

REVIEW ARTICLE OPEN



Implementation of multi-omics in diagnosis of pediatric rare diseases

Sara S. Ali¹, Qifei Li¹ and Pankaj B. Agrawal¹✉

© The Author(s) 2024

The rapid and accurate diagnosis of rare diseases is paramount in directing clinical management. In recent years, the integration of multi-omics approaches has emerged as a potential strategy to overcome diagnostic hurdles. This review examines the application of multi-omics technologies, including genomics, epigenomics, transcriptomics, proteomics, and metabolomics, in relation to the diagnostic journey of rare diseases. We explore how these combined approaches enhance the detection of pathogenic genetic variants and decipher molecular mechanisms. This review highlights the groundbreaking potential of multi-omics in advancing the precision medicine paradigm for rare diseases, offering insights into future directions and clinical applications.

Pediatric Research (2025) 97:1337–1344; <https://doi.org/10.1038/s41390-024-03728-w>

IMPACT:

- This review discusses using current tests and emerging technologies to diagnose pediatric rare diseases.
- We describe the next steps after inconclusive molecular testing and a structure for using multi-omics in further investigations.
- The use of multi-omics is expanding, and it is essential to incorporate it into clinical practice to enhance individualized patient care.

INTRODUCTION

Approximately 7000 rare disorders affect over 30 million Americans, many of which lack a genetic diagnosis.¹ Notably, around 70% of these rare genetic disorders exhibit symptoms in early childhood.²

These rare disorders inflict significant economic and social burden. Yang and colleagues estimated the total economic burden of 379 rare diseases in 2019 at \$997 billion. This comprised 45% of direct medical costs, 44% in indirect costs due to lost productivity, 7% in non-medical costs, and 4% in uninsured healthcare costs.³

Beyond financial implications, rare diseases also decrease families' quality of life. Understanding the genetic cause of the disease can improve management strategies and inform families about potential risks for future generations, prompting consideration of alternate family planning options.

Advancements in genomic technologies have enhanced diagnostic capabilities through whole exome, genome, epigenome, transcriptome, and proteome testing. Additionally, the rapid decline in sequencing costs over the past decade has allowed for large-scale studies.

This review aims to inform clinicians, trainees, and researchers about available multi-omics technologies, their application in rare genetic disorders, and their limitations in clinical settings.

Understanding the need for a Multi-omics Approach

The complex relationship between DNA, RNA, and proteins forms the foundation of genetic information transfer. DNA serves as the

genetic material, carrying instructions that are transcribed into RNA, which is crucial for protein synthesis. A comprehensive understanding of molecular processes requires integrating genomics, epigenomics (DNA), transcriptomics (RNA), proteomics (proteins), and metabolomics (metabolites). These multi-omics approaches help decipher the molecular mechanisms that connect genotype to phenotype.⁴

Introduction to Genomics

An organism's complete DNA set is called its genome. Nearly every cell in the human body contains a full copy of approximately 3 billion DNA base pairs that make up this genome. Genomics is the comprehensive study of the structure, function, mapping, evolution, and editing of the information encoded in our genomes.

DNA sequencing determines the exact order of bases in a DNA strand. This process is crucial for identifying genetic variations or mutations that may contribute to disease development or progression. These variations can range from small substitutions, deletions, or additions of single base pairs to large deletions of thousands of bases.

In recent decades, genomic medicine has achieved significant milestones. These include the first draft of the human genome by the National Human Genome Research Institute, the use of microarrays for detecting DNA deletions or duplications, and the emergence of whole exome and genome sequencing (WES/WGS) for identifying pathogenic variants at the nucleotide level.^{5,6}

¹Division of Neonatology, Department of Pediatrics, University of Miami Miller School of Medicine and Holtz Children's Hospital, Jackson Health System, Miami, FL, USA.

✉email: pagrawal@miami.edu

Received: 20 March 2024 Revised: 24 October 2024 Accepted: 28 October 2024

Published online: 19 November 2024

Continued advancements in genome-based research are improving diagnostics, therapeutic strategies, evidence-based clinical efficacy, and decision-making tools for patients and providers.

WHOLE EXOME OR GENOME SEQUENCING (WES/WGS)

WES and WGS are increasingly used as first-line genetic tests for evaluating rare diseases.^{7–9} WES focuses on 2% of the genome, primarily the coding regions, while WGS covers the entire genome, including non-coding regions. WGS is superior as it analyzes an individual's entire DNA, identifying single nucleotide variants (SNVs), copy number variants (CNVs), small insertions or deletions, and chromosomal and mitochondrial changes.¹⁰ WES and WGS are often performed as trio tests with the proband and biological parents, which enhances variant interpretation and diagnostic yield, thus shortening the diagnostic process.

Utility of WES/WGS

Traditional diagnostic methods such as single gene sequencing or gene panels, often fail to identify the genetic basis of many rare diseases compared to WES and WGS.¹¹ For instance, a study by the Australian National Acute Care Genomics program offered WGS to 290 families with critically ill infants suspected of genetic conditions. Results were obtained in an average of 2.9 days, with a diagnostic yield of 47%.^{12,13} Similarly, a recent meta-analysis of 23 studies in 2022 involving 1567 critically ill infants found a pooled diagnostic utility of 42% for rapid genome sequencing ($p < 0.1$).¹⁴ Additionally, a systematic review of 21 prospective studies with 1654 infants under one year old found a mean of 46% positive genetic test results, with 37% benefitting from rapid genome testing.¹⁵

WGS, which sequences both intronic and exonic regions, has shown a higher diagnostic yield compared to WES. Studies like NISIGHT1 and NICUseq have highlighted WGS's ability to identify various genomic variants, including CNVs and SNVs.¹⁶ A study in NEJM indicated that WGS can identify 8% of cases that WES might miss.^{17,18}

Rapid WES/WGS

With rapid technological advancements, the turnaround time for obtaining clinical reports has significantly decreased. Tests are considered rapid if results are available in less than two weeks.¹⁹ This urgency is particularly critical in ICU settings, as high care costs necessitate prompt decision-making. Many studies have demonstrated the cost-effectiveness and actionable insights of rapid WES and WGS in NICU and PICU.^{19–23} From 2012 to 2021, 33 clinical studies documented the diagnostic and clinical utility of first-tier rapid whole genome sequencing in those settings.^{20,21}

Turnaround time can be substantially reduced with the availability of adequate resources. For example, in 2012, disease-causing mutations in critically ill newborns were identified within 50 hours of sample collection. A recent study involving 12 patients yielded a diagnosis in as little as 5 hours for one case.^{24,25}

Role of WES/WGS in Newborn Screening

Universal newborn screening (NBS) has been pivotal for early diagnosis of various conditions, including rare inborn errors of metabolism present at birth. Certain conditions like phenylketonuria, may not exhibit symptoms initially but may lead to permanent damage if left untreated. Traditional newborn screening methods have limitations, capturing only a limited number of conditions. The emergence of WES/WGS could enhance newborn screening by enabling a comprehensive and precise diagnosis of a wider range of genetic conditions.²⁶

Current NBS methods yield few false-negative results but produce many false positives results, leading to emotional and financial strain on families. For instance, a study screening 176,186 specimens by mass spectrometry (MS/MS) revealed 51

true positives, 2 false negatives, and 454 false positives. Utilizing WES/WGS could improve the specificity of existing tests and assist in diagnosing ambiguous biochemical profiles.^{26,27}

Moreover, sequencing technologies could validate conditions identified through other methods, assist in prognosis and treatment decisions, and provide families with valuable information about pathogenic variants. This approach could also deepen our understanding of genotype-phenotype correlations and clarify factors contributing to false positives.

A pivotal study by Chen et al. demonstrated the advantages of genomic sequencing as a primary screening tool. The authors designed a targeted gene panel of 142 genes linked to 128 diseases, included in China's NBS program. By enrolling nearly 30,000 newborns, they found 59 cases undetected by traditional methods but identified by gene panel sequencing, indicating that 1 in every 500 newborns could benefit from this technology. The study suggests that a genomics-based approach to newborn screening could significantly enhance detection rates.²⁸

However, integrating sequencing into newborn screening presents challenges, including gaps in understanding genetic variants, complexities in data interpretation, and ethical concerns regarding data handling and consent. Addressing these challenges is crucial to obtain the benefits of genomic newborn screening while ensuring equitable access and outcomes.

Limitations of WES/WGS

Despite their increasing accessibility, cost-effectiveness, and speed, sequencing technologies face various challenges. Results may include 'variant of uncertain significance' (VUS), necessitating further bioinformatic analyses and transcriptome sequencing for accurate diagnosis. Reanalysis of genome data may reveal pathogenic variants years later, particularly as new gene discoveries emerge.^{29–31} A systematic review identified a 15% increase in diagnostic yield across 27 studies and recommended reanalysis approximately 18 months after the initial analysis.³²

Methods of Next Generation Sequencing (NGS) including WES/WGS

Next-generation sequencing (NGS) for WES/WGS and RNA sequencing utilizes unique sequencing chemistries and sophisticated bioinformatics to enable rapid, parallel sequencing of various DNA or RNA fragments.³³

Short-read sequencing. Short-read sequencing produces reads ranging from 50 to 300 base pairs and is widely used for high-throughput genomic analyses due to its efficiency and cost-effectiveness. Techniques such as sequencing by synthesis (SBS) and sequencing by binding (SBB) use polymerase enzymes to replicate DNA fragments to determine the nucleotide sequence of a sample.³⁴

It is typically utilized for applications in WES, WGS, gene panels, and single-gene testing. However, the limited read length can pose challenges in complex genomic regions, leading to potential disparities.³⁵

Long-read sequencing. Long-read sequencing processes DNA fragments spanning thousands of base pairs and can be categorized into 'true' and 'synthetic' long-read technologies. True long-read platforms such as PacBio and Nanopore, directly sequence long DNA strands, while synthetic long-read methods reconstruct long sequences from shorter reads.^{33,35,36}

Comparison with Optical Genome Mapping (OGM)

Optical Genome Mapping (OGM) is a non-sequencing technology by Bionano that analyzes large genomes using fluorophore-labeled DNA molecules.^{34,36} OGM can map entire chromosome arms and detect structural variations like insertions, deletions,

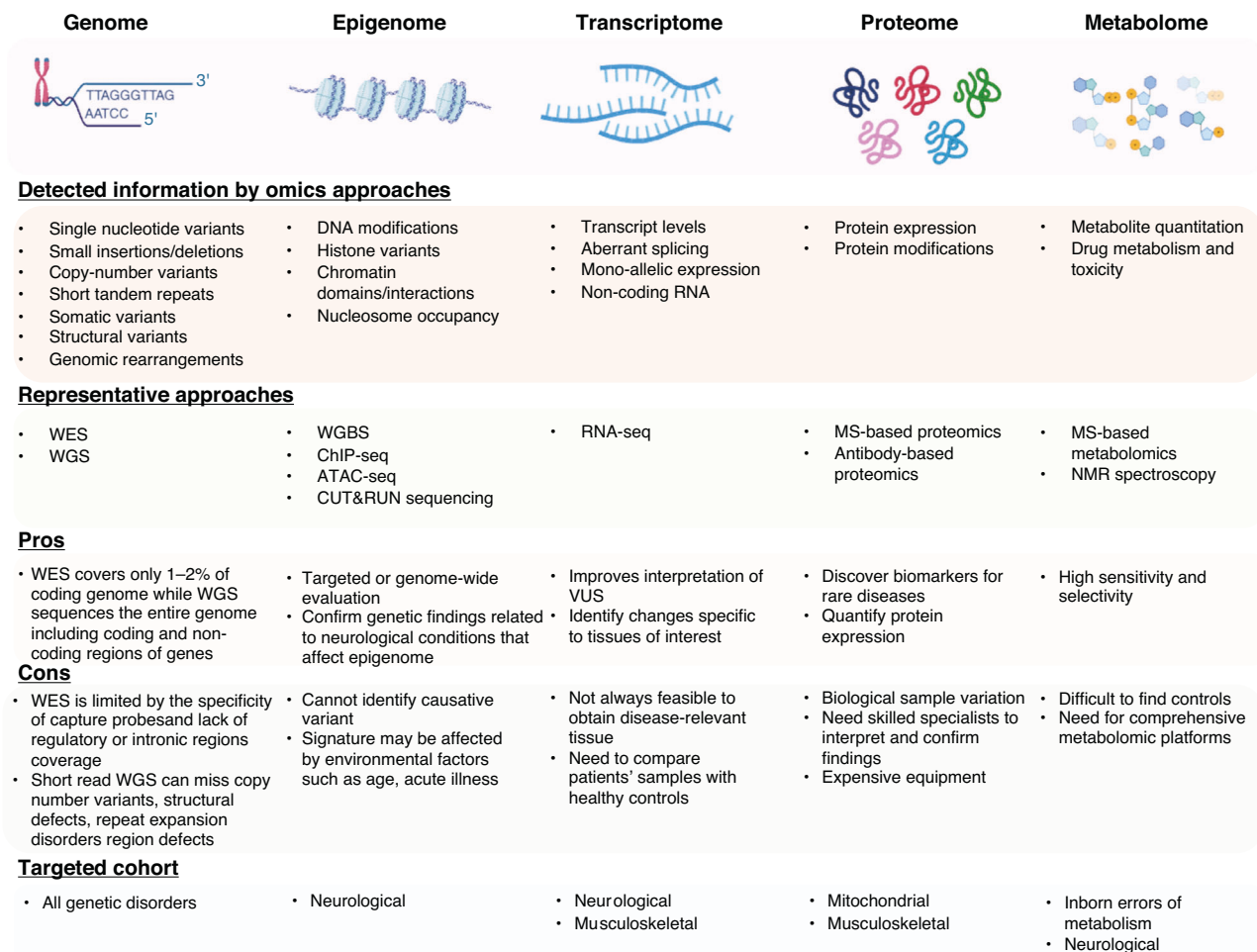


Fig. 1 Summarizes and Compares the key multi-omics branches – genomics, epigenomics, transcriptomics, proteomics, and metabolomics. It highlights the information each detects, their advantages and disadvantages, and the types of patient cohorts they typically target.

duplications, inversions, translocations, and complex rearrangements more effectively.³⁶ Several studies have shown the utility of this sequencing approach. For example, Cope and colleagues identified a mosaic deletion and inversion in *CDKL5* in a patient that remained undiagnosed despite chromosomal microarray, an epilepsy panel, and exome sequencing.³⁷ Additionally, a study comparing OGM and a 54-gene NGS panel in myeloid cancers found that OGM matched the performance of NGS in detecting cytogenetic abnormalities.³⁸ This validates that optical mapping enhances genome assembly and gap closure when integrated with NGS.

Role of Multi-omics when Genomic Sequencing is Inconclusive

When DNA sequencing is inconclusive, a multi-omics approach can provide a comprehensive understanding of molecular processes. This includes integrating epigenomics, transcriptomics, proteomics, and metabolomics (Fig. 1). Epigenomics explores genome-wide DNA modifications, transcriptomics analyzes RNA expression across all RNAs, proteomics uncovers protein structures, functions, interactions, and levels on a large scale, and metabolomics examines identification and quantification of metabolites in biological specimens.

EPIGENOMICS

Epigenetic mechanisms involve chemical modifications to DNA itself or changes to proteins, such as chromatin, which interacts

closely with DNA. Key modifications include DNA methylation, chromatin remodeling, histone modification, and noncoding RNA (ncRNA)-associated mechanisms.³⁹ Understanding these dynamic mechanisms is crucial, as the epigenome can vary between different cell types and even among individual cells, influencing gene expression in various ways – by altering nuclear architecture, modulating transcription factor access, and directly mediating gene expression.³⁹

Types of Epigenetic Modifications

DNA Methylation. DNA methylation involves adding a methyl group to cytosine residues in CpG dinucleotides, primarily leading to gene silencing. This process is regulated by three types of DNA methyltransferases (DNMT): DNMT1, DNMT3A, and DNMT3B.⁴⁰ Mutations in these enzymes are associated with various neuropathies and syndromes. For instance, *DNMT1* mutations are linked to hereditary sensory neuropathy type IE (HSANIE), mutations in *DNMT3A* are associated with overgrowth like Tatton-Brown-Rahman syndrome; while *DNMT3B* mutations are involved in immunodeficiency and intellectual disability syndromes such as immunodeficiency-centromeric instability-facial anomalies syndrome 1 (ICF 1).⁴¹

Histone Modifications. In eukaryotic cells, DNA wraps around nucleosomes formed by histone proteins, creating chromatin. The N-terminal tails of histone proteins are rich in arginine and lysine, and they undergo various post-transcriptional modifications that influence gene expression. These modifications-like methylation

and acetylation – can activate or repress transcription.³⁹ Dysregulation of these processes is linked to several rare diseases. For example, lysine methylation, which is crucial for gene expression, can lead to conditions such as Kabuki syndrome type 1 (*KMT2D*) and Sotos syndrome (*EHMT1*) when affected by mutations in specific genes.⁴²

Chromatin Remodeling. Chromatin remodeling alters the accessibility of DNA to transcription factors, influencing gene expression. This process is energy-dependent, utilizing ATP hydrolysis to modify nucleosome arrangements.⁴³ For instance, mutation in *SMARCB1*, which encodes a subunit of the SWI/SNF (switch/sucrose non-fermenting) chromatin remodeler, can reduce promoter accessibility, seen in disorders like the Coffin-Siris syndrome.⁴³

Noncoding RNA (ncRNA)-associated Mechanisms. Recent studies indicate that up to 90% of eukaryotic genomic DNA is transcribed as ncRNAs, which regulate transcription by recruiting chromatin-remodeling complexes and facilitating epigenetic changes.^{40,44} Dysregulated ncRNAs are implicated in various conditions, including cancers and neurological disorders. For instance, deletions in the long ncRNA *CHASERR* have been associated with developmental and epileptic encephalopathy while deletions in the long ncRNA *TBX2-AS1* are linked to hearing loss.⁴⁴ Another study showed that the long ncRNA *ENSG00000257522* is recurrently disrupted in individuals with microcephaly.⁴⁵

Types of Epigenomic Profiling Methods

DNA Methylation Profiling. Whole-genome bisulfite sequencing (WGBS): Combines sodium bisulfite treatment with high-throughput sequencing to map DNA methylation across the genome at single-base resolution.⁴⁶

Methylated DNA immunoprecipitation sequencing (MeDIP-seq): Enriches and sequences methylated DNA fragments to provide a genome-wide view of DNA methylation patterns.⁴⁷

Reduced representation bisulfite sequencing (RRBS): Targets CpG-rich regions for high-resolution DNA methylation analysis.⁴⁷

Histone Modification Profiling. Chromatin immunoprecipitation sequencing (ChIP-seq): Identifies genomic regions associated with specific histone modifications or DNA-binding proteins by sequencing immunoprecipitated DNA fragments.⁴⁷

Assay for cleavage under targets and release using nuclease (CUT&RUN): Isolates protein-DNA complexes for mapping chromatin proteins.⁴⁸

Chromatin immunoprecipitation with exonuclease treatment (ChIP-exo): Enhances resolution of ChIP-seq by incorporating exonuclease digestion to map protein-DNA interactions precisely.⁴⁷

Chromatin Accessibility Profiling. Assay for transposase-accessible chromatin using sequencing (ATAC-seq): Utilizes Tn5 transposase to insert sequencing adapters into open chromatin regions, providing insights into active regulatory elements.⁴⁹

DNase-seq: Digests DNA at open chromatin regions using DNase 1, followed by sequencing to identify accessible chromatin regions.⁴⁷

RNA Modification Profiling. N6-methyladenosine sequencing (m6A-seq): Profiles RNA modifications by immunoprecipitating m6A-modified RNA fragments followed by sequencing.⁴⁷

Single-cell Epigenomic Profiling. Single-cell ATAC-seq (scATAC-seq): evaluates chromatin accessibility at the single-cell level, revealing epigenetic heterogeneity within cell populations.⁴⁹

These methods contribute to our understanding of how epigenetic modifications regulate gene expression and cellular

function, providing insights into development, disease mechanisms, and potential therapeutic targets.^{49,50}

Advancements and Limitations

Recent literature demonstrates that DNA methylation testing can diagnose up to 30% of individuals with rare neurodevelopmental conditions.⁵¹ This yield is comparable to the solve rates reported for chromosomal microarray (15–20%) and exome sequencing (30–40%).⁵¹ This suggests integrating methylation profiling into initial diagnostic assessments for specific phenotypes, such as neurodevelopmental disorders, suspected imprinting disorders, repeat expansion disorders, or a VUS in a known methylation gene.⁵²

While histone modifications are imperative in neurogenesis, their roles in pediatric neurodevelopmental diseases remain poorly understood.^{53,54} Epigenetic profiles can vary significantly across different tissues, meaning blood sample findings may not always apply to the affected tissues. Additionally, the heterogeneity between phenotype and genotype further complicates the analysis of experimental results.⁴¹ Lastly, achieving consensus in DNA methylation analysis requires adjusting for factors such as age, sex, environmental influences, and other variables that could influence outcomes.⁴²

TRANSCRIPTOMICS

Transcriptomics studies the expression of all RNAs in a cell population, providing insights into molecular changes induced by environmental factors or pathogens. RNA sequencing has emerged as a complementary assay for rare disease diagnosis.⁵⁵

There are four ways to analyze RNA sequencing data: expression outliers, aberrant splicing, allele-specific expression, and transcriptomic structural variants.

Approaches to analyze RNA Sequencing Data

Expression Outliers. Expression outliers identify genes with unusually high or low expression levels in a sample. This analysis together with the patient's phenotype, helps pinpoint strong candidate variants for clinical interpretation.⁵⁶

Aberrant Splicing. Aberrant splicing occurs when errors disrupt the natural splicing of pre-mRNA, leading to rare diseases. Typically, noncoding sequences (introns) are removed, and certain exons are included or excluded from processed mRNA. Errors can involve exon skipping, inclusion of pseudoexons, exon extension, and intron retention, among others.⁵⁷

Allele-specific Expression. Allele-specific expression (ASE) occurs when one allele is expressed at significantly higher levels than the other. In WES/WGS, single heterozygous rare variants are usually separated; however, some may exhibit ASE.⁵²

Transcriptomic Structural Variants. Structural variants (SVs) such as translocations, duplications, inversions, and deletions can join or separate genomic regions. This can lead to gene fusions (exons from two or more distinct genes transcribed together) or altered gene functions when non-transcribed regions are included in a gene. Modifications in transcribed mRNA due to these genomic SVs are referred to as transcriptomic structural variants.^{56–58}

Advancements and Limitations

RNA sequencing can increase the diagnostic yield when WES/WGS is inconclusive. In a cohort of 113 subjects at the University of California, the diagnostic rate for WES/WGS was 31%, and incorporating RNA sequencing increased this rate to 38%.⁵⁹ In patients with autism spectrum disorder, blood RNA sequencing identified intronic mutations and deregulated expressions in genes such as *PTEN* and *MECP2*, which had been missed by WES.⁶⁰

As RNA sequencing technology and algorithms improve, its application in rare disease research will continue to accelerate. A comprehensive analysis of the transcriptome in each cell could uncover novel cellular and molecular components in tissues affected by rare diseases.

PROTEOMICS

Proteins are the end product of the central dogma of molecular biology, playing a pivotal role in various cellular functions. While proteomics has historically had a lower output compared to other 'omics technologies, it remains a valuable tool for uncovering abnormalities in protein synthesis, stability, degradation, and signaling.

Proteomics can be divided into three dimensions: expression, structural, and functional. Expression proteomics focuses on the qualitative and quantitative changes in protein composition. Structural proteomics aims on characterizing protein structures in specific cell types or organelles, while functional proteomics assesses the biological functions and mechanisms of proteins.⁶¹

Recent technological advancements have significantly evolved proteomics methods. Traditional techniques, such as immunohistochemistry (IHC) staining, western blotting, and enzyme-linked immunosorbent assay (ELISA), have been supplemented by high-throughput methods like tissue microarray (TMA), protein pathway array (PPA), and mass spectrometry (MS). These modern techniques not only reduce analysis time but also increase the accuracy and depth of proteome coverage.⁶²

High-Throughput Proteomic Techniques

Mass Spectrometry (MS). Mass spectrometry (MS) has emerged as one of the most essential tools in proteomics. It is crucial for identifying proteins and their isoforms, as well as quantifying post-translational modifications. MS can directly detect intact proteins or specific peptide fragments, which is often challenging with traditional immunoassays. By integrating MS with various separation and pre-fractionation techniques, researchers can significantly enhance identification accuracy and yield.^{62,63}

Protein Pathway Array (PPA). Protein pathway array (PPA) is a gel-based high-throughput platform that uses antibody mixtures to detect antigens in protein samples extracted from biopsies or tissue. Immunofluorescence signals from antibody-antigen reactions are converted into numerical data, which provides insights into protein expression levels. This data can be further analyzed to explore biomarkers and proteomic networks.^{62,63}

Next generation Tissue Microarrays (TMAs). Tissue Microarrays (TMAs) consist of numerous small tissue cores from formalin-fixed paraffin-embedded (FFPE) or frozen blocks, arranged on a single histologic slide. This preparation allows for large-scale antibody-based molecular analysis of multiple samples simultaneously. TMAs are valuable for validating new biomarkers identified through PPA or MS, as well as for locating target proteins within cellular compartments such as the cell membrane, cytoplasm, or nucleus.^{62,63}

Advancements and Limitations

In the last few decades, there have been notable advancements in the field of proteomics. The UK Biobank Pharma Proteomics Project (UKB-PPP) used plasma protein signatures to develop prediction models for over 200 diseases in 41,931 patients. For 67 of these diseases, models using just 5 to 20 proteins outperformed traditional clinical models, demonstrating a median delta C-index of 0.07 (range = 0.02–0.31).⁶⁴

Moreover, Kopajtic et al. utilized proteomics, to identify the genetic origins of rare mitochondrial diseases. Using tandem mass tag-labeled proteomics to fibroblast cell lines from 145 individuals,

they identified approximately 8,000 proteins per sample, covering over 50% of Mendelian disease-associated genes.⁶⁵ Similarly, in Spain researchers are increasingly using MS-based strategies and other proteomics methods to investigate rare inherited metabolic diseases such as methylmalonic aciduria, Fabry disease, various mucopolysaccharidoses.⁶¹

A significant advancement in this field is the launch of The Human Proteome Project (HPP), which seeks to map the entire human proteome using both current and emerging techniques. This initiative is expected to further enhance our understanding of rare diseases at the cellular level and open new avenues for developing therapeutic and diagnostic solutions. Recently updated guidelines for interpreting proteomic data have been published.⁶⁶

Despite these advancements, challenges remain. Issues such as pre-analytical variables, analytical variability, and biological sample variation must be addressed to improve the performance and reproducibility of proteomics across laboratories. Additionally, the reliance on specialized operators and expensive equipment limits the routine use of proteomic tools in clinical settings. Nevertheless, as technology progresses, proteomics is expected to become a pivotal method for identifying disease biomarkers and developing more precise biochemical and immunological tests.⁶¹

METABOLOMICS

Metabolomics is a comprehensive analytical approach used to study metabolites within biological specimens. Modern metabolomic technologies surpass traditional clinical chemistry techniques, enabling precise analyzes of hundreds to thousands of metabolites. Integrating metabolomics data with WES/WGS data is crucial for identifying genes associated with disease mechanisms, especially in rare and unexplained metabolic disorders in children. Unlike genes, transcripts, and proteins, metabolites have diverse physiochemical properties, necessitating multiple bioanalytical techniques for measurement, as no single technique can capture metabolite types.

Techniques in Metabolomics

Nuclear Magnetic Resonance (NMR). NMR is a spectroscopic technique that observes energetic transition of nuclear spins within a strong magnetic field. It is essential for identifying and interpreting the structures of organic molecules and metabolites, studying the dynamics of macromolecules like proteins and nucleic acids, and advancing metabolomics research. An important advantage of NMR is its reproducibility among laboratories. Standardizing procedures has become increasingly feasible, particularly for clinical applications such as the analyzing human urine, blood serum, and plasma. For instance, NMR has been used to analyze blood plasma samples from approximately 121,000 participants in the UK Biobank, resulting in a clinical chemistry panel of 249 biomarkers and ratios based on signals from lipoproteins, lipids, amino acids, and glycolysis intermediates.^{67,68}

Mass Spectrometry (MS). MS measures the mass-to-charge ratio (m/z) of ions to identify and quantify molecules in both simple and complex mixtures. This highly sensitive method can simultaneously detect and quantify thousands of metabolite features. MS is effective for the detecting, quantifying, and elucidating the structure of several hundred metabolites in a single measurement. The sensitivity and accuracy of MS depend on various experimental conditions and instrumental settings, including metabolite extraction, separation, ionization, and detection. Given the complexity of biological matrices, separating metabolites of interest before MS analysis is often necessary. Thus, analytical techniques that combine separation methods with MS, such as high-performance liquid chromatography (HPLC), gas chromatography (GS), and capillary electrophoresis (CE), have become

highly effective for small-molecule analysis. However, achieving high resolution and sensitivity in a single MS detection mode is challenging, as higher sensitivity often compromises resolution, and vice versa. While MS can detect approximately 1000 metabolites, NMR can measure up to 100.⁶⁹

Advances and Limitations

Recent studies highlight the advancements in metabolomics. For example, Abela et al. used a combined genetic-metabolic approach to investigate early-onset epileptic encephalopathies in a cohort of 63 patients. Their untargeted metabolomic analysis identified two novel potential plasma biomarkers for Snyder-Robinson Syndrome and infantile cerebellar retinal degeneration.⁷⁰ The development of processing tools and metabolomic databases has significantly improved the diagnosis of inherited errors of metabolism (IEM), with identified biomarkers now incorporated into newborn IEM screenings.⁷¹ Another study analyzed data from 19,994 patients, including plasma levels of over 900 metabolites, revealing 2599 variant-metabolite associations across 330 genomic regions.⁷²

The NIH Common Fund Undiagnosed Diseases Network (UDN) developed a metabolomics program responsible for improving metabolomic methods and making them more accessible to the scientific community. This program developed the National Metabolomics Data Repository (NMDR), which houses 2397 studies from over 40 countries, containing 75,000 samples and 32,000 known metabolites. It also provides access to protocols and analytic tools that facilitate the analysis of metabolomic data through cloud computing.⁷³

Biochemical changes often precede anatomical changes, suggesting that these biomarkers might predict disease conditions even in asymptomatic stages. This allows for early and effective treatment, potentially reducing complications and mortality rates. As metabolomics advances, it holds promise for enhancing our understanding of disease pathophysiology and contributing to novel treatment approaches. However, the field is still relatively new, highlighting the need for improved specialized training, reduced reliance on high-cost analytical instrumentation, and standardized operating procedures to decrease pre-analytical errors.⁷⁴

Challenges in Diagnosing Rare Diseases

Despite these technological advancements, diagnosing rare diseases poses several challenges. Many clinicians struggle to interpret these novel findings, creating a demand for specialists who can translate results into clinical insights. Testing is often unavailable in low-resource medical settings, necessitating referrals to specialized centers, which can lead to increased time and costs. Advanced diagnostic methods, including WES/WGS, DNA methylation, RNA sequencing, proteomics, and metabolomics require sophisticated technologies, data repositories, bioinformatic pipelines, and age/tissue-matched controls, further complicating accessibility.

Continuous funding for existing research programs dedicated to diagnosing rare and undiagnosed diseases is essential. While ongoing advancements may lead to a decline in the cost of advanced technologies, targeted sequencing, state-of-the-art computational algorithms, data-sharing platforms, and machine learning approaches will be critical in overcoming challenges related to variant detection and interpretation.

CONCLUSION

Over the past two decades, gene panels, microarrays, and exome/genome sequencing have identified causal mutations in many rare disease patients, but a significant proportion remain undiagnosed. This review explores various approaches to enhance

diagnostic yield and uncover the molecular mechanisms behind these diseases.

Although challenges such as cost and the need for specialists persist; there is optimism that continuous improvements will help close the diagnostic gap for undiagnosed patients; ultimately providing actionable findings. Successful implementation of multi-omics approaches in rare diseases is anticipated to revolutionize how undiagnosed patients are diagnosed and treated.

DATA AVAILABILITY

This review article does not contain any primary data collection, and therefore, no new data were generated or analyzed as part of this study. The article is based solely on a comprehensive literature review of existing publications, which are cited within the text.

REFERENCES

- Haendel, M. et al. How many rare diseases are there? *Nat. Rev. Drug Discov.* **19**, 77–78 (2020).
- Batshaw, M. L., Groft, S. C. & Krischer, J. P. Research into rare diseases of childhood. *JAMA* **311**, 1729–1730 (2014).
- Yang, G. et al. The national economic burden of rare disease in the United States in 2019. *Orphanet J. Rare Dis.* **17**, 163 (2022).
- Wekesa, J. S. & Kimwele, M. A Review of Multi-Omics Data Integration through Deep Learning Approaches for Disease Diagnosis, Prognosis, and Treatment. *Frontiers in Genetics* **14** (2023).
- Collins, F. S. & Fink, L. The human genome project. *Alcohol Health Res World* **19**, 190–195 (1995).
- Caspar, S. M. et al. Clinical sequencing: from raw data to diagnosis with lifetime value. *Clin. Genet.* **93**, 508–519 (2018).
- Morton, J. L. U. et al. Multicenter consensus approach to evaluation of neonatal hypotonia in the genomic era: a review. *JAMA Neurol.* **79**, 405–413 (2022).
- Vinkškel, M., Writzl, K., Maver, A. & Peterlin, B. Improving diagnostics of rare genetic diseases with Ngs approaches. *J. Community Genet.* **12**, 247–256 (2021).
- Ewans, L. J. et al. Whole exome and genome sequencing in mendelian disorders: a diagnostic and health economic analysis. *Eur. J. Hum. Genet.* **30**, 1121–1131 (2022).
- Satam, H. et al. Next-generation sequencing technology: current trends and advancements. *Biol. (Basel)* **12**, 997 (2023).
- Maron, J. L. U. et al. Rapid whole-genomic sequencing and a targeted neonatal gene panel in infants with a suspected genetic disorder. *Jama* **330**, 161–169 (2023).
- Lunke, S. et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat. Med.* **29**, 1681–1691 (2023).
- Vears, D. F., Lynch, F., Nisselle, A., Ayres, S. & Stark, Z. Rapid genomic testing in critically ill patients with genetic conditions: position statement by the human genetics society of Australasia. *Eur. J. Hum. Genet.* **32**, 150–154 (2024).
- Xiao, F. et al. Diagnostic utility of rapid sequencing in critically ill infants: a systematic review and meta-analysis. *Expert Rev. Mol. Diagn.* **22**, 833–840 (2022).
- Callahan, K. P., Mueller, R., Flibotte, J., Largent, E. A. & Feudtner, C. Measures of utility among studies of genomic medicine for critically ill infants: a systematic review. *JAMA Netw. Open* **5**, e2225980 (2022).
- Petrikin, J. E. et al. The nsight1-randomized controlled trial: rapid whole-genome sequencing for accelerated etiologic diagnosis in critically ill infants. *npj Genom. Med.* **3**, 6 (2018).
- Belkadi, A. et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl Acad. Sci. USA* **112**, 5473–5478 (2015).
- Wojcik, M. H. et al. Genome sequencing for diagnosing rare diseases. *N. Engl. J. Med.* **390**, 1985–1997 (2024).
- Kingsmore, S. F., Nofsinger, R. & Ellsworth, K. Rapid genomic sequencing for genetic disease diagnosis and therapy in intensive care units: a review. *npj Genom. Med.* **9**, 17 (2024).
- Kingsmore, S. F. & Cole, F. S. The role of genome sequencing in neonatal intensive care units. *Annu Rev. Genomics Hum. Genet.* **23**, 427–448 (2022).
- Clark, M. M. et al. Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Sci. Transl. Med.* **11**, eaat6177 (2019).
- Wu, B. et al. Application of full-spectrum rapid clinical genome sequencing improves diagnostic rate and clinical outcomes in critically ill infants in the China neonatal genomics project. *Crit. Care Med.* **49**, 1674–1683 (2021).
- Gubbels, C. S. et al. Prospective, phenotype-driven selection of critically ill neonates for rapid exome sequencing is associated with high diagnostic yield. *Genet. Med.* **22**, 736–744 (2020).

24. Seydel, C. Baby's first genome. *Nat. Biotechnol.* **40**, 636–640 (2022).
25. Saunders, C. J. et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci. Transl. Med.* **4**, 154ra135 (2012).
26. Berg, J. S. et al. Newborn sequencing in genomic medicine and public health. *Pediatrics* **139**, e20162252 (2017).
27. Jiang, S., Wang, H. & Gu, Y. Genome sequencing for newborn screening—an effective approach for tackling rare diseases. *JAMA Netw. Open* **6**, e2331141 (2023).
28. Chen, T. et al. Genomic sequencing as a first-tier screening test and outcomes of newborn screening. *JAMA Netw. Open* **6**, e2331162–e2331162 (2023).
29. Hills, S. et al. High number of candidate gene variants are identified as disease-causing in a period of 4 years. *Am. J. Med Genet A* **194**, e63509 (2024).
30. Li, Q. et al. Reanalysis of clinical exome identifies the second variant in two individuals with recessive disorders. *Eur. J. Hum. Genet* **31**, 712–715 (2023).
31. Schmitz-Abe, K. et al. Unique bioinformatic approach and comprehensive reanalysis improve diagnostic yield of clinical exomes. *Eur. J. Hum. Genet* **27**, 1398–1405 (2019).
32. Tan, N. B. et al. Evaluating systematic reanalysis of clinical genomic data in rare disease from single center experience and literature review. *Mol. Genet Genom. Med* **8**, e1508 (2020).
33. Qin, D. Next-generation sequencing and its clinical application. *Cancer Biol. Med* **16**, 4–10 (2019).
34. Slatko, B. E., Gardner, A. F. & Ausubel, F. M. Overview of next-generation sequencing technologies. *Curr. Protoc. Mol. Biol.* **122**, e59 (2018).
35. Conlin, L. K., Aref-Eshghi, E., McEldrew, D. A., Luo, M. & Rajagopalan, R. Long-read sequencing for molecular diagnostics in constitutional genetic disorders. *Hum. Mutat.* **43**, 1531–1544 (2022).
36. Dremsek, P. et al. Optical genome mapping in routine human genetic diagnostics—its advantages and limitations. *Genes (Basel)* **12**, 1958 (2021).
37. Cope, H. et al. Detection of a mosaic Cdkl5 deletion and inversion by optical genome mapping ends an exhaustive diagnostic odyssey. *Mol. Genet Genom. Med* **9**, e1665 (2021).
38. Sahajpal, N., Mondal, A., Jillella, A. P., Kota, V. & Kolhe, R. Optical genome mapping and comprehensive next generation sequencing panel for myeloid cancers: new era of complimentary profiling. *Blood* **140**, 9175–9176 (2022).
39. Wang, K. C. & Chang, H. Y. Epigenomics: technologies and applications. *Circ. Res* **122**, 1191–1199 (2018).
40. Bure, I. V., Nemtsova, M. V. & Kuznetsova, E. B. Histone modifications and non-coding RNAs: mutual epigenetic regulation and role in pathogenesis. *Int. J. Mol. Sci.* **23**, 5801 (2022).
41. Fu, M. P. et al. Rare diseases of epigenetic origin: challenges and opportunities. *Front Genet* **14**, 1113086 (2023).
42. Martinez-Delgado, B. & Barrero, M. J. Epigenomic Approaches for the Diagnosis of Rare Diseases. *Epigenomes* **6** (2022).
43. Gourisankar, S., Krokhotin, A., Wenderski, W. & Crabtree, G. R. Context-specific functions of chromatin remodellers in development and disease. *Nat. Rev. Genet.* **25**, 340–361 (2024).
44. Kaikkonen, M. U., Lam, M. T. & Glass, C. K. Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res* **90**, 430–440 (2011).
45. Andersen, R. E. et al. Chromosomal structural rearrangements implicate long non-coding RNAs in rare germline disorders. *Hum. Genet* **143**, 921–938 (2024).
46. Gong, T. et al. Analysis and performance assessment of the whole genome bisulfite sequencing data workflow: currently available tools and a practical guide to advance dna methylation studies. *Small Methods* **6**, e2101251 (2022).
47. Mehrohamadi, M., Sepehri, M. H., Nazer, N. & Norouzi, M. R. A Comparative Overview of Epigenomic Profiling Methods. *Frontiers in Cell and Developmental Biology* **9** (2021).
48. Hainer, S. J. & Fazio, T. G. High-resolution chromatin profiling using cut&run. *Curr. Protoc. Mol. Biol.* **126**, e85 (2019).
49. Grandi, F. C., Modi, H., Kampman, L. & Corces, M. R. Chromatin accessibility profiling by Atac-seq. *Nat. Protoc.* **17**, 1518–1552 (2022).
50. Cao, B. et al. The performance of whole genome bisulfite sequencing on Dnaseq-Tx platform examined by different library preparation strategies. *Heliyon* **9**, e16571 (2023).
51. Srivastava, S. et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med* **21**, 2413–2421 (2019).
52. Wojcik, M. H. et al. Beyond the exome: what's next in diagnostic testing for mendelian conditions. *Am. J. Hum. Genet* **110**, 1229–1248 (2023).
53. Park, J., Lee, K., Kim, K. & Yi, S.-J. The role of histone modifications: from neurodevelopment to neurodegeneration. *Signal Transduct. Target. Ther.* **7**, 217 (2022).
54. Maze, I., Noh, K.-M. & Allis, C. D. Histone regulation in the CNS: basic principles of epigenetic plasticity. *Neuropsychopharmacology* **38**, 3–22 (2013).
55. Montgomery, S. B., Bernstein, J. A. & Wheeler, M. T. Toward transcriptomics as a primary tool for rare disease investigation. *Cold Spring Harb. Mol. Case Stud.* **8**, a006198 (2022).
56. Marwaha, S., Knowles, J. W. & Ashley, E. A. A guide for the diagnosis of rare and undiagnosed disease: beyond the exome. *Genome Med* **14**, 23 (2022).
57. Black, D. L. Mechanisms of alternative pre-messenger Rna splicing. *Annu. Rev. Biochem* **72**, 291–336 (2003).
58. Tung, J. et al. Allele-specific gene expression in a wild nonhuman primate population. *Mol. Ecol.* **20**, 725–739 (2011).
59. Lee, H. et al. Diagnostic utility of transcriptome sequencing for rare mendelian diseases. *Genet Med* **22**, 490–499 (2020).
60. Codina-Solà, M. et al. Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. *Mol. Autism* **6**, 21 (2015).
61. Chantada-Vázquez, M. D. P., Bravo, S. B., Barbosa-Gouveia, S., Alvarez, J. V. & Couce, M. L. Proteomics in inherited metabolic disorders. *Int. J. Mol. Sci.* **23**, 14744 (2022).
62. Cui, M., Cheng, C. & Zhang, L. High-throughput proteomics: a methodological mini-review. *Lab. Invest.* **102**, 1170–1181 (2022).
63. Aslam, B., Basit, M., Nisar, M. A., Khurshid, M. & Rasool, M. H. Proteomics: technologies and their applications. *J. Chromatogr. Sci.* **55**, 182–196 (2017).
64. Carrasco-Zanini, J. et al. Proteomic Prediction of Common and Rare Diseases. *medRxiv*, 2023.2007.2018.23292811 (2023).
65. Kopajtic, R. et al. Integration of Proteomics with Genomics and Transcriptomics Increases the Diagnostic Rate of Mendelian Disorders. *medRxiv*, 2021.2003.2009.21253187 (2021).
66. Adhikari, S. et al. A high-stringency blueprint of the human proteome. *Nat. Commun.* **11**, 5301 (2020).
67. Moco, S. Studying metabolism by Nmr-based metabolomics. *Front Mol. Biosci.* **9**, 882487 (2022).
68. Julkunen, H. et al. Atlas of plasma Nmr biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat. Commun.* **14**, 604 (2023).
69. Gowda, G. A. & Djukovic, D. Overview of mass spectrometry-based metabolomics: opportunities and challenges. *Methods Mol. Biol.* **1198**, 3–12 (2014).
70. Abela, L. et al. A Combined Metabolic–Genetic Approach to Early-Onset Epileptic Encephalopathies: Results from a Swiss Study Cohort. *Neuropediatrics* **47** (2016).
71. Cossu, M. et al. Metabolomic studies in inborn errors of metabolism: last years and future perspectives. *Metabolites* **13**, 447 (2023).
72. Surendran, P. et al. Rare and common genetic determinants of metabolic individuality and their effects on human health. *Nat. Med* **28**, 2321–2332 (2022).
73. Sud, M. et al. Metabolomics workbook: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res* **44**, D463–D470 (2016).
74. Di Minno, A. et al. Challenges in metabolomics-based tests, biomarkers revealed by metabolomic analysis, and the promise of the application of metabolomics in precision medicine. *Int. J. Mol. Sci.* **23**, 5213 (2022).

ACKNOWLEDGEMENTS

No funding sources have supported this scoping review.

AUTHOR CONTRIBUTIONS

S.S.A drafted the manuscript. P.B.A edited and reviewed the manuscript. Q.L designed the figure and contributed to manuscript editing. All authors approved the final version of the manuscript for publication.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Pankaj B. Agrawal.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the

article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024