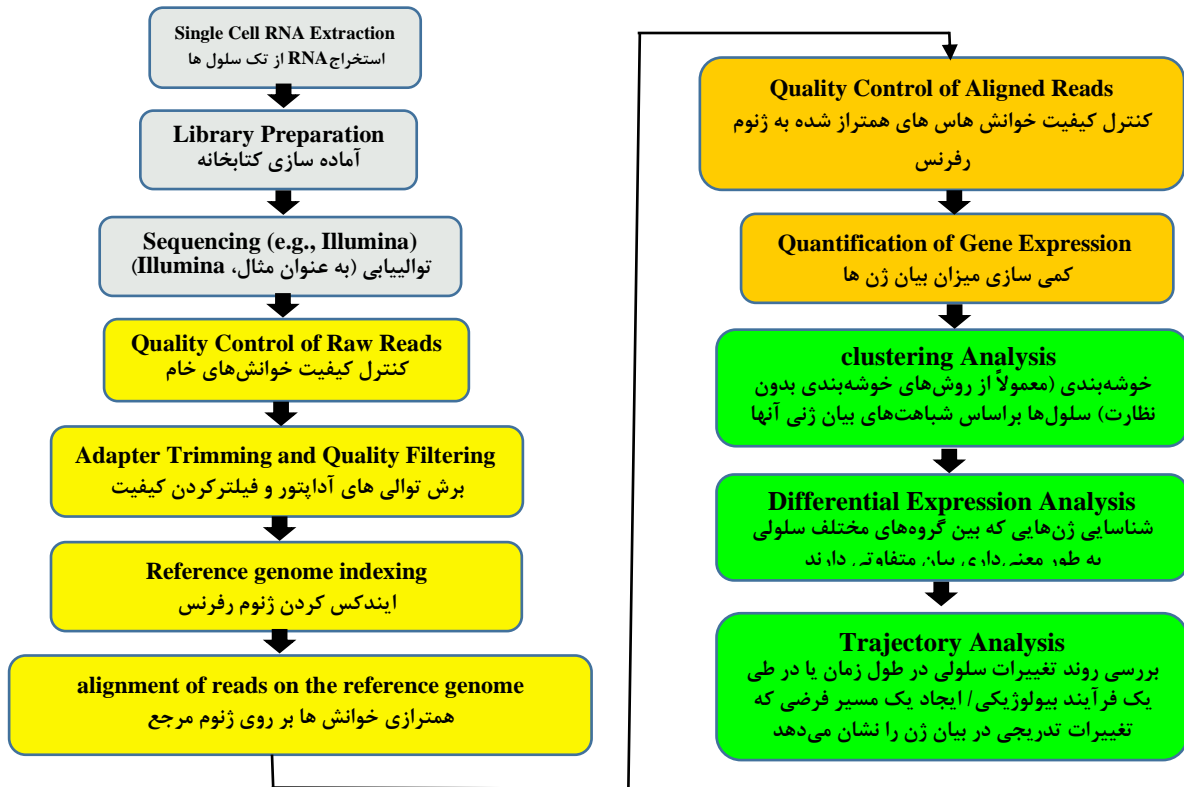
¹⁰⁸ Single-cell RNA-seq

(خوشه‌بندی سلول‌ها)^{۱۱۲}، بررسی روند تغییرات سلولی در طول زمان یا در طی یک فرایند بیولوژیکی و ایجاد یک مسیر فرضی که تغییرات تدریجی در بیان ژن را نشان می‌دهد (تجزیه و تحلیل مسیرهای تمایزی)^{۱۱۳}، شناسایی ژن‌هایی که بین گروه‌های مختلف سلولی به‌طور معنی‌داری بیان متفاوتی دارند (تجزیه و تحلیل تغییرات بیان)^{۱۱۴} و همچنین پیش‌بینی عملکرد ژن‌ها (حاشیه نویسی عملکردی)^{۱۱۵} انجام می‌شود. این تجزیه و تحلیلها با استفاده از ابزارهای تخصصی نظیر Scanpy, Seurat, MAST, edgeR, Deseq2, RaceID, SC3, CIDR, PAGA, Wishbonet, Slingshot, Monocle, Velocity صورت می‌پذیرد (Kiselev *et al.*, 2019). در مجموع، مراحل اصلی تجزیه و تحلیل داده‌های scRNA-Seq شامل پیش‌پردازش توالی‌ها، هم‌ترازی به ژنوم مرجع، کمی‌سازی بیان ژن‌ها و سپس تجزیه و تحلیل‌های آماری و بیوانفورماتیکی پیشرفته‌تر است که با استفاده از ابزارهای قدرتمند این حوزه انجام می‌شود. روش scRNA-Seq به دلیل قابلیت شناسایی و کمی‌سازی دقیق بیان ژن‌ها در سطح تک‌سلولی، کاربردهای گسترده‌ای در زمینه‌های مختلف علوم زیستی و پزشکی دارد. از جمله این کاربردها می‌توان به شناسایی زیرجمعیت‌های سلولی و روابط بین آنها، بررسی سازوکارهای تمایز سلولی، رهیافت‌های جدید در تشخیص و درمان بیماری‌ها، کشف سلول‌های بنیادی و پرتوان،

استفاده از نرم‌افزارها و ابزارهای بیوانفورماتیکی تخصصی انجام می‌شود (Haque *et al.*, 2017). در ابتدا، مرحله پیش‌پردازش توالی‌های خام انجام می‌گیرد. در این مرحله، توالی‌های با کیفیت پایین، توالی‌های کوتاه و آداپتورها حذف می‌شوند. همچنین در برخی موارد، حذف توالی‌های متعلق به ریبوزوم یا سایر آلودگی‌ها نیز ضروری است. برای انجام این مراحل از ابزارهایی مانند Fastp و Cutadapt استفاده می‌شود (Chen *et al.*, 2018). پس از پیش‌پردازش، توالی‌های خام به ژنوم مرجع هم‌تراز^{۱۰۹} می‌شوند. این مرحله به منظور تعیین منشأ هر یک از توالی‌ها و ارتباط آنها با ژن‌ها انجام می‌شود. برای این منظور از ابزارهایی مانند STAR, CellRanger و Alevin استفاده می‌شود که توانایی هم‌ترازی توالی‌ها به ژنوم مرجع را با دقت و سرعت بالا دارند (Haque *et al.*, 2017). پس از هم‌ترازی توالی‌ها، مرحله کمی‌سازی بیان ژن‌ها در هر سلول انجام می‌شود. در این مرحله، تعداد توالی‌های منسوب به هر ژن در هر سلول محاسبه شده و معیارهایی مانند UMI^{۱۱۰} و TPM^{۱۱۱} برای بیان ژن‌ها به کار می‌روند. این مرحله به منظور شناسایی الگوهای بیانی متفاوت در جمعیت سلولی انجام می‌شود (Haque *et al.*, 2017). در ادامه، تجزیه و تحلیل‌های آماری و بیوانفورماتیکی پیشرفته‌ای برای شناسایی زیرجمعیت‌های سلولی و سلول‌ها براساس شباهت‌های بیان ژنی آنها

¹¹² Cell clustering¹¹³ Trajectory Analysis¹¹⁴ Differential expression analysis¹¹⁵ Functional annotation¹⁰⁹ Align¹¹⁰ Unique Molecular Identifier¹¹¹ Transcript per million

درک تنوع ژنتیکی در جمعیت‌های سلولی و همچنین شناسایی نشانگرهای سلولی اختصاصی اشاره کرد (Stuart & Satija, 2019). در مجموع، این روش به عنوان یک ابزار قدرتمند در علوم زیستی، قابلیت بررسی ترنسکریپتوم را در سطح تک‌سلولی فراهم می‌کند. این امر به درک عمیق‌تر تنوع و خصوصیات سلول‌ها، شناسایی زیرجمعیت‌های سلولی و سازوکارهای سلولی کمک شایانی می‌کند و کاربردهای گسترده‌ای در حوزه‌های مختلف پزشکی و زیست‌شناسی دارد. در ادامه، به مراحل تجزیه و تحلیل داده‌های scRNAseq حاصل از توالی‌یابی تک سلول‌ها^{۱۱۶} اشاره می‌شود (شکل ۱۵). اسکرپ‌های تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (Stuart *et al.*, 2019; Luecken & Theis, 2018; Hwang *et al.*, 2018).



شکل ۱۵- مراحل تجزیه و تحلیل داده‌های Single cell RNA-seq

Figure 15. Single cell RNA-seq data analysis steps

ریز RNAها نقش مهمی در تنظیم بیان ژن ایفا می‌کنند و در فرایندهای بیولوژیکی متنوعی مانند تمایز سلولی، تنظیم رشد و توسعه، دفاع ضد ویروسی و پاسخ ایمنی درگیر هستند (Bartel, 2018). شناسایی و بررسی این ریز RNAها می‌تواند بینش‌های مهمی در مورد مکانیزم‌های مولکولی حیاتی ایجاد کند. Small RNA Sequencing به محققان اجازه می‌دهد تا پروفایل بیان ریز RNAها را در نمونه‌های مختلف مورد مطالعه قرار دهند و تغییرات بیان این ریز RNAها را در پاسخ به شرایط فیزیولوژیکی، بیماری‌ها یا مداخلات درمانی بررسی

مطالعه RNAهای کوچک یا Small RNA Sequencing

توالی‌یابی RNAهای کوچک یا Small RNAseq¹¹⁷ یک روش کارآمد در زیست‌شناسی مولکولی است که به محققان اجازه می‌دهد تا طیف گسترده‌ای از ریز RNAهای سلول را شناسایی و تجزیه و تحلیل کنند. این روش اساساً مبتنی بر تعیین توالی نسل جدید (NGS) است و امکان شناسایی و کمی‌سازی انواع مختلف ریز RNAها مانند miRNAها، siRNAها، piRNAها و سایر ریز RNAهای مهم را فراهم می‌کند (Bartel, 2018).

¹¹⁷ Small RNA Sequencing

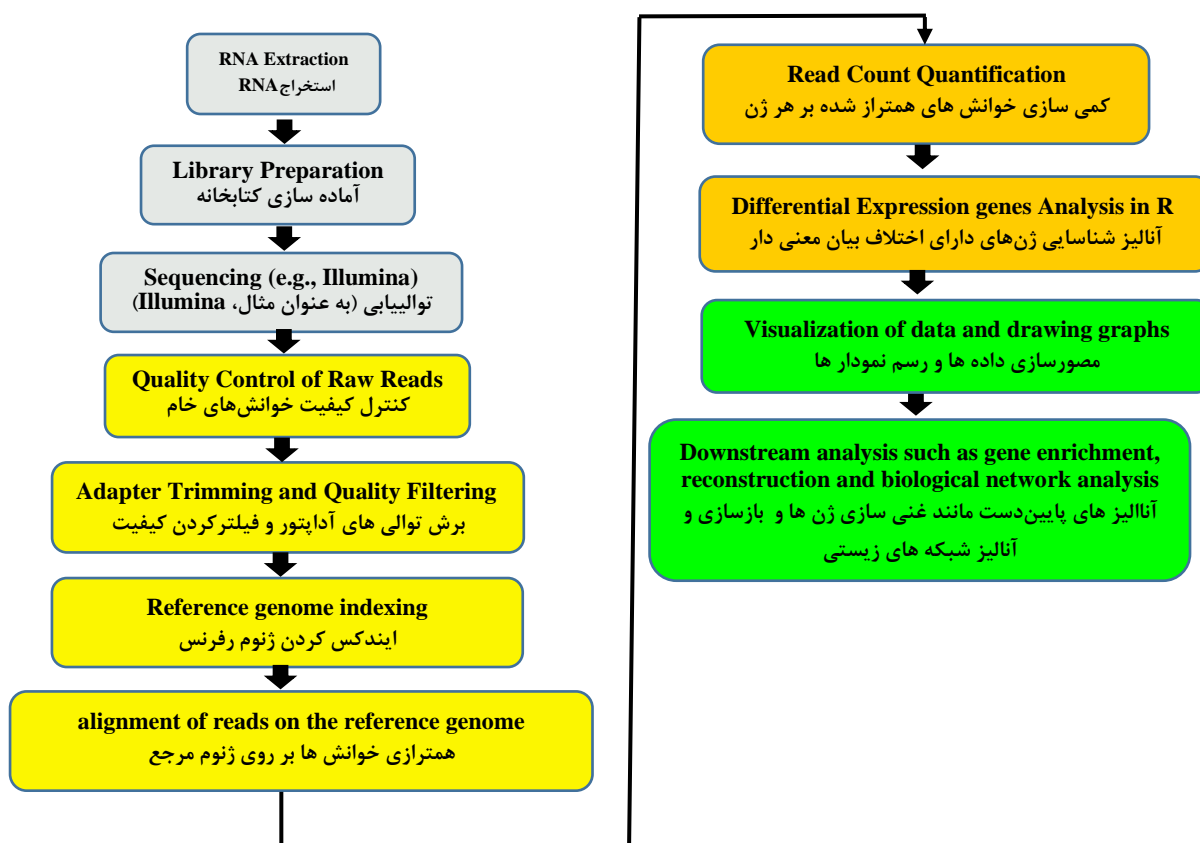
به انتهای ۳' و ۵' cDNA هم متصل شده و سپس توالی‌های آنها با استفاده از دستگاه‌های توالی‌یاب نسل جدید (NGS) تعیین می‌شوند. این مرحله می‌تواند با استفاده از پلتفرم‌های مختلفی مثل Illumina، Ion Torrent یا Ion Proton انجام شود (Pritchard *et al.*, 2012). پس از انجام مراحل آزمایشگاهی استاندارد برای تهیه و تعیین توالی ریز RNAها، مرحله بسیار مهم دیگری که نقش کلیدی در استخراج اطلاعات مفید از داده‌ها ایفا می‌کند، تجزیه و تحلیل داده‌های Small RNA Sequencing است. این فرایند شامل چندین مرحله تخصصی است که به کمک آنها می‌توان به درک عمیق‌تری از پروفایل بیان ریز RNAها و نقش آنها در فرایندهای بیولوژیکی دست یافت. اولین گام در تجزیه و تحلیل داده‌های Small RNA Sequencing، پیش‌پردازش و آماده‌سازی داده-هاست. در این مرحله، توالی‌های خام به دست آمده از تعیین توالی، ابتدا باید از توالی‌های آداپتور و سایر آلودگی‌های احتمالی پاکسازی شوند. همچنین، توالی‌های نامعتبر و با کیفیت پایین باید از مجموعه داده حذف گردند (Friedländer *et al.*, 2012). این کار به کمک نرم‌افزارهای تخصصی مانند Cutadapt و FastQC انجام می‌شود. پس از مرحله پیش‌پردازش، توالی‌های تمیز شده باید به ژنوم مرجع هم‌تراز شوند. این فرایند به کمک ابزارهای هم‌ترازی مانند Bowtie2 یا STAR انجام می‌شود و به محققان امکان می‌دهد تا توالی‌های ریز RNA را به موقعیت‌های ژنومی مربوطه نسبت دهند (Friedländer *et al.*, 2012). این مرحله همچنین می-

کنند. این روش شامل مراحل آزمایشگاهی استاندارد و منسجمی است که برای به دست آوردن نتایج قابل اعتماد و دقیق بسیار مهم است. در ادامه مراحل اصلی Wet-lab برای انجام روش Small RNA Sequencing بررسی می‌شود. اولین قدم در فرایند Small RNA Sequencing، استخراج RNA کل از نمونه مورد مطالعه است. این مرحله بسیار حساس بوده و باید با دقت زیاد انجام شود تا RNA با کیفیت و غلظت مناسب به دست آید (Umu *et al.*, 2018). پس از استخراج RNA، مرحله بعدی جداسازی ریز RNAهای موردنظر است. این کار به‌طور معمول با استفاده از ستون‌های کروماتوگرافی و یا از طریق اتصال ریز RNAها به آداپتورهای متصل شونده به مهره‌های مغناطیسی^{۱۱۸} انجام می‌شود. در این مرحله، ریز RNAهای کمتر از ۲۰۰ نوکلئوتید از سایر انواع RNAها مانند mRNA و rRNA جدا می‌شوند (Pritchard *et al.*, 2012). پس از جداسازی ریز RNA^{۱۱۹}ها، در مرحله بعدی، این مولکول‌ها به کمک آنزیم‌های رونویسی معکوس به cDNA تبدیل می‌شوند. این مرحله به محققان اجازه می‌دهد تا توالی‌های ریز RNA را به راحتی تعیین کنند (Pritchard *et al.*, 2012). در طی این فرایند، اغلب از آداپتورهای اختصاصی برای اتصال به انتهای ۳' و ۵' ریز RNAها استفاده می‌شود تا امکان تعیین توالی هر دو انتها فراهم شود. پس از مراحل قبلی، مولکول‌های cDNA آماده سازی و توالی‌یابی می‌شوند. در این مرحله، آداپتورهای اختصاصی

¹¹⁸ Bead-based retrieval using adapter balloons
¹¹⁹ microRNA

در اختیار محققان می‌گذارد. برخی از این کاربردهای کلیدی عبارتند از: کشف و مطالعه ریز RNAهای جدید و نقش آنها در فرایندهای سلولی و بیماری‌ها، بررسی تغییرات بیان ریز RNAها در پاسخ به شرایط فیزیولوژیکی، بیماری‌ها و مداخلات درمانی، شناسایی بیومارکرهای ریز RNAی برای تشخیص و پیش‌بینی بیماری‌ها، درک نقش ریز RNAها در تنظیم بیان ژن و مکانیزم‌های مولکولی مختلف، طراحی و ارزیابی داروهای مبتنی بر ریز RNAها برای درمان بیماری‌ها، مطالعه نقش ریز RNAها در تکامل، رشد و توسعه ارگانیسم‌ها، بررسی تنوع ریز RNAها در جوامع میکروبی و نقش آنها در تعامل میزبان-پاتوژن. در مجموع، Small RNA Sequencing یک ابزار قدرتمند در زیست‌شناسی مولکولی است که به محققان اجازه می‌دهد تا به درک عمیق‌تری از نقش ریز RNAها در فرایندهای زیستی و بیماری‌های مختلف دست یابند. این روش مجموعه‌ای از اطلاعات ارزشمند را فراهم می‌کند که می‌تواند به توسعه روش‌های تشخیصی و درمانی نوآورانه در پزشکی کمک کند. در ادامه، به مراحل تجزیه و تحلیل داده‌های Small RNAseq که حاصل از توالی‌یابی RNAهای کوچک (small RNAs) است اشاره می‌شود (شکل ۱۶). اسکرپ‌های تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (Tam *et al.*, 2015; Friedländer *et al.*, 2012; Love *et al.*, 2014).

تواند به شناسایی انواع مختلف ریز RNAها مانند miRNA، siRNA و piRNAها کمک کند. مرحله بعدی، کمی‌سازی و تحلیل الگوهای بیان ریز RNAهاست. در این مرحله، با استفاده از ابزارهای آماری و بیوانفورماتیکی مثل HT-seq و featureCounts میزان بیان هر ریز RNA در نمونه‌ها محاسبه می‌شود و در ادامه الگوهای بیان ریز RNAها بین گروه‌های مختلف (به عنوان مثال، گروه‌های بیمار و سالم) از طریق ابزارهایی مانند DESeq2 و edgeR مقایسه می‌شود تا ریز RNAهای متمایز شناسایی گردند (Liang *et al.*, 2016). پس از این مراحل، برای درک بهتر نقش بیولوژیکی ریز RNAهای شناسایی شده، می‌توان از تجزیه و تحلیل‌های تکمیلی مانند پیش‌بینی ژن‌ها، تحلیل مسیرهای بیولوژیکی و شبکه‌های تنظیمی استفاده کرد. این تجزیه و تحلیل‌ها به کمک نرم‌افزارهای تخصصی مانند miRTarBase، TargetScan و Cytoscape انجام می‌شوند (Agarwal *et al.*, 2015). در نهایت، نتایج حاصل از تجزیه و تحلیل داده‌های Small RNA Sequencing می‌تواند به شناسایی ریز RNAهای مرتبط با بیماری‌ها، فرایندهای زیستی و پاسخ‌های سلولی به مداخلات مختلف منجر شوند. روش Small RNA Sequencing به دلیل قدرت شناسایی و تجزیه و تحلیل ریز RNAها، اطلاعات ارزشمندی که کاربردهای متنوعی در زیست‌شناسی، پزشکی و سایر زمینه‌های علمی دارند را



شکل ۱۶- مراحل تجزیه و تحلیل داده‌های Small RNA-seq

Figure 16. Small RNA-seq data analysis steps

هر موجود زنده شامل اطلاعات ژنتیکی منحصر به فرد و غنی است که نقش کلیدی در تعیین خصوصیات فنوتیپی و رفتاری آن ایفا می‌کند. شناخت کامل ساختار ژنومی موجودات از اهمیت بسزایی برخوردار است و می‌تواند به درک بهتر فرایندهای حیاتی، تکامل گونه‌ها و توسعه رویکردهای درمانی و تشخیصی کمک کند. در مواردی که ژنوم مرجع موجود نیست یا ناکافی است، *de novo* Genome Assembly به عنوان یک راه حل مؤثر مطرح می‌شود (Sedlazeck et al., 2018). فرایند *de novo* Genome Assembly شامل چندین مرحله آزمایشگاهی استاندارد و پیچیده‌ای است که در آزمایشگاه‌های تخصصی

مطالعه ژنوم از طریق *de novo* Genome

Assembly

مطالعه ژنومیک از طریق مونتاژ و سرهم‌بندی ژنوم^{۱۲۰} یک روش مهم و پیشرفته در زمینه ژنومیک است که به محققان امکان می‌دهد تا توالی ژنوم یک موجود زنده را بدون استفاده از یک ژنوم مرجع، از ابتدا بازسازی کنند. این رویکرد به دلیل قابلیت شناسایی و مجموعه‌سازی توالی‌های ژنومی ناشناخته و منحصر به فرد، نقش حیاتی در کشف و فهم عمیق‌تر ساختار ژنتیکی ارگانیزم‌های مختلف ایفا می‌کند (Nagarajan & Pop, 2013). ژنوم

¹²⁰ *de novo* Genome assembly

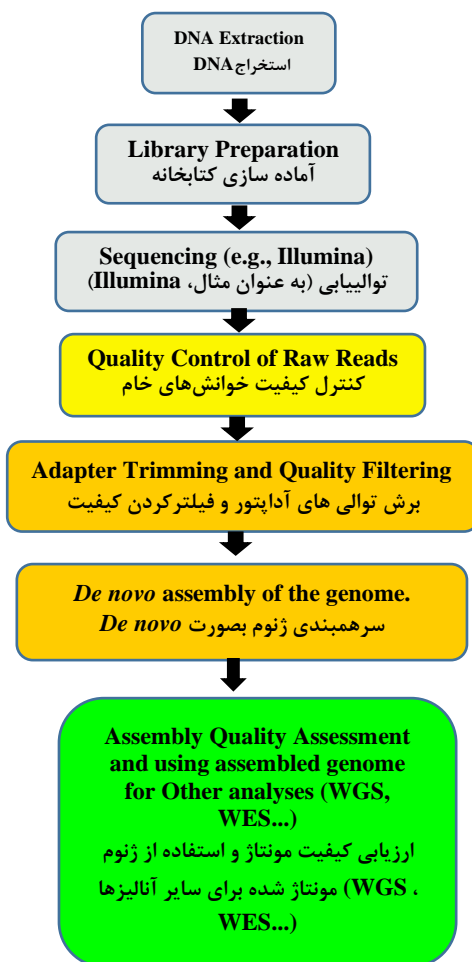
ژنومیک انجام می‌شود و باید با دقت و کنترل کیفیت بالا انجام شوند تا نتایج قابل اعتماد به دست آید. اولین قدم در فرایند *de novo Genome Assembly* استخراج DNA ژنومی با کیفیت و غلظت مطلوب از نمونه مورد مطالعه است. این مرحله باید با استفاده از پروتکل‌های استاندارد مانند فنل-کلروفرم یا ستون‌های سیلیکایی انجام شود تا DNA با ویژگی‌های مناسب (از جمله طول مولکولی بالا و قطعات کامل) به دست آید (Eklom & Wolf, 2014). پس از استخراج DNA، در مرحله بعدی این مولکول‌ها به قطعات کوچک‌تر شکسته می‌شوند. این کار می‌تواند با استفاده از آنزیم‌های محدودکننده، روش‌های مکانیکی مانند سونیکاسیون^{۱۲۱} یا تزریق میکروسپال^{۱۲۲} انجام شود. قطعات DNA حاصل از مرحله قبل، سپس توسط فناوری‌های توالی‌یابی نسل جدید (NGS) مورد تعیین توالی قرار می‌گیرند. پلتفرم‌های متعددی مانند Illumina, Ion Torrent و PacBio برای این منظور در دسترس هستند که هر یک ویژگی‌ها و کاستی‌های خاص خود را دارند (Eklom & Wolf, 2014). انتخاب مناسب‌ترین پلتفرم بسته به نیاز پروژه (مانند طول قرائت مورد نیاز و هزینه) صورت می‌گیرد. در ادامه، استراتژی‌های Boinformatics برای مجموعه‌سازی و مونتاژ این توالی‌های خرد شده به کار گرفته می‌شوند. پس از انجام مراحل آزمایشگاهی استاندارد برای استخراج DNA، تعیین توالی و مجموعه‌سازی ژنوم، مرحله بسیار مهم دیگر در فرایند *de novo Genome Assembly*

تجزیه و تحلیل داده‌های حاصل است و تجزیه و تحلیل دقیق داده‌های حاصل از *de novo Genome Assembly* نقش مهمی در دستیابی به ژنوم مونتاژ شده با کیفیت بالا ایفا می‌کند. این بخش نقش کلیدی در استخراج اطلاعات مفید و قابل اعتماد از داده‌های تولید شده ایفا می‌کند. اولین گام در تجزیه و تحلیل داده‌های *de novo Genome Assembly* پیش پردازش و آماده سازی داده‌هاست. در این مرحله، توالی‌های خام به دست آمده از تعیین توالی باید از آلودگی‌ها و قطعات نامعتبر پاکسازی شوند. همچنین، توالی‌های با کیفیت پایین و با طول نامناسب باید از مجموعه داده حذف گردند (Bolger et al., 2014). این کار به کمک نرم‌افزارهای تخصصی مانند Trimmomatic, MultiQC و FastQC انجام می‌شود. پس از مرحله پیش پردازش توالی‌ها، جهت ساخت یک ژنوم پیوسته، توالی‌های خرد با استفاده از الگوریتم‌های پیچیده و قدرتمند مونتاژ همچون De Bruijn graph و Overlap-Layout-Consens جمع می‌شوند. این فرایند شامل چندین مرحله است که به‌طور معمول با استفاده از ابزارهای Boinformatics انجام می‌شود. ابتدا، توالی‌ها بر اساس همپوشانی‌های موجود در آنها در کانتیگ‌ها^{۱۲۳} (قطعات بزرگتر) مجموعه‌سازی می‌شوند. سپس، این کانتیگ‌ها با استفاده از اطلاعات جفت شدن توالی‌ها یا توالی‌های پل زن، به اسکافولدها^{۱۲۴} (قطعات بزرگتر) ادغام می‌شوند (Nagarajan & Pop, 2013). نرم‌افزارهای تخصصی مانند SPAdes.

¹²³ Contigs¹²⁴ Scaffolds¹²¹ Sonication¹²² Microfluidic injection

و فهم سازوکار های تکاملی و جهش ژنتیکی، طراحی و توسعه داروها و واکسن‌های مبتنی بر ژنوم، تجزیه و تحلیل ژنوم میکروارگانیزم‌ها و شناسایی ژن‌های مرتبط با بیماری‌زایی، شناسایی نشانگرهای ژنتیکی مفید در به‌نژادی گیاهان و حیوانات، مطالعه ساختار ژنوم سلول-های سرطانی و مکانیزم‌های مولکولی آنها دارد. به‌طور کلی *de novo* Genome Assembly یک روش قدرتمند در ژنومیک است که به محققان اجازه می‌دهد تا به درک عمیق‌تری از ساختار و تنوع ژنتیکی موجودات زنده دست یابند. این امر می‌تواند منجر به کشفیات مهمی در زمینه‌های مختلف، از زیست‌شناسی تا پزشکی و کشاورزی شود. در ادامه، به مراحل تجزیه و تحلیل *de novo* Genome Assembly از طریق توالی‌یابی کل ژنوم اشاره می‌شود (شکل ۱۷). اسکریپ‌های تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (Sohn & Nam, 2018; Ekblom & Wolf, 2014; Badouin *et al.*, 2017).

MEGAHIT و Canu در اجرای این فرایندها نقش اساسی ایفا می‌کنند (Bankevich *et al.*, 2012). در مرحله بعدی، پس از مجموعه‌سازی ژنوم، کیفیت و دقت آن باید مورد ارزیابی قرار گیرد. این کار با استفاده از شاخص‌هایی مانند N50 طول میانه قطعات، پوشش ژنومی و تعداد قطعات مونتاژ شده انجام می‌شود. ابزارهایی مانند BUSCO و QUAST برای انجام این ارزیابی مورد استفاده قرار می‌گیرند (Gurevich *et al.*, 2013). در نهایت، ژنوم مجموعه‌سازی شده می‌تواند به عنوان ژنوم مرجع برای مطالعات بعدی مورد شناسایی و حاشیه‌نویسی ژن‌ها مورد استفاده قرار گیرد. که این اطلاعات می‌تواند به درک بهتر ساختار ژنتیکی موجود مورد مطالعه و کشف ژن‌های جدید کمک کند. روش *de novo* Genome Assembly به دلیل قدرت کشف و مجموعه‌سازی توالی‌های ژنومی جدید و منحصر به فرد، کاربردهای گسترده‌ای در زمینه‌های مختلف علمی مانند: کشف و شناخت ساختار ژنتیکی گونه‌های جدید و نادر، مطالعه تنوع ژنتیکی میان گونه‌ها و درون گونه‌ای، بررسی

شکل ۱۷- مراحل تجزیه و تحلیل *de novo* Genome AssemblyFigure 17. *de novo* Genome Assembly analysis steps

ترنسکریپتوم را بازسازی می‌کند. مراحل آزمایشگاهی و Wet-lab این روش شامل چندین گام حیاتی است که دقت و کیفیت بالا در هر مرحله برای موفقیت نهایی پروژه ضروری است. این فرایند با جمع‌آوری نمونه‌های بافتی یا سلولی آغاز می‌شود که باید با دقت و سرعت انجام شود تا از تخریب RNA جلوگیری شود (Conesa *et al.*, 2016). نمونه‌ها به‌طور معمول بلافاصله در نیتروژن مایع منجمد می‌شوند یا در محلول‌های پایدارکننده RNA مانند RNAlater قرار می‌گیرند. استخراج RNA کل مرحله بعدی است که می‌تواند با استفاده از کیت‌های

مطالعه ترنسکریپتوم و انجام تجزیه و تحلیل *de novo* Transcriptome از طریق RNAseq assembly روش مطالعه ترنسکریپتوم از طریق ترنسکریپتوم سرهم‌بندی شده^{۱۲۵} یک روش قدرتمند برای بازسازی و تجزیه و تحلیل ترنسکریپتوم در گونه‌هایی است که ژنوم مرجع ندارند یا ژنوم آن‌ها به خوبی شناخته نشده است (Haas *et al.*, 2013). این روش با استفاده از داده‌های توالی‌یابی RNA (RNA-seq) و بدون نیاز به ژنوم مرجع،

¹²⁵ *de novo* Transcriptome assembly

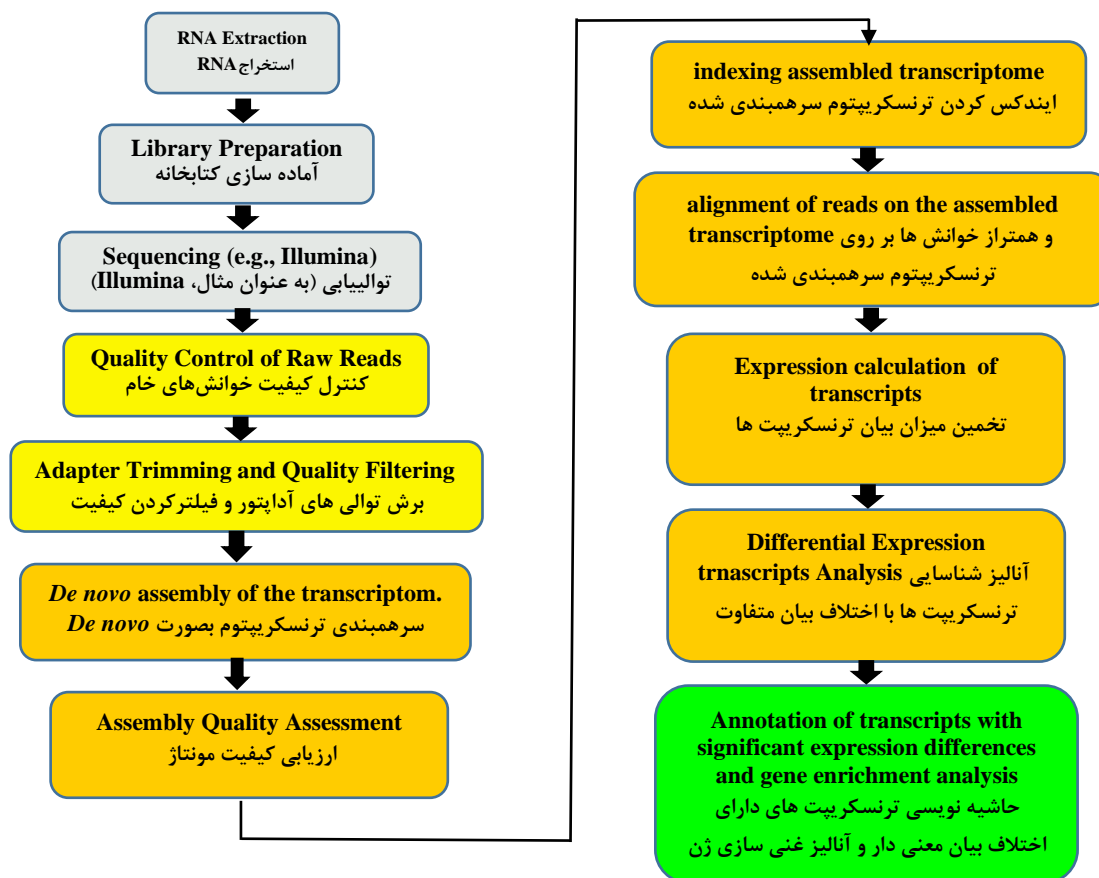
تجاری مانند TRIzol یا RNeasy انجام شود (Rio *et al.*, 2010). کیفیت و کمیت RNA استخراج شده با روش‌هایی مانند الکتروفورز ژل آگارز، اسپکتروفتومتری (نسبت A260/A280) و تجزیه و تحلیل Bioanalyzer ارزیابی می‌شود. RNA با کیفیت بالا باید نسبت A260/A280 حدود ۲/۰ و عدد یکپارچگی RNA (RIN) بالای هشت داشته باشد (Schroeder *et al.*, 2006). پس از تأیید کیفیت RNA، مرحله ساخت کتابخانه cDNA آغاز می‌شود. این مرحله شامل سنتز cDNA، افزودن آداپتورها و تکثیر است. برای سنتز cDNA، به‌طور معمول از آغازگرهای Oligo(dT) یا آغازگرهای Random hexamer استفاده می‌شود تا اطمینان حاصل شود که تمام انواع RNA از جمله mRNA و lncRNA در کتابخانه نهایی نمایندگی می‌شوند (Zhong *et al.*, 2011). سپس، cDNA دو رشته‌ای سنتز می‌شود و آداپتورهای خاص پلتفرم توالی‌یابی به انتهای قطعات متصل می‌شوند. این آداپتورها شامل توالی‌های شناسایی برای اتصال به فلوسل و آغازگرهای توالی‌یابی هستند. در صورت نیاز به تکثیر کتابخانه، به‌طور معمول از PCR با تعداد چرخه کم استفاده می‌شود تا از ایجاد اربیبی در نمایندگی ترانسکرپت‌ها جلوگیری شود (Head *et al.*, 2014). کیفیت و کمیت کتابخانه نهایی با روش‌هایی مانند Bioanalyzer و qPCR ارزیابی می‌شود. در نهایت، کتابخانه آماده شده با استفاده از پلتفرم‌های توالی‌یابی نسل بعدی (NGS) مانند Illumina NovaSeq یا

HiSeq توالی‌یابی می‌شود (Wang *et al.*, 2009). این پلتفرم‌ها قادر به تولید میلیون‌ها خوانش کوتاه (به‌طور معمول ۳۰-۷۵ جفت باز) با کیفیت بالا هستند. برای اطمینان از پوشش کافی ترنسکرپتوم، به‌طور معمول عمق توالی‌یابی ۳۰-۵۰ میلیون خوانش برای هر نمونه توصیه می‌شود، اگرچه این مقدار بسته به پیچیدگی ترنسکرپتوم و اهداف مطالعه می‌تواند متفاوت باشد (Sims *et al.*, 2014). پس از توالی‌یابی، داده‌های خام به صورت فایل‌های FASTQ ذخیره می‌شوند که حاوی توالی خوانش‌ها و امتیازات کیفیت مربوطه هستند. این داده‌ها سپس وارد مرحله تجزیه و تحلیل بیوانفورماتیکی می‌شوند که شامل کنترل کیفیت، پیرایش خوانش‌ها، سرهم‌بندی *de novo* و تجزیه و تحلیل‌های بعدی است. پس از توالی‌یابی، مرحله حیاتی تجزیه و تحلیل داده‌ها آغاز می‌شود. ابتدا، کیفیت خوانش‌ها با ابزارهایی مانند FastQC بررسی و در صورت نیاز، خوانش‌ها با استفاده از نرم افزار Trimmomatic پیرایش می‌شوند. سپس، خوانش‌ها با استفاده از نرم‌افزارهای تخصصی مانند Trinity، Velvet-Oases یا SOAPdenovo-Trans سرهم‌بندی می‌شوند (Xie *et al.*, 2014). این نرم‌افزارها از الگوریتم‌های پیچیده‌ای برای شناسایی و ترکیب خوانش‌های همپوشان استفاده می‌کنند تا کانتیگ‌ها و در نهایت ترانسکرپت‌های کامل را بازسازی کنند. پس از سرهم‌بندی، ارزیابی کیفیت ترنسکرپتوم بازسازی شده با معیارهایی مانند N50، طول متوسط کانتیگ‌ها و درصد خوانش‌های هم‌تراز شده انجام می‌شود. در ادامه، تجزیه و

استفاده می‌شود که تفاوت‌های معنی‌دار در بیان ژن‌ها را بین شرایط مختلف آزمایشی تعیین می‌کند (Robinson *et al.*, 2010). پس از شناسایی DEGs، تجزیه و تحلیل‌های عملکردی مانند غنی‌سازی GO^{۱۲۶} و تجزیه و تحلیل مسیر انجام می‌شود از طریق نرم‌افزارهایی مانند Blast2GO و Trinotate انجام می‌شود تا درک عمیق‌تری از فرایندهای بیولوژیکی درگیر به دست آید (Bryant *et al.*, 2017; Conesa *et al.*, 2005; Conesa *et al.*, 2008). این رویکرد یکپارچه از *de novo* Transcriptome Assembly و تجزیه و تحلیل RNA-seq امکان مطالعه جامع الگوهای بیان ژن و عملکرد آنها را در گونه‌های غیرمدل فراهم می‌کند و به درک بهتر سازوکارهای مولکولی در پاسخ به شرایط محیطی، بیماری‌ها و سایر فاکتورهای بیولوژیکی کمک می‌کند. با پیشرفت فناوری‌های توالی‌یابی و ابزارهای بیوانفورماتیکی، انتظار می‌رود که این روش‌ها در آینده نقش مهم‌تری در تحقیقات زیست‌شناسی و پزشکی ایفا کنند و به کشف‌های جدیدی در زمینه عملکرد ژن‌ها و تنظیم بیان آن‌ها منجر شوند. در ادامه به مراحل تجزیه و تحلیل *de novo* transcriptome Assembly- RNAseq از طریق توالی‌یابی کل ترنسکرپتوم اشاره می‌شود (شکل ۱۸). اسکریپ‌های تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (Haas *et al.*, 2013; Conesa *et al.*, 2016; Love *et al.*, 2014; Grabherr *et al.*, 2011; Martin, 2011).

تحلیل‌های بیوانفورماتیکی پیشرفته‌تری مانند شناسایی ORF‌ها، پیش‌بینی ژن‌ها و حاشیه‌نویسی عملکردی با استفاده از پایگاه‌های داده مانند UniProt، KEGG و GO صورت می‌گیرد (Conesa *et al.*, 2016). این روش کاربردهای گسترده‌ای در زمینه‌های مختلف زیست‌شناسی دارد. در اکولوژی مولکولی، برای مطالعه تنوع ژنتیکی و سازگاری‌های گونه‌های غیرمدل استفاده می‌شود. در کشاورزی، برای شناسایی ژن‌های مرتبط با صفات مهم اقتصادی در گیاهان و حیوانات اهلی به کار می‌رود. در زمینه پزشکی، برای کشف نشانگرهای زیستی جدید و اهداف دارویی در بیماری‌های مختلف استفاده می‌شود (Čalounová, 2021; Eldem *et al.*, 2017; Ungaro *et al.*, 2017). علاوه بر این، *de novo* transcriptome Assembly در مطالعات تکاملی برای بررسی تغییرات ژنومی در طول زمان و در میان گونه‌های مختلف کاربرد دارد. پس از سرهم‌بندی ترنسکرپتوم، تجزیه و تحلیل RNA-seq با استفاده از ترنسکرپتوم سرهم‌بندی شده انجام می‌شود. این مرحله شامل هم-ترازی خوانش‌های RNA-seq به ترنسکرپتوم سرهم‌بندی شده، کمی‌سازی بیان ژن‌ها و شناسایی ژن‌های با بیان متفاوت (DEGs) است (Love *et al.*, 2014). برای هم‌ترازی خوانش‌ها، نرم‌افزارهایی مانند Bowtie2 یا STAR استفاده می‌شوند. سپس، با استفاده از ابزارهایی مانند RSEM یا Kallisto، سطح بیان هر ترانسکرپت محاسبه می‌شود. در نهایت، برای شناسایی DEGs، از بسته‌های آماری مانند DESeq2 یا edgeR

¹²⁶ Gene ontology enrichment



شکل ۱۸- مراحل تجزیه و تحلیل *de novo* Transcriptome Assembly-RNA-seq

Figure 18. De Novo transcriptome assembly-RNA-seq analysis steps

می‌شوند. این مرحله شامل انتخاب ژن‌ها، آگزون‌ها یا مناطق تنظیمی خاص بر اساس اهداف مطالعه است (Mertes *et al.*, 2011). مراحل آزمایشگاهی و-Wet lab روش توالی‌یابی هدفمند شامل چندین گام اساسی است که با دقت و ظرافت خاصی انجام می‌شوند. در فرایند توالی‌یابی هدفمند، ابتدا نواحی مورد نظر از ژنوم انتخاب و طراحی می‌شوند. این مرحله شامل انتخاب ژن‌ها، آگزون‌ها یا مناطق تنظیمی خاص بر اساس اهداف مطالعه است (Mertes *et al.*, 2011). در مرحله بعد استخراج DNA با کیفیت بالا از نمونه‌های بیولوژیکی مانند خون، بافت یا سلول‌های کشت داده شده انجام می‌شود (Psifidi *et al.*,

توالی‌یابی هدفمند (TGS)

روش توالی‌یابی هدفمند یا TGS^{۱۲۷} در مطالعه مناطق خاصی از ژنوم، امکان بررسی دقیق و عمیق نواحی مورد نظر را فراهم می‌کند (Mamanova *et al.*, 2010). این روش با تمرکز بر توالی‌یابی مناطق خاص ژنومی، امکان شناسایی واریانت‌های ژنتیکی مرتبط با بیماری‌ها، مطالعه تکامل مولکولی و بررسی تنوع ژنتیکی را فراهم می‌آورد (Katsanis & Katsanis, 2013). در فرایند توالی‌یابی هدفمند، ابتدا نواحی مورد نظر از ژنوم انتخاب و طراحی

¹²⁷ Targeted sequencing

2015). سپس، کیفیت و کمیت DNA استخراج شده با استفاده از روش‌هایی مانند الکتروفورز ژل آگارز و اسپکتروفتومتری ارزیابی می‌گردد (Desjardins & Conklin, 2010). در مرحله بعد، DNA به قطعات کوچک‌تر (به‌طور معمول ۳۰۰-۱۵۰ جفت باز) شکسته می‌شود که این کار می‌تواند با روش‌های مکانیکی مانند سونیکاسیون یا روش‌های آنزیمی انجام شود (Head et al., 2014). پس از شکستن DNA، مرحله ترمیم انتهاها و اتصال آداپتورها انجام می‌شود. در این مرحله، انتهای قطعات DNA صاف شده و آداپتورهای اختصاصی به آنها متصل می‌شوند که برای مراحل بعدی توالی‌یابی ضروری هستند (Kozarewa et al., 2015). سپس، مرحله کلیدی غنی‌سازی مناطق هدف انجام می‌شود که می‌تواند با روش‌های مختلفی مانند دورگه‌شدن در محلول، کپچر بر پایه آرایه یا PCR مبتنی بر آمپلیکون صورت گیرد (Mamanova et al., 2010). در روش دورگه‌شدن در محلول، کاوشگرهای اولیگونوکلوئوتیدی طراحی شده برای مناطق هدف با کتابخانه DNA هیبرید می‌شوند و سپس با استفاده از مهره‌های مغناطیسی جدا می‌گردند (Gnirke et al., 2009). پس از غنی‌سازی، کتابخانه DNA تکثیر می‌شود که به‌طور معمول با استفاده از PCR مولسیون یا تکثیر پل انجام می‌شود (Metzker, 2010). در نهایت، کتابخانه غنی‌شده و تکثیر شده برای توالی‌یابی با استفاده از پلتفرم‌های توالی‌یابی نسل بعدی (NGS) مانند Ion Torrent، Oxford Nanopore آماده می‌شود که شامل تعیین غلظت دقیق، نرمال‌سازی و در برخی موارد ترکیب نمونه‌های مختلف با استفاده از شاخص‌های منحصر به فرد است (Buermans

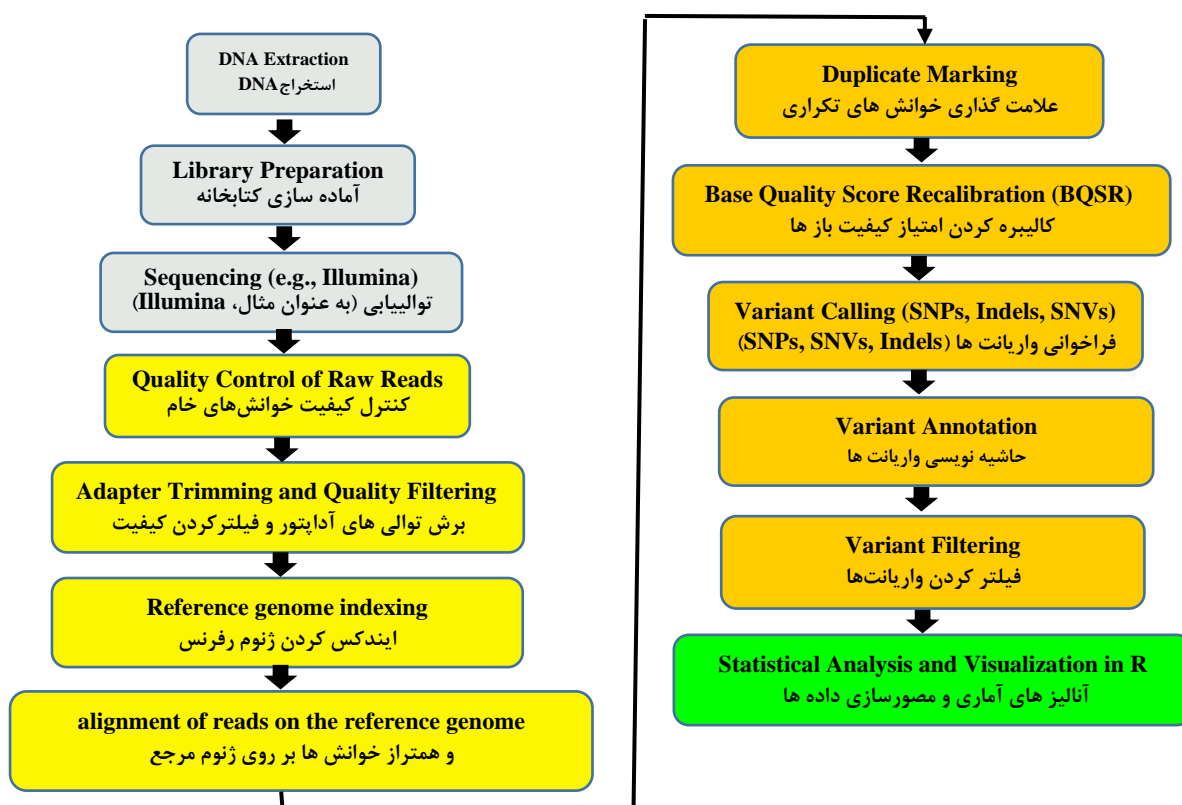
2016; Goodwin et al., 2016; den Dunnen & Goodwin, 2016). کنترل کیفیت در تمام مراحل فرایند بسیار مهم است و به‌طور معمول با استفاده از روش‌هایی مانند qPCR و الکتروفورز موینه انجام می‌شود تا از کیفیت و کمیت مناسب کتابخانه نهایی اطمینان حاصل شود (Robin et al., 2016). پیشرفت‌های اخیر در روش‌های آزمایشگاهی مانند استفاده از روبات‌های خودکار برای آماده‌سازی کتابخانه و بهبود پروتکل‌های غنی‌سازی، باعث افزایش دقت، تکرارپذیری و کارایی فرایند توالی‌یابی هدفمند شده است (Ballester et al., 2016). تجزیه و تحلیل داده‌های توالی‌یابی هدفمند شامل چندین مرحله پیچیده و مهم است که با دقت و ظرافت خاصی انجام می‌شود. ابتدا، کنترل کیفیت داده‌های خام با استفاده از ابزارهایی مانند FastQC یا MultiQC انجام می‌شود تا کیفیت خوانش‌ها، توزیع نوکلئوتیدها و وجود توالی‌های آداپتور بررسی شود (Ewels et al., 2016; Andrews, 2010). سپس، پیش‌پردازش داده‌ها با استفاده از ابزارهایی مانند Trimmomatic یا Cutadapt انجام می‌شود تا آداپتورها و توالی‌های با کیفیت پایین حذف شوند (Bolger et al., 2011; Martin, 2014). پس از آن، خوانش‌ها به ژنوم مرجع هم‌تراز می‌شوند که به‌طور معمول با استفاده از نرم‌افزارهایی مانند BWA-MEM یا Bowtie2 انجام می‌شود (Li & Durbin, 2009; Li, 2013). پس از هم‌ترازی، مرحله پس‌پردازش شامل حذف دوپلیکیت‌ها و بازکالیبراسیون نمرات کیفیت با استفاده از ابزارهایی مانند Picard و GATK انجام می‌شود (McKenna et al., 2010). سپس، شناسایی واریانت‌ها با استفاده از ابزارهایی

دارای کاربردهای گسترده‌ای در زمینه‌های مختلف علوم زیستی و پزشکی است. یکی از مهم‌ترین کاربردهای آن در تشخیص بیماری‌های ژنتیکی است، جایی که می‌تواند به شناسایی جهش‌های مسئول بیماری‌های نادر یا پیچیده کمک کند (Rehm, 2013). در زمینه سرطان‌شناسی، این روش برای شناسایی جهش‌های سوماتیک در تومورها و انتخاب درمان‌های هدفمند استفاده می‌شود (Frampton *et al.*, 2013). در مطالعات جمعیتی، توالی‌یابی هدفمند امکان بررسی تنوع ژنتیکی در مقیاس بزرگ را فراهم می‌کند و می‌تواند به شناسایی واریانت‌های مرتبط با صفات پیچیده کمک کند (Koboldt *et al.*, 2013). در زمینه تکامل مولکولی، این روش برای مطالعه تغییرات ژنتیکی در طول زمان و بین گونه‌های مختلف استفاده می‌شود (Burbano *et al.*, 2010). علاوه بر این، توالی‌یابی هدفمند در مطالعات فارماکوژنومیکس^{۱۲۸} برای پیش‌بینی پاسخ به داروها و بهینه‌سازی درمان‌ها کاربرد دارد (Rodén *et al.*, 2011). یکی از مزایای اصلی توالی‌یابی هدفمند، هزینه‌ی پایین‌تر و عمق پوشش بالاتر نسبت به توالی‌یابی کل ژنوم است که امکان شناسایی واریانت‌های نادر را افزایش می‌دهد (Sims *et al.*, 2014). همچنین، این روش امکان بررسی تعداد زیادی نمونه را در یک آزمایش فراهم می‌کند که برای مطالعات بزرگ مقیاس مناسب است. با این حال، کاستی‌هایی نیز وجود دارد، از جمله عدم توانایی در شناسایی تغییرات ساختاری بزرگ یا واریانت‌های خارج از مناطق هدف (Bamshad *et al.*, 2011). در سال‌های اخیر، پیشرفت‌های قابل توجهی در فناوری‌های مرتبط با

مانند FreeBayes، GATK Haplotype Caller یا SAMtools/BCFtools انجام می‌شود (Garrison & Marth, 2012; Li, 2011; McKenna *et al.*, 2010; Li *et al.*, 2009). واریانت‌های شناسایی شده با استفاده از پایگاه‌های داده و ابزارهای بیوانفورماتیکی مانند VEP، ANNOVAR، SnpEff، HASHIE و تفسیر می‌شوند تا اثرات احتمالی آن‌ها بر عملکرد پروتئین یا فنوتیپ مشخص شود (McLaren *et al.*, 2016; Wang *et al.*, 2012; Cingolani *et al.*, 2010). پس از حاشیه نویسی واریانت‌ها، مرحله فیلتر کردن و اعتبارسنجی آن‌ها با ابزارهایی مانند GATK، VCFtools و BCFtools انجام می‌شود تا واریانت‌های با کیفیت بالا از خطاهای توالی‌یابی جدا شوند. این مرحله به‌طور معمول با استفاده از معیارهایی مانند عمق خوانش، کیفیت همترازی و فرکانس آلل انجام می‌شود (DePristo *et al.*, 2011). در نهایت، در طول فرایند تجزیه و تحلیل، کنترل کیفیت مداوم با استفاده از ابزارهایی مانند Qualimap و GATK انجام می‌شود تا از صحت و دقت نتایج اطمینان حاصل شود (Okonechnikov *et al.*, 2016). همچنین، استفاده از پایپلاین‌های خودکار مانند Snakemake یا Nextflow برای استانداردسازی و تکرارپذیری فرایند تجزیه و تحلیل توصیه می‌شود (Köster & Rahmann, 2012; Di Tommaso *et al.*, 2017). در نهایت، نتایج تجزیه و تحلیل با استفاده از ابزارهای تجسم داده مانند IGV یا UCSC Genome Browser بررسی می‌شوند تا صحت واریانت‌های شناسایی شده تأیید شود (Robinson *et al.*, 2002; Kent *et al.*, 2011).

پزشکی شخصی سازی شده و تحقیقات ژنومیک ایفا کند. در ادامه به مراحل تجزیه و تحلیل داده‌های Targeted sequencing (TGS) حاصل از توالی‌یابی نواحی خاص از ژنوم اشاره می‌شود (شکل ۱۹). اسکریپ‌های تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (Van der Auwera & O'Connor, 2020; DePristo *et al.*, 2011; Li & Durbin, 2009; McKenna *et al.*, 2010; Cock *et al.*, 2010).

توالی‌یابی هدفمند صورت گرفته است. استفاده از روش‌های جدید غنی‌سازی مانند CRISPR-Cas9 برای انتخاب مناطق هدف، افزایش دقت و کارایی این روش را به همراه داشته است (Gu *et al.*, 2016). همچنین، توسعه روش‌های تجزیه و تحلیل داده مبتنی بر هوش مصنوعی و یادگیری ماشین، امکان تفسیر دقیق‌تر و سریع‌تر نتایج را فراهم کرده است (Zou *et al.*, 2019). با توجه به پیشرفت‌های مداوم در این زمینه، انتظار می‌رود که توالی‌یابی هدفمند نقش مهمی در آینده



شکل ۱۳- مراحل تجزیه و تحلیل داده‌های Targeted sequencing; TGS
Figure 13. Targeted sequencing (TGS) data analysis steps

مراحل واسرشت‌سازی DNA، تیمار با بیسولفیت و خالص سازی DNA تیمار شده است. پس از تیمار با بیسولفیت، DNA باید به سرعت تکثیر شود زیرا DNA تک رشته‌ای حاصل از این تیمار ناپایدار است. تکثیر به‌طور معمول با استفاده از PCR و آغازگرهای اختصاصی برای توالی‌های تیمار شده با بیسولفیت انجام می‌شود (Laird, 2010). در مرحله بعد، کتابخانه‌های توالی‌یابی آماده می‌شوند. این فرایند شامل شکستن DNA به قطعات کوچکتر، اتصال آداپتورها و تکثیر است. روش‌های مختلفی برای آماده‌سازی کتابخانه وجود دارد، از جمله Whole Genome Bisulfite Sequencing; WGBS که کل ژنوم را پوشش می‌دهد، و Representation Bisulfite Sequencing; RRBS که بر نواحی غنی از CpG تمرکز می‌کند (Meissner *et al.*, 2005). پس از آماده‌سازی کتابخانه، نمونه‌ها برای توالی‌یابی آماده می‌شوند. توالی‌یابی به‌طور معمول با استفاده از پلتفرم‌های توالی‌یابی نسل بعدی (NGS) مانند Illumina انجام می‌شود که امکان توالی‌یابی موازی میلیون‌ها قطعه DNA را فراهم می‌کند (Metzker, 2010). در طول این فرایند، کنترل کیفیت در هر مرحله بسیار مهم است. این شامل استفاده از کنترل‌های مثبت و منفی، ارزیابی کارایی تیمار با بیسولفیت و بررسی کیفیت کتابخانه‌های تهیه شده است. علاوه بر این، برای اطمینان از صحت نتایج، تکرارهای زیستی و روشی به‌طور معمول در طراحی آزمایش گنجانده می‌شوند (Bock, 2012). پس از اتمام توالی‌یابی، داده‌های خام وارد مرحله

مطالعه تغییرات اپی‌ژنتیک -متیلاسیون (Methyl-seq)

روش مطالعه تغییرات اپیژنتیکی از طریق مطالعه متیلاسیون DNA¹²⁹ بر اساس تعیین توالی DNA متیله شده عملکرده و اطلاعات ارزشمندی در مورد الگوهای متیلاسیون DNA در سراسر ژنوم فراهم می‌کند (Jones, 2012). متیلاسیون DNA یک تغییر اپیژنتیکی مهم است که نقش کلیدی در تنظیم بیان ژن، خاموش کردن ژن و تمایز سلولی ایفا می‌کند. روش Epigenomics by DNA Methylation Sequencing شامل چندین مرحله آزمایشگاهی دقیق و حساس است که برای بررسی الگوهای متیلاسیون DNA در سطح ژنوم طراحی شده اند. این فرایند با استخراج DNA ژنومی از نمونه مورد نظر آغاز می‌شود که می‌تواند شامل سلول‌های کشت شده، بافت یا حتی مایعات بدن باشد (Kurdyukov & Bullock, 2016). پس از استخراج، کیفیت و کمیت DNA با استفاده از روش‌هایی مانند اسپکتروفتومتری و الکتروفورز ژل آگارز ارزیابی می‌شود تا از مناسب بودن آن برای مراحل بعدی اطمینان حاصل شود. مرحله کلیدی بعدی، تیمار DNA با بیسولفیت سدیم است که باعث تبدیل سیتوزین‌های غیرمتیله به یوراسیل می‌شود، در حالی که سیتوزین‌های متیله بدون تغییر باقی می‌مانند (Li & Tollefsbol, 2011). این مرحله به‌طور معمول با استفاده از کیت‌های تجاری انجام می‌شود که شامل

¹²⁹ Epigenomics by DNA methylation sequencing

اختصاصی مانند Bismark یا BSMAP استفاده می‌شود که برای هم‌ترازی خوانش‌های تیمار شده با بیسولفیت طراحی شده‌اند (Li, 2009). پس از هم‌ترازی، وضعیت متیلاسیون هر سایت CpG با مقایسه توالی خوانش‌ها با توالی ژنوم مرجع تعیین می‌شود. این کار به‌طور معمول با استفاده از نرم‌افزارهایی مانند MethylDackel یا بخشی از پایپلاین Bismark انجام می‌شود. در مرحله بعد، تجزیه و تحلیل آماری برای شناسایی مناطق با تفاوت معنی‌دار در سطح متیلاسیون بین نمونه‌های مختلف انجام می‌شود. این کار به‌طور معمول با استفاده از بسته‌های نرم‌افزاری R مانند methylKit یا DSS انجام می‌شود (Akalin *et al.*, 2012; Feng *et al.*, 2014). این بسته‌ها امکان مقایسه سطوح متیلاسیون بین گروه‌های مختلف، شناسایی مناطق با تفاوت معنی‌دار در متیلاسیون (DMRs) و تجسم^{۱۳۱} نتایج را فراهم می‌کنند. علاوه بر این، تجزیه و تحلیل‌های پیشرفته‌تر مانند تجزیه و تحلیل غنی‌سازی مسیرهای زیستی^{۱۳۲} و شبکه‌های تنظیمی نیز انجام می‌شود تا اهمیت زیستی تغییرات متیلاسیون مشاهده شده بهتر درک شود. این کار به‌طور معمول با استفاده از ابزارهایی مانند GREAT یا GOseq انجام می‌شود (McLean *et al.*, 2010; Young *et al.*, 2010). برای بررسی ارتباط بین تغییرات متیلاسیون و بیان ژن، داده‌های متیلوم اغلب با داده‌های ترنسکریپتوم ادغام می‌شوند. این کار می‌تواند با استفاده از ابزارهایی مانند

تجزیه و تحلیل بیوانفورماتیکی می‌شوند که شامل کنترل کیفیت خوانش‌ها، هم‌ترازی به ژنوم مرجع و تعیین وضعیت متیلاسیون هر سایت CpG¹³⁰ است. با پیشرفت فناوری، روش‌های جدیدی نیز برای بهبود این روش در حال توسعه هستند. به عنوان مثال، روش‌های توالی‌یابی تک مولکولی مانند Nanopore sequencing امکان تعیین مستقیم متیلاسیون DNA بدون نیاز به تیمار با بیسولفیت را فراهم می‌کنند که می‌تواند برخی از کاستی‌های روش‌های سنتی را برطرف کند (Simpson *et al.*, 2017). پس از توالی‌یابی، داده‌های خام حاصل وارد مرحله تجزیه و تحلیل می‌شوند. تجزیه و تحلیل داده‌های Epigenomics by DNA Methylation Sequencing شامل چندین گام کلیدی است که نیازمند ترکیبی از دانش بیوانفورماتیک، آمار و زیست‌شناسی مولکولی است. این فرایند با کنترل کیفیت اولیه داده‌های خام توالی‌یابی آغاز می‌شود. در این مرحله، از نرم‌افزارهایی مانند FastQC استفاده می‌شود تا کیفیت خوانش‌ها، توزیع طول خوانش‌ها و محتوای GC بررسی شود (Andrews, 2010). خوانش‌های با کیفیت پایین و آداپتورها حذف می‌شوند که به‌طور معمول با استفاده از ابزارهایی مانند Trimmomatic انجام می‌شود (Bolger *et al.*, 2014). پس از پاکسازی داده‌ها، خوانش‌ها به ژنوم مرجع هم‌تراز می‌شوند. این مرحله با چالش‌هایی همراه است زیرا تیمار با بیسولفیت باعث کاهش پیچیدگی توالی می‌شود. برای غلبه بر این مشکل، از نرم‌افزارهای

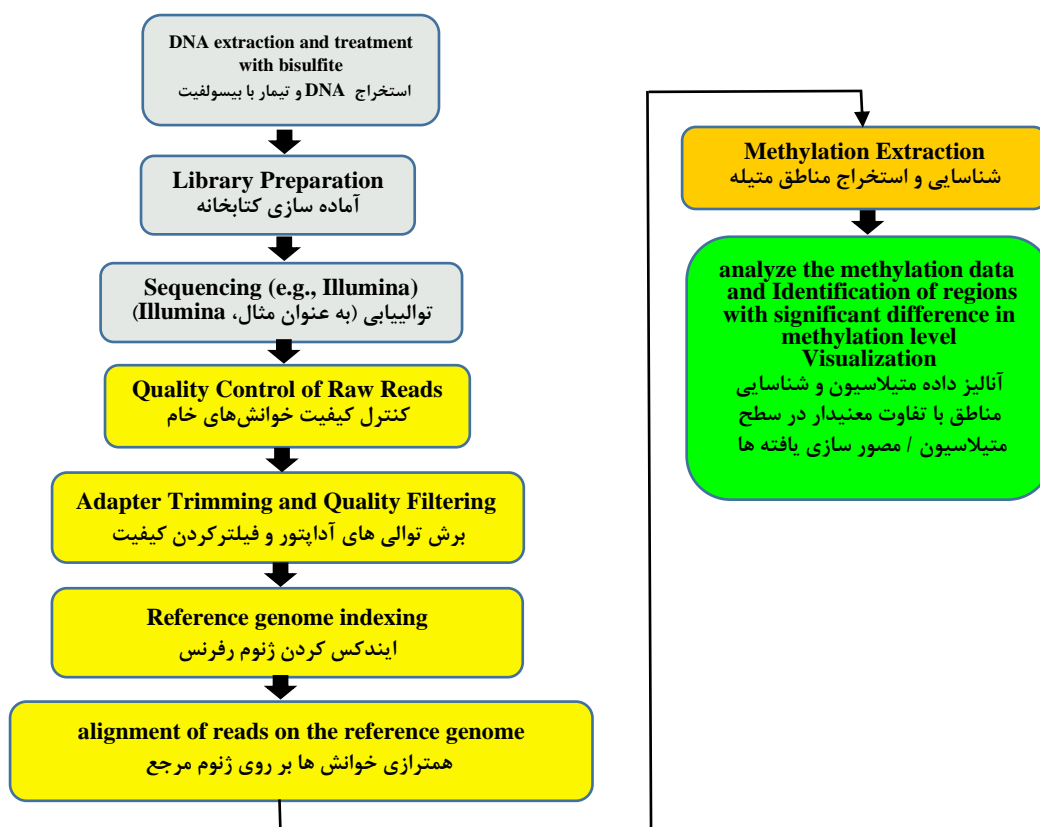
¹³¹ Visualization¹³² Pathway enrichment analysis¹³⁰ Cytosine-phosphate-guanine (CpG)

روش برای مطالعه تغییرات اپیژنتیکی در بیماری‌های مختلف از جمله اختلالات عصبی، بیماری‌های خودایمنی و بیماری‌های متابولیک استفاده می‌شود (Portela & Esteller, 2010). در زمینه زیست‌شناسی تکاملی، این روش برای بررسی نقش تغییرات اپیژنتیکی در سازگاری و تکامل گونه‌ها مورد استفاده قرار می‌گیرد (Mendizabal et al., 2014). همچنین در مطالعات مربوط به پیری، این روش برای بررسی تغییرات الگوهای متیلاسیون DNA در طول عمر و ارتباط آن با فرایند پیری استفاده می‌شود (Horvath & Raj, 2018). در زمینه تحقیقات سلول‌های بنیادی و برنامه‌ریزی مجدد سلولی، این روش برای درک بهتر تغییرات اپیژنتیکی در طی تمایز و تبدیل سلولی به کار می‌رود (Meissner, 2010). علاوه بر این، از این روش در مطالعات اپیژنتیک محیطی برای بررسی تأثیرات عوامل محیطی مانند رژیم غذایی، تنش و آلودگی بر الگوهای متیلاسیون DNA استفاده می‌شود (Baccarelli & Bollati, 2009). با پیشرفت فناوری‌های توالی‌یابی و روش‌های تجزیه و تحلیل داده، انتظار می‌رود که کاربردهای این روش در آینده گسترش یابد. به عنوان مثال، توسعه روش‌های توالی‌یابی تک سلولی امکان بررسی تغییرات متیلاسیون DNA در سطح تک سلول را فراهم کرده است که می‌تواند به درک بهتر ناهمگنی سلولی در بافت‌ها و تومورها کمک کند (Smallwood et al., 2014). همچنین، ادغام داده‌های حاصل از این روش با سایر داده‌های Omics مانند ترنسکریپتومیکس و پروتئومیکس می‌تواند

MethylMix انجام شود (Gevaert, 2015). در نهایت، نتایج به صورت گرافیکی با استفاده از ابزارهایی مانند Circos یا R ggplot2 نمایش داده می‌شوند (Wickham, 2016; Krzywinski et al., 2009). یکی از چالش‌های اصلی در تجزیه و تحلیل داده‌های متیلوم، حجم بالای داده‌ها و نیاز به قدرت محاسباتی بالاست. برای غلبه بر این مشکل، اغلب از محاسبات توزیع شده و پردازش موازی استفاده می‌شود. علاوه بر این، تفسیر زیستی نتایج نیز چالش برانگیز است و نیاز به دانش عمیق زیست‌شناسی و اپیژنتیک دارد. با پیشرفت فناوری، روش‌های جدیدی نیز برای بهبود تجزیه و تحلیل داده‌های متیلوم در حال توسعه هستند. به عنوان مثال، استفاده از یادگیری ماشین و هوش مصنوعی برای پیش‌بینی اثرات عملکردی تغییرات متیلاسیون در حال افزایش است (Angermueller et al., 2017). همچنین، روش‌های جدید برای تجزیه و تحلیل داده‌های متیلوم تک سلولی در حال توسعه هستند که می‌توانند بینش‌های جدیدی در مورد ناهمگنی سلولی در الگوهای متیلاسیون ارائه دهند (Farlik et al., 2015). این روش کاربردهای گسترده‌ای در زمینه‌های مختلف زیست پزشکی دارد. یکی از مهمترین کاربردهای آن در مطالعات سرطان است، جایی که تغییرات گسترده در الگوهای متیلاسیون DNA اغلب مشاهده می‌شود. این روش امکان شناسایی بیومارکرهای اپیژنتیکی برای تشخیص زودهنگام، پیش‌آگهی و پاسخ به درمان در انواع مختلف سرطان را فراهم می‌کند (Baylin & Jones, 2016). علاوه بر این، از این

قدرتمندترین ابزارها برای مطالعه تغییرات اپی‌ژنتیکی در سطح ژنوم است و نقش مهمی در پیشرفت درک ما از تنظیم ژن و بیماری‌های مختلف ایفا می‌کند. در ادامه به مراحل تجزیه و تحلیل داده‌های Methylation sequencing (Methyl-seq) حاصل از توالی‌یابی ژنوم در رابطه با متیله بودن نواحی مختلف DNA اشاره می‌شود (شکل ۲۰). اسکرپ‌های تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (Krueger & Andrews, 2011; Akalin *et al.*, 2012; Andrews, 2010; Martin, 2011; Feng *et al.*, 2014; Assenov *et al.*, 2014).

دید جامع‌تری از تنظیم ژن و عملکرد سلولی ارائه دهد (Bock, 2012). با این حال، این روش با چالش‌هایی نیز روبرو است. یکی از مهمترین چالش‌ها، حجم بالای داده‌های تولید شده و نیاز به قدرت محاسباتی بالا برای پردازش و تجزیه و تحلیل این داده‌ها است. علاوه بر این، تفسیر زیستی تغییرات مشاهده شده در الگوهای متیلاسیون DNA همچنان چالش برانگیز است و نیاز به توسعه روش‌های تجزیه و تحلیل پیشرفته‌تر دارد (Bock, 2012). با وجود این چالش‌ها، روش Epigenomics by DNA methylation sequencing همچنان یکی از



شکل ۲۰- مراحل تجزیه و تحلیل داده‌های Methylation sequencing; Methyl-seq
Figure 20. Methylation Sequencing (Methyl-seq) data analysis steps

توالی‌یابی کروماتین با ایمونوپرسیپیتاسیون (ChIP-seq)

روش ChIP-Seq یا ایمونوپرسیپیتاسیون کروماتین همراه با توالی‌یابی^{۱۳۳}، در مطالعه تعاملات پروتئین-DNA در مقیاس ژنومی کاربرد دارد. این روش ترکیبی از ایمونوپرسیپیتاسیون کروماتین (ChIP) و توالی‌یابی نسل بعدی (NGS) است که امکان شناسایی مکان‌های اتصال فاکتورهای رونویسی، هیستون‌ها و سایر پروتئین‌های متصل به DNA را در سراسر ژنوم فراهم می‌کند (Furey, 2012). روش ChIP-Seq برای مطالعه تعاملات پروتئین-DNA شامل مراحل آزمایشگاهی دقیق و پیچیده‌ای است. فرایند با تثبیت سلول‌ها توسط فرمالدئید آغاز می‌شود که باعث ایجاد پیوندهای عرضی بین پروتئین‌ها و DNA می‌گردد (Landt *et al.*, 2012). این مرحله برای حفظ تعاملات طبیعی پروتئین-DNA ضروری است. پس از تثبیت، سلول‌ها لیز شده و کروماتین استخراج می‌شود. سپس، کروماتین با استفاده از روش‌های مکانیکی (مانند سونیکاسیون) یا آنزیمی (مانند نوکلئاز میکروکوکوی) به قطعات کوچک‌تر (به‌طور معمول ۲۰۰-۶۰۰ جفت باز) شکسته می‌شود (Furey, 2012). این مرحله برای دستیابی به وضوح مکانی مناسب در نتایج نهایی حیاتی است. در ادامه، ایمونوپرسیپیتاسیون با استفاده از آنتی‌بادی‌های اختصاصی علیه پروتئین مورد نظر انجام می‌شود. کیفیت و اختصاصیت آنتی‌بادی در این

مرحله بسیار مهم است و می‌تواند بر نتایج نهایی تأثیر قابل توجهی داشته باشد (Kidder *et al.*, 2011). پس از ایمونوپرسیپیتاسیون^{۱۳۴}، کمپلکس‌های پروتئین-DNA جدا شده و شستشو می‌شوند تا اتصال‌های غیراختصاصی حذف شوند. سپس، پیوندهای عرضی با استفاده از حرارت یا پروتئیناز K برداشته می‌شوند و DNA خالص‌سازی می‌گردد (Park, 2009). در مرحله بعد، کتابخانه‌های توالی‌یابی ایجاد می‌شوند که شامل ترمیم انتهای DNA، افزودن نوکلئوتیدهای A به انتهای^۳، اتصال آداپتورها و تکثیر PCR است (Meyer & Liu, 2014). کیفیت کتابخانه با روش‌هایی مانند Bioanalyzer ارزیابی می‌شود. در نهایت، کتابخانه‌ها با استفاده از پلتفرم‌های توالی‌یابی نسل بعدی مانند Illumina توالی‌یابی می‌شوند. برای اطمینان از صحت نتایج، به‌طور معمول از نمونه‌های کنترل منفی مانند IgG غیراختصاصی و کنترل ورودی DNA شکسته شده بدون ایمونوپرسیپیتاسیون استفاده می‌شود (Bailey *et al.*, 2013). پیشرفت‌های اخیر در روش‌های مرتبط مانند CUT&RUN و CUT&Tag، حساسیت و ویژگی بالاتری نسبت به ChIP-Seq سنتی ارائه می‌دهند و می‌توانند با تعداد سلول کمتری انجام شوند (Kaya-Okur *et al.*, 2019). این روش‌ها از آنزیم‌های برش DNA متصل به آنتی‌بادی استفاده می‌کنند که امکان برش دقیق‌تر و اختصاصی‌تر DNA را فراهم می‌سازند. علاوه بر این، روش‌های جدید مانند ChIP-exo و ChIP-nexus با افزودن مراحل اضافی

¹³³ Chromatin immunoprecipitation sequencing (ChIP-seq)

¹³⁴ Immunoprecipitation

مناطق با غنی‌سازی قابل توجه خوانش‌ها (قله‌ها) که نشان‌دهنده مکان‌های احتمالی اتصال پروتئین هستند، استفاده می‌شوند. MACS2 یکی از پرکاربردترین ابزارهای قله‌یابی است (Zhang *et al.*, 2008). این الگوریتم‌ها از روش‌های آماری پیشرفته برای تمایز بین سیگنال‌های واقعی و نویز پس‌زمینه استفاده می‌کنند. پس از شناسایی قله‌ها، تحلیل‌های بیشتری انجام می‌شود. تجزیه و تحلیل موتیف با استفاده از ابزارهایی مانند MEME-ChIP و حاشیه‌نویسی برای شناسایی توالی‌های اتصال اختصاصی در مناطق قله از طریق ابزارهایی مانند ChIPseeker

TxDb.Hsapiens.UCSC.hg19.knownGene

clusterProfiler انجام می‌شود (Machanic & Bailey, 2011). همچنین، مطالعات همبستگی با داده‌های بیان ژن مثلاً با استفاده از RNA-Seq می‌تواند ارتباط بین اتصال پروتئین و تنظیم بیان ژن را نشان دهد. تجزیه و تحلیل غنی‌سازی هستی‌شناسی ژن (GO)¹³⁵ با استفاده از ابزارهایی مانند GREAT برای درک عملکرد زیستی مناطق اتصال انجام می‌شود (McLean *et al.*, 2010). تجسم داده‌ها با استفاده از نرم‌افزارهایی مانند IGV، UCSC Genome Browser و پکیج‌های نرم‌افزاری در R برای بررسی چشمی قله^{۱۳۶}ها و مقایسه با سایر داده‌های ژنومی انجام می‌شود (Robinson *et al.*, 2002; Kent *et al.*, 2011). در سال‌های اخیر، روش‌های یادگیری ماشین و هوش مصنوعی نیز برای

برش آنزیمی، وضوح مکانی بالاتری را امکان‌پذیر می‌سازند (Skene & Henikoff, 2015). بهینه‌سازی پروتکل‌های ChIP-Seq برای انواع مختلف سلول‌ها و بافت‌ها، از جمله نمونه‌های بالینی منجمد و فیکس شده، همچنان یک زمینه فعال تحقیقاتی است (Savic *et al.*, 2015). در مجموع، مراحل آزمایشگاهی ChIP-Seq نیازمند دقت بالا، بهینه‌سازی برای هر نوع نمونه و پروتئین هدف، و کنترل کیفیت دقیق در هر مرحله است تا داده‌های قابل اعتماد و با کیفیت بالا تولید شود. تجزیه و تحلیل داده‌های ChIP-Seq شامل مراحل متعدد و پیچیده‌ای است که برای استخراج اطلاعات معنی‌دار از داده‌های خام توالی‌یابی ضروری هستند. فرایند تجزیه و تحلیل با کنترل کیفیت اولیه داده‌های خام آغاز می‌شود که به‌طور معمول با استفاده از ابزارهایی مانند FastQC انجام می‌شود (Andrews, 2010). این مرحله شامل بررسی کیفیت خوانش‌ها، شناسایی توالی‌های آداپتور و حذف خوانش‌های با کیفیت پایین است. سپس، خوانش‌های با کیفیت بالا به ژنوم مرجع همتراز می‌شوند. این کار به‌طور معمول با استفاده از نرم‌افزارهایی مانند Bowtie2 یا BWA انجام می‌شود (Langmead & Salzberg, 2012; Li & Durbin, 2009). پس از هم‌ترازی، فایل‌های SAM یا BAM تولید می‌شوند که موقعیت هر خوانش روی ژنوم را نشان می‌دهند. در مرحله بعد، خوانش‌های تکراری که ممکن است ناشی از تکثیر PCR باشند، با استفاده از ابزارهایی مانند Picard حذف می‌شوند (Toolkit, 2019). سپس، الگوریتم‌های قله‌یابی برای شناسایی

¹³⁵ Gene ontology enrichment

¹³⁶ Peak

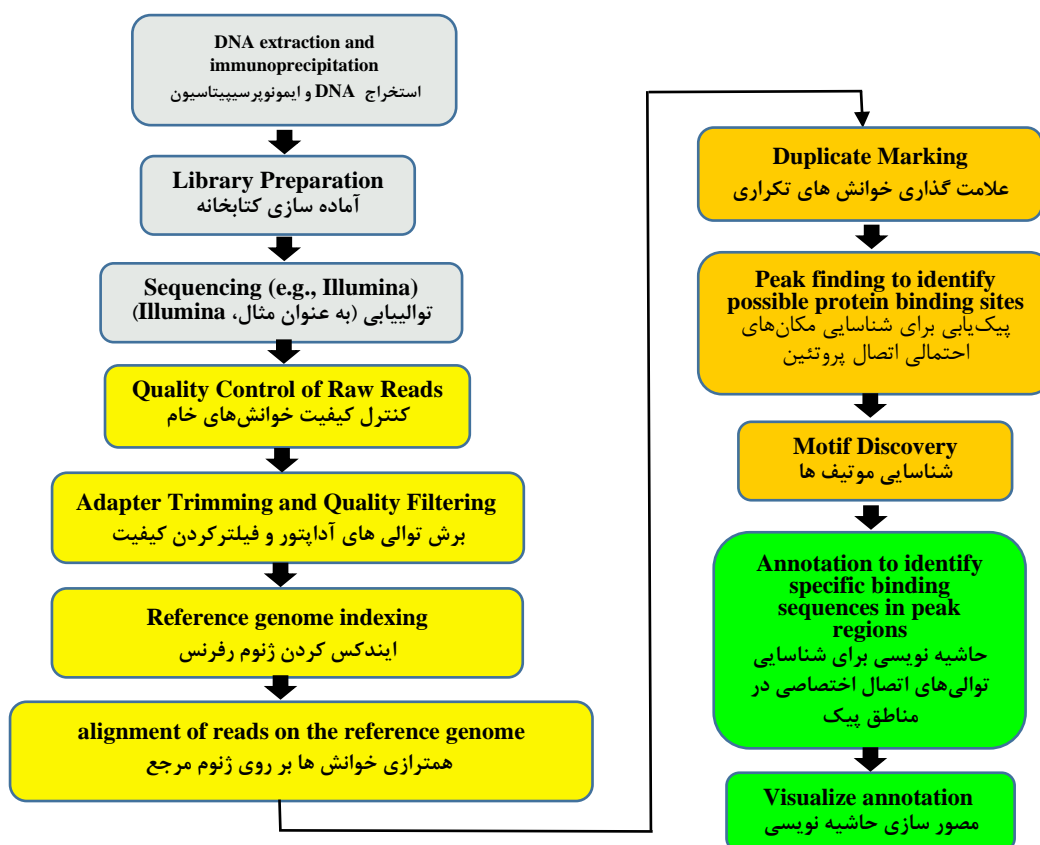
مختلف و درک تکامل شبکه‌های تنظیمی ژن به کار می‌رود (Schmidt *et al.*, 2010). در مطالعات سلول‌های بنیادی و تمایز سلولی، ChIP-Seq برای بررسی تغییرات در الگوهای اتصال فاکتورهای رونویسی و تغییرات اپی‌ژنتیکی طی فرایند تمایز استفاده می‌شود (Gifford *et al.*, 2013). علاوه بر این، ChIP-Seq در توسعه داروها و درمان‌های هدفمند نیز کاربرد دارد، زیرا می‌تواند اهداف مولکولی داروها را در سطح ژنوم شناسایی کند و سازوکار عمل آنها را روشن سازد (Stellrecht & Chen, 2014). با وجود مزایای زیاد، ChIP-Seq با چالش‌هایی نیز روبرو است. یکی از مهم‌ترین کاستی‌ها، نیاز به آنتی‌بادی‌های با کیفیت بالا و اختصاصی است که همیشه در دسترس نیستند. علاوه بر این، تفسیر نتایج ChIP-Seq می‌تواند پیچیده باشد، زیرا اتصال یک پروتئین به DNA ناگزیر به معنای عملکرد زیستی نیست و نیاز به تأیید عملکردی دارد (Kidder *et al.*, 2011). پیشرفت‌های اخیر در روش‌های مرتبط مانند CUT&RUN و CUT&Tag، که حساسیت و ویژگی بالاتری نسبت به ChIP-Seq سنتی دارند، در حال گسترش امکانات مطالعه تعاملات پروتئین-DNA هستند (Kaya-Okur *et al.*, 2019). ادغام داده‌های ChIP-Seq با سایر داده‌های امیکس مانند RNA-Seq، ATAC-Seq و Hi-C، به پژوهشگران این امکان را می‌دهد که درک عمیق‌تری از تنظیم ژن و ساختار کروماتین به دست آورند. این ترکیب داده‌ها نه تنها به شناسایی الگوهای بیان ژن کمک می‌کند، بلکه تعاملات بین کروماتین و

بهبود دقت و قدرت تفسیر داده‌های ChIP-Seq مورد استفاده قرار گرفته‌اند (Chen *et al.*, 2021). این روش‌ها می‌توانند الگوهای پیچیده در داده‌ها را شناسایی کنند و پیش‌بینی‌های دقیق‌تری از مکان‌های اتصال و عملکرد آنها ارائه دهند. علاوه بر این، ادغام داده‌های ChIP-Seq با سایر داده‌های امیکس مانند ATAC-Seq و Hi-C، دیدگاه‌های جامع‌تری از تنظیم ژن و ساختار کروماتین ارائه می‌دهد (Angelini & Costa, 2014). در نهایت، تفسیر زیست‌شناختی نتایج نیازمند دانش تخصصی در زمینه زیست‌شناسی مولکولی و تنظیم ژن است. این مرحله شامل بررسی نقش ژن‌های نزدیک به قله‌ها، ارتباط با مسیرهای سیگنالینگ، و مقایسه با مطالعات قبلی است. مهم است که توجه داشته باشیم که تجزیه و تحلیل داده‌های ChIP-Seq یک فرایند تکراری است و ممکن است نیاز به تنظیم پارامترها و تکرار برخی مراحل برای دستیابی به نتایج بهینه باشد.

کاربردهای ChIP-Seq بسیار گسترده و متنوع است. این روش در مطالعه سازوکارهای تنظیم رونویسی، شناسایی عناصر تنظیمی ژنوم مانند افزایشنده‌ها و خاموش‌کننده‌ها، و درک تغییرات اپی‌ژنتیکی مانند تغییرات هیستونی نقش کلیدی دارد (Nakato & Shirahige, 2017). در تحقیقات سرطان، ChIP-Seq برای شناسایی تغییرات در الگوهای اتصال فاکتورهای رونویسی و تغییرات اپی‌ژنتیکی مرتبط با بیماری استفاده می‌شود (Zhang *et al.*, 2019). در زمینه زیست‌شناسی تکاملی، این روش برای مقایسه مکان‌های اتصال فاکتورهای رونویسی بین گونه‌های

ChIP-seq هستند (Chen *et al.*, 2021). این پیشرفت‌ها نوید بخش کشف‌های جدید در زمینه تنظیم ژن و عملکرد ژنوم در سال‌های آینده است. در ادامه به مراحل تجزیه و تحلیل داده‌های ChIP-seq حاصل از توالی‌یابی ژنوم در رابطه با تعاملات پروتئین-DNA اشاره می‌شود (شکل ۲۱). اسکرپ‌های تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (Andrews, 2010; Bolger *et al.*, 2014; Langmead & Salzberg, 2012; Li *et al.*, 2009; Broad Institute, N.D.; Zhang *et al.*, 2008; Machanick & Bailey, 2011; Yu *et al.*, 2015).

پروتئین‌ها را نیز روشن می‌سازد و به تحلیل دینامیک‌های ساختاری در سطح ژنوم می‌پردازد. به‌علاوه، این رویکرد چندبعدی می‌تواند به شناسایی نقاط کلیدی تنظیمی و مسیرهای سیگنال‌دهی مرتبط با بیماری‌ها و فرآیندهای بیولوژیکی کمک کند و به توسعه درمان‌های هدفمند و استراتژی‌های نوین در تحقیقات ژنتیکی و بیولوژیکی منجر شود (Angelini & Costa, 2014). در نهایت، پیشرفت‌های مداوم در فناوری‌های توالی‌یابی و روش‌های تحلیل داده، از جمله استفاده از یادگیری ماشین و هوش مصنوعی، در حال افزایش دقت و قدرت تفسیر داده‌های



شکل ۲۱- مراحل تجزیه و تحلیل داده‌های ChIP-seq
Figure 21. Chromatin immunoprecipitation sequencing (ChIP-seq) data analysis steps

توالی‌یابی کل متازنوم (WMS)

توالی‌یابی کل متازنوم یا WMS¹³⁷ یک روش قدرتمند در زمینه ژنومیکس محیطی است که امکان مطالعه جامع جوامع میکروبی را در محیط‌های مختلف فراهم می‌کند (Quince *et al.*, 2017). این روش با استخراج و توالی‌یابی تمام DNA موجود در یک نمونه محیطی، بدون نیاز به کشت میکروارگانیسم‌ها، اطلاعات ارزشمندی درباره ترکیب، تنوع و عملکرد جوامع میکروبی ارائه می‌دهد (Rausch *et al.*, 2019). مراحل آزمایشگاهی و Wet-lab در روش توالی‌یابی کل متازنوم (WMS) شامل چندین گام اساسی است که با دقت و ظرافت خاصی انجام می‌شوند. این فرایند با جمع‌آوری نمونه آغاز می‌شود که می‌تواند از منابع مختلفی مانند خاک، آب، رسوبات، یا نمونه‌های بالینی باشد (Quince *et al.*, 2017). نمونه‌ها باید به سرعت پردازش شوند یا در شرایط مناسب (به‌طور معمول در دمای -۸۰ درجه سانتیگراد) نگهداری شوند تا از تخریب DNA جلوگیری شود. مرحله بعدی، استخراج DNA است که یکی از مهم‌ترین و چالش‌برانگیزترین مراحل است. روش‌های استخراج باید به‌گونه‌ای انتخاب شوند که DNA با کیفیت بالا از تمام میکروارگانیسم‌های موجود در نمونه از جمله باکتری‌ها، آرکی‌ها، قارچ‌ها و ویروس‌ها، استخراج شود (Knudsen *et al.*, 2016). به‌طور معمول از کیت‌های تجاری یا روش‌های دستی مانند استخراج فنل-کلروفرم استفاده می‌شود. پس از استخراج، کیفیت و کمیت DNA با روش‌هایی مانند

الکتروفورز ژل آگارز، اسپکتروفتومتری مانند NanoDrop و فلورومتری مانند Qubit ارزیابی می‌شود (Rausch *et al.*, 2019). در صورت نیاز، DNA استخراج شده ممکن است تغلیظ یا تصفیه شود. مرحله بعدی، تهیه کتابخانه توالی‌یابی است که شامل قطعه‌کردن¹³⁸ DNA، اتصال آداپتورها و تکثیر است. قطعه‌کردن به‌طور معمول با روش‌های مکانیکی (مانند سونیکاسیون) یا آنزیمی انجام می‌شود تا قطعات DNA با اندازه مناسب برای پلتفرم توالی‌یابی مورد نظر تولید شود (Van Dijk *et al.*, 2014). سپس، آداپتورهای اختصاصی به انتهای قطعات DNA متصل می‌شوند که شامل توالی‌های شناسایی برای اتصال به فلوسل و آغازگرهای توالی‌یابی هستند. در برخی پروتکل‌ها، مرحله غنی‌سازی PCR نیز انجام می‌شود تا مقدار کافی DNA برای توالی‌یابی فراهم شود. کیفیت کتابخانه‌های تهیه شده با روش‌هایی مانند Bioanalyzer یا TapeStation ارزیابی می‌شود (Head *et al.*, 2014). در نهایت، کتابخانه‌ها برای توالی‌یابی آماده می‌شوند که به‌طور معمول با استفاده از پلتفرم‌های توالی‌یابی نسل بعدی (NGS) مانند Illumina، Ion Torrent یا Oxford Nanopore انجام می‌شود (Goodwin *et al.*, 2016). انتخاب پلتفرم توالی‌یابی بستگی به اهداف پروژه، بودجه و نیازهای خاص تجزیه و تحلیل دارد. به عنوان مثال، پلتفرم Illumina خوانش‌های کوتاه با دقت بالا تولید می‌کند، در حالی که Oxford Nanopore امکان تولید خوانش‌های بلند را فراهم

¹³⁸ Fragmentation¹³⁷ Whole metagenome sequencing

می‌شود (Langmead & Salzberg, 2012; Li & Durbin, 2009). در مواردی که ژنوم مرجع در دسترس نیست یا هدف بازسازی ژنوم‌های جدید است، از روش‌های مونتاژ *de novo* استفاده می‌شود. ابزارهایی مانند MetaSPAdes یا MEGAHIT برای این منظور کاربرد دارند (Nurk et al., 2017; Li et al., 2015). پس از هم‌ترازی یا مونتاژ، تجزیه و تحلیل‌های تاکسونومیک و عملکردی انجام می‌شود. برای تجزیه و تحلیل تاکسونومیک، از ابزارهایی مانند MetaPhlan3 یا Kraken2 استفاده می‌شود که امکان شناسایی و تعیین فراوانی نسبی گونه‌های میکروبی را فراهم می‌کنند (Beghini et al., 2021; Wood et al., 2019). تجزیه و تحلیل عملکردی با استفاده از ابزارهایی مانند HUMAnN3، KEGG Orthology-Based eggNOG-Annotation System؛ KOBAS mapper انجام می‌شود که پتانسیل متابولیکی و عملکردی جامعه میکروبی را بررسی می‌کنند (Beghini et al., 2021; Huerta-Cepas et al., 2017; Xie et al., 2011; Abubucker et al., 2012). این تجزیه و تحلیل‌ها اطلاعاتی درباره مسیرهای متابولیکی، ژن‌های عملکردی و پتانسیل‌های بیوشیمیایی جامعه میکروبی ارائه می‌دهند. در ادامه، مرحله تجزیه و تحلیل‌های آماری و مقایسه‌ای برای شناسایی تفاوت‌های معنی‌دار بین نمونه‌ها یا شرایط مختلف با استفاده از نرم‌افزارهای آماری مانند R و پکیج‌هایی مانند phyloseq یا vegan برای این منظور استفاده می‌شود.

می‌کند که برای مونتاژ ژنوم مفید است. در طول تمام این مراحل، رعایت اصول کنترل کیفیت و جلوگیری از آلودگی بسیار مهم است. استفاده از کنترل‌های منفی و مثبت، کار در محیط استریل و استفاده از تجهیزات و مواد عاری از نوکلئاز ضروری است (Eisenhofer et al., 2019). همچنین، ثبت دقیق تمام مراحل و پارامترها برای تکرارپذیری و مقایسه نتایج در مطالعات مختلف اهمیت دارد. با اتمام مرحله توالی‌یابی، داده‌های خام تولید شده برای تجزیه و تحلیل‌های بیوانفورماتیکی آماده می‌شوند که شامل کنترل کیفیت، پردازش و تحلیل داده‌ها است. تجزیه و تحلیل داده‌های حاصل از توالی‌یابی کل متاژنوم (WMS) شامل مراحل متعدد و پیچیده‌ای است که با استفاده از ابزارهای بیوانفورماتیکی متنوع انجام می‌شود. این فرایند با کنترل کیفیت داده‌های خام آغاز می‌شود که به‌طور معمول با استفاده از نرم‌افزارهایی مانند FastQC یا MultiQC انجام می‌شود (Andrews, 2010; Ewels et al., 2016). در این مرحله، کیفیت خوانش‌ها، توزیع طول، محتوای GC و حضور توالی‌های آداپتور بررسی می‌شود. سپس، پیش‌پردازش داده‌ها با استفاده از ابزارهایی مانند Trimmomatic یا Cutadapt انجام می‌شود تا خوانش‌های با کیفیت پایین و توالی‌های آداپتور حذف شوند (Bolger et al., 2014; Martin, 2011). پس از پیش‌پردازش، خوانش‌ها به‌طور معمول به یک ژنوم مرجع یا دیتابیس ژنومی هم‌تراز می‌شوند. برای این منظور، از ابزارهایی مانند Bowtie2 یا BWA استفاده

کاربرد دارد. به عنوان مثال، این روش برای بررسی تغییرات میکروبیوم روده در بیماری‌های التهابی روده یا سرطان روده بزرگ استفاده شده است (Lloyd-Price *et al.*, 2019). در زمینه کشاورزی، WMS برای مطالعه میکروبیوم خاک و ریزوسفر و تأثیر آن بر رشد گیاه و مقاومت به بیماری‌ها کاربرد دارد (Zarraonaindia *et al.*, 2015). در صنعت، این روش برای بهینه‌سازی فرایندهای تخمیر و تصفیه فاضلاب استفاده می‌شود (Vanwonderghem *et al.*, 2014). در زمینه زیست‌پالایی، WMS برای شناسایی میکروارگانیسم‌ها و مسیرهای متابولیکی دخیل در تجزیه آلاینده‌ها کاربرد دارد (Kimes *et al.*, 2014). علاوه بر این، WMS در اکتشاف زیستی برای یافتن آنزیم‌ها و ترکیبات جدید با پتانسیل دارویی یا صنعتی استفاده می‌شود (Nesme *et al.*, 2014).

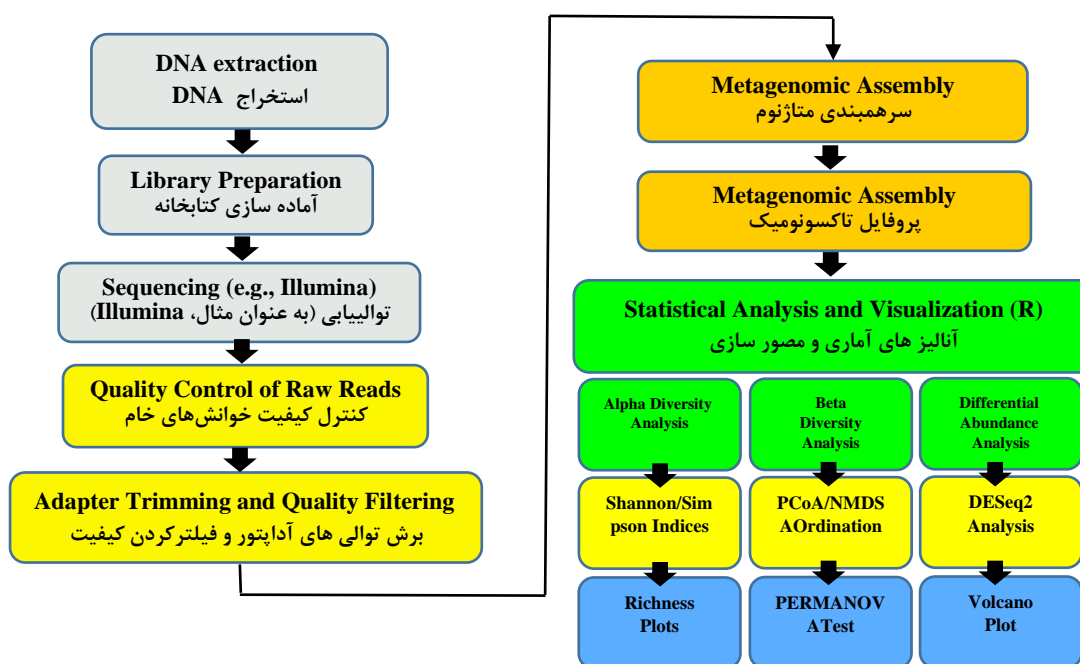
با وجود مزایای زیاد، WMS با چالش‌هایی نیز روبرو است. حجم بالای داده‌های تولید شده نیازمند زیرساخت‌های محاسباتی قوی و الگوریتم‌های کارآمد برای پردازش و تجزیه و تحلیل است. علاوه بر این، تفسیر نتایج WMS نیازمند دانش تخصصی در زمینه‌های مختلف از جمله میکروبیولوژی، ژنومیکس و بیوانفورماتیک است. با این حال، پیشرفت‌های مداوم در فناوری‌های توالی‌یابی و روش‌های تجزیه و تحلیل داده در حال بهبود کارایی و دقت این روش هستند. به طور کلی، WMS یک ابزار قدرتمند برای درک جامع اکوسامانه‌های میکروبی است که کاربردهای گسترده‌ای در علوم زیستی، پزشکی و صنعت دارد و انتظار

(McMurdie & Holmes, 2013; Oksanen *et al.*, 2020). تجزیه و تحلیل‌هایی مانند تنوع آلفا و بتا، آزمون‌های مقایسه‌ای مانند PERMANOVA و روش‌های یادگیری ماشین برای شناسایی الگوها و ارتباطات در داده‌ها به کار می‌روند. تجسم داده‌ها نیز بخش مهمی از تجزیه و تحلیل WMS است که با استفاده از ابزارهایی مانند ggplot2 در R یا Matplotlib در Python انجام می‌شود (Wickham, 2016; Hunter, 2007). نمودارهای مختلف مانند نمودارهای میله‌ای، نمودارهای جعبه‌ای، نمودارهای پراکنندگی و نقشه‌های حرارتی برای نمایش نتایج استفاده می‌شوند. در نهایت، تفسیر نتایج و استخراج معنای بیولوژیکی از داده‌ها انجام می‌شود که نیازمند دانش تخصصی در زمینه‌های مختلف از جمله میکروبیولوژی، ژنومیکس و اکولوژی است. لازم به ذکر است که انتخاب ابزارها و پارامترهای دقیق برای هر مرحله بستگی به نوع داده، هدف مطالعه و ویژگی‌های خاص پروژه دارد. همچنین، با توجه به پیشرفت سریع در زمینه بیوانفورماتیک، ابزارها و روش‌های جدید به طور مداوم در حال توسعه هستند که می‌توانند کارایی و دقت تجزیه و تحلیل‌ها را بهبود بخشند.

کاربردهای WMS بسیار گسترده و متنوع است. در زمینه اکولوژی میکروبی، این روش برای مطالعه تنوع و عملکرد جوامع میکروبی در محیط‌های مختلف مانند اقیانوس‌ها، خاک و زیستگاه‌های انسانی استفاده می‌شود (Sunagawa *et al.*, 2015). در پزشکی، WMS برای مطالعه میکروبیوم انسان و ارتباط آن با سلامت و بیماری

تجزیه و تحلیل نیز در فایل تکمیلی آورده شده‌اند (Andrews, 2010; Krueger, 2015; Li *et al.*, 2015; Segata *et al.*, 2012; Franzosa *et al.*, 2018; McMurdie & Holmes, 2013; Love *et al.*, 2014; Oksanen *et al.*, 2019).

می‌رود در آینده نقش مهمی در پیشرفت‌های علمی و فناوری ایفا کند. در ادامه به مراحل تجزیه و تحلیل داده Whole metagenome sequencing; WMS حاصل از ژنومیکس محیطی اشاره می‌شود (شکل ۲۲). اسکرپ‌های



شکل ۲۲- مراحل تجزیه و تحلیل داده‌های WGS Whole metagenome sequencing; WGS
Figure 22. Whole metagenome sequencing (WGS) data analysis steps

فناوری در تشخیص پیش از تولد غیرتهاجمی (NIPT) نیز مورد استفاده قرار می‌گیرد و در شناسایی ناهنجاری‌های کروموزومی جنین دقت بالایی نشان داده است (Chitty *et al.*, 2021). در حوزه سرطان‌شناسی، NGS برای شناسایی جهش‌های سوماتیک، تحلیل پروفایل بیان ژنی تومورها و توسعه درمان‌های هدفمند به کار می‌رود. این فناوری امکان پایش بیماری باقی‌مانده و

کاربرد های توالی‌یابی نسل جدید در حوزه‌های مختلف
توالی‌یابی نسل جدید (NGS) تحولی بنیادین در علوم زیستی و پزشکی ایجاد کرده و کاربردهای وسیعی در زمینه‌های گوناگون یافته است. در ژنتیک پزشکی، NGS با دقت و سرعت بالا به شناسایی جهش‌های نادر و تشخیص بیماری‌های ژنتیکی کمک کرده است. این

دارو و کاهش هزینه‌های آن کمک می‌کند (Slatko *et al.*, 2018). در علوم قضایی، NGS برای تحلیل نمونه‌های پیچیده DNA و تعیین هویت افراد از طریق توالی‌یابی کل ژنوم استفاده می‌شود و حتی با مقادیر بسیار کم DNA نیز امکان شناسایی فراهم می‌کند (Børsting & Morling, 2015). در مطالعات تکاملی و فیلوژنتیک، NGS به بازسازی دقیق‌تر روابط تکاملی بین گونه‌ها و مطالعه سازوکارهای تکاملی و درک بهتر تاریخ تکاملی انسان و سایر گونه‌ها کمک کرده است (Jarvis, 2016). در صنعت غذا، NGS برای شناسایی و ردیابی پاتوژن‌ها، تعیین اصالت محصولات و مطالعه فرایندهای تخمیر استفاده شده و در نتیجه بهبود ایمنی و کیفیت مواد غذایی را به همراه داشته است (Jagadeesan *et al.*, 2019). در علوم اعصاب، NGS برای مطالعه بیان ژن در سطح تک‌سلولی در مغز و شناسایی تغییرات ژنتیکی مرتبط با بیماری‌های عصبی به کار می‌رود و به درک بهتر پیچیدگی‌های مغز و سازوکارهای بیماری‌های عصبی کمک کرده است (Tasic *et al.*, 2018). همچنین، در مطالعات اپی‌ژنتیک، NGS برای بررسی تغییرات متیلاسیون DNA، تغییرات هیستونی و ساختار کروماتین استفاده شده و به درک بهتر نقش تنظیم اپی‌ژنتیکی در سلامت و بیماری یاری رسانده است (Dirks *et al.*, 2016). با وجود مزایای فراوان، چالش‌هایی نیز در استفاده از NGS وجود دارد. یکی از این چالش‌ها، حجم عظیم داده‌های تولید شده و نیاز به زیرساخت‌های محاسباتی قوی و متخصصان بیوانفورماتیک

شناسایی زود هنگام عود را از طریق تحلیل DNA تومور در گردش خون (ctDNA) فراهم می‌کند (Ignatiadis *et al.*, 2021). در زمینه بیماری‌های عفونی، NGS برای شناسایی و تعیین توالی پاتوژن‌ها، مطالعه مقاومت آنتی‌بیوتیکی و ردیابی شیوع بیماری‌ها استفاده می‌شود. این فناوری در مقابله با همه‌گیری COVID-19 نقشی اساسی ایفا کرده و برای توالی‌یابی ویروس SARS-CoV-2 و شناسایی واریانت‌های جدید مورد استفاده قرار گرفته است (Chiara *et al.*, 2021). در حوزه میکروبیولوژی، NGS از طریق متاژنومیکس به مطالعه جامع جمعیت‌های میکروبی پرداخته و به درک بهتر نقش میکروبیوم در سلامت انسان، اکوسامانه‌های طبیعی و فرایندهای صنعتی کمک کرده است (Knight *et al.*, 2018). در کشاورزی، NGS برای به‌نژادی گیاهان، توسعه نشانگرهای مولکولی و مطالعه تنوع ژنتیکی گیاهان به کار گرفته می‌شود که این امر منجر به توسعه محصولات مقاوم به آفات و بیماری‌ها و سازگار با تغییرات اقلیمی شده است (Varshney *et al.*, 2021). در علوم محیطی، NGS برای مطالعه تنوع زیستی، پایش اکوسامانه‌ها و ارزیابی اثرات تغییرات اقلیمی بر جوامع زیستی استفاده می‌شود و از طریق DNA محیطی (eDNA) امکان شناسایی گونه‌های نادر و در معرض خطر را فراهم می‌کند (Ruppert *et al.*, 2019). در صنعت داروسازی، NGS برای کشف داروهای جدید، مطالعه سازوکارهای اثر دارو و توسعه درمان‌های شخصی‌سازی شده به کار می‌رود و به بهبود فرایند توسعه

عمق خوانش^{۱۴۵}، جفت‌خوانش^{۱۴۶}، خوانش‌های تقسیم‌شده^{۱۴۷} و روش‌های ترکیبی هستند (Kosugi *et al.*, 2019). روش عمق خوانش بر اساس تغییرات در تعداد خوانش‌های همتراز شده به مناطق مختلف ژنوم عمل می‌کند و برای شناسایی تغییرات تعداد کپی (CNVs) مؤثر است (Yoon *et al.*, 2009). روش جفت‌خوانش از فاصله و جهت‌گیری غیرطبیعی جفت‌خوانش‌ها برای شناسایی SVs استفاده می‌کند و برای تشخیص حذف‌ها، درج‌ها و معکوس‌شدگی‌ها مفید است (Chen *et al.*, 2009). روش خوانش‌های تقسیم‌شده بر اساس تجزیه و تحلیل خوانش‌هایی که در چندین موقعیت ژنومی همتراز می‌شوند، عمل می‌کند و می‌تواند نقاط شکست SVs را با دقت بالا تعیین کند (Ye *et al.*, 2009). روش‌های ترکیبی از ترکیب چندین روش برای افزایش حساسیت و دقت شناسایی SVs استفاده می‌کنند (Rausch *et al.*, 2012). پیشرفت‌های اخیر در فناوری‌های توالی‌یابی طولانی‌خوانش مانند Oxford Nanopore و PacBio، امکان شناسایی دقیق‌تر SVs، به‌ویژه در مناطق تکراری ژنوم را فراهم کرده‌اند (Sedlazeck *et al.*, 2018). این فناوری‌ها می‌توانند خوانش‌های بسیار طولانی (تا چندین کیلوباز) تولید کنند که برای پوشش کامل SVs و تعیین دقیق نقاط شکست آنها مفید هستند. علاوه بر این، روش‌های مبتنی بر نقشه‌برداری نوری ژنوم (Optical genome mapping) مانند Bionano Genomics

برای تحلیل داده‌هاست. علاوه بر این، تفسیر نتایج NGS و ترجمه آنها به کاربردهای بالینی و عملی نیازمند دانش تخصصی و همکاری بین رشته‌ای است (Gargis *et al.*, 2016). با این حال، پیشرفت‌های مداوم در فناوری NGS و روش‌های تحلیل داده، چشم‌انداز امیدوارکننده‌ای را برای آینده ترسیم می‌کند. توسعه روش‌های توالی‌یابی با طول خوانش بلندتر و دقت بالاتر، امکان مطالعه دقیق‌تر ساختارهای پیچیده ژنومی را فراهم می‌کند. همچنین، ادغام داده‌های NGS با سایر فناوری‌های omics و استفاده از هوش مصنوعی و یادگیری ماشین در تحلیل داده‌ها، می‌تواند به کشف‌های جدید و کاربردهای نوآورانه در حوزه‌های مختلف منجر شود (Stark *et al.*, 2019b).

شناسایی تنوعات ساختاری ژنومی (SVs) از طریق

توالی‌یابی نسل جدید (NGS)

توالی‌یابی نسل جدید (NGS) نقش مهمی در شناسایی تنوعات ساختاری ژنومی (SVs)¹³⁹ ایفا می‌کند. SVs شامل تغییرات بزرگ مقیاس در ساختار ژنوم مانند حذف‌ها^{۱۴۰}، درج‌ها^{۱۴۱}، معکوس‌شدگی‌ها^{۱۴۲}، جابجایی‌ها^{۱۴۳} و تغییرات تعداد کپی (CSVs)¹⁴⁴ هستند که نقش مهمی در تنوع ژنتیکی و بیماری‌های انسانی دارند (Sudmant *et al.*, 2015). روش‌های مبتنی بر NGS برای شناسایی SVs شامل روش‌های مبتنی بر

¹³⁹ Structural Variations

¹⁴⁰ Deletions

¹⁴¹ Insertions

¹⁴² Inversions

¹⁴³ Translocations

¹⁴⁴ Copy number variations

¹⁴⁵ Read depth

¹⁴⁶ Paired-end read

¹⁴⁷ Split-read

SVs در تشخیص پیش از تولد غیرتهاجمی (NIPT) و آزمایش‌های ژنتیکی پیش از لانه‌گزینی (PGT) در حال گسترش است (Vermeesch *et al.*, 2016). همچنین، ادغام داده‌های SVs با سایر داده‌های ژنومی مانند بیان ژن و اطلاعات اپی‌ژنتیکی، امکان درک جامع‌تری از تأثیر SVs بر عملکرد ژنوم و فنوتیپ را فراهم می‌کند (Spielmann *et al.*, 2018). در آینده، انتظار می‌رود که پیشرفت‌های بیشتر در فناوری‌های NGS و روش‌های تجزیه و تحلیل داده، امکان شناسایی طیف کامل‌تری از SVs، از جمله تغییرات بسیار کوچک تا بازآرایی‌های بزرگ ژنومی را فراهم کند و به درک عمیق‌تر از نقش این تغییرات در سلامت و بیماری انسان کمک کند.

کاربرد توالی‌یابی نسل جدید (NGS) در مطالعه

تغییرات اپی‌ژنتیکی

در مطالعه تغییرات اپی‌ژنتیکی، توالی‌یابی نسل جدید (NGS) امکان بررسی جامع و دقیق سازوکارهای تنظیمی ژنوم را فراهم آورده است. تغییرات اپی‌ژنتیکی، که شامل تغییرات در بیان ژن بدون تغییر در توالی DNA می‌شوند، نقش مهمی در تکامل، رشد و بیماری‌ها دارند (Cavalli & Heard, 2019). توالی‌یابی نسل جدید (NGS) با ارائه روش‌های مختلف مانند توالی‌یابی بیسولفیت کل ژنوم (WGBS)، توالی‌یابی ایمونوپرسیپیتاسیون کروماتین (ChIP-seq) و توالی‌یابی دسترسی کروماتین (ATAC-seq)، امکان مطالعه گسترده تغییرات اپی‌ژنتیکی را فراهم کرده است. WGBS به‌طور خاص برای مطالعه متیلاسیون DNA در سطح ژنوم استفاده می‌شود و امکان

امکان شناسایی SVs بزرگ در مقیاس ژنوم را فراهم می‌کند (Levy-Sakin *et al.*, 2019). روش‌های محاسباتی پیشرفته و الگوریتم‌های یادگیری ماشین نیز برای بهبود دقت و حساسیت شناسایی SVs توسعه یافته‌اند (Heller & Vingron, 2019). این روش‌ها می‌توانند الگوهای پیچیده در داده‌های NGS را شناسایی کنند و SVs را با دقت بیشتری تشخیص دهند. کاربردهای بالینی شناسایی SVs با استفاده از NGS شامل تشخیص بیماری‌های ژنتیکی، شناسایی عوامل خطر ژنتیکی برای بیماری‌های پیچیده و تشخیص و طبقه‌بندی سرطان‌ها است (Stankiewicz & Lupski, 2010). در زمینه تحقیقات ژنومی، شناسایی SVs با استفاده از NGS به درک بهتر تکامل ژنوم، تنوع ژنتیکی جمعیت‌ها و سازوکارهای مولکولی بیماری‌ها کمک کرده است (Weischenfeldt *et al.*, 2013). چالش‌های باقی‌مانده در شناسایی SVs با استفاده از NGS شامل دشواری در شناسایی SVs در مناطق تکراری ژنوم، تشخیص SVs پیچیده و موزاییک، و تفسیر اهمیت عملکردی SVs شناسایی شده است (Huddleston & Eichler, 2016). با این حال، پیشرفت‌های مداوم در فناوری‌های توالی‌یابی، روش‌های تجزیه و تحلیل داده و الگوریتم‌های محاسباتی، امکان شناسایی جامع‌تر و دقیق‌تر SVs را فراهم می‌کند. این پیشرفت‌ها منجر به درک عمیق‌تر از نقش SVs در تنوع ژنتیکی و بیماری‌های انسانی شده و کاربردهای بالینی گسترده‌تری را امکان‌پذیر می‌سازد. به عنوان مثال، استفاده از NGS برای شناسایی

تغییرات اپی‌ژنتیکی در سازگاری و تکامل گونه‌ها کمک کرده است (Skvortsova *et al.*, 2018). در زمینه بیماری‌های پیچیده، NGS امکان شناسایی تغییرات اپی‌ژنتیکی مرتبط با بیماری‌هایی مانند دیابت، بیماری‌های قلبی-عروقی و اختلالات عصبی را فراهم کرده است (Bernstein *et al.*, 2020). با وجود پیشرفت‌های قابل توجه، چالش‌هایی در استفاده از NGS برای مطالعه تغییرات اپی‌ژنتیکی وجود دارد. یکی از این چالش‌ها، حجم زیاد داده‌های تولید شده و نیاز به روش‌های پیشرفته تحلیل داده است (Bock, 2012). توسعه الگوریتم‌های محاسباتی پیشرفته و روش‌های یادگیری ماشین برای تحلیل داده‌های NGS در حال پیشرفت است و به بهبود دقت و کارایی مطالعات اپی‌ژنتیک کمک می‌کند (Holder *et al.*, 2017; Rauschert *et al.*, 2020). چالش دیگر، نیاز به تفسیر عملکردی تغییرات اپی‌ژنتیکی شناسایی شده است که نیازمند مطالعات تجربی گسترده است (Stricker *et al.*, 2017). با این حال، پیشرفت‌های اخیر در روش‌های ویرایش اپی‌ژنوم مانند CRISPR-dCas9 امکان مطالعه عملکردی تغییرات اپی‌ژنتیکی را فراهم کرده است (Liu *et al.*, 2016). آینده کاربرد NGS در مطالعات اپی‌ژنتیک بسیار امیدوارکننده است. توسعه روش‌های جدید مانند توالی‌یابی تک سلولی امکان مطالعه تغییرات اپی‌ژنتیکی در سطح سلول‌های منفرد را فراهم کرده که به درک بهتر ناهمگونی سلولی و سازوکارهای تنظیمی دقیق کمک می‌کند (Kelsey *et al.*, 2017). همچنین، ادغام

شناسایی الگوهای متیلاسیون با دقت تک نوکلئوتیدی را فراهم می‌کند (Ziller *et al.*, 2015). این روش به درک بهتر نقش متیلاسیون DNA در خاموشی ژن و تنظیم بیان ژن کمک کرده است. ChIP-seq برای مطالعه تعاملات پروتئین-DNA و تغییرات هیستونی استفاده می‌شود و امکان نقشه‌برداری از مناطق تنظیمی ژنوم را فراهم می‌کند (Nakato & Shirahige, 2017). این روش به شناسایی انواع مختلف تغییرات هیستونی مانند متیلاسیون، استیلاسیون و فسفریلاسیون کمک کرده است که نقش مهمی در تنظیم بیان ژن دارند. ATAC-seq روشی نوین برای مطالعه ساختار کروماتین و دسترسی‌پذیری DNA است که امکان شناسایی مناطق تنظیمی فعال در ژنوم را فراهم می‌کند (Buenrostro *et al.*, 2015). این روش به درک بهتر از چگونگی تأثیر ساختار کروماتین بر بیان ژن کمک کرده است. علاوه بر این، روش‌های NGS امکان مطالعه RNAهای غیرکدکننده (ncRNAs) مانند microRNAها و Long non-coding RNAها را فراهم کرده‌اند که نقش مهمی در تنظیم اپی‌ژنتیکی دارند (Cech & Steitz, 2014). کاربرد NGS در مطالعات اپی‌ژنتیک به پیشرفت‌های قابل توجهی در زمینه‌های مختلف پزشکی و زیست‌شناسی منجر شده است. در سرطان‌شناسی، NGS امکان شناسایی تغییرات اپی‌ژنتیکی خاص تومور را فراهم کرده که می‌تواند به تشخیص زودهنگام، پیش‌آگهی دقیق‌تر و درمان‌های هدفمند منجر شود (Baylin & Jones, 2016). در مطالعات تکاملی، NGS به درک بهتر نقش

کمک کرده است (Ranjan *et al.*, 2016). به عنوان مثال، مطالعات متاژنومیکس بر روی میکروبیوم انسانی منجر به کشف ارتباطات مهم بین ترکیب میکروبی روده و سلامت انسان شده است، که کاربردهای بالقوه‌ای در تشخیص و درمان بیماری‌ها دارد (Lloyd-Price *et al.*, 2017). در حوزه اکولوژی میکروبی، NGS امکان مطالعه جامع جمعیت‌های میکروبی در اکوسامانه‌های مختلف مانند اقیانوس‌ها، خاک و مناطق قطبی را فراهم کرده است. این مطالعات به درک بهتر نقش میکروب‌ها در چرخه‌های بیوژئوشیمیایی و تغییرات اقلیمی کمک کرده‌اند (Sunagawa *et al.*, 2015). علاوه بر این، NGS در مطالعه تکامل میکروبی و انتقال ژن افقی کاربرد گسترده‌ای دارد و به درک بهتر سازوکارهای سازگاری و تنوع ژنتیکی در جمعیت‌های میکروبی کمک می‌کند (Brito *et al.*, 2016). در زمینه بیماری‌های عفونی، NGS نقش مهمی در شناسایی سریع و دقیق عوامل بیماری‌زا، ردیابی شیوع بیماری‌ها و مطالعه مقاومت آنتی‌بیوتیکی ایفا می‌کند. این فناوری امکان تشخیص همزمان چندین پاتوژن و بررسی تغییرات ژنتیکی آنها را در طول زمان فراهم می‌کند (Gardy & Loman, 2018). در صنعت غذا و نوشیدنی، NGS برای مطالعه میکروب‌های مفید در تخمیر و همچنین شناسایی عوامل فساد مواد غذایی استفاده می‌شود (De Filippis *et al.*, 2017). در حوزه بیوفناوری، NGS به کشف آنزیم‌ها و متابولیت‌های جدید با کاربردهای صنعتی و دارویی از طریق مطالعه متاژنوم محیط‌های مختلف کمک کرده

داده‌های اپی‌ژنومیک با سایر داده‌های Omics مانند ژنومیکس، ترنسکریپتومیکس و پروتئومیکس، امکان درک جامع‌تری از تنظیم ژن و سازوکارهای بیماری را فراهم می‌کند (Hasin *et al.*, 2017). در نهایت، کاربرد NGS در مطالعه تغییرات اپی‌ژنتیکی نه تنها دانش ما را در مورد سازوکارهای تنظیمی ژنوم گسترش داده، بلکه چشم‌اندازهای جدیدی را برای تشخیص، پیشگیری و درمان بیماری‌ها گشوده است. با پیشرفت مداوم در فناوری‌ها و روش‌های تحلیلی، انتظار می‌رود که NGS نقش حیاتی در کشف و درک تغییرات اپی‌ژنتیکی در آینده داشته باشد و به پیشرفت‌های قابل توجهی در پزشکی دقیق و درمان‌های شخصی‌سازی شده منجر شود.

تجزیه و تحلیل جمعیت میکروبی از طریق توالی‌یابی

نسل جدید (NGS)

توالی‌یابی نسل جدید (NGS) از طریق فراهم کردن درک عمیق‌تر از تنوع، عملکرد و اکولوژی میکروارگانیسم‌ها و همچنین با ارائه روش‌های سریع، دقیق و مقرون به صرفه برای تجزیه و تحلیل ژنوم و ترنسکریپتوم میکروب‌ها، به پیشرفت‌های چشمگیری در حوزه‌های مختلف علوم زیستی، پزشکی و محیط زیست منجر شده است (Quince *et al.*, 2017). یکی از مهم‌ترین کاربردهای NGS در این زمینه، مطالعات متاژنومیکس است که امکان بررسی جامع جمعیت‌های میکروبی در محیط‌های پیچیده را بدون نیاز به کشت فراهم می‌کند. این روش به شناسایی گونه‌های جدید و ناشناخته و درک بهتر از تعاملات پیچیده بین میکروب‌ها و محیط زیست آنها

NGS با سایر فناوری‌های Omics مانند متابولومیکس و پروتئومیکس، می‌تواند به درک جامع‌تری از عملکرد و تعاملات میکروبی منجر شود (Wörheide *et al.*, 2021). در نهایت، کاربرد NGS در مطالعه جمعیت‌های میکروبی نه‌تنها به پیشرفت دانش پایه در مورد تنوع و اکولوژی میکروبی کمک می‌کند، بلکه کاربردهای مهمی در حوزه‌های مختلف از جمله پزشکی شخصی‌سازی شده، کشاورزی پایدار، حفاظت از محیط زیست و توسعه فناوری‌های زیستی دارد.

توالی‌یابی نسل جدید (NGS) در کشاورزی

توالی‌یابی نسل جدید (NGS) کاربردهای گسترده‌ای در بخش‌های گیاهی و دامی دارد. در زمینه به‌نژادی گیاهان، NGS امکان شناسایی سریع و دقیق نشانگرهای مولکولی را فراهم کرده که به بهبود انتخاب ژنومی و اصلاح به کمک نشانگر (MAS) منجر شده است (Varshney *et al.*, 2021). این فناوری همچنین در توسعه نقشه‌های ژنتیکی با تراکم بالا، که برای شناسایی QTLs و ژن‌های کنترل‌کننده صفات مهم زراعی ضروری هستند، نقش کلیدی دارد (Rasheed *et al.*, 2017). در مطالعات ژنومیکس جمعیت گیاهی، NGS به درک بهتر تنوع ژنتیکی، ساختار جمعیت و سازگاری محلی در گونه‌های زراعی و وحشی کمک کرده است (Huang & Han, 2014). این اطلاعات برای حفاظت از منابع ژنتیکی گیاهی و بهره‌برداری از تنوع موجود در برنامه‌های اصلاحی بسیار ارزشمند است. در زمینه مهندسی ژنتیک گیاهان، NGS در توسعه و ارزیابی محصولات تراریخته، از جمله

است (Alves *et al.*, 2018). علاوه بر متاژنومیکس، NGS در مطالعات متاترنسکرپتومیکس نیز کاربرد دارد که امکان بررسی فعالیت ژنی در سطح جامعه میکروبی را فراهم می‌کند. این روش به درک بهتر پاسخ‌های فیزیولوژیکی جمعیت‌های میکروبی به تغییرات محیطی کمک می‌کند (Shakya *et al.*, 2019). در زمینه ایمنی زیستی و نظارت بر محیط زیست، NGS برای ردیابی میکروارگانیسم‌های بیماری‌زا و ژن‌های مقاومت به آنتی‌بیوتیک در محیط‌های مختلف استفاده می‌شود (Hendriksen *et al.*, 2019). این فناوری همچنین در مطالعه میکروپلاستیک‌ها و تأثیر آنها بر اکوسامانه‌های میکروبی کاربرد دارد (Zhu *et al.*, 2018). با وجود مزایای فراوان، چالش‌هایی نیز در استفاده از NGS برای مطالعه جمعیت‌های میکروبی وجود دارد. یکی از این چالش‌ها، پیچیدگی تحلیل داده‌های حجیم تولید شده و نیاز به روش‌های بیوانفورماتیکی پیشرفته است (Quince *et al.*, 2017). علاوه بر این، کاستی‌هایی در شناسایی دقیق گونه‌ها و تعیین عملکرد ژن‌ها بر اساس داده‌های NGS وجود دارد که نیاز به تکمیل با روش‌های تجربی دارد (Escobar-Zepeda *et al.*, 2018). با این حال، پیشرفت‌های مداوم در فناوری NGS و روش‌های تحلیل داده، چشم‌انداز امیدوارکننده‌ای را برای آینده مطالعات میکروبی ترسیم می‌کند. توسعه روش‌های توالی‌یابی با طول خوانش بلندتر و دقت بالاتر، امکان بازسازی دقیق‌تر ژنوم‌های میکروبی از داده‌های متاژنومی را فراهم می‌کند (Pollard *et al.*, 2018). همچنین، ادغام داده‌های

شناسایی دقیق محل درج ژن و بررسی اثرات غیرهدف، نسل جدید در مطالعات ژنومیک، ترنسکریپتوم و کاربرد دارد در جدول ۱ چند مثال از کاربرد توالی‌یابی اپیزنومیک غلات ذکر شده است (Xie et al., 2019).

جدول ۱ - مطالعات ژنومیکس، ترنسکریپتومیکس و اپیزنومیکس صورت گرفته در غلات از طریق انواع روش‌های توالی‌یابی نسل جدید

Table 1. Genomics, transcriptomics and epigenomics studies conducted in cereals through various new generation sequencing methods

منبع	نتیجه	هدف و متد تحقیق	نوع توالی‌یابی NGS	نام گیاه
(Xu et al., 2014)	۵۲۴ nsSNP مرتبط با ۲۷۱ ژن کاندید در فرایندهای مهمی مانند تنظیم هورمون‌های گیاهی، متابولیسم کربوهیدرات و قند، تنظیم مولکول‌های سیگنالینگ، واکنش‌های اکسایش-کاهش و سازگاری فتوسنتز با محیط/ nsSNP ها در منطقه bin 1.07 خوشه‌بندی شدند که قبلاً شش QTL مرتبط با تحمل خشکی در آن گزارش شده بود/تجزیه و تحلیل RNAseq مشخص کرد که ۷۰٪ ژن‌های کاندید تغییرات بیان معنی‌داری داشتند	شناسایی ژن‌های کاندید برای تحمل خشکی در ذرت با استفاده از توالی‌یابی مجدد کل ژنوم/ تجزیه و تحلیل واریانت‌های مشترک (cv) و خوشه‌بندی، جهت شناسایی SNPهای غیرمتراقد (nsSNPs) و ژن‌های کاندید مرتبط با تحمل خشکی	Whole genome sequencing (WGS)	ذرت (<i>Zea mays</i>)
(Massel et al., 2016)	تنوع ژنتیکی در اثر اهلی‌سازی کاهش یافته و ۱۲۸ ژن نشانه‌های انتخاب خالص‌کننده را نشان دادند/ ژن‌های ناقل آمونیم (AMT) عموماً تنوع آلی پایینی داشتند، در حالی که تعداد قابل توجهی از ژن‌های ناقل نیترات/ پپتید ۱ (NRT1/PTR) در ژرم پلاسما اهلی شده تنوع نوکلئوتیدی بالاتری نشان دادند/ اعضای نژاد متمایز <i>Guinea margaritifera</i> حاوی تعدادی آلل منحصر به فرد بودند و همراه با گونه‌های وحشی سورگوم، منبع غنی از تنوع جدید برای بهبود NUE در سورگوم را نشان دادند/ اولین ارزیابی جامع از تنوع ژنتیکی در ژن‌های مرتبط با NUE در سورگوم ارائه شد.	شناسایی اهداف ژنتیکی جدید برای بهبود کارایی استفاده از نیتروژن (NUE) در سورگوم و سایر غلات/ توالی‌یابی کامل و تحلیل ۲۳۰ ژن احتمالی درگیر در جذب و استفاده از نیتروژن در یک پائل متنوع از لاین‌های سورگوم	Whole genome sequencing (WGS)	سورگوم (<i>Sorghum bicolor</i>)
(Ichikawa et al., 2023)	در مجموع ۳۳۸ واریانت نسبت به توالی گیاه مادری در نتاج شناسایی شد/ میانگین تعداد واریانت‌ها از ۹۰ در گیاهان حاصل از بذر تا ۳۷.۴ در گیاهان باززایی شده از جنین‌های بالغ متغیر بود/ نرخ جهش طبیعی محاسبه شده (۱۰-۸ × ۱.۲) با مقادیر گزارش شده قبلی سازگار بود/ نسبت واریانت‌های تک نوکلئوتیدی (SNVs) در گیاهان حاصل از بذر ۹۱.۱٪ بود که بالاتر از مقدار ۵۶.۱٪ گزارش شده قبلی بود/ ترانزیشن به ترانسورژن در SNV ها در گیاهان باززایی شده کمتر بود. گیاهان باززایی شده از جنین‌های بالغ به‌طور معنی‌داری واریانت‌های بیشتری داشتند/ استفاده از زیگوت‌ها و جنین‌های نابالغ می‌تواند تغییرات سوماکلونال را در دستکاری‌های ژنتیکی برنج کاهش دهد.	بررسی تغییرات سوماکلونال در گیاهان برنج (<i>Oryza sativa</i>)، رقم Nipponbare باززایی شده از منابع مختلف سلولی/ توالی‌یابی کل ژنوم گیاه مادری، نتاج حاصل از بذر و گیاهان باززایی شده از سه منبع سلولی مختلف	Whole genome sequencing (WGS)	برنج (<i>Oryza sativa</i>)
(Brenchley et al., 2012)	. نتایج نشان داد که بین ۹۴,۰۰۰ تا ۹۶,۰۰۰ ژن شناسایی شد و دو سوم آنها به سه ژنوم تشکیل‌دهنده گندم هگزاپلوئید A، B و D نسبت داده شدند/ نقشه‌های سینتتی با وضوح بالا، اختلالات کوچک زیادی را در ترتیب حفاظت‌شده ژن‌ها نشان دادند. محققان دریافتند که ژنوم هگزاپلوئید بسیار پویاست و کاهش قابل توجهی در اعضای خانواده	هدف اصلی، ایجاد منابع ژنومی برای تسریع پیشرفت در افزایش تولید گندم از طریق گسترش تنوع ژنتیکی و تحلیل صفات کلیدی/ شامل توالی‌یابی ژنوم	Whole genome sequencing (WGS)	گندم (<i>Triticum aestivum</i> L.)

		هگزاپلوئید ۱۷ گیگابازی گندم با استفاده از فناوری پایروسکوئنسینگ ۴۵۴ و مقایسه آن با ژنوم‌های اجدادی دیپلوئید	ژنی در طی پلی‌پلوئیدی شدن و اهلی‌سازی رخ داده است. / چندین کلاس از ژن‌های درگیر در برداشت انرژی، متابولیسم و رشد در میان خانواده‌های ژنی گسترش‌یافته بودند که می‌توانند با بهره‌وری محصول مرتبط باشند/ شناسایی تنوع ژنتیکی گسترده، منبعی برای تسریع کشف ژن و بهبود این محصول.	
گندم (<i>Triticum aestivum</i> L.)	Whole genome sequencing (WGS)	بررسی تغییرات ژنتیکی در گندم طی فرایند اصلاح مدرن در چین و ایالات متحده/ درک چگونگی تأثیر اصلاح گندم بر ساختار ژنومی این محصول/ تحلیل مقایسه‌ای جامع داده‌های توالی‌یابی کل ژنوم ۳۵۵ نمونه گندم معمولی (شامل توده‌های بومی و ارقام مدرن از چین و آمریکا) در سطوح فنوتیپی و ژنومیکی	تنوع ژنتیکی ارقام مدرن گندم در مقایسه با توده‌های بومی ^{۱۴۸} به وضوح کاهش یافته است/ با این تغییرات ژنتیکی، اکثر فنوتیپ‌های ارقام چینی و آمریکایی تغییرات قابل توجهی داشتند. از ۲۱ صفت زراعی مورد بررسی، ۸ صفت تغییرات همگرا بین دو کشور نشان دادند/، از ۲۰۷ جایگاه ژنی مرتبط با این ۲۱ صفت، بیش از نیمی با مناطق ژنومی که شواهد انتخاب را نشان می‌دادند، همپوشانی داشتند/ توزیع جایگاه‌های انتخاب شده بین ارقام چینی و آمریکایی نشان داد که اصلاح برای افزایش بهره‌وری در این دو منطقه از طریق تجمع واریانت‌های مشترک و خاص هر منطقه انجام شده است / درک ساختار ژنتیکی سازگاری گندم با محیط‌های متنوع تولید کشاورزی و راهنمایی برای بهینه‌سازی استراتژی‌های اصلاحی جهت طراحی ارقام بهتر گندم	(Niu et al., 2023)
گندم (<i>Triticum aestivum</i> L.)	Whole exome sequencing (WES)	بررسی توالی‌یابی کل اگزوم در رگه‌های درون‌زاد ^{۱۵۱} نوترکیب انتخاب شده گندم نان / هدف اصلی، ایجاد یک منبع مفید برای استخراج آلل‌ها و تحلیل جمعیت‌های تفکیک شده (BSA) ^{۱۵۲} / ترکیب روش‌های توالی‌یابی هدفمند و غنی‌سازی با توالی‌یابی نسل جدید (NGS) برای بررسی ۴۶ نمونه گندم نان از یک جمعیت رگه‌های درون‌زاد نوترکیب.	که تعداد زیادی پلی‌مورفیسم تک نوکلئوتیدی (SNP) ^{۱۴۹} شناسایی و طبقه‌بندی و . برای اعتبارسنجی فنی، هشت SNP به صورت تصادفی انتخاب و به نشانگرهای KASP ^{۱۵۰} تبدیل شدند. / این مطالعه یک مجموعه ابزار مولکولی قابل دسترس و قابل استفاده مجدد برای استخراج داده‌های آلی ایجاد کرد/ که استفاده از روش‌های پیشرفته توالی‌یابی می‌تواند به شناسایی واریانت‌های ژنتیکی مرتبط با فنوتیپ‌های هدف در گندم کمک کند و منابع ارزشمندی برای مطالعات ژنتیکی و اصلاحی فراهم آورد.	(Esposito et al., 2014)
برنج (<i>Oryza sativa</i>)	Single cell RNA-seq (scRNA seq)	به بررسی سازوکار مقاومت برنج به آفت زنجره قهوه‌ای (<i>Nilaparvata lugens</i>) با استفاده از فناوری توالی‌یابی RNA تک سلولی ^{۱۵۳} /scRNAseq/ هدف اصلی، شناسایی انواع سلول‌ها و سازوکارهای مولکولی درگیر در مقاومت به این آفت بود. روش تحقیق شامل مقایسه پاسخ غلاف برگ‌های دو واریته	که ۱۴,۶۹۹ و ۱۶,۲۳۷ سلول به ترتیب در TN1 و YHY15 شناسایی و به ۹ خوشه سلولی طبقه‌بندی شدند/ تفاوت‌های قابل توجهی در انواع سلول‌ها (مانند سلول‌های غلاف آوندی، سلول‌های نگهبان، سلول‌های مزوفیل، سلول‌های آوند چوبی، سلول‌های بولیفورم و سلول‌های آوند آبکش) بین دو واریته مشاهده شد/ هر نوع سلول از سازوکار مولکولی متفاوتی برای مقاومت استفاده می‌کند. برای مثال، سلول‌های مزوفیل در تنظیم بیان ژن‌های مرتبط با تولید وانیلین، کپسایسین و ROS، سلول‌های آوند آبکش در تنظیم ژن‌های مرتبط با گسترش دیواره سلولی، و سلول‌های آوند چوبی در کنترل بیان ژن‌های مرتبط با کیتین و پکتین نقش دارند./ که مقاومت برنج به زنجره قهوه‌ای یک فرایند پیچیده شامل عوامل متعدد مقاومت به	(Zha et al., 2023)

¹⁴⁸ Landrace

¹⁴⁹ Single nucleotide polymorphisms

¹⁵⁰ Kompetitive allele-specific PCR

¹⁵¹ Inbred line

¹⁵² Bulk segregant analysis

¹⁵³ Single cell RNA-seq

		حساس (TN1) و مقاوم (YHY15) برنج به حمله زنجره قهوه‌ای، ۴۸ ساعت پس از آلودگی، با استفاده از توالی‌یابی RNA تک سلولی	حشرات است	
ذرت (<i>Zea mays</i>)	Single cell RNA-seq (scRNA seq)	به بررسی توسعه بلال ذرت در سطح تک سلولی با استفاده از فناوری توالی‌یابی RNA تک سلولی (scRNA-seq) / هدف اصلی، ایجاد یک اطلس توسعه‌های جامع برای درک بهتر تنوع انواع سلول‌ها، دامنه‌های توسعه‌های و شبکه‌های ژنی مرتبط با آنها در بلال ذرت/ بررسی پروفایل‌های رونویسی ۱۲,۵۲۵ سلول منفرد از بلال‌های در حال توسعه ذرت با استفاده از scRNA-seq	نتایج با استفاده از دوره‌شدن درجا mRNA و توالی‌یابی RNA سلول‌های جداسازی شده با فلوسایتومتری (FACS RNA-seq) اعتبارسنجی شدند. / این مطالعه منجر به ایجاد اولین نقشه scRNA-seq از یک گل‌آذین شد/ داده‌ها می‌توانند به مطالعات ژنتیکی کمک کنند، از جمله پیش‌بینی افزونگی ژنتیکی، یکپارچه‌سازی شبکه‌های رونویسی و شناسایی ژن‌های کاندید مرتبط با صفات عملکرد محصول. / این تحقیق یک منبع ارزشمند برای درک بهتر سازوکارهای مولکولی درگیر در توسعه بلال ذرت فراهم می‌کند و می‌تواند به بهبود عملکرد محصول از طریق اصلاح نباتات هدفمند کمک کند.	(Xu et al., 2021)
برنج (<i>Oryza sativa</i>)	Single cell RNA-seq (scRNA seq)	بررسی سازوکار تنظیم توسعه جوانه رأسی (پلومول) در دانه برنج با استفاده از توالی‌یابی ترنسکریپتوم تک سلولی / هدف اصلی، شناسایی الگوهای بیان ژن در سطح فضایی در بافت‌های مختلف جوانه رأسی در حال توسعه / توالی‌یابی ترنسکریپتوم تک سلولی جوانه رأسی دانه برنج رقم Zhongjiazao 17 یک رقم محبوب برنج اندیکا در چین .	۱۵ خوشه سلولی شناسایی شد که ۱۳ خوشه با استفاده از ژن‌های نشانگر و ژن‌های اختصاصی خوشه به انواع سلولی مختلف نسبت داده شدند / بیان ژن‌های نشانگر چندین نوع سلولی در چندین خوشه نشان‌دهنده یک سامانه توسعه‌ای پیچیده بود / برخی ژن‌های مرتبط با سیگنال‌دهی فیتوهورمون‌ها مانند اسید آبسزیک در خوشه‌های خاصی بیان بالایی داشتند / عناصر تنظیمی سپس در پروموتور ژن‌های بیان شده در خوشه‌های سلولی خاص شناسایی شدند و برخی موتیف‌های کلیدی مرتبط با هورمون‌ها در خوشه‌های خاصی فراوان بودند / الگوهای بیان فضایی ژن‌های درگیر در جوانه‌زنی سریع بذر، رشد گیاهچه و توسعه شناسایی شدند / این یافته‌ها درک ما از تنوع سلولی و تخصصی شدن در جوانه رأسی برنج، به عنوان یک گیاه مدل تک‌لپه‌ای، را افزایش داد.	(Zhu et al., 2024)
گندم (<i>Triticum aestivum L.</i>)	Bulk RNA-seq	بررسی سازوکارهای مقاومت در برابر بیماری سفیدک پودری گندم در ژنوتیپ مقاوم BSR-Seq / YD588 از روش (توالی‌یابی RNA گروه‌های تفکیک شده) برای شناسایی ژن‌های مرتبط با مقاومت استفاده شد /	تجزیه و تحلیل ژنتیکی نشان داد که یک ژن غالب منفرد به نام PmYD588 مسئول مقاومت است/ با استفاده از BSR-Seq ^{۱۵۴} و تحلیل SNP، دو ناحیه کاندید روی بازوی کروموزوم 2B شناسایی شد/ نشانگرهای مولکولی توسعه یافته، PmYD588 را در یک فاصله ۴/۲ سانتی‌مورگان نقشه‌یابی کرد که احتمالاً یک ژن Pm جدید است/ تحلیل ژن‌های بیان شده متفاوت (DEGs) بین والدین و گروه‌های تفکیک شده انجام و ۲۳ ژن احتمالی تنظیم‌کننده در مسیر برهمکنش گیاه-پاتوزن انتخاب و الگوهای بیان زمانی آنها بررسی شد/ شش ژن مرتبط با بیماری، پروفایل‌های بیان متمایزی پس از تهاجم پاتوزن نشان دادند که می‌توانند به عنوان کاندیداهای کلیدی برای درک سازوکارهای مقاومت و بهبود مقاومت پایدار به سفیدک پودری گندم مورد استفاده قرار گیرند	(Ma et al., 2021)
برنج	Bulk	بررسی اثرات فیزیولوژیکی و	۴۴۷ ژن با افزایش بیان بیش از چهار برابر شناسایی شدند/ تجزیه و	(Chandr

<i>(Oryza sativa)</i>	RNA-seq	سازوکارهای مولکولی تنش شوری خاک بر ریشه‌های برنج / روش RNA-Seq برای تجزیه و تحلیل ترنسکریپتوم ریشه‌های برنج تحت تنش شوری خاک (۲۵۰ میلی‌مولار NaCl در زمان‌های مختلف) استفاده شد	تحلیل غنی‌سازی GO نشان داد که فرایندهای متابولیکی کربوهیدرات و اسیدهای آمینه به‌طور قابل توجهی تحت تأثیر تنش شوری قرار می‌گیرند/ تجزیه و تحلیل MapMan نشان داد که ژن‌های مرتبط با متابولیت‌های ثانویه تحت تنش شوری القا می‌شوند. /تجزیه و تحلیل پروفایل متابولیت‌ها تجمع فنل‌ها و فلاونوئیدها در ریشه تحت تنش شوری را تأیید کرد/ در نهایت، یک شبکه عملکردی متشکل از ژن‌های تنظیمی بر اساس برهمکنش‌های پروتئین-پروتئین پیش‌بینی شده ساخته شد که می‌تواند برای کاربردهای آینده مفید باشد.	an et al., 2019)
جو (<i>Hordeum vulgare</i>)	Bulk RNA-seq	هدف شناسایی ژن‌ها و مسیرهای کاندید دخیل در پاسخ به تنش شوری در جو/ روش RNA-seq برای تجزیه و تحلیل ترنسکریپتوم ریشه و برگ یک توده بومی تونس (Boulifa) متحمل به شوری تحت تیمار نمک در زمان‌های ۲، ۸، و ۲۴ ساعت استفاده شد	نتایج نشان داد که در برگ‌ها بیان ۳۵۸۵ ژن افزایش و ۵۵۸۶ ژن کاهش یافت، در حالی که در ریشه‌ها بیان ۱۳۲۰۰ ژن افزایش و ۱۰۵۷۵ ژن کاهش یافت/ تنظیم بیان ژن در ریشه‌ها به شدت تحت تأثیر قرار گرفت که پیچیدگی سازوکارهای پاسخ به تنش شوری در این بافت را نشان می‌دهد/ تجزیه و تحلیل‌های عملکردی در هر دو بافت نشان داد که پاسخ به تنش شوری عمدتاً از طریق مسیرهای حسگری و پیام‌رسانی، برنامه‌ریزی مجدد رونویسی قوی، تثبیت هومئوستازی هورمون‌ها و اسمولیت‌ها و یون‌ها، افزایش پاکسازی گونه‌های فعال اکسیژن و فعال‌سازی سامانه‌های انتقال و فتوسنتز انجام می‌شود/ تعدادی ژن کاندید دخیل در مسیرهای پیام‌رسانی هورمونی و کینازی، همچنین چندین خانواده فاکتور رونویسی و ناقل‌ها شناسایی شدند/ اطلاعات ارزشمندی در مورد ژن‌های پاسخ‌دهنده اولیه به تنش شوری در ریشه و برگ جو فراهم شد.	(Nefissi Ouertan i et al., 2021)
گندم دروم (<i>Triticum durum</i>)	Small RNA-seq	بررسی اثرات فراوراثتی تنش کم‌آبی قبل از گلدهی بر نسل اول گندم دوروم/ از روش‌های توالی‌یابی نسل جدید برای تجزیه و تحلیل جامع small RNA-omics، ترنسکریپتومیکس ^{۱۵۵} و mRNA دگرگادومیکس ^{۱۵۶} در نسل اول استفاده شد/ عملکرد، ویژگی‌های کیفی دانه و صفات فیزیولوژیکی به‌طور دقیق مورد بررسی قرار گرفتند	نتایج نشان داد که تنش کم‌آبی والدین تأثیرات مثبتی بر نسل بعد، به ویژه در رقم کمتر متحمل به تنش، از جمله بهبود شاخص برداشت و محتوای پروتئین داشت / تجزیه و تحلیل small RNA-seq منجر به شناسایی ۱۷۳۹ میکرو RNA حفاظت شده و ۷۷۴ میکرو RNA جدید شد/ توالی‌یابی ترنسکریپتوم بیان ۶۶،۵۵۹ ژن را مشخص کرد، در حالی که توالی‌یابی دگرگادوم دینامیک شکست mRNA هدایت شده توسط میکرو RNA را نشان داد/ میکرو RNA ها و ژن‌های با بیان متفاوت شناسایی شدند که الگوهای تنظیمی قابل توجهی را در پاسخ به تنش فراوراثتی و بین نسلی نشان دادند. / تجزیه و تحلیل یکپارچه سه پلتفرم امیکس، برهمکنش‌های بیولوژیکی مهمی را بین میکرو RNA ها و اهداف پاسخگو به تنش آشکار کرد، که نشان می‌دهد تحمل فراوراثتی به تنش احتمالاً از طریق مسیرهایی مانند سیگنالینگ هورمونی و متابولیسم مواد مغذی ایجاد می‌شود. / این مطالعه اولین تأیید اثرات فراوراثتی تنش کم‌آبی در گندم دوروم را فراهم می‌کند و بینش‌های جدیدی در سطح مولکولی ارائه می‌دهد که نشان می‌دهد ماژول‌های کلیدی میکرو RNA-mRNA کاندیداهایی برای بهبود تحمل به تنش فراوراثتی هستند	(Liu et al., 2021)
گندم (<i>Triticum aestivum L.</i>)	Small RNA-seq	بررسی پروفایل microRNA و سازوکارهای احتمالی مرتبط با تجمع کادمیوم در گندم/ از دو رقم گندم با تجمع بالای کادمیوم (Annon9267) و	نتایج منجر به شناسایی microRNAهای مهم مانند tae-miR9663-5p و tae-miR6201 و سازوکارهای احتمالی مرتبط با تجمع کادمیوم شد/ محققان الگوهای خاص پروفایل microRNA و ژن‌های هدف آنها را خلاصه کردند. در ریشه‌ها و برگ‌های گندم، microRNAهای با بیان متفاوت مرتبط با تجمع کادمیوم در	(Liu et al., 2021)

¹⁵⁵ mRNA Transcriptomics

¹⁵⁶ mRNA Degradomics

		<p>تجمع پایین کادمیوم (Qian 102032) به عنوان مدل‌های آزمایشی استفاده شد/ هر دو رقم به مدت ۲ ساعت با کادمیوم تیمار شدند و سپس از روش توالی‌یابی RNA کوچک برای بررسی پروفایل microRNA در بافت‌های ریشه و برگ استفاده شد.</p>	<p>بافت‌های مختلف گیاه (ریشه یا برگ) شناسایی شدند / تجزیه و تحلیل غنی‌سازی عملکردی بر روی ژن‌های هدف microRNA های با بیان متفاوت در ارقام گندم با تجمع کم و زیاد کادمیوم در بافت‌های مختلف گیاه (ریشه یا برگ) منجر به شناسایی برخی microRNA های بالغ شناخته شده و microRNA های جدید شد / این مطالعه اولین گزارش جامع از پروفایل microRNA در پاسخ به تنش کادمیوم در گندم است و microRNA های شناسایی شده می‌توانند به عنوان نشانگرهای زیستی بالقوه برای غربالگری گندم با تجمع کم کادمیوم در نظر گرفته شوند.</p>	
گندم (<i>Triticum aestivum</i> L.)	de novo genome assembly	<p>توالی‌یابی و سرهم‌بندی ژنوم رقم ژاپنی گندم نان Norin 61 و بررسی تنوع ژنتیکی آن در مقایسه با سایر ارقام گندم / روش‌های توالی‌یابی نسل جدید و سرهم‌بندی ژنوم برای ایجاد یک سرهم‌بندی کروموزومی با کیفیت بالا استفاده شد /</p>	<p>نتایج نشان داد که ژنوم Norin 61 حدود ۱۵ گیگابایت طول دارد با ۱۵۰ اسکافولد خام ۲۲ مگابایت / تجزیه و تحلیل عناصر تکراری مناطق کروموزومی منحصر به فردی را در Norin 61 شناسایی کرد که شامل آرایه‌های پشت سر هم از خانواده ژنی PR13 مرتبط با بیماری‌زایی است / تغییرات جدیدی در تعداد کپی ژن FT1/VRN3 (فلوریژن) در هومئولوگ B، شبه‌ژن‌شدگی^{۱۵۷} هومئولوگ D و ارتباط آلل‌های هومئولوگ A با عادت رشد بهاره/ زمستانه گزارش شد / همچنین، ژنوم Norin 61 حاوی واریانت‌های عملکردی خاص شرق آسیا متفاوت از رقم Chinese Spring است، از جمله لوکوس Fhb1 (مقاومت به فوزاریوم)، Ppd-D1a (گلدهی زودرس)، Glu-D1f (کیفیت نودل آسیایی) و Rht-D1b (نیمه پاکوتاهی) / این مطالعه اولین سرهم‌بندی با کیفیت از یک رقم گندم شرق آسیا را فراهم می‌کند و می‌تواند به عنوان ژنوم مرجع برای مطالعات عملکردی و تکاملی استفاده شود تا تنوع گندم نان آسیایی که کمتر مورد بررسی قرار گرفته است را به‌طور جامع مشخص کند.</p>	(Shimizu et al., 2021)
جو (<i>Hordeum vulgare</i>)	de novo transcriptome assembly	<p>بررسی پاسخ‌های مختلف ژنتیکی به تنش شوری در ریشه دو ژنوتیپ جو با استفاده از تجزیه و تحلیل ترنسکرپتوم / از روش توالی‌یابی RNA و سرهم‌بندی de novo ترنسکرپتوم برای ایجاد دو مجموعه مرجع ترنسکرپتومی جدید برای یک توده بومی جو (Sahara) و یک رقم مالتینگ (Clipper) با پاسخ‌های متفاوت به شوری استفاده شد/ از مدل‌های خطی تعمیم‌یافته برای ارزیابی آماری پاسخ‌های مکانی، مرتبط با تیمار و اختصاصی ژنوتیپ استفاده شد.</p>	<p>شناسایی یک شیب بیان ژنی در طول ریشه جو/ تشخیص رونوشت‌های با بیان متفاوت بیشتر بین مناطق مختلف ریشه نسبت به بین تیمارها / مشاهده یک گذار تدریجی از رونوشت‌های مرتبط با سیگنالینگ قندی در منطقه مریستمی ریشه به رونوشت‌های درگیر در متابولیسم دیواره سلولی در منطقه طویل شدن، و مسیرهای مرتبط با پاسخ دفاعی به سمت منطقه بلوغ / شناسایی تفاوت‌های قابل توجه بین دو ژنوتیپ در پاسخ به تنش شوری/ این مطالعه دو مجموعه مرجع ترنسکرپتومی جدید برای جو ارائه داد که می‌تواند به عنوان منبعی ارزشمند برای جامعه تحقیقاتی غلات مورد استفاده قرار گیرد و ممکن است صفات ارزشمندی را برای کمک به برنامه‌های اصلاحی شناسایی کند.</p>	(Hill et al., 2016)
برنج (<i>Oryza sativa</i>)	Targeted sequencing (TGS)	<p>شناسایی هدفمند تنوع آلی ژن‌های مرتبط با دانه برنج از طریق القای جهش، توالی‌یابی هدفمند و توالی‌یابی کل ژنوم همراه با استراتژی نمونه‌های</p>	<p>نتایج نشان داد که تنوع گسترده‌ای در شکل دانه وجود دارد، با طول دانه بین ۸ تا ۱۰/۲۲ میلی‌متر و عرض دانه بین ۱/۵۴ تا ۲/۸۷ میلی‌متر. با استفاده از توالی‌یابی هدفمند ۴۸۴ نمونه مخلوط (۸:۱)، ۱۲ نمونه مخلوط حاوی جهش‌های بالقوه و ۱۵ SNP شناسایی شد/ پس از تحلیل ساختار و عملکرد پروتئین، ۶ واریانت آلی منجر به</p>	(Sun et al., 2022)

		مخلوط / محققان از روش‌های پرتوتابی یون سنگین برای ایجاد جهش، توالی‌یابی هدفمند برای شناسایی تنوع آلی در ژن‌های GS3 و GW5، توالی‌یابی سنگر برای تأیید جهش‌ها، و توالی‌یابی کل ژنوم برای شناسایی واریانت‌های آلی جدید استفاده شد / در مجموع ۳۸۷۲ بذر و برگ نسل M2 برنج جهش‌یافته مورد بررسی قرار گرفت	تغییر شکل دانه و جهش‌یافته‌های مربوطه شناسایی شدند، از جمله جهش‌یافته‌های دانه بلند (GS3-2 و GS3-7)، دانه کوتاه (GS3-3 و GS3-5)، دانه پهن (GW5-1) و دانه باریک (GW5-4) همچنین، توالی‌یابی کل ژنوم منجر به شناسایی واریانت‌های آلی جدید ژن طول دانه GS3-G1، GS3-G2 و GS3-G3 شد / این مطالعه در مجموع شش جهش‌یافته شکل دانه و ۹ واریانت آلی مرتبط با شکل دانه را شناسایی کرد که می‌تواند برای اصلاح مولکولی و تحلیل ارتباط ژنوتیپ-فنوتیپ مورد استفاده قرار گیرد. روش ارائه شده در این مطالعه برای شناسایی هدفمند تنوع آلی ژن‌های شکل دانه برنج، دارای مزایایی چون هزینه پایین، دوره کوتاه و انعطاف‌پذیری در بررسی صفات و ژن‌های مختلف است	
گندم (<i>Triticum aestivum</i> L.)	chromatin immunoprecipitation sequencing (ChIP-seq)	بررسی شبکه تنظیمی فاکتور رونویسی PnPf2 در کنترل بیماری‌زایی قارچ <i>Parastagonospora nodorum</i> در گندم / از روش‌های ایمونوپرسیپیتاسیون کروماتین (ChIP-seq) و موتاژنز برای شناسایی اهداف ژنومی PnPf2 و سایر تنظیم‌کننده‌های درگیر استفاده شد	نتایج نشان داد که دو موتیف اتصال متمایز مرتبط با تنظیم مثبت ژن توسط PnPf2 شناسایی شد. ژن‌های هدف مستقیم PnPf2 شامل ژن‌های کدکننده افکتورهای اصلی و سایر اجزای مرتبط با سبک زندگی بیماری‌زای <i>P. nodorum</i> مانند آنزیم‌های فعال کربوهیدراتی و جذب‌کننده‌های مواد مغذی بودند. این یافته‌ها نقش مستقیم PnPf2 در هماهنگ‌سازی بیماری‌زایی در گندم را تأیید می‌کند. فاکتورهای رونویسی دیگر نیز از اهداف برجسته PnPf2 بودند که نشان می‌دهد PnPf2 در یک شبکه رونویسی عمل می‌کند / چندین فاکتور رونویسی از جمله PnPro1، PnAda1، PnEbr1 و PnCreA مورد بررسی عملکردی قرار گرفتند و نقش‌های متمایز متابولیکی و تکاملی آنها مشخص شد / در مجموع، این نتایج PnPf2 را به عنوان تنظیم‌کننده اصلی رونویسی ژن‌های مؤثر در بیماری‌زایی گندم تأیید می‌کند و بینش مکانیستیکی در مورد نحوه عملکرد آن فراهم می‌کند.	(John ET AL., 2022)
برنج (<i>Oryza sativa</i>)	Epigenomics by DNA Methylation Sequencing (Methyl-seq)	بررسی نقش متیلاسیون DNA در پاسخ به تنش‌های خشکی و شوری در سه رقم برنج IR64 حساس به تنش، Nagina 22 متحمل به خشکی و Pokkali متحمل به شوری انجام شد / از روش توالی‌یابی بیسولفیت برای تجزیه و تحلیل الگوهای متیلاسیون DNA استفاده شد	نتایج نشان داد که متیلاسیون در بافت CG درون ژن و متیلاسیون در بافت CHH در مناطق پرموتر دیستال با بیان ژن همبستگی مثبت دارد / هیپومتیلاسیون در Nagina 22 در پاسخ به تنش خشکی و هیپرمتیلاسیون در Pokkali در پاسخ به تنش شوری با بیان بالاتر برخی ژن‌های پاسخ به تنش غیرزیستی همبستگی داشت / اکثر ژن‌های با متیلاسیون و بیان متفاوت ^{۱۵۸} (DMR-DEGs) اختصاصی رقم بودند که نشان‌دهنده نقش مهم متیلاسیون DNA در پاسخ‌های تنش غیرزیستی در برنج به صورت اختصاصی رقم است / همچنین، DMR-DEG‌هایی که دارای سیتوزین‌های متیله شده متفاوت به دلیل پلی‌مورفیسم‌های DNA بین ارقام حساس و متحمل در مناطق پرموتر و/یا کدکننده خود بودند، شناسایی شدند که نقش اپی‌آل‌ها را در پاسخ‌های تنش غیرزیستی پیشنهاد می‌کند / این مطالعه بینش جدیدی در مورد نقش تنظیمی متیلاسیون DNA در پاسخ به تنش‌های غیرزیستی در برنج فراهم می‌کند و می‌تواند به توسعه استراتژی‌های اصلاحی برای بهبود تحمل به تنش کمک	(Rajku mar et al., 2020)
ذرت (<i>Zea mays</i>)	Whole metagenome sequencing (WMS)	بررسی تأثیر کاربرد کودهای مختلف بر تنوع عملکردی میکروبیوم‌های اندوفیت ریشه ذرت انجام شد / از روش متاژنومیکس شاتگان برای تجزیه و تحلیل کل DNA	نتایج نشان داد که تفاوت معنی‌داری در فراوانی نسبی ۲۸ گروه عملکردی شناسایی شده در میکروبیوم‌های اندوفیت بین سایت‌ها وجود ندارد / گروه‌های عملکردی و مسیرهای متابولیکی مرتبط با ترویج رشد گیاه مانند متابولیسم کربوهیدرات، متابولیسم ثانویه، متابولیسم نیتروژن، جذب و متابولیسم آهن و متابولیسم فسفر در اندوفیت‌ها مشاهده شدند / مطالعه تنوع آلفا تفاوت معنی‌داری بین	(Fadiji et al., 2021)

	<p>استخراج شده از ریشه‌های ذرت کشت شده در سه شرایط مختلف کوددهی: کود آلی FK، کود غیرآلی NK و بدون کود CK استفاده شد/ داده‌های توالی‌یابی با استفاده از پایگاه داده آنلاین MG-RAST تحلیل شدند و پیش‌بینی عملکرد میکروبیوم‌های اندوفیت با استفاده از سامانه SEED انجام شد</p>	<p>گروه‌های عملکردی اندوفیت‌ها نشان نداد / اما مطالعه تنوع بتا تفاوت معنی‌داری را بین سایت‌های کوددهی نشان داد. فراوانی بالای گروه‌های عملکردی، به ویژه در نمونه‌های FK، نشان می‌دهد که شیوه‌های مختلف کشاورزی بر عملکرد میکروبیوم‌های اندوفیت تأثیر می‌گذارند / این مطالعه پیشنهاد می‌کند که تحقیقات بیشتری برای بررسی ژن‌های عملکردی در میکروبیوم‌های اندوفیت با هدف ارزیابی مفید بودن آنها در ترویج کشاورزی پایدار انجام شود.</p>
--	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

ارزیابی کیفیت و ایمنی مواد غذایی کاربرد دارد (Staats *et al.*, 2016). در بخش مدیریت آفات و بیماری‌های گیاهی، NGS در شناسایی سریع و دقیق عوامل بیماری‌زا، مطالعه سازوکارهای مقاومت گیاهان و توسعه استراتژی‌های کنترل آفات مؤثر است (Maree *et al.*, 2018). در زمینه تحقیقات ژنومیکس عملکردی، NGS امکان مطالعه جامع بیان ژن (RNA-seq) در گیاهان و دام‌ها را فراهم کرده که به درک بهتر سازوکارهای مولکولی کنترل‌کننده صفات مهم اقتصادی کمک می‌کند (Wang *et al.*, 2009). این فناوری همچنین در مطالعات اپی‌ژنومیکس، از جمله متیلاسیون DNA و تغییرات هیستونی، که نقش مهمی در تنظیم ژن و سازگاری به تنش‌های محیطی دارند، کاربرد دارد (Zhang *et al.*, 2018). در زمینه حفاظت از تنوع زیستی کشاورزی، NGS در مطالعات فیلوژنتیک و فیلوژئوگرافی گونه‌های زراعی و وحشی، که برای درک تکامل و تنوع ژنتیکی آنها ضروری است، استفاده می‌شود (Allendorf *et al.*, 2010). در بخش توسعه واکسن‌های دامی، NGS در شناسایی آنتی‌ژن‌های مناسب و طراحی

در بخش دامپروری، NGS نقش مهمی در بهبود دقت ارزیابی ژنومی دام‌ها داشته است. این فناوری امکان شناسایی SNPها و نشانگرهای مولکولی مرتبط با صفات اقتصادی مهم مانند تولید شیر، کیفیت گوشت و مقاومت به بیماری‌ها را فراهم کرده است (Georges *et al.*, 2019). NGS همچنین در مطالعات اپی‌ژنومیکس دام‌ها، که به درک بهتر تنظیم ژن و سازگاری به شرایط محیطی کمک می‌کند، کاربرد دارد (Ibeagha-Awemu & Zhao, 2015). در زمینه سلامت دام، NGS در تشخیص سریع و دقیق بیماری‌های عفونی، شناسایی پاتوژن‌های جدید و مطالعه اپیدمیولوژی مولکولی بیماری‌های دامی نقش مهمی ایفا می‌کند (Mwangi *et al.*, 2019). هر دو بخش گیاهی و دامی، NGS در مطالعات متاژنومیکس برای بررسی جوامع میکروبی مرتبط با گیاهان و دام‌ها کاربرد دارد. این مطالعات به درک بهتر نقش میکروبیوم در سلامت، تغذیه و عملکرد گیاهان و دام‌ها کمک می‌کند (Buermans & den Dunnen, 2014). در زمینه امنیت غذایی، NGS در ردیابی منشأ محصولات کشاورزی، شناسایی تقلبات در مواد غذایی و

آزمایش می‌شود. این مزایا، NGS را به روشی کارآمد و مقرون به صرفه برای مطالعات گسترده ژنومی تبدیل کرده است. با وجود این مزایا، NGS با چالش‌های قابل توجهی نیز روبرو است. یکی از بزرگترین چالش‌ها، حجم عظیم داده‌های تولید شده و نیاز به تجزیه و تحلیل پیچیده این داده‌هاست. اگرچه در سطح تحقیقاتی راهکارهایی برای این مسئله وجود دارد، اما در سطح بالینی هنوز نیاز به توسعه روش‌های سریع‌تر و کارآمدتر برای پردازش و تفسیر داده‌ها احساس می‌شود. این امر به خصوص در ترجمه نتایج توالی‌یابی (جفت بازهای توالی) به اطلاعات قابل استفاده در تشخیص بالینی و درمان بیماری‌ها اهمیت دارد. چالش دیگر، کاستی‌های زیرساخت‌های محاسباتی است. حجم داده‌های تولید شده توسط سامانه‌های NGS از ظرفیت پردازشی بسیاری از سامانه‌های کامپیوتری موجود فراتر رفته است. این مسئله نیاز به سرمایه‌گذاری در زیرساخت‌های محاسباتی قوی‌تر و توسعه الگوریتم‌های کارآمدتر برای پردازش داده‌ها را برجسته می‌کند. علاوه بر چالش‌های فنی، مسائل حقوقی و اخلاقی مرتبط با NGS نیز بسیار پیچیده و بحث‌برانگیز است. پیامدهای قانونی و اخلاقی تعیین توالی ژنوم انسان گسترده بوده و هنوز به‌طور کامل مورد بررسی قرار نگرفته است. به عنوان مثال، اشتراک‌گذاری اطلاعات ژنتیکی افراد با کارفرمایان یا شرکت‌های بیمه می‌تواند منجر به تبعیض ژنتیکی شود. این مسئله می‌تواند پیامدهای قانونی جدی داشته باشد و نیاز به تدوین قوانین و مقررات جدید برای حفاظت از حریم خصوصی ژنتیکی افراد را برجسته

واکسن‌های مؤثرتر نقش دارد (Dellagostin *et al.*, 2017). علاوه بر این، NGS در مطالعات مربوط به تغییرات اقلیمی و تأثیر آن بر کشاورزی، از جمله شناسایی ژن‌های مرتبط با تحمل به تنش‌های محیطی در گیاهان و دام‌ها، کاربرد دارد (Varshney *et al.*, 2018). با پیشرفت فناوری‌های NGS و کاهش هزینه‌های آن، انتظار می‌رود که کاربردهای این فناوری در کشاورزی گسترش یابد و به توسعه روش‌های نوین در تولید محصولات کشاورزی پایدار و سازگار با تغییرات اقلیمی کمک کند.

نتیجه‌گیری و بیان مزایا و چالش‌های پیشروی

توالی‌یابی نسل جدید (NGS)

توالی‌یابی نسل جدید (NGS) علاوه بر کاربردهای گسترده‌ای که دارد، با مزایا و چالش‌های قابل توجهی همراه است. از جمله مزایای برجسته این فناوری می‌توان به قابلیت پوشش‌دهی عمیق (تا حد 99x) اشاره کرد که دقت بالایی در شناسایی واریانت‌های ژنتیکی فراهم می‌کند. همچنین، NGS با نیاز به مقدار بسیار کمی از نمونه ژنومی (حدود ۵۰ نانوگرم)، امکان بررسی نمونه‌های محدود را میسر می‌سازد. توالی‌یابی تک مرحله‌ای و مقیاس‌پذیری بالا از دیگر ویژگی‌های مهم این روش است که امکان بررسی همزمان تعداد زیادی نمونه را فراهم می‌کند. قابلیت کمی‌سازی RNA با کیفیت بسیار بالا، NGS را به ابزاری قدرتمند برای مطالعات ترنسکریپتومیکس تبدیل کرده است. علاوه بر این، پروتکل‌های آماده‌سازی نمونه در GS سریع‌تر و مستقیم‌تر شده‌اند که منجر به کاهش چشمگیر زمان و هزینه‌های

می‌کند. در مجموع، در حالی که NGS فرصت‌های بی‌نظیری برای پیشرفت در زمینه‌های مختلف علوم زیستی و پزشکی فراهم کرده است، مواجهه با چالش‌های فنی، اخلاقی و قانونی آن نیازمند همکاری گسترده میان متخصصان حوزه‌های مختلف از جمله زیست‌شناسان، متخصصان انفورماتیک، حقوقدانان و سیاست‌گذاران است تا بتوان از پتانسیل کامل این فناوری بهره برد و در عین حال، از حقوق و منافع افراد و جامعه محافظت کرد.

References

- Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., Rodriguez-Mueller, B., Zucker, J., Thiagarajan, M., Henriessat, B., White, O., Kelley, S. T., Methé, B., Schloss, P. D., Gevers, D., Mitreva, M., & Huttenhower, C. 2012. Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Computational Biology*, 8(6), e1002358. <https://doi.org/10.1371/journal.pcbi.1002358>
- Adessi, C. 2000. Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Research*, 28(20), 87e-887. <https://doi.org/10.1093/nar/28.20.e87>
- Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. 2015. Predicting effective microRNA target sites in mammalian mRNAs. *ELife*, 4. CLOCKSS. <https://doi.org/10.7554/elife.05005>
- Ahmadian, A., Ehn, M., & Hober, S. 2006. Pyrosequencing: history, biochemistry and future. *Clinica Chimica Acta*, 363(1-2), 83-94. <https://doi.org/10.1016/j.cccn.2005.04.038>
- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. 2012. MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, 13(10), R87. <https://doi.org/10.1186/gb-2012-13-10-r87>
- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. 2010. Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11(10), 697-709. <https://doi.org/10.1038/nrg2844>
- Alles, J., Karaiskos, N., Praktiknjo, S. D., Grosswendt, S., Wahle, P., Ruffault, P.-L., Ayoub, S., Schreyer, L., Boltengagen, A., Birchmeier, C., Zinzen, R., Kocks, C., & Rajewsky, N. 2017. Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biology*, 15(1). <https://doi.org/10.1186/s12915-017-0383-5>
- Alves, L. de F., Westmann, C. A., Lovate, G. L., de Siqueira, G. M. V., Borelli, T. C., & Guazzaroni, M.-E. 2018. Metagenomic approaches for understanding new concepts in microbial science. *International Journal of Genomics*, 2018, 1-15. <https://doi.org/10.1155/2018/2312987>
- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biology*, 21(1). <https://doi.org/10.1186/s13059-020-1935-5>
- Ambardar, S., Gupta, R., Trakroo, D., Lal, R., & Vakhlu, J. 2016. High throughput sequencing: an overview of sequencing chemistry. *Indian Journal of Microbiology*, 56(4), 394-404. <https://doi.org/10.1007/s12088-016-0606-4>
- Anaparthi, N., Ho, Y.-J., Martelotto, L., Hammell, M., & Hicks, J. 2019. Single-cell applications of next-generation sequencing. *Cold Spring Harbor Perspectives in Medicine*, 9(10), a026898. <https://doi.org/10.1101/cshperspect.a026898>
- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. <https://doi.org/10.7554/elife.33070.019>

- Angelini, C., & Costa, V. 2014. Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. *Frontiers in Cell and Developmental Biology*, 2. <https://doi.org/10.3389/fcell.2014.00051>
- Angermueller, C., Lee, H. J., Reik, W., & Stegle, O. 2017. DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning. *Genome Biology*, 18(1). <https://doi.org/10.1186/s13059-017-1189-z>
- Ansorge, W. J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology*, 25(4), 195–203. <https://doi.org/10.1016/j.nbt.2008.12.009>
- Ansorge, W., Zimmermann, J., Erfze, H., Hewitt, N., Rupp, T., Schwager, C., ... & Voss, H. 1993. Sequencing reactions for ALF (EMBL) automated DNA sequencer. *DNA sequencing protocols*, 317-356. <https://doi.org/10.1385/0-89603-248-5:317>
- Ardui, S., Ameer, A., Vermeesch, J. R., & Hestand, M. S. 2018. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. *Nucleic Acids Research*, 46(5), 2159–2168. <https://doi.org/10.1093/nar/gky066>
- Assenov, Y., Müller, F., Lutsik, P., Walter, J., Lengauer, T., & Bock, C. 2014. Comprehensive analysis of DNA methylation data with RnBeads. *Nature Methods*, 11(11), 1138-1140. <https://doi.org/10.1038/nmeth.3115>
- Auer, H., Newsom, D. L., & Kornacker, K. 2009. Expression profiling using affymetrix genechip microarrays. *microchip methods in diagnostics*, 35–46. https://doi.org/10.1007/978-1-59745-372-1_3
- Avery, O. T., Macleod, C. M., & McCarty, M. 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *The Journal of Experimental Medicine*, 79(2), 137–158. <https://doi.org/10.1084/jem.79.2.137>
- Baccarelli, A., & Bollati, V. 2009. Epigenetics and environmental chemicals. *Current Opinion in Pediatrics*, 21(2), 243–251. <https://doi.org/10.1097/mop.0b013e32832925cc>
- Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., ... & Langlade, N. B. 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, 546(7656), 148-152. <https://doi.org/10.1038/nature22380>
- Bailey, T., Krajewski, P., Ladunga, I., Lefebvre, C., Li, Q., Liu, T., Madrigal, P., Taslim, C., & Zhang, J. 2013. Practical guidelines for the comprehensive analysis of ChIP-seq data. *PLoS Computational Biology*, 9(11), e1003326. <https://doi.org/10.1371/journal.pcbi.1003326>
- Bao, R., Huang, L., Andrade, J., Tan, W., Kibbe, W. A., Jiang, H., & Feng, G. 2014. Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer Informatics*, 13s2, CIN.S13779. <https://doi.org/10.4137/cin.s13779>
- Ballester, L. Y., Luthra, R., Kanagal-Shamanna, R., & Singh, R. R. 2016. Advances in clinical next-generation sequencing: target enrichment and sequencing technologies. *Expert Review of Molecular Diagnostics*, 16(3), 357–372. <https://doi.org/10.1586/14737159.2016.1133298>
- Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., & Shendure, J. 2011. Exome sequencing as a tool for mendelian disease gene discovery. *Nature Reviews Genetics*, 12(11), 745–755. <https://doi.org/10.1038/nrg3031>
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Bartel, D. P. 2018. Metazoan microRNAs. *Cell*, 173(1), 20–51. <https://doi.org/10.1016/j.cell.2018.03.006>

- Baylin, S. B., & Jones, P. A. 2016. Epigenetic determinants of cancer. *Cold Spring Harbor Perspectives in Biology*, 8(9), a019505. <https://doi.org/10.1101/cshperspect.a019505>
- Becker, C., Hammerle-Fickinger, A., Riedmaier, I., & Pfaffl, M. W. 2010. mRNA and microRNA quality control for RT-qPCR analysis. *Methods*, 50(4), 237–243. <https://doi.org/10.1016/j.ymeth.2010.01.010>
- Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *ELife*, 10. CLOCKSS. <https://doi.org/10.7554/elife.65088>
- Behjati, S., & Tarpey, P. S. 2013. What is next generation sequencing? *Archives of Disease in Childhood - Education & Practice Edition*, 98(6), 236–238. <https://doi.org/10.1136/archdischild-2013-304340>
- Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., Irving, L. J., ... Smith, A. J. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*, 456(7218), 53–59. <https://doi.org/10.1038/nature07517>
- Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M. A., Beaudet, A. L., Ecker, J. R., Farnham, P. J., Hirst, M., Lander, E. S., Mikkelsen, T. S., & Thomson, J. A. 2010. The NIH roadmap epigenomics mapping consortium. *Nature Biotechnology*, 28(10), 1045–1048. <https://doi.org/10.1038/nbt1010-1045>
- Bizouarn, F. 2014. Clinical applications using digital PCR. *Quantitative Real-Time PCR*, 189–214. https://doi.org/10.1007/978-1-4939-0733-5_16
- Bock, C. 2012. Analysing and interpreting DNA methylation data. *Nature Reviews Genetics*, 13(10), 705–719. <https://doi.org/10.1038/nrg3273>
- Bolger, A. M., Lohse, M., & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Børsting, C., & Morling, N. 2015. Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics*, 18, 78–89. <https://doi.org/10.1016/j.fsigen.2015.02.002>
- Bowers, J., Mitchell, J., Beer, E., Buzby, P. R., Causey, M., Efcavitch, J. W., Jarosz, M., Krzymanska-Olejnik, E., Kung, L., Lipson, D., Lowman, G. M., Marappan, S., McInerney, P., Platt, A., Roy, A., Siddiqi, S. M., Steinmann, K., & Thompson, J. F. 2009. Virtual terminator nucleotides for next-generation DNA sequencing. *Nature Methods*, 6(8), 593–595. <https://doi.org/10.1038/nmeth.1354>
- Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. V., Ramsey, J. M., ... Schloss, J. A. 2008. The potential and challenges of nanopore sequencing. *Nature Biotechnology*, 26(10), 1146–1153. <https://doi.org/10.1038/nbt.1495>
- Braslavsky, I., Hebert, B., Kartalov, E., & Quake, S. R. 2003. Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences*, 100(7), 3960–3964. <https://doi.org/10.1073/pnas.0230489100>
- Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L. A., D'Amore, R., Allen, A. M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M.-C., Sehgal, S., Gill, B., ... Hall, N. 2012. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*, 491(7426), 705–710. <https://doi.org/10.1038/nature11650>

- Brito, I. L., Yilmaz, S., Huang, K., Xu, L., Jupiter, S. D., Jenkins, A. P., Naisilisili, W., Tamminen, M., Smillie, C. S., Wortman, J. R., Birren, B. W., Xavier, R. J., Blainey, P. C., Singh, A. K., Gevers, D., & Alm, E. J. 2016. Mobile genes in the human microbiome are structured from global to individual scales. *Nature*, 535(7612), 435–439. <https://doi.org/10.1038/nature18927>
- Broad Institute. n.d. Picard tools. <http://broadinstitute.github.io/picard/>
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G., Bateman, J., Bryant, S., Guzikowski, A. R., Tsai, S. L., Coyne, S., Ye, W. W., Freeman, R. M., Peshkin, L., Tabin, C. J., ... Whited, J. L. 2017. A tissue-mapped axolotl De novo transcriptome enables identification of limb regeneration factors. *Cell Reports*, 18(3), 762–776. <https://doi.org/10.1016/j.celrep.2016.12.063>
- Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 109(1). Portico. <https://doi.org/10.1002/0471142727.mb2129s109>
- Buermans, H. P. J., & den Dunnen, J. T. 2014. Next generation sequencing technology: Advances and applications. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1842(10), 1932–1941. <https://doi.org/10.1016/j.bbadis.2014.06.015>
- Burbano, H. A., Hodges, E., Green, R. E., Briggs, A. W., Krause, J., Meyer, M., Good, J. M., Maricic, T., Johnson, P. L. F., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F. W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G. J., & Pääbo, S. 2010. Targeted investigation of the neandertal genome by array-based sequence capture. *Science*, 328(5979), 723–725. <https://doi.org/10.1126/science.1188046>
- Caldwell, C. C., & Spies, M. 2017. Helicase SPRNTing through the nanopore. *Proceedings of the National Academy of Sciences*, 114(45), 11809–11811. <https://doi.org/10.1073/pnas.1716866114>
- ČALOUNOVÁ, Tereza. 2021. De novo transcriptomics and its use in non-model organisms. Bakalářská práce, vedoucí Pluskal, Tomáš. Praha: Univerzita Karlova, Přírodovědecká fakulta, Katedra buněčné biologie. <http://hdl.handle.net/20.500.11956/149281>
- Casbon, J. A., Osborne, R. J., Brenner, S., & Lichtenstein, C. P. 2011. A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic Acids Research*, 39(12), e81–e81. <https://doi.org/10.1093/nar/gkr217>
- Castro-Wallace, S. L., Chiu, C. Y., John, K. K., Stahl, S. E., Rubins, K. H., McIntyre, A. B. R., Dworkin, J. P., Lupisella, M. L., Smith, D. J., Botkin, D. J., Stephenson, T. A., Juul, S., Turner, D. J., Izquierdo, F., Federman, S., Stryke, D., Somasekar, S., Alexander, N., Yu, G., ... Burton, A. S. 2017. Nanopore DNA sequencing and genome assembly on the international space station. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-18364-0>
- Cavalli, G., & Heard, E. 2019. Advances in epigenetics link genetics to the environment and disease. *Nature*, 571(7766), 489–499. <https://doi.org/10.1038/s41586-019-1411-0>
- Cech, T. R., & Steitz, J. A. 2014. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell*, 157(1), 77-94.
- Chaisson, M. J. P., Huddleston, J., Dennis, M. Y., Sudmant, P. H., Malig, M., Hormozdiari, F., Antonacci, F., Surti, U., Sandstrom, R., Boitano, M., Landolin, J. M., Stamatoyannopoulos, J. A., Hunkapiller, M. W., Korlach, J., & Eichler, E. E. 2014. Resolving the complexity of the human genome using single-molecule sequencing. *Nature*, 517(7536), 608–611. <https://doi.org/10.1038/nature13907>
- Chandran, A. K. N., Kim, J.-W., Yoo, Y.-H., Park, H. L., Kim, Y.-J., Cho, M.-H., & Jung, K.-H. 2019. Transcriptome analysis of rice-seedling roots under soil-salt stress using RNA-Seq method. *Plant Biotechnology Reports*, 13(6), 567–578. <https://doi.org/10.1007/s11816-019-00550-3>
- Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L.,

- & Mardis, E. R. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature Methods*, 6(9), 677–681. <https://doi.org/10.1038/nmeth.1363>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chen, Y., Li, Y., Narayan, R., Subramanian, A., & Xie, X. 2016. Gene expression inference with deep learning. *Bioinformatics*, 32(12), 1832–1839. <https://doi.org/10.1093/bioinformatics/btw074>
- Chiara, M., Horner, D. S., Gissi, C., & Pesole, G. 2021. Comparative genomics reveals early emergence and biased spatiotemporal distribution of SARS-CoV-2. *Molecular Biology and Evolution*, 38(6), 2547–2565. <https://doi.org/10.1093/molbev/msab049>
- Choi, M., Scholl, U. I., Ji, W., Liu, T., Tikhonova, I. R., Zumbo, P., Nayir, A., Bakkaloğlu, A., Özen, S., Sanjad, S., Nelson-Williams, C., Farhi, A., Mane, S., & Lifton, R. P. 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences*, 106(45), 19096–19101. <https://doi.org/10.1073/pnas.0910672106>
- Chomczynski, P., & Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, 162(1), 156-159. [https://doi.org/10.1016/0003-2697\(87\)90021-2](https://doi.org/10.1016/0003-2697(87)90021-2)
- Church, G. M., & Gilbert, W. 1984. Genomic sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 81(7), 1991–1995. <https://doi.org/10.1073/pnas.81.7.1991>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... & Ruden, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80-92. <https://doi.org/10.4161/fly.19695>
- Clark, T. A., Murray, I. A., Morgan, R. D., Kislyuk, A. O., Spittle, K. E., Boitano, M., ... & Korlach, J. 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Research*, 40(4), e29-e29. <https://doi.org/10.1093/nar/gkr1146>
- Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. 2010. The sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, 38(6), 1767-1771. <https://doi.org/10.1093/nar/gkp1137>
- Conesa, A., & Götz, S. 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *International Journal of Plant Genomics*, 2008, 619832. <https://doi.org/10.1155/2008/619832>
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21(18), 3674-3676. <https://doi.org/10.1093/bioinformatics/bti610>
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., ... & Mortazavi, A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biology*, 17(1), 13. <https://doi.org/10.1186/s13059-016-0881-8>
- Crick, F. 1970. Central dogma of molecular biology. *Nature*, 227(5258), 561-563. <https://doi.org/10.1038/227561a0>
- De Filippis, F., Parente, E., & Ercolini, D. 2017. Metagenomics insights into food fermentations. *Microbial Biotechnology*, 10(1), 91-102. <https://doi.org/10.1111/1751-7915.12421>
- Deamer, D., Akeson, M., & Branton, D. 2016. Three decades of nanopore sequencing. *Nature Biotechnology*, 34(5), 518-524. <https://doi.org/10.1038/nbt.3423>
- Dellagostin, O. A., Grassmann, A. A., Rizzi, C., Schuch, R. A., Jorge, S., Oliveira, T. L., ... & Hartwig, D. D. 2017. Reverse vaccinology: an approach for identifying leptospiral vaccine

- candidates. *International Journal of Molecular Sciences*, 18(1), 158. <https://doi.org/10.3390/ijms18010158>
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... & Daly, M. J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491-498. <https://doi.org/10.1038/ng.806>
- Desjardins, P., & Conklin, D. 2010. NanoDrop microvolume quantitation of nucleic acids. *JoVE (Journal of Visualized Experiments)*, (45), e2565. <https://doi.org/10.3791/2565>
- Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. 2017. Nextflow enables reproducible computational workflows. *Nature Biotechnology*, 35(4), 316-319. <https://doi.org/10.1038/nbt.3820>
- Dirks, R. A. M., Stunnenberg, H. G., & Marks, H. 2016. Genome-wide epigenomic profiling for biomarker discovery. *Clinical Epigenetics*, 8(1). <https://doi.org/10.1186/s13148-016-0284-4>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... & Gingeras, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15-21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dressman, D., Yan, H., Traverso, G., Kinzler, K. W., & Vogelstein, B. 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences*, 100(15), 8817-8822. <https://doi.org/10.1073/pnas.1133470100>
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., ... & Turner, S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science*, 323(5910), 133-138. <https://doi.org/10.1126/science.1162986>
- Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. 2019. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends in Microbiology*, 27(2), 105-117. <https://doi.org/10.1016/j.tim.2018.11.003>
- Eklom, R., & Wolf, J. B. 2014. A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, 7(9), 1026-1042. <https://doi.org/10.1111/eva.12178>
- Eldem, V., Zararsiz, G., Taşçi, T., Duru, I. P., Bakir, Y., & Erkan, M. 2017. Transcriptome analysis for non-model organism: current status and best-practices. *Applications of RNA-Seq and omics strategies - from microorganisms to human health*. <https://doi.org/10.5772/intechopen.68983>
- Escalante, A. E., Jardón Barbolla, L., Ramírez-Barahona, S., & Eguiarte, L. E. 2014. The study of biodiversity in the era of massive sequencing. *Revista Mexicana de Biodiversidad*, 85(4), 1249-1264. <https://doi.org/10.7550/rmb.43498>
- Escobar-Zepeda, A., Vera-Ponce de León, A., & Sanchez-Flores, A. 2015. The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics*, 6, 348. <https://doi.org/10.3389/fgene.2015.00348>
- Esposito, S., D'Agostino, N., Taranto, F., Sonnante, G., Sestili, F., Lafiandra, D., & De Vita, P. 2022. Whole-exome sequencing of selected bread wheat recombinant inbred lines as a useful resource for allele mining and bulked segregant analysis. *Frontiers in Genetics*, 13, 1058471. <https://doi.org/10.3389/fgene.2022.1058471>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047-3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Fuentes-Pardo, A. P., & Ruzzante, D. E. 2017. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, 26(20), 5369-5406. <https://doi.org/10.1111/mec.14264>

- Fadji, A. E., Ayangbenro, A. S., & Babalola, O. O. 2021. Shotgun metagenomics reveals the functional diversity of root-associated endophytic microbiomes in maize plant. *Current Plant Biology*, 25, 100195. <https://doi.org/10.1016/j.cpb.2021.100195>
- Farlik, M., Sheffield, N. C., Nuzzo, A., Datlinger, P., Schönegger, A., Klughammer, J., & Bock, C. 2015. Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Reports*, 10(8), 1386-1397. <https://doi.org/10.1016/j.celrep.2015.02.001>
- Fedurco, M. 2006. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Research*, 34(3), e22-e22. <https://doi.org/10.1093/nar/gnj023>
- Feng, H., Conneely, K. N., & Wu, H. 2014. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Research*, 42(8), e69-e69. <https://doi.org/10.1093/nar/gku154>
- Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., ... & Turner, S. W. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*, 7(6), 461-465. <https://doi.org/10.1038/nmeth.1459>
- Fragouli, E., Alfarawati, S., Spath, K., Babariya, D., Tarozzi, N., Borini, A., & Wells, D. 2017. Analysis of implantation and ongoing pregnancy rates following the transfer of mosaic diploid-aneuploid blastocysts. *Human Genetics*, 136(7), 805-819. <https://doi.org/10.1007/s00439-017-1797-4>
- Frampton, G. M., Fichtenholtz, A., Otto, G. A., Wang, K., Downing, S. R., He, J., ... & Yelensky, R. 2013. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nature Biotechnology*, 31(11), 1023-1031. <https://doi.org/10.1038/nbt.2696>
- Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., ... & Huttenhower, C. 2018. Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, 15(11), 962-968. <https://doi.org/10.1038/s41592-018-0176-y>
- Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. 2012. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Research*, 40(1), 37-52. <https://doi.org/10.1093/nar/gkr688>
- Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., ... & Paul, C. L. 1992. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences*, 89(5), 1827-1831. <https://doi.org/10.1073/pnas.89.5.1827>
- Furey, T. S. 2012. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nature Reviews Genetics*, 13(12), 840-852. <https://doi.org/10.1038/nrg3306>
- Garalde, D. R., Snell, E. A., Jachimowicz, D., Sipos, B., Lloyd, J. H., Bruce, M., ... & Turner, D. J. 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, 15(3), 201-206. <https://doi.org/10.1038/nmeth.4577>
- Gardy, J. L., & Loman, N. J. 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nature Reviews Genetics*, 19(1), 9-20. <https://doi.org/10.1038/nrg.2017.88>
- Gargis, A. S., Kalman, L., Bick, D. P., Da Silva, C., Dimmock, D. P., Funke, B. H., ... & Lubin, I. M. 2015. Good laboratory practice for clinical next-generation sequencing informatics pipelines. *Nature Biotechnology*, 33(7), 689-693. <https://doi.org/10.1038/nbt.3237>
- Garrison, E., & Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907. <https://doi.org/10.48550/arXiv.1207.3907>

- Georges, M., Charlier, C., & Hayes, B. 2019. Harnessing genomic information for livestock improvement. *Nature Reviews Genetics*, 20(3), 135-156. <https://doi.org/10.1038/s41576-018-0082-2>
- Gevaert, O. 2015. MethylMix: an R package for identifying DNA methylation-driven genes. *Bioinformatics*, 31(11), 1839-1841. <https://doi.org/10.1093/bioinformatics/btv020>
- Gifford, C. A., Ziller, M. J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., ... & Meissner, A. 2013. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell*, 153(5), 1149-1163. <https://doi.org/10.1016/j.cell.2013.04.037>
- Gilbert, W., & Maxam, A. 1973. The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences*, 70(12), 3581-3584. <https://doi.org/10.1073/pnas.70.12.3581>
- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., ... & Nusbaum, C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology*, 27(2), 182-189. <https://doi.org/10.1038/nbt.1523>
- Goldberg, B., Sichtig, H., Geyer, C., Ledebor, N., & Weinstock, G. M. 2015. Making the leap from research laboratory to clinic: challenges and opportunities for next-generation sequencing in infectious disease diagnostics. *MBio*, 6(6), e01888-15. <https://doi.org/10.1128/mBio.01888-15>
- Goodwin, S., McPherson, J. D., & McCombie, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333-351. <https://doi.org/10.1038/nrg.2016.49>
- Goyal, P., Krasteva, P. V., Van Gerven, N., Gubellini, F., Van den Broeck, I., Troupiotis-Tsaïlaki, A., ... & Remaut, H. 2014. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature*, 516(7530), 250-253. <https://doi.org/10.1038/nature13768>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644-652. <https://doi.org/10.1038/nbt.1883>
- Gu, W., Crawford, E. D., O'Donovan, B. D., Wilson, M. R., Chow, E. D., Retallack, H., & DeRisi, J. L. 2016. Depletion of abundant sequences by hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biology*, 17, 41. <https://doi.org/10.1186/s13059-016-0904-5>
- Gupta, P. K. 2008. Single-molecule DNA sequencing technologies for future genomics research. *Trends in Biotechnology*, 26(11), 602-611. <https://doi.org/10.1016/j.tibtech.2008.07.003>
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072-1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., ... & Regev, A. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494-1512. <https://doi.org/10.1038/nprot.2013.084>
- Haque, A., Engel, J., Teichmann, S. A., & Lönnberg, T. 2017. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Medicine*, 9, 75. <https://doi.org/10.1186/s13073-017-0467-4>
- Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., ... & Xie, Z. 2008. Single-molecule DNA sequencing of a viral genome. *Science*, 320(5872), 106-109. <https://doi.org/10.1126/science.1150427>
- Hasin, Y., Seldin, M., & Lusis, A. 2017. Multi-omics approaches to disease. *Genome Biology*, 18, 83. <https://doi.org/10.1186/s13059-017-1215-1>

- Hanssen, F., Garcia, M. U., Folkersen, L., Pedersen, A. S., Lescai, F., Jodoin, S., Miller, E., Seybold, M., Wacker, O., Smith, N., Gabernet, G., & Nahnsen, S. 2024. Scalable and efficient DNA sequencing analysis on different compute infrastructures aiding variant discovery. *NAR Genomics and Bioinformatics*, 6(2). <https://doi.org/10.1093/nargab/lqae031>
- Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. 2014. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*, 56(2), 61-77. <https://doi.org/10.2144/000114133>
- Heller, D., & Vingron, M. 2019. SVIM: structural variant identification using mapped long reads. *Bioinformatics*, 35(17), 2907-2915. <https://doi.org/10.1093/bioinformatics/btz041>
- Hendriksen, R. S., Munk, P., Njage, P., Van Bunnik, B., McNally, L., Lukjancenka, O., ... & Aarestrup, F. M. 2019. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nature Communications*, 10(1), 1124. <https://doi.org/10.1038/s41467-019-08853-3>
- Hill, C. B., Cassin, A., Keeble-Gagnère, G., Doblin, M. S., Bacic, A., & Roessner, U. 2016. De novo transcriptome assembly and analysis of differentially expressed genes of two barley genotypes reveal root-zone-specific responses to salt exposure. *Scientific Reports*, 6, 31558. <https://doi.org/10.1038/srep31558>
- Holder, L. B., Haque, M. M., & Skinner, M. K. 2017. Machine learning for epigenetics and future medical applications. *Epigenetics*, 12(7), 505-514. <https://doi.org/10.1080/15592294.2017.1329068>
- Horvath, S., & Raj, K. 2018. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, 19(6), 371-384. <https://doi.org/10.1038/s41576-018-0004-3>
- Huang, X., & Han, B. 2014. Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology*, 65(1), 531-551. <https://doi.org/10.1146/annurev-arplant-050213-035715>
- Huddleston, J., & Eichler, E. E. 2016. An incomplete understanding of human genetic variation. *Genetics*, 202(4), 1251-1254. <https://doi.org/10.1534/genetics.115.180539>
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., & Bork, P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, 34(8), 2115-2122. <https://doi.org/10.1093/molbev/msx148>
- Hunter, J. D. 2007. Matplotlib: a 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95. <https://doi.org/10.1109/MCSE.2007.55>
- Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., & Welch, D. M. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, 8, R143. <https://doi.org/10.1186/gb-2007-8-7-r143>
- Hwang, B., Lee, J. H., & Bang, D. 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine*, 50(8), 1-14. <https://doi.org/10.1038/s12276-018-0071-8>
- Hwang, S., Kim, E., Lee, I., & Marcotte, E. M. 2015. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Scientific Reports*, 5, 17875. <https://doi.org/10.1038/srep17875>
- Ibeagha-Awemu, E. M., & Zhao, X. 2015. Epigenetic marks: regulators of livestock phenotypes and conceivable sources of missing variation in livestock improvement programs. *Frontiers in Genetics*, 6, 302. <https://doi.org/10.3389/fgene.2015.00302>
- Ichikawa, M., Kato, N., Toda, E., Kashihara, M., Ishida, Y., Hiei, Y., Isobe, S. N., Shirasawa, K., Hirakawa, H., Okamoto, T., & Komari, T. 2023. Whole-genome sequence analysis of mutations

- in rice plants regenerated from zygotes, mature embryos, and immature embryos. *Breeding Science*, 73(3), 349–353. <https://doi.org/10.1270/jsbbs.22100>
- Ignatiadis, M., Sledge, G. W., & Jeffrey, S. S. 2021. Liquid biopsy enters the clinic—implementation issues and future challenges. *Nature Reviews Clinical Oncology*, 18(5), 297-312. <https://doi.org/10.1038/s41571-020-00457-x>
- Illumina. 2021. NovaSeq 6000 System Specifications. Retrieved from <https://www.illumina.com/systems/sequencing-platforms/novaseq/specifications.html>
- Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., Leuillet, S., Winkler, A., Xiao, Y., ... & Grant, K. 2019. The use of next generation sequencing for improving food safety: Translation into practice. *Food Microbiology*, 79, 96-115. <https://doi.org/10.1016/j.fm.2018.11.005>
- Jain, M., Olsen, H. E., Paten, B., & Akeson, M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biology*, 17, 239. <https://doi.org/10.1186/s13059-016-1103-0>
- Japrung, D., Bahrami, A., Nadzeyka, A., Peto, L., Bauerdick, S., Edel, J. B., & Albrecht, T. 2014. SSB binding to single-stranded DNA probed using solid-state nanopore sensors. *The journal of physical chemistry. B*, 118(40), 11605–11612. <https://doi.org/10.1021/jp506832u>
- Jarvis, E. D. 2016. Perspectives from the avian phylogenomics project: questions that can be answered with sequencing all genomes of a vertebrate class. *Annual Review of Animal Biosciences*, 4(1), 45-59. <https://doi.org/10.1146/annurev-animal-021815-111216>
- John, E., Singh, K. B., Oliver, R. P., Soyer, J. L., Muria-Gonzalez, J., Soo, D., Jacques, S., & Tan, K.-C. 2022. Chromatin-immunoprecipitation reveals the PnP2 transcriptional network controlling effector-mediated virulence in a fungal pathogen of wheat. <https://doi.org/10.1101/2022.06.16.496517>
- Jones, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, 13(7), 484-492. <https://doi.org/10.1038/nrg3230>
- Jou, W. M., Haegeman, G., Ysebaert, M., & Fiers, W. 1972. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature*, 237(5350), 82-88. <https://doi.org/10.1038/237082a0>
- Jiang, G., Zheng, J.-Y., Ren, S.-N., Yin, W., Xia, X., Li, Y., & Wang, H.-L. 2024. A comprehensive workflow for optimizing RNA-seq data analysis. *BMC Genomics*, 25(1). <https://doi.org/10.1186/s12864-024-10414-y>
- Kanagal-Shamanna, R. 2016. Digital PCR: Principles and Applications. *Clinical Applications of PCR*, 43–50. https://doi.org/10.1007/978-1-4939-3360-0_5
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., ... & MacArthur, D. G. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809), 434-443. <https://doi.org/10.1038/s41586-020-2308-7>
- Katsanis, S. H., & Katsanis, N. 2013. Molecular genetic testing and the future of clinical genomics. *Nature Reviews Genetics*, 14(6), 415-426. <https://doi.org/10.1038/nrg3493>
- Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., ... & Henikoff, S. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications*, 10(1), 1930. <https://doi.org/10.1038/s41467-019-09982-5>
- Kelsey, G., Stegle, O., & Reik, W. 2017. Single-cell epigenomics: Recording the past and predicting the future. *Science*, 358(6359), 69-75. <https://doi.org/10.1126/science.aan6826>
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. 2002. The human genome browser at UCSC. *Genome Research*, 12(6), 996-1006. <https://doi.org/10.1101/gr.229102>

- Kidder, B. L., Hu, G., & Zhao, K. 2011. ChIP-Seq: technical considerations for obtaining high-quality data. *Nature Immunology*, 12(10), 918-922. <https://doi.org/10.1038/ni.2117>
- Kimes, N. E., Callaghan, A. V., Aktas, D. F., Smith, W. L., Sunner, J., Golding, B., ... & Morris, P. J. 2013. Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Frontiers in Microbiology*, 4, 50. <https://doi.org/10.3389/fmicb.2013.00050>
- Kiselev, V. Y., Andrews, T. S., & Hemberg, M. 2019. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nature Reviews Genetics*, 20(5), 273-282. <https://doi.org/10.1038/s41576-018-0088-9>
- Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., & Taipale, J. 2012. Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*, 9(1), 72-74. <https://doi.org/10.1038/nmeth.1778>
- Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., ... & Dorrestein, P. C. 2018. Best practices for analysing microbiomes. *Nature Reviews Microbiology*, 16(7), 410-422. <https://doi.org/10.1038/s41579-018-0029-9>
- Knudsen, B. E., Bergmark, L., Munk, P., Lukjancenko, O., Priemé, A., Aarestrup, F. M., & Pamp, S. J. 2016. Impact of sample type and DNA isolation procedure on genomic inference of microbiome composition. *mSystems*, 1(5), e00095-16. <https://doi.org/10.1128/mSystems.00095-16>
- Koboldt, D. C. 2020. Best practices for variant calling in clinical sequencing. *Genome Medicine*, 12(1). <https://doi.org/10.1186/s13073-020-00791-w>
- Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. 2013. The next-generation sequencing revolution and its impact on genomics. *Cell*, 155(1), 27-38. <https://doi.org/10.1016/j.cell.2013.09.006>
- Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. 2015. The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell*, 58(4), 610-620. <https://doi.org/10.1016/j.molcel.2015.04.005>
- Korlach, J., Bibillo, A., Wegener, J., Peluso, P., Pham, T. T., Park, I., Clark, S., Otto, G. A., & Turner, S. W. 2008. Long, processive enzymatic dna synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. *Nucleosides, Nucleotides and Nucleic Acids*, 27(9), 1072-1082. <https://doi.org/10.1080/15257770802260741>
- Korlach, J., Marks, P. J., Cicero, R. L., Gray, J. J., Murphy, D. L., Roitman, D. B., ... & Turner, S. W. 2008. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proceedings of the National Academy of Sciences*, 105(4), 1176-1181. <https://doi.org/10.1073/pnas.0710982105>
- Köster, J., & Rahmann, S. 2012. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19), 2520-2522. <https://doi.org/10.1093/bioinformatics/bts480>
- Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., & Kamatani, Y. 2019. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome Biology*, 20, 117. <https://doi.org/10.1186/s13059-019-1720-5>
- Kozarewa, I., Armisen, J., Gardner, A. F., Slatko, B. E., & Hendrickson, C. L. 2015. Overview of target enrichment strategies. *Current Protocols in Molecular Biology*, 112(1). Portico. <https://doi.org/10.1002/0471142727.mb0721s112>
- Krueger, F. 2015. Trim Galore! available online at: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
- Krueger, F., & Andrews, S. R. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11), 1571-1572. <https://doi.org/10.1093/bioinformatics/btr167>

- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... & Marra, M. A. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research*, 19(9), 1639-1645. <https://doi.org/10.1101/gr.092759.109>
- Kukurba, K. R., & Montgomery, S. B. 2015. RNA sequencing and analysis. *Cold Spring Harbor Protocols*, 2015(11), pdb-top084970. <https://doi.org/10.1101/pdb.top084970>
- Kurdyukov, S., & Bullock, M. 2016. DNA methylation analysis: choosing the right method. *Biology*, 5(1), 3. <https://doi.org/10.3390/biology5010003>
- Lafzi, A., Moutinho, C., Picelli, S., & Heyn, H. 2018. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nature Protocols*, 13(12), 2742–2757. <https://doi.org/10.1038/s41596-018-0073-y>
- Laird, P. W. 2010. Principles and challenges of genome-wide DNA methylation analysis. *Nature Reviews Genetics*, 11(3), 191-203. <https://doi.org/10.1038/nrg2732>
- Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., ... & Snyder, M. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research*, 22(9), 1813-1831. <https://doi.org/10.1101/gr.136184.111>
- Langmead, B., & Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357-359. <https://doi.org/10.1038/nmeth.1923>
- Laszlo, A. H., Derrington, I. M., Ross, B. C., Brinkerhoff, H., Adey, A., Nova, I. C., ... & Gundlach, J. H. 2014. Decoding long nanopore sequencing reads of natural DNA. *Nature Biotechnology*, 32(8), 829-833. <https://doi.org/10.1038/nbt.2950>
- Leggett, R. M., & Clark, M. D. 2017. A world of opportunities with nanopore sequencing. *Journal of Experimental Botany*, 68(20), 5419-5429. <https://doi.org/10.1093/jxb/erx289>
- Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., & Webb, W. W. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, 299(5607), 682-686. <https://doi.org/10.1126/science.1079700>
- Levy-Sakin, M., Pastor, S., Mostovoy, Y., Li, L., Leung, A. K., McCaffrey, J., ... & Kwok, P. Y. 2019. Genome maps across 26 human populations reveal population-specific patterns of structural variation. *Nature Communications*, 10(1), 1025. <https://doi.org/10.1038/s41467-019-08992-7>
- Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10), 1674-1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987-2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*. <https://doi.org/10.48550/arXiv.1303.3997>
- Li, H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*, 30(20), 2843-2851. <https://doi.org/10.1093/bioinformatics/btu356>
- Li, H., & Durbin, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, Y., & Tollefsbol, T. O. 2011. DNA methylation detection: bisulfite genomic sequencing analysis. *Epigenetics Protocols*, 11-21. https://doi.org/10.1007/978-1-61779-316-5_2

- Liang, Y., Ridzon, D., Wong, L., & Chen, C. 2007. Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*, 8, 166. <https://doi.org/10.1186/1471-2164-8-166>
- Libbrecht, M. W., & Noble, W. S. 2015. Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6), 321-332. <https://doi.org/10.1038/nrg3920>
- Liu, H., Able, A. J., & Able, J. A. 2021. Small RNAs and their targets are associated with the transgenerational effects of water-deficit stress in durum wheat. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-83074-7>
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., ... & Law, M. 2012. Comparison of next-generation sequencing systems. *BioMed Research International*, 2012, 251364. <https://doi.org/10.1155/2012/251364>
- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. 2012. Comparison of next-generation sequencing systems. *Journal of Biomedicine & Biotechnology*, 2012, 251364. <https://doi.org/10.1155/2012/251364>
- Liu, X. S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., ... & Jaenisch, R. 2016. Editing DNA methylation in the mammalian genome. *Cell*, 167(1), 233-247. <https://doi.org/10.1016/j.cell.2016.08.056>
- Liu, Y., Wang, X., Yuan, L., Liu, Y., Shen, T., & Zhang, Y. 2021. Comparative small RNA profiling and functional exploration on wheat with high-and low-cadmium accumulation. *Frontiers in Genetics*, 12, 635599. <https://doi.org/10.3389/fgene.2021.635599>
- Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., ... & Huttenhower, C. 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, 569(7758), 655-662. <https://doi.org/10.1038/s41586-019-1237-9>
- Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., ... & Huttenhower, C. 2017. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*, 550(7674), 61-66. <https://doi.org/10.1038/nature23889>
- Logsdon, G. A., Vollger, M. R., & Eichler, E. E. 2020. Long-read human genome sequencing and its applications. *Nature Reviews Genetics*, 21(10), 597-614. <https://doi.org/10.1038/s41576-020-0236-x>
- Loose, M., Malla, S., & Stout, M. 2016. Real-time selective sequencing using nanopore technology. *Nature Methods*, 13(9), 751-754. <https://doi.org/10.1038/nmeth.3930>
- Love, M. I., Huber, W., & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Luecken, M. D., & Theis, F. J. 2019. Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology*, 15(6), e8746. <https://doi.org/10.15252/msb.20188746>
- Luo, C., Hajkova, P., & Ecker, J. R. 2018. Dynamic DNA methylation: In the right place at the right time. *Science*, 361(6409), 1336-1340. <https://doi.org/10.1126/science.aat6806>
- Ma, P., Wu, L., Xu, Y., Xu, H., Zhang, X., Wang, W., ... & Wang, B. 2021. Bulk segregant RNA-Seq provides distinctive expression profile against powdery mildew in the wheat genotype YD588. *Frontiers in Plant Science*, 12, 764978. <https://doi.org/10.3389/fpls.2021.764978>
- Machanic, P., & Bailey, T. L. 2011. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, 27(12), 1696-1697. <https://doi.org/10.1093/bioinformatics/btr189>
- Maes, R. K., Langohr, I. M., Wise, A. G., Smedley, R. C., Thaiwong, T., & Kiupel, M. 2014. Beyond H&E: Integration of nucleic acid-based analyses into diagnostic pathology. *Veterinary Pathology*, 51(1), 238-256. <https://doi.org/10.1177/0300985813505878>
- Magli, M. C., Pomante, A., Cafueri, G., Valerio, M., Crippa, A., Ferraretti, A. P., & Gianaroli, L. 2016. Preimplantation genetic testing: polar bodies, blastomeres, trophectoderm cells, or

- blastocoelic fluid? Fertility and Sterility, 105(3), 676-683.e5.
<https://doi.org/10.1016/j.fertnstert.2015.11.018>
- Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., ... & Reich, D. 2016. The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature*, 538(7624), 201-206. <https://doi.org/10.1038/nature18964>
- Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., ... & Turner, D. J. 2010. Target-enrichment strategies for next-generation sequencing. *Nature Methods*, 7(2), 111-118. <https://doi.org/10.1038/nmeth.1419>
- Manrao, E. A., Derrington, I. M., Laszlo, A. H., Langford, K. W., Hopper, M. K., Gillgren, N., ... & Gundlach, J. H. 2012. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nature Biotechnology*, 30(4), 349-353. <https://doi.org/10.1038/nbt.2171>
- Mardis, E. R. 2008. Next-generation DNA sequencing methods. *Annual Review of Genomics and Human Genetics*, 9(1), 387-402. <https://doi.org/10.1146/annurev.genom.9.081307.164359>
- Mardis, E. R. 2008. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24(3), 133-141. <https://doi.org/10.1016/j.tig.2007.12.007>
- Mardis, E. R. 2011. A decade's perspective on DNA sequencing technology. *Nature*, 470(7333), 198-203. <https://doi.org/10.1038/nature09796>
- Mardis, E. R. 2013. Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, 6(1), 287-303. <https://doi.org/10.1146/annurev-anchem-062012-092628>
- Mardis, E. R. 2017. DNA sequencing technologies: 2006–2016. *Nature Protocols*, 12(2), 213-218. <https://doi.org/10.1038/nprot.2016.182>
- Maree, H. J., Fox, A., Al Rwahnih, M., Boonham, N., & Candresse, T. 2018. Application of HTS for routine plant virus diagnostics: state of the art and challenges. *Frontiers in Plant Science*, 9, 1082. <https://doi.org/10.3389/fpls.2018.01082>
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bembien, L. A., ... & Rothberg, J. M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376-380. <https://doi.org/10.1038/nature03959>
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal*, 17(1), 10-12. <https://doi.org/10.14806/ej.17.1.200>
- Massel, K., Campbell, B. C., Mace, E. S., Tai, S., Tao, Y., Worland, B. G., Jordan, D. R., Botella, J. R., & Godwin, I. D. 2016. Whole genome sequencing reveals potential new targets for improving nitrogen uptake and utilization in Sorghum bicolor. *Frontiers in Plant Science*, 7, 1544. <https://doi.org/10.3389/fpls.2016.01544>
- Mathur, M., Prajapat, R. K., Upadhyay, T. K., Lal, D., Khatik, N., & Sharma, D. 2021. Advances in Genomics and Proteomics in Agriculture. *Crop Improvement*, 23–35. <https://doi.org/10.1201/9781003099079-2>
- Maxam, A. M., & Gilbert, W. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560-564. <https://doi.org/10.1073/pnas.74.2.560>
- McGuire, A. L., Caulfield, T., & Cho, M. K. 2008. Research ethics and the challenge of whole-genome sequencing. *Nature Reviews Genetics*, 9(2), 152-156. <https://doi.org/10.1038/nrg2302>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., ... & DePristo, M. A. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297-1303. <https://doi.org/10.1101/gr.107524.110>
- McKernan, K. J., Peckham, H. E., Costa, G. L., McLaughlin, S. F., Fu, Y., Tsung, E. F., ... & Blanchard, A. P. 2009. Sequence and structural variation in a human genome uncovered by short-

- read, massively parallel ligation sequencing using two-base encoding. *Genome Research*, 19(9), 1527-1541. <https://doi.org/10.1101/gr.091868.109>
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., ... & Cunningham, F. 2016. The ensembl variant effect predictor. *Genome Biology*, 17, 122. <https://doi.org/10.1186/s13059-016-0974-4>
- McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., ... & Bejerano, G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, 28(5), 495-501. <https://doi.org/10.1038/nbt.1630>
- McMurdie, P. J., & Holmes, S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS One*, 8(4), e61217. <https://doi.org/10.1371/journal.pone.0061217>
- Meienberg, J., Bruggmann, R., Oexle, K., & Matyas, G. 2016. Clinical sequencing: is WGS the better WES? *Human Genetics*, 135(3), 359–362. <https://doi.org/10.1007/s00439-015-1631-9>
- Meissner, A. 2010. Epigenetic modifications in pluripotent and differentiated cells. *Nature Biotechnology*, 28(10), 1079-1088. <https://doi.org/10.1038/nbt.1684>
- Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research*, 33(18), 5868-5877. <https://doi.org/10.1093/nar/gki901>
- Mendizabal, I., Keller, T. E., Zeng, J., & Yi, S. V. 2014. Epigenetics and evolution. *Integrative and Comparative Biology*, 54(1), 31-42. <https://doi.org/10.1093/icb/icu040>
- Merriman, B., R&D Team, I. T., & Rothberg, J. M. 2012. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, 33(23), 3397-3417. <https://doi.org/10.1002/elps.201200424>
- Mertes, F., ElSharawy, A., Sauer, S., van Helvoort, J. M., Van Der Zaag, P. J., Franke, A., ... & Brookes, A. J. 2011. Targeted enrichment of genomic DNA regions for next-generation sequencing. *Briefings in Functional Genomics*, 10(6), 374-386. <https://doi.org/10.1093/bfpg/elr033>
- Metzker, M. L. 2010. Sequencing technologies—the next generation. *Nature Reviews Genetics*, 11(1), 31-46. <https://doi.org/10.1038/nrg2626>
- Meyer, C. A., & Liu, X. S. 2014. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nature Reviews Genetics*, 15(11), 709-721. <https://doi.org/10.1038/nrg3788>
- Michael, T. P., & VanBuren, R. 2020. Building near-complete plant genomes. *Current Opinion in Plant Biology*, 54, 26–33. <https://doi.org/10.1016/j.pbi.2019.12.009>
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7), 621-628. <https://doi.org/10.1038/nmeth.1226>
- Mwangi, W., de Figueiredo, P., & Criscitiello, M. F. 2016. One health: addressing global challenges at the nexus of human, animal, and environmental health. *PLOS Pathogens*, 12(9), e1005731. <https://doi.org/10.1371/journal.ppat.1005731>
- Nagarajan, N., & Pop, M. 2013. Sequence assembly demystified. *Nature Reviews Genetics*, 14(3), 157-167. <https://doi.org/10.1038/nrg3367>
- Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., ... & Kanaya, S. 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Research*, 39(13), e90. <https://doi.org/10.1093/nar/gkr344>
- Nakato, R., & Shirahige, K. 2017. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. *Briefings in Bioinformatics*, 18(2), 279-290. <https://doi.org/10.1093/bib/bbw023>

- Nefissi Ouertani, R., Arasappan, D., Abid, G., Ben Chikha, M., Jardak, R., Mahmoudi, H., ... & Jansen, R. K. 2021. Transcriptomic analysis of salt-stress-responsive genes in barley roots and leaves. *International Journal of Molecular Sciences*, 22(15), 8155. <https://doi.org/10.3390/ijms22158155>
- Nekrutenko, A., & Taylor, J. 2012. Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nature Reviews Genetics*, 13(9), 667-672. <https://doi.org/10.1038/nrg3305>
- Nesme, J., Achouak, W., Agathos, S. N., Bailey, M., Baldrian, P., Brunel, D., ... & Simonet, P. 2016. Back to the future of soil metagenomics. *Frontiers in Microbiology*, 7, 73. <https://doi.org/10.3389/fmicb.2016.00073>
- Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., ... & Shendure, J. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461(7261), 272-276. <https://doi.org/10.1038/nature08250>
- Nielsen, R., Paul, J. S., Albrechtsen, A., & Song, Y. S. 2011. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, 12(6), 443-451. <https://doi.org/10.1038/nrg2986>
- Niu, J., Ma, S., Zheng, S., Zhang, C., Lu, Y., Si, Y., Tian, S., Shi, X., Liu, X., Naeem, M. K., Sun, H., Hu, Y., Wu, H., Cui, Y., Chen, C., Long, W., Zhang, Y., Gu, M., Cui, M., ... Ling, H.-Q. 2023. Whole-genome sequencing of diverse wheat accessions uncovers genetic changes during modern breeding in China and the United States. *The Plant Cell*, 35(12), 4199-4216. <https://doi.org/10.1093/plcell/koad229>
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, 27(5), 824-834. <https://doi.org/10.1101/gr.213959.116>
- Okonechnikov, K., Conesa, A., & García-Alcalde, F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 32(2), 292-294. <https://doi.org/10.1093/bioinformatics/btv566>
- Oksanen, J. 2010. Vegan: community ecology package. Retrieved from <http://vegan.r-forge.r-project.org/>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... & Wagner, H. 2019. vegan: Community Ecology Package. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>
- Olsvik, Ø., Wahlberg, J., Petterson, B., Uhlen, M., Popovic, T., Wachsmuth, I. K., & Fields, P. I. 1993. Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *Journal of Clinical Microbiology*, 31(1), 22-25. <https://doi.org/10.1128/jcm.31.1.22-25.1993>
- Ozsolak, F., & Milos, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, 12(2), 87-98. <https://doi.org/10.1038/nrg2934>
- Ozsolak, F., Goren, A., Gymrek, M., Guttman, M., Regev, A., Bernstein, B. E., & Milos, P. M. 2010. Digital transcriptome profiling from attomole-level RNA samples. *Genome Research*, 20(4), 519-525. <https://doi.org/10.1101/gr.102129.109>
- Ozsolak, F., Platt, A. R., Jones, D. R., Reifenger, J. G., Sass, L. E., McInerney, P., ... & Milos, P. M. 2009. Direct RNA sequencing. *Nature*, 461(7265), 814-818. <https://doi.org/10.1038/nature08390>
- Pandey, V., Nutter, R. C., & Prediger, E. 2008. Applied Biosystems SOLiD™ System: Ligation-Based Sequencing. *Next Generation Genome Sequencing*, 29-42. Portico. <https://doi.org/10.1002/9783527625130.ch3>

- Pareek, C. S., Smoczynski, R., & Tretyn, A. 2011. Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, 52(4), 413-435. <https://doi.org/10.1007/s13353-011-0057-x>
- Pereira, R., Oliveira, J., & Sousa, M. 2020. Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. *Journal of Clinical Medicine*, 9(1), 132. <https://doi.org/10.3390/jcm9010132>
- Park, P. J. 2009. ChIP-seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics*, 10(10), 669-680. <https://doi.org/10.1038/nrg2641>
- Payne, A., Holmes, N., Rakyan, V., & Loose, M. 2018. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics*, 35(13), 2193-2198. <https://doi.org/10.1093/bioinformatics/bty841>
- Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, 11(9), 1650-1667. <https://doi.org/10.1038/nprot.2016.095>
- Pettersson, E., Lundeberg, J., & Ahmadian, A. 2009. Generations of sequencing technologies. *Genomics*, 93(2), 105-111. <https://doi.org/10.1016/j.ygeno.2008.10.003>
- Pinxten, W., & Howard, H. C. 2014. Ethical issues raised by whole genome sequencing. *Best Practice & Research Clinical Gastroenterology*, 28(2), 269-279. <https://doi.org/10.1016/j.bpg.2014.02.004>
- Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. 2018. Long reads: their purpose and place. *Human Molecular Genetics*, 27(R2), R234-R241. <https://doi.org/10.1093/hmg/ddy177>
- Poplin, R., Chang, P. C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., ... & DePristo, M. A. 2018. A universal SNP and small-indel variant caller using deep neural networks. *Nature Biotechnology*, 36(10), 983-987. <https://doi.org/10.1038/nbt.4235>
- Portela, A., & Esteller, M. 2010. Epigenetic modifications and human disease. *Nature Biotechnology*, 28(10), 1057-1068. <https://doi.org/10.1038/nbt.1685>
- Price, V., Ngwira, L. G., Lewis, J. M., Baker, K. S., Peacock, S. J., Jauneikaite, E., & Feasey, N. 2023. A systematic review of economic evaluations of whole-genome sequencing for the surveillance of bacterial pathogens. *Microbial Genomics*, 9(2), 000947. <https://doi.org/10.1099/mgen.0.000947>
- Priestley, P., Baber, J., Lolkema, M. P., Steeghs, N., de Bruijn, E., Shale, C., ... & Cuppen, E. 2019. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature*, 575(7781), 210-216. <https://doi.org/10.1038/s41586-019-1689-y>
- Pritchard, C. C., Cheng, H. H., & Tewari, M. 2012. MicroRNA profiling: approaches and considerations. *Nature Reviews Genetics*, 13(5), 358-369. <https://doi.org/10.1038/nrg3198>
- Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., ... & Baumeister, K. 1987. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science*, 238(4825), 336-341. <https://doi.org/10.1126/science.2443975>
- Psifidi, A., Dovas, C. I., Bramis, G., Lazou, T., Russel, C. L., Arsenos, G., & Banos, G. 2015. Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and long-term DNA banking using blood samples. *PLOS ONE*, 10(1), e0115960. <https://doi.org/10.1371/journal.pone.0115960>
- Purugganan, M. D., & Jackson, S. A. 2021. Advancing crop genomics from lab to field. *Nature Genetics*, 53(5), 595-601. <https://doi.org/10.1038/s41588-021-00866-3>
- Pushkarev, D., Neff, N. F., & Quake, S. R. 2009. Single-molecule sequencing of an individual human genome. *Nature Biotechnology*, 27(9), 847-850. <https://doi.org/10.1038/nbt.1561>
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., ... & Gu, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences

- and Illumina MiSeq sequencers. *BMC Genomics*, 13, 341. <https://doi.org/10.1186/1471-2164-13-341>
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. 2017. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35(9), 833-844. <https://doi.org/10.1038/nbt.3935>
- Rabbani, B., Nakaoka, H., Akhondzadeh, S., Tekin, M., & Mahdih, N. 2016. Next generation sequencing: implications in personalized medicine and pharmacogenomics. *Molecular BioSystems*, 12(6), 1818–1830. <https://doi.org/10.1039/c6mb00115g>
- Rabbani, B., Tekin, M., & Mahdih, N. 2014. The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*, 59(1), 5-15. <https://doi.org/10.1038/jhg.2013.114>
- Rajkumar, M. S., Shankar, R., Garg, R., & Jain, M. 2020. Bisulphite sequencing reveals dynamic DNA methylation under desiccation and salinity stresses in rice cultivars. *Genomics*, 112(5), 3537–3548. <https://doi.org/10.1016/j.ygeno.2020.04.005>
- Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., ... & Pybus, O. G. 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5(11), 1403-1407. <https://doi.org/10.1038/s41564-020-0770-5>
- Rang, F. J., Kloosterman, W. P., & de Ridder, J. 2018. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biology*, 19(1), 90. <https://doi.org/10.1186/s13059-018-1462-9>
- Ranjan, R., Rani, A., Metwally, A., McGee, H. S., & Perkins, D. L. 2016. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and Biophysical Research Communications*, 469(4), 967-977. <https://doi.org/10.1016/j.bbrc.2015.12.083>
- Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., & He, Z. 2017. Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Molecular Plant*, 10(8), 1047-1064. <https://doi.org/10.1016/j.molp.2017.06.008>
- Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., Domin, H., Fraune, S., von Frieling, J., Hentschel, U., Heinsen, F.-A., Höppner, M., Jahn, M. T., Jaspers, C., Kissoyan, K. A. B., Langfeldt, D., Rehman, A., Reusch, T. B. H., Roeder, T., ... Baines, J. F. 2019. Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome*, 7(1). <https://doi.org/10.1186/s40168-019-0743-1>
- Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 28(18), i333-i339. <https://doi.org/10.1093/bioinformatics/bts378>
- Rauschert, S., Raubenheimer, K., Melton, P. E., & Huang, R. C. 2020. Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification. *Clinical Epigenetics*, 12(1). <https://doi.org/10.1186/s13148-020-00842-4>
- Regier, A. A., Farjoun, Y., Larson, D. E., Krasheninina, O., Kang, H. M., Howrigan, D. P., ... & Hall, I. M. 2018. Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. *Nature Communications*, 9(1), 4038. <https://doi.org/10.1038/s41467-018-06159-4>
- Rehm, H. L. 2013. Disease-targeted sequencing: a cornerstone in the clinic. *Nature Reviews Genetics*, 14(4), 295-300. <https://doi.org/10.1038/nrg3463>
- Reinert, K., Langmead, B., Weese, D., & Evers, D. J. 2015. Alignment of next-generation sequencing reads. *Annual Review of Genomics and Human Genetics*, 16(1), 133-151. <https://doi.org/10.1146/annurev-genom-090413-025358>

- Reuter, J. A., Spacek, D. V., & Snyder, M. P. 2015. High-throughput sequencing technologies. *Molecular Cell*, 58(4), 586-597. <https://doi.org/10.1016/j.molcel.2015.05.004>
- Rhoads, A., & Au, K. F. 2015. PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics*, 13(5), 278-289. <https://doi.org/10.1016/j.gpb.2015.08.002>
- Rio, D. C., Ares, M., Hannon, G. J., & Nilsen, T. W. 2010. Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harbor Protocols*, 2010(6), pdb. prot5439. <https://doi.org/10.1101/pdb.prot5439>
- Robin, J. D., Ludlow, A. T., LaRanger, R., Wright, W. E., & Shay, J. W. 2016. Comparison of DNA quantification methods for next generation sequencing. *Scientific Reports*, 6(1), 24067. <https://doi.org/10.1038/srep24067>
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. 2011. Integrative genomics viewer. *Nature Biotechnology*, 29(1), 24-26. <https://doi.org/10.1038/nbt.1754>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. <https://doi.org/10.1093/bioinformatics/btp616>
- Roden, D. M., Wilke, R. A., Kroemer, H. K., & Stein, C. M. 2011. Pharmacogenomics: the genetics of variable drug responses. *Circulation*, 123(15), 1661-1670. <https://doi.org/10.1161/CIRCULATIONAHA.109.914820>
- Ronaghi, M. 2001. Pyrosequencing sheds light on DNA sequencing. *Genome Research*, 11(1), 3-11. <https://doi.org/10.1101/gr.11.1.3>
- Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., & Nyrén, P. 1996. Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry*, 242(1), 84-89. <https://doi.org/10.1006/abio.1996.0432>
- Rothberg, J. M., & Leamon, J. H. 2008. The development and impact of 454 sequencing. *Nature Biotechnology*, 26(10), 1117-1124. <https://doi.org/10.1038/nbt1485>
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., ... & Bustillo, J. 2011. An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348-352. <https://doi.org/10.1038/nature10242>
- Royer-Bertrand, B., Cisarova, K., Niel-Butschi, F., Mittaz-Crettol, L., Fodstad, H., & Superti-Furga, A. 2021. CNV detection from exome sequencing data in routine diagnostics of rare genetic disorders: opportunities and limitations. *Genes*, 12(9), 1427. <https://doi.org/10.3390/genes12091427>
- Ruppert, K. M., Kline, R. J., & Rahman, M. S. 2019. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*, 17, e00547. <https://doi.org/10.1016/j.gecco.2019.e00547>
- Salk, J. J., Schmitt, M. W., & Loeb, L. A. 2018. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nature Reviews Genetics*, 19(5), 269-285. <https://doi.org/10.1038/nrg.2017.117>
- Sanger, F., & Coulson, A. R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), 441-448. [https://doi.org/10.1016/0022-2836\(75\)90213-2](https://doi.org/10.1016/0022-2836(75)90213-2)
- Sanger, F., Nicklen, S., & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467. <https://doi.org/10.1073/pnas.74.12.5463>

- Sanger, F., Nicklen, S., & Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467. <https://doi.org/10.1073/pnas.74.12.5463>
- Savic, D., Gertz, J., Jain, P., Cooper, G. M., & Myers, R. M. 2013. Mapping genome-wide transcription factor binding sites in frozen tissues. *Epigenetics & Chromatin*, 6, 1-10. <https://doi.org/10.1186/1756-8935-6-30>
- Schadt, E. E., Turner, S., & Kasarskis, A. 2010. A window into third-generation sequencing. *Human Molecular Genetics*, 19(R2), R227-R240. <https://doi.org/10.1093/hmg/ddq416>
- Schmidt, D., Wilson, M. D., Ballester, B., Schwalie, P. C., Brown, G. D., Marshall, A., ... & Odom, D. T. 2010. Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science*, 328(5981), 1036-1040. <https://doi.org/10.1126/science.1186176>
- Schneider, G. F., & Dekker, C. 2012. DNA sequencing with nanopores. *Nature Biotechnology*, 30(4), 326-328. <https://doi.org/10.1038/nbt.2181>
- Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., ... & Ragg, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, 7, 3. <https://doi.org/10.1186/1471-2199-7-3>
- Schuster, S. C. 2008. Next-generation sequencing transforms today's biology. *Nature Methods*, 5(1), 16-18. <https://doi.org/10.1038/nmeth1156>
- Schwarze, K., Buchanan, J., Taylor, J. C., & Wordsworth, S. 2018. Are whole-exome and whole-genome sequencing approaches cost-effective? A systematic review of the literature. *Genetics in Medicine*, 20(10), 1122-1130. <https://doi.org/10.1038/gim.2017.247>
- Sedlazeck, F. J., Lee, H., Darby, C. A., & Schatz, M. C. 2018. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nature Reviews Genetics*, 19(6), 329-346. <https://doi.org/10.1038/s41576-018-0003-4>
- Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., & Schatz, M. C. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods*, 15(6), 461–468. <https://doi.org/10.1038/s41592-018-0001-7>
- Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., & Huttenhower, C. 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, 9(8), 811-814. <https://doi.org/10.1038/nmeth.2066>
- Shakya, M., Lo, C. C., & Chain, P. S. 2019. Advances and challenges in metatranscriptomic analysis. *Frontiers in Genetics*, 10, 904. <https://doi.org/10.3389/fgene.2019.00904>
- Shendure, J., & Ji, H. 2008. Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135-1145. <https://doi.org/10.1038/nbt1486>
- Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. 2017. DNA sequencing at 40: past, present and future. *Nature*, 550(7676), 345-353. <https://doi.org/10.1038/nature24286>
- Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., ... & Church, G. M. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309(5741), 1728-1732. <https://doi.org/10.1126/science.1117389>
- Shimizu, K. K., Copetti, D., Okada, M., Wicker, T., Tameshige, T., Hatakeyama, M., Shimizu-Inatsugi, R., Aquino, C., Nishimura, K., Kobayashi, F., Murata, K., Kuo, T., Delorean, E., Poland, J., Haberer, G., Spannagl, M., Mayer, K. F. X., Gutierrez-Gonzalez, J., Muehlbauer, G. J., ... Handa, H. 2020. De Novo genome assembly of the japanese wheat cultivar norin 61 highlights functional variation in flowering time and fusarium-resistant genes in east asian genotypes. *Plant and Cell Physiology*, 62(1), 8–27. <https://doi.org/10.1093/pcp/pcaa152>

- Shokralla, S., Spall, J. L., Gibson, J. F., & Hajibabaei, M. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, 21(8), 1794-1805. <https://doi.org/10.1111/j.1365-294X.2012.05538.x>
- Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. 2017. Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, 14(4), 407-410. <https://doi.org/10.1038/nmeth.4184>
- Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics*, 15(2), 121-132. <https://doi.org/10.1038/nrg3642>
- Skene, P. J., & Henikoff, S. 2015. A simple method for generating high-resolution maps of genome-wide protein binding. *Elife*, 4, e09225. <https://doi.org/10.7554/eLife.09225>
- Skvortsova, K., Iovino, N., & Bogdanović, O. 2018. Functions and mechanisms of epigenetic inheritance in animals. *Nature Reviews Molecular Cell Biology*, 19(12), 774-790. <https://doi.org/10.1038/s41580-018-0074-2>
- Slatko, B. E., Gardner, A. F., & Ausubel, F. M. 2018. Overview of next-generation sequencing technologies. *Current Protocols in Molecular Biology*, 122(1), e59. <https://doi.org/10.1002/cpmb.59>
- Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., ... & Kelsey, G. 2014. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods*, 11(8), 817-820. <https://doi.org/10.1038/nmeth.3035>
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., ... & Hood, L. E. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature*, 321(6071), 674-679. <https://doi.org/10.1038/321674a0>
- Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., ... & Herndl, G. J. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences*, 103(32), 12115-12120. <https://doi.org/10.1073/pnas.0605127103>
- Sohn, J. I., & Nam, J. W. 2018. The present and future of de novo whole-genome assembly. *Briefings in Bioinformatics*, 19(1), 23-40. <https://doi.org/10.1093/bib/bbw096>
- Soni, V., Akhade, A. S., Bhattacharya, C., Chawla, Y., Bukhari, Z., Gupta, S. L., Basu, S., Jain, S., & Soni, Y. 2023. Genomic surveillance of bacterial pathogens. *Genomic Surveillance and Pandemic Preparedness*, 71-117. <https://doi.org/10.1016/b978-0-443-18769-8.00011-8>
- Spielmann, M., Lupiáñez, D. G., & Mundlos, S. 2018. Structural variation in the 3D genome. *Nature Reviews Genetics*, 19(7), 453-467. <https://doi.org/10.1038/s41576-018-0007-0>
- Staats, M., Arulandhu, A. J., Gravendeel, B., Holst-Jensen, A., Scholtens, I., Peelen, T., ... & Kok, E. 2016. Advances in DNA metabarcoding for food and wildlife forensic species identification. *Analytical and Bioanalytical Chemistry*, 408, 4615-4630. <https://doi.org/10.1007/s00216-016-9595-8>
- Ståhlberg, A., & Kubista, M. 2014. The workflow of single-cell expression profiling using quantitative real-time PCR. *Expert Review of Molecular Diagnostics*, 14(3), 323-331. <https://doi.org/10.1586/14737159.2014.901154>
- Stankiewicz, P., & Lupski, J. R. 2010. Structural variation in the human genome and its role in disease. *Annual Review of Medicine*, 61(1), 437-455. <https://doi.org/10.1146/annurev-med-100708-204735>
- Stapleton, J. A., Kim, J., Hamilton, J. P., Wu, M., Irber, L. C., Maddamsetti, R., ... & Whitehead, T. A. 2016. Haplotype-phased synthetic long reads from short-read sequencing. *PLOS ONE*, 11(1), e0147229. <https://doi.org/10.1371/journal.pone.0147229>

- Stark, R., Grzelak, M., & Hadfield, J. 2019a. RNA sequencing: the teenage years. *Nature Reviews Genetics*, 20(11), 631-656. <https://doi.org/10.1038/s41576-019-0150-2>
- Stark, Z., Dolman, L., Manolio, T. A., Ozenberger, B., Hill, S. L., Caulfield, M. J., ... & North, K. N. 2019b. Integrating genomics into healthcare: a global responsibility. *The American Journal of Human Genetics*, 104(1), 13-20. <https://doi.org/10.1016/j.ajhg.2018.11.014>
- Steinbock, L. J., & Radenovic, A. 2015. The emergence of nanopores in next-generation sequencing. *Nanotechnology*, 26(7), 074003. <https://doi.org/10.1088/0957-4484/26/7/074003>
- Stellrecht, C. M., & Chen, L. S. 2011. Transcription inhibition as a therapeutic target for cancer. *Cancers*, 3(4), 4170-4190. <https://doi.org/10.3390/cancers3044170>
- Stricker, S. H., Köferle, A., & Beck, S. 2017. From profiles to function in epigenomics. *Nature Reviews Genetics*, 18(1), 51-66. <https://doi.org/10.1038/nrg.2016.138>
- Structure, function and diversity of the healthy human microbiome. 2012. *Nature*, 486(7402), 207-214. <https://doi.org/10.1038/nature11234>
- Stuart, T., & Satija, R. 2019. Integrative single-cell analysis. *Nature Reviews Genetics*, 20(5), 257-272. <https://doi.org/10.1038/s41576-019-0093-7>
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck III, W. M., ... & Satija, R. 2019. Comprehensive integration of single-cell data. *Cell*, 177(7), 1888-1902. <https://doi.org/10.1016/j.cell.2019.05.031>
- Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., ... & Korbel, J. O. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature*, 526(7571), 75-81. <https://doi.org/10.1038/nature15394>
- Sugden, P. H., & Clerk, A. 1998. "Stress-responsive" mitogen-activated protein kinases (c-Jun N-terminal kinases and p38 mitogen-activated protein kinases) in the myocardium. *Circulation Research*, 83(4), 345-352. <https://doi.org/10.1161/01.RES.83.4.345>
- Sun, K., Li, D., Xia, A., Zhao, H., Wen, Q., Jia, S., Wang, J., Yang, G., Zhou, D., Huang, C., Wang, H., Chen, Z., & Guo, T. 2022. Targeted Identification of Rice Grain-Associated Gene Allelic Variation Through Mutation Induction, Targeted Sequencing, and Whole Genome Sequencing Combined with a Mixed-Samples Strategy. *Rice*, 15(1). <https://doi.org/10.1186/s12284-022-00603-2>
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., ... & Velayoudon, D. 2015. Structure and function of the global ocean microbiome. *Science*, 348(6237), 1261359. <https://doi.org/10.1126/science.1261359>
- Suvakov, M., Panda, A., Diesh, C., Holmes, I., & Abyzov, A. 2021. CNVpytor: a tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. *GigaScience*, 10(11), giab074. <https://doi.org/10.1093/gigascience/giab074>
- Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., ... & Stimp, A. M. 2021. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, 590(7845), 290-299. <https://doi.org/10.1038/s41586-021-03205-y>
- Tam, S., Tsao, M. S., & McPherson, J. D. 2015. Optimization of miRNA-seq data preprocessing. *Briefings in Bioinformatics*, 16(6), 950-963. <https://doi.org/10.1093/bib/bbv019>
- Tasic, B., Yao, Z., Graybiel, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., ... & Zeng, H. 2018. Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729), 72-78. <https://doi.org/10.1038/s41586-018-0654-5>
- Thermo Fisher Scientific. 2021. 5500 Series Genetic Analysis Systems. Retrieved from <https://www.thermofisher.com/order/catalog/product/4474867>
- Thompson, J. F., & Milos, P. M. 2011. The properties and applications of single-molecule DNA sequencing. *Genome Biology*, 12, 217. <https://doi.org/10.1186/gb-2011-12-2-217>

- Thompson, J. F., & Steinmann, K. E. 2010. Single molecule sequencing with a HeliScope genetic analysis system. *Current Protocols in Molecular Biology*, 92(1), 7-10. <https://doi.org/10.1002/0471142727.mb0710s92>
- Toolkit, P. 2019. GitHub repository. Broad Institute. Available online at: <http://broadinstitute.github.io/picard>.
- Toumazou, C., Shepherd, L. M., Reed, S. C., Chen, G. I., Patel, A., Garner, D. M., ... & Zhang, L. 2013. Simultaneous DNA amplification and detection using a pH-sensing semiconductor system. *Nature Methods*, 10(7), 641-646. <https://doi.org/10.1038/nmeth.2520>
- Travers, K. J., Chin, C. S., Rank, D. R., Eid, J. S., & Turner, S. W. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic acids research*, 38(15), e159-e159. <https://doi.org/10.1093/nar/gkq543>
- Treangen, T. J., & Salzberg, S. L. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, 13(1), 36-46. <https://doi.org/10.1038/nrg3117>
- Turcatti, G., Romieu, A., Fedurco, M., & Tairi, A.-P. 2008. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Research*, 36(4), e25-e25. <https://doi.org/10.1093/nar/gkn021>
- Umu, S. U., Langseth, H., Bucher-Johannessen, C., Fromm, B., Keller, A., Meese, E., Lauritzen, M., Leithaug, M., Lyle, R., & Rounge, T. B. 2017. A comprehensive profile of circulating RNAs in human serum. *RNA Biology*, 15(2), 242-250. <https://doi.org/10.1080/15476286.2017.1403003>
- Ugur Sezerman, O., Ulgen, E., Seymen, N., & Melis Durasi, I. 2019. Bioinformatics workflows for genomic variant discovery, interpretation and prioritization. *bioinformatics tools for detection and clinical interpretation of genomic variations*. <https://doi.org/10.5772/intechopen.85524>
- Ungaro, A., Pech, N., Martin, J. F., McCairns, R. S., Mévy, J. P., Chappaz, R., & Gilles, A. 2017. Challenges and advances for transcriptome assembly in non-model species. *PloS one*, 12(9), e0185020. <https://doi.org/10.1371/journal.pone.0185020>
- Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., ... & Johnson, S. M. 2008. A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome research*, 18(7), 1051-1063. <https://doi.org/10.1101/gr.076463.108>
- Van der Auwera, G. A., & O'Connor, B. D. 2020. *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. O'Reilly Media. ISBN:1491975164, 9781491975169
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... & DePristo, M. A. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43(1), 11-10. <https://doi.org/10.1002/0471250953.bi1110s43>
- Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. 2014. Ten years of next-generation sequencing technology. *Trends in genetics*, 30(9), 418-426. <https://doi.org/10.1016/j.tig.2014.07.001>
- Vanwonderghem, I., Jensen, P. D., Dennis, P. G., Hugenholtz, P., Rabaey, K., & Tyson, G. W. 2014. Deterministic processes guide long-term synchronised population dynamics in replicate anaerobic digesters. *The ISME Journal*, 8(10), 2015-2028. <https://doi.org/10.1038/ismej.2014.50>
- Varshney, R. K., Bohra, A., Yu, J., Graner, A., Zhang, Q., & Sorrells, M. E. 2021. Designing future crops: genomics-assisted breeding comes of age. *Trends in Plant Science*, 26(6), 631-649. <https://doi.org/10.1016/j.tplants.2021.03.010>
- Varshney, R. K., Singh, V. K., Kumar, A., Powell, W., & Sorrells, M. E. 2018. Can genomics deliver climate-change ready crops? *Current Opinion in Plant Biology*, 45, 205-211. <https://doi.org/10.1016/j.pbi.2018.03.007>

- Vermeesch, J. R., Voet, T., & Devriendt, K. 2016. Prenatal and pre-implantation genetic diagnosis. *Nature Reviews Genetics*, 17(10), 643–656. <https://doi.org/10.1038/nrg.2016.97>
- Voelkerding, K. V., Dames, S. A., & Durtschi, J. D. 2009. Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry*, 55(4), 641-658. <https://doi.org/10.1373/clinchem.2008.112789>
- Wagner, G. P., Kin, K., & Lynch, V. J. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in biosciences*, 131, 281-285. <https://doi.org/10.1007/s12064-012-0162-3>
- Wang, K., Li, M., & Hakonarson, H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), e164-e164. <https://doi.org/10.1093/nar/gkq603>
- Wang, Z., Gerstein, M., & Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57-63. <https://doi.org/10.1038/nrg2484>
- Watson, J. D., & Crick, F. H. 1953. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356), 737-738. <https://doi.org/10.1038/171737a0>
- Weischenfeldt, J., Symmons, O., Spitz, F., & Korbel, J. O. 2013. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nature Reviews Genetics*, 14(2), 125-138. <https://doi.org/10.1038/nrg3373>
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., Hall, R. J., Concepcion, G. T., ... & Hunkapiller, M. W. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, 37(10), 1155-1162. <https://doi.org/10.1038/s41587-019-0217-9>
- Wheeler, D. A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., ... & Rothberg, J. M. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature*, 452(7189), 872-876. <https://doi.org/10.1038/nature06884>
- Wick, R. R., Judd, L. M., & Holt, K. E. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biology*, 20, 1-10. <https://doi.org/10.1186/s13059-019-1727-y>
- Wickham, H., & Wickham, H. 2016. Data analysis (pp. 189-201). Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4_7
- Wickham, H., & Wickham, H. 2016. Getting Started with ggplot2. *ggplot2: Elegant graphics for data analysis*, 11-31. https://doi.org/10.1007/978-3-319-24277-4_2
- Wood, D. E., & Salzberg, S. L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15(3). <https://doi.org/10.1186/gb-2014-15-3-r46>
- Wood, D. E., Lu, J., & Langmead, B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology*, 20, 1-13. <https://doi.org/10.1186/s13059-019-1891-0>
- Wörheide, M. A., Krumsiek, J., Kastenmüller, G., & Arnold, M. 2021. Multi-omics integration in biomedical research – A metabolomics-centric review. *Analytica Chimica Acta*, 1141, 144–162. <https://doi.org/10.1016/j.aca.2020.10.038>
- Xi, Y., & Li, W. 2009. BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics*, 10, 1-9. <https://doi.org/10.1186/1471-2105-10-232>
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., ... & Wei, L. 2011. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, 39(suppl_2), W316-W322. <https://doi.org/10.1093/nar/gkr483>
- Xie, K., Wu, S., Li, Z., Zhou, Y., Zhang, D., Dong, Z., An, X., Zhu, T., Zhang, S., Liu, S., Li, J., & Wan, X. 2018. Map-based cloning and characterization of Zea mays male sterility33 (ZmMs33)

- gene, encoding a glycerol-3-phosphate acyltransferase. *Theoretical and Applied Genetics*, 131(6), 1363–1378. <https://doi.org/10.1007/s00122-018-3083-9>
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., ... & Wang, J. 2014. SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics*, 30(12), 1660-1666. <https://doi.org/10.1093/bioinformatics/btu077>
- Xu, J., Yuan, Y., Xu, Y., Zhang, G., Guo, X., Wu, F., ... & Lu, Y. 2014. Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. *BMC Plant Biology*, 14, 1-15. <https://doi.org/10.1186/1471-2229-14-83>
- Xu, X., Crow, M., Rice, B. R., Li, F., Harris, B., Liu, L., Demesa-Arevalo, E., Lu, Z., Wang, L., Fox, N., Wang, X., Drenkow, J., Luo, A., Char, S. N., Yang, B., Sylvester, A. W., Gingeras, T. R., Schmitz, R. J., Ware, D., ... Jackson, D. 2021. Single-cell RNA sequencing of developing maize ears facilitates functional analysis and trait candidate gene discovery. *Developmental Cell*, 56(4), 557-568.e6. <https://doi.org/10.1016/j.devcel.2020.12.015>
- Ye, K., Schulz, M. H., Long, Q., Apweiler, R., & Ning, Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*, 25(21), 2865-2871. <https://doi.org/10.1093/bioinformatics/btp394>
- Ying, Y.-L., Hu, Z.-L., Zhang, S., Qing, Y., Fragasso, A., Maglia, G., Meller, A., Bayley, H., Dekker, C., & Long, Y.-T. 2022. Nanopore-based technologies beyond DNA sequencing. *Nature Nanotechnology*, 17(11), 1136–1146. <https://doi.org/10.1038/s41565-022-01193-2>
- Yoon, S., Xuan, Z., Makarov, V., Ye, K., & Sebat, J. 2009. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Research*, 19(9), 1586–1592. <https://doi.org/10.1101/gr.092981.109>
- Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, 11, 1-12. <https://doi.org/10.1186/gb-2010-11-2-r14>
- Yu, G., Wang, L. G., & He, Q. Y. 2015. ChIPseeker: an R/bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, 31(14), 2382-2383. <https://doi.org/10.1093/bioinformatics/btv145>
- Zarraonaindia, I., Owens, S. M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., Bokulich, N. A., Mills, D. A., Martin, G., Taghavi, S., van der Lelie, D., & Gilbert, J. A. 2015. The soil microbiome influences grapevine-associated microbiota. *MBio*, 6(2). <https://doi.org/10.1128/mbio.02527-14>
- Zha, W., Li, C., Wu, Y., Chen, J., Li, S., Sun, M., ... & You, A. 2023. Single-Cell RNA sequencing of leaf sheath cells reveals the mechanism of rice resistance to brown planthopper (*Nilaparvata lugens*). *Frontiers in Plant Science*, 14, 1200014. <https://doi.org/10.3389/fpls.2023.1200014>
- Zhang, X., Wang, Y., Chiang, H.-C., Hsieh, Y.-P., Lu, C., Park, B. H., Jatoi, I., Jin, V. X., Hu, Y., & Li, R. 2019. BRCA1 mutations attenuate super-enhancer function and chromatin looping in haploinsufficient human breast epithelial cells. *Breast Cancer Research*, 21(1). <https://doi.org/10.1186/s13058-019-1132-1>
- Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W. L., Chen, H., ... & Ecker, J. R. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell*, 126(6), 1189-1201. <https://doi.org/10.1016/j.cell.2006.08.003>
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., ... & Liu, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9), R137. <https://doi.org/10.1186/gb-2008-9-9-r137>
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., ... & Liu, X. S. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9, 1-9. <https://doi.org/10.1186/gb-2008-9-9-r137>

- Zheng, G. X., Lau, B. T., Schnall-Levin, M., Jarosz, M., Bell, J. M., Hindson, C. M., ... & Ji, H. P. 2016. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology*, 34(3), 303-311. <https://doi.org/10.1038/nbt.3432>
- Zhong, S., Joung, J. G., Zheng, Y., Chen, Y. R., Liu, B., Shao, Y., ... & Giovannoni, J. J. 2011. High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harbor Protocols*, 2011(8), pdb-prot5652. <https://doi.org/10.1101/pdb.prot5652>
- Zhu, D., Chen, Q. L., An, X. L., Yang, X. R., Christie, P., Ke, X., ... & Zhu, Y. G. 2018. Exposure of soil collembolans to microplastics perturbs their gut microbiota and alters their isotopic composition. *Soil Biology and Biochemistry*, 116, 302-310. <https://doi.org/10.1016/j.soilbio.2017.10.027>
- Zhu, M., Zhang, M., Huang, K., Lu, F., Wang, H., Zhao, S., Yu, Y., Tang, S., Wu, H., Hu, P., & Wei, X. 2024. Single-cell transcriptome sequencing reveals the mechanism regulating rice plumule development. *The Crop Journal*, 12(3), 688–697. <https://doi.org/10.1016/j.cj.2024.04.009>
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., ... & Enard, W. 2017. Comparative analysis of single-cell RNA sequencing methods. *Molecular Cell*, 65(4), 631-643. <https://doi.org/10.1016/j.molcel.2017.01.023>
- Ziller, M. J., Hansen, K. D., Meissner, A., & Aryee, M. J. 2015. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nature Methods*, 12(3), 230-232. <https://doi.org/10.1038/nmeth.3152>
- Zook, J. M., McDaniel, J., Olson, N. D., Wagner, J., Parikh, H., Heaton, H., ... & Salit, M. 2019. An open resource for accurately benchmarking small variant and reference calls. *Nature Biotechnology*, 37(5), 561-566. <https://doi.org/10.1038/s41587-019-0074-6>
- Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. 2019. A primer on deep learning in genomics. *Nature Genetics*, 51(1), 12-18. <https://doi.org/10.1038/s41588-018-0295-5>