Review Article

J Genet Med 2023;20(1):1-5
https://doi.org/10.5734/JGM.2023.20.1.1
ISSN 1226-1769 (Print) 2383-8442 (Online)

JGM Journal of Genetic Medicine

Check for updates

Multi-omics techniques for the genetic and epigenetic analysis of rare diseases

Yeonsong Choi[1,2], David Whee-Young Choi[1,2], and Semin Lee[1,2,*]

[1]Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Korea
[2]Korean Genomics Center, Ulsan National Institute of Science and Technology, Ulsan, Korea

Until now, rare disease studies have mainly been carried out by detecting simple variants such as single nucleotide substitutions and short insertions and deletions in protein-coding regions of disease-associated gene panels using diagnostic next-generation sequencing in association with patient phenotypes. However, several recent studies reported that the detection rate hardly exceeds 50% even when whole-exome sequencing is applied. Therefore, the necessity of introducing whole-genome sequencing is emerging to discover more diverse genomic variants and examine their association with rare diseases. When no diagnosis is provided by whole-genome sequencing, additional omics techniques such as RNA-seq also can be considered to further interrogate causal variants. This paper will introduce a description of these multi-omics techniques and their applications in rare disease studies.

Key words: Rare diseases, Multi-omics, Exome sequencing, Whole genome sequencing, RNA sequencing, Bisulfite sequencing, ATAC-seq.

Introduction

Globally, 3.5-5.9% of the general population is affected by rare diseases. Proportionately, this percentage seems insignificant, but the estimated number of patients with rare diseases adds up to 263-446 million people [1]. Thus, the total number of cases is not negligible and necessitates further research in this field of study for quick and accurate diagnoses.

A rare disease is defined as a condition affecting fewer than 200,000 people in the United States [2] and fewer than 1 in 2,000 people in Europe [3]. In Korea, a rare disease is defined as a condition affecting fewer than 20,000 people in the general population. In 2020, the number of rare disease cases in Korea was a total of 52,069.

Rare disease patients take an average of 6-8 years to receive an accurate diagnosis [4]. According to statistics in Korea, almost 80% of patients visited two or more hospitals before receiving a rare disease diagnosis, making it difficult to get an accurate and quick diagnosis [5]. If a patient is diagnosed with a causative mutation, they can receive economic support such as the Exempted Calculation of Health Insurance. However, a significant number of patients remain undiagnosed even after a long period of time.

Genetic causes are known to account for 80% of patients with rare diseases [6,7]. Currently, targeted sequencing or whole-exome sequencing (WES) technologies are mainly used to detect mutations that cause rare diseases. However, WES covers only protein-coding regions which comprise less than 2% of the genome, so there is a clear limitation in detecting mutations in intronic and intergenic regions, large-scale structural variants,

Received: 23 November 2022, Revised: 16 February 2023, Accepted: 22 February 2023, Published: 30 June 2023
*Corresponding author: Semin Lee, Ph.D. https://orcid.org/0000-0002-9015-6046
Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Eonyang-eup, Ulsan 44919, Korea.
Tel: +82-52-217-2663, Fax: +82-52-217-3582, E-mail: seminlee@unist.ac.kr
Conflict of interest: The authors declare that they do not have any conflicts of interest.

and repeat expansions. As a result, the WES-based diagnosis rate is about 25-41% [8]. For this reason, there have been efforts to overcome this limitation of WES by adopting additional omics technologies such as whole-genome sequencing (WGS), RNA sequencing (RNA-seq), bisulfite sequencing, and assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Fig. 1). As the continuous development of next-generation sequencing (NGS) techniques has lowered the cost and time for these various omics techniques [9], they now can be more readily applied to rare disease diagnosis and study. In this review, we introduce some of the multi-omics techniques and studies harnessing them for rare disease studies.

Whole-Exome Sequencing vs. Whole-Genome Sequencing

Whole-exome sequencing (WES) is a method of sequencing protein-coding regions, which takes about 2% of our genome and has been widely used for the diagnosis of rare diseases. However, WES has limitations in being able to detect mutations occurring in non-coding regions, such as intronic/intergenic variants, splicing variants, and complex structural variants. To overcome these limitations, efforts have recently been made to introduce WGS technology to the diagnosis of rare diseases.

A recent study by Burdick et al. [10] reported that 15 of 54 (28%) diagnoses for Undiagnosed Diseases Network participants were not able to be solved by WES and required WGS or other omics techniques because WES failed to identify pathogenic non-coding variants, copy number variations, and repeat expansions. The UK100K project also identified novel pathogenic non-coding variants disrupting the transcription of disease-associated genes such as *ARPC1B, GATA1, LRBA*, and *MPL* using WGS [11]. One interesting reported case in their study is that a boy with autism spectrum disorder and thrombocytopenia turned out to carry a hemizygous deletion of a *GATA1* enhancer, which explained his persistently low platelet count, elevated mean platelet volume, and normal RBC parameters except for mild dyserythropoietic that are typical in patients with a pathogenic *GATA1* mutation [12].

RNA Sequencing

RNA-seq is a technology for analyzing gene expression patterns using NGS [13]. Compared to conventional microarray-based methods, it is possible to detect gene expression levels more precisely at the base-pair level [14]. RNA-seq also has the advantage of being able to detect alternative splicing patterns and gene fusions, which are hard to be identified by WES and WGS. Although it should be considered that gene expression patterns are tissue-specific, there are recent efforts to diagnose and

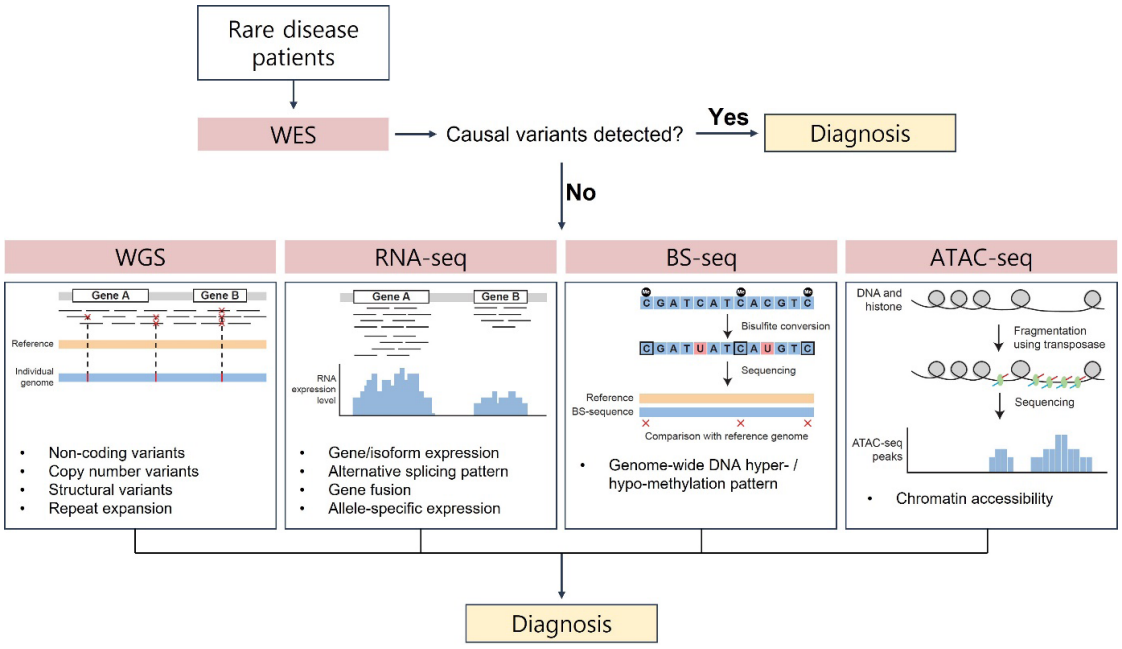

Fig. 1. Schematic diagram of multiple omics techniques for identifying various genomic and epigenomic features in rare diseases. WES, whole-exome sequencing; WGS, whole-genome sequencing; RNA-seq, RNA sequencing; BS-seq, Bisulfite sequencing; ATAC-seq, Assay for transposase-accessible chromatin using sequencing.

analyze rare diseases using RNA-seq data from blood samples.

Frésard et al. [15] analyzed RNA-seq data from 94 individuals with undiagnosed rare diseases and compared them with publicly available RNA-seq data from healthy individuals and tissues to identify outlier expression of genes that are potentially implicated in rare diseases. They found that 1) under-expression outliers were more enriched in the genes sensitive to loss-of-function mutations, 2) the number of splicing outliers was higher in patients, and 3) a large number of rare variants show allelic-specific expression (ASE) biased toward the deleterious allele.

Ferraro et al. [16] also characterized transcriptomic abnormalities such as gene expression, ASE, and alternative splicing from RNA-seq data of multiple different tissue types and developed a statistical model for predicting their impact by integrating more than 800 genomes matched with tissue-specific transcriptomes. They reported that outliers having aberrant gene expression, ASE, and splicing patterns tend to have a higher chance to carry a rare pathogenic variant near the corresponding gene.

Furthermore, a recent study from Oliver et al. [17] analyzed 47 individuals with undiagnosed rare genetic diseases using RNA-seq and reported 11 potentially pathogenic fusion transcripts such as *SAMD12-EXT1* fusion in a patient with multiple exostoses and *ATM-SLC35F2* fusion in a patient with severe combined immunodeficiency.

Bisulfite Sequencing

In addition to genetic mutations, epigenomic changes can also cause rare diseases. In particular, given that mutations in DNA methyltransferases have been reported in various rare diseases such as Heyn–Sproul–Jackson syndrome and immunodeficiency-centromeric instability-facial anomalies syndrome 1 (ICF1), it is necessary to accurately determine how these mutations actually affect genome-wide methylation patterns. There have been various different techniques developed to profile genomic DNA methylation, and most of them are based on bisulfite treatment converting unmethylated cytosines to uracil by deamination while leaving methylated cytosines unconverted [18]. After bisulfite conversion, NGS can be used to distinguish unmethylated cytosines from methylated ones.

Sun et al. [19] interrogated genome-wide DNA methylation by whole-genome bisulfite sequencing of hereditary sensory and autonomic neuropathy type 1 with dementia and hearing loss (HSAN1E) patients with *DNMT1* mutations and their siblings. They found that all chromosomes are generally hypomethylated, and genes associated with differentially methylated regions were significantly enriched in NAD+/NADH metabolism pathways, which are implicated in diverse neurological disorders.

Gatto et al. [20] interrogated the effects of *DNMT3B* dysfunction on the genome-wide DNA methylation profiles in ICF1 by performing reduced representation bisulfite sequencing of patient-derived B-cell lines. They found that pathogenic rare variants in *DNMT3B* can induce catalytic inactivation of *DNMT3B* and eventually lead to DNA hypomethylation, and the genes affected by the *DNMT3B* mutation-induced DNA hypomethylation were mostly direct targets of *DNMT3B*.

Assay for Transposase-Accessible Chromatin Using Sequencing

Chromatin accessibility is highly dynamic and a key epigenomic feature for defining cellular identity because gene expression is also regulated by physical accessibility to its regulatory elements such as enhancers, promoters, and insulators [21]. The genome-wide profiles of DNA accessibility can be characterized by various molecular techniques such as DNase I hypersensitive sites sequencing [22], formaldehyde-assisted identification of regulatory elements followed by sequencing [23], and ATAC-seq. Among them, ATAC-seq is the most recently developed chromatin accessibility assay and the fastest and most sensitive of the available assays [24].

A recent study by Luperchio et al. [25] adopted ATAC-seq to investigate shared epigenetic alterations in mouse models of Kabuki type 1 and 2 and Rubinstein-Taybi type 1 syndromes. They found that disruption of chromatin accessibility at promoters frequently dysregulates downstream gene expression, and a considerable number of dysregulated genes were shared among the three rare disease mouse models, which may explain the shared disease manifestations.

Conclusion

With the recent rapid development of NGS technology, causal variants have been identified for many rare diseases. However, in a significant number of rare diseases, pathogenic variants still have not been discovered, and studies on underlying mechanisms are also lacking. Here, we introduced recent efforts harnessing multi-omics approaches to improve the diagnostic yield and to better understand the molecular mechanism of rare diseases.

WGS can detect various genomic variants such as non-coding mutations, structural variants, and repeat expansions, which

Table 1. Summary table of sequencing techniques and their applications

Sequencing technique	Description	Detectable variants	Reference
Whole-exome sequencing (WES)	- Covering exonic (protein-coding) regions. - Much lower cost than WGS. - Higher sequencing depth than WGS. - Faster sequencing and bioinformatic analysis than WGS.	- Single nucleotide variants (SNVs) and short indels (indels) in exonic regions. - Copy number variants in exonic regions.	- Suwinski et al. [26] - Rabbani et al. [27]
Whole-genome sequencing (WGS)	- Covering the entire genomic region including intronic and intergenic regions. - Identifying more complex genomic variants than WES. - Higher cost than WES.	- SNVs and indels in the entire genome including non-coding regions. - Copy number variants. - Complex structural variants. - Repeat expansions.	- Austin-Tse et al. [28] - Ng and Kirkness [29]
RNA sequencing (RNA-seq)	- Covering transcriptome. - Quantifying RNA expression levels - Identifying differentially expressed genes.	- Gene/isoform expression. - Allele-specific gene expression. - Alternative splicing patterns. - Gene fusions.	- Stark et al. [30] - Hong et al. [31]
Bisulfite sequencing (BS-seq)	- Detecting methylated cytosine in genomic DNA at single-base resolution.	- Genome-wide DNA methylation profiles. - Hyper/hypo-methylated CpG islands.	- Feng and Lou [32] - Wreczycka et al. [33]
Assay for transposase-accessible chromatin using sequencing (ATAC-seq)	- Detecting chromatin accessibility along the genome. - Identifying differentially accessible regions.	- Genome-wide DNA accessibility profiles - Enriched transcription factor binding sites	- Yan et al. [34] - Grandi et al. [35]

cannot be accurately covered by WES. RNA-seq can be also very useful not only for understanding the downstream impact of genomic variants on gene expression profiles but also for detecting additional variant types such as alternative splicing and gene fusions implicated in the pathogenesis of rare diseases. As transcriptomic features can be heavily affected by various epigenomic features such as DNA methylation, histone modification, and DNA accessibility, additional epigenomic approaches such as bisulfite sequencing and ATAC-seq can be useful for understanding the underlying mechanisms of pathogenic variants (Table 1, Fig. 1) [26-35].

Overall, by integrating and analyzing these various omics techniques, it is expected that disease-associated variants will be more precisely identified, and pathogenesis will be better understood, thereby increasing the diagnosis rate of diseases and ultimately contributing to the development of novel treatment technologies.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2020M3E5D7115320). This research was also partly supported by Basic Science Research Program through the NRF funded by the Ministry of Education (NRF-2021R1A6A3A13045998 to Y.C.).

Authors' Contributions

Conception and design: SL. Acquisition of data: YC, DWYC. Analysis and interpretation of data: YC, DWYC. Drafting the article: YC, DWYC, SL. Critical revision of the article: YC, DWYC, SL. Final approval of the version to be published: SL.

References

1. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet 2020;28:165-73.
2. U.S. Food and Drug Administration. An Act to amend the Federal Food, Drug, and Cosmetic Act to facilitate the development of drugs for rare diseases and conditions, and for other purposes. Public Law 97-414. 96 STAT. 2049. Washington, D.C.: U.S. Congress; 1983.
3. European Commission. Directorate-General for Research and Innovation. Collaboration: a key to unlock the challenges of rare diseases research. Luxembourg: Publications Office of the European Union; 2021.
4. Global Genes. RARE disease facts. [https://globalgenes.org/learn/rare-disease-facts/]
5. Ko KP. Analyzing the status of rare diseases and ways to improve support for rare diseases patients in Korea. Cheongju: Korea Centers for Disease Control and Prevention; 2018 Nov.
6. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). Nucleic Acids Res 2009;37(Database issue):D793-6.

7. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. Nucleic Acids Res 2015;43(Database issue):D789-98.
8. Zastrow DB, Kohler JN, Bonner D, Reuter CM, Fernandez L, Grove ME, et al. A toolkit for genetics providers in follow-up of patients with non-diagnostic exome sequencing. J Genet Couns 2019;28:213-28.
9. Pervez MT, Hasnain MJU, Abbas SH, Moustafa MF, Aslam N, Shah SSM. A comprehensive review of performance of next-generation sequencing platforms. Biomed Res Int 2022;2022:3457806.
10. Burdick KJ, Cogan JD, Rives LC, Robertson AK, Koziura ME, Brokamp E, et al.; Undiagnosed Diseases Network. Limitations of exome sequencing in detecting rare and undiagnosed diseases. Am J Med Genet A 2020;182:1400-6.
11. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, et al. Whole-genome sequencing of patients with rare diseases in a national health system. Nature 2020;583:96-102.
12. Freson K, Devriendt K, Matthijs G, Van Hoof A, De Vos R, Thys C, et al. Platelet characteristics in patients with X-linked macrothrombocytopenia because of a novel GATA1 mutation. Blood 2001;98:85-92.
13. Chu Y, Corey DR. RNA sequencing: platform selection, experimental design, and data interpretation. Nucleic Acid Ther 2012;22:271-4.
14. Brahamsha B, Haselkorn R. Isolation and characterization of the gene encoding the principal sigma factor of the vegetative cell RNA polymerase from the cyanobacterium Anabaena sp. strain PCC 7120. J Bacteriol 1991;173:2442-50.
15. Frésard L, Smail C, Ferraro NM, Teran NA, Li X, Smith KS, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. Nat Med 2019;25:911-9.
16. Ferraro NM, Strober BJ, Einson J, Abell NS, Aguet F, Barbeira AN, et al. Transcriptomic signatures across human tissues identify functional rare genetic variation. Science 2020;369:eaaz5900.
17. Oliver GR, Tang X, Schultz-Rogers LE, Vidal-Folch N, Jenkinson WG, Schwab TL, et al. A tailored approach to fusion transcript identification increases diagnosis of rare inherited disease. PLoS One 2019;14:e0223337.
18. Fraga MF, Esteller M. DNA methylation: a profile of methods and applications. Biotechniques 2002;33:632, 634, 636-49.
19. Sun Z, Wu Y, Ordog T, Baheti S, Nie J, Duan X, et al. Aberrant signature methylome by DNMT1 hot spot mutation in hereditary sensory and autonomic neuropathy 1E. Epigenetics 2014;9:1184-93.
20. Gatto S, Gagliardi M, Franzese M, Leppert S, Papa M, Cammisa M, et al. ICF-specific DNMT3B dysfunction interferes with intragenic regulation of mRNA transcription and alternative splicing. Nucleic Acids Res 2017;45:5739-56.
21. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet 2019;20:207-20.
22. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell 2008;132:311-22.
23. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. Genome Res 2007;17:877-85.
24. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 2013;10:1213-8.
25. Luperchio TR, Boukas L, Zhang L, Pilarowski G, Jiang J, Kalinousky A, et al. Leveraging the Mendelian disorders of the epigenetic machinery to systematically map functional epigenetic variation. Elife 2021;10:e65884.
26. Suwinski P, Ong C, Ling MHT, Poh YM, Khan AM, Ong HS. Advancing personalized medicine through the application of whole exome sequencing and big data analytics. Front Genet 2019;10:49.
27. Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet 2014;59:5-15.
28. Austin-Tse CA, Jobanputra V, Perry DL, Bick D, Taft RJ, Venner E, et al.; Medical Genome Initiative. Best practices for the interpretation and reporting of clinical whole genome sequencing. NPJ Genom Med 2022;7:27.
29. Ng PC, Kirkness EF. Whole genome sequencing. Methods Mol Biol 2010;628:215-26.
30. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nat Rev Genet 2019;20:631-56.
31. Hong M, Tao S, Zhang L, Diao LT, Huang X, Huang S, et al. RNA sequencing: new technologies and applications in cancer research. J Hematol Oncol 2020;13:166.
32. Feng L, Lou J. DNA methylation analysis. Methods Mol Biol 2019;1894:181-227.
33. Wreczycka K, Gosdschan A, Yusuf D, Grüning B, Assenov Y, Akalin A. Strategies for analyzing bisulfite sequencing data. J Biotechnol 2017;261:105-15.
34. Yan F, Powell DR, Curtis DJ, Wong NC. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. Genome Biol 2020;21:22.
35. Grandi FC, Modi H, Kampman L, Corces MR. Chromatin accessibility profiling by ATAC-seq. Nat Protoc 2022;17:1518-52.