Received: 13 March 2023 | Revised: 26 July 2023 | Accepted: 27 July 2023

DOI: 10.1002/jimd.12663

REVIEW

JIMD SSIEM WILEY

Integrative omics approaches to advance rare disease diagnostics

Dmitrii Smirnov[1,2] | **Nikita Konstantinovskiy**[1] | **Holger Prokisch**[1,2]

[1]School of Medicine, Institute of Human Genetics, Technical University of Munich, Munich, Germany

[2]Institute of Neurogenomics, Computational Health Center, Helmholtz Munich, Neuherberg, Germany

Correspondence
Holger Prokisch, Helmholtz Munich, Institute of Neurogenomics, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany.
Email: prokisch@helmholtz-muenchen.de

Funding information
Bayerisches Staatsministerium für Gesundheit und Pflege (DE), Grant/Award Number: DMB-1805-0002; Bundesministerium für Bildung und Forschung, Grant/Award Number: 01GM1906B; PerMiM Personalized Mitochondrial Medicine, Grant/Award Number: 01KU2016A; E-Rare project GENOMIT, Grant/Award Number: 01GM1207; European Union's Horizon 2020 research and innovation program project, Grant/Award Number: 101080997

Abstract

Over the past decade high-throughput DNA sequencing approaches, namely whole exome and whole genome sequencing became a standard procedure in Mendelian disease diagnostics. Implementation of these technologies greatly facilitated diagnostics and shifted the analysis paradigm from variant identification to prioritisation and evaluation. The diagnostic rates vary widely depending on the cohort size, heterogeneity and disease and range from around 30% to 50% leaving the majority of patients undiagnosed. Advances in omics technologies and computational analysis provide an opportunity to increase these unfavourable rates by providing evidence for disease-causing variant validation and prioritisation. This review aims to provide an overview of the current application of several omics technologies including RNA-sequencing, proteomics, metabolomics and DNA-methylation profiling for diagnostics of rare genetic diseases in general and inborn errors of metabolism in particular.

KEYWORDS

episignatures, methylomics, multi-omics, proteomics, rare genetic disorders, RNA sequencing

1 | INTRODUCTION

Historically, molecular diagnosis began with phenotype. Clinical specialists catalogued multiple symptoms, performed biochemical analyses and collected molecular data in order to narrow down the number of potential disorders and associated genes to screen for pathogenic variants. In some conditions, a specific metabolic change has a high sensitivity for a corresponding disease and is indicative of a single gene molecular test, but this is the exception rather than the rule. More comprehensive genetic testing is often required. Since whole-exome sequencing (WES) and whole-genome sequencing (WGS) have become more affordable, they are routinely used in many centres worldwide. The power of these methods is particularly evident when there are no unique candidate genes or when no pathogenic variants have been found in disease-related genes. They enable genome-wide screening for disease-causing variants, which changes the diagnostic paradigm from variant identification to interpretation.

However, in large cohorts that do not apply specific case selection, the diagnostic yield of WES reaches around 35% with an upper limit of about 50%.[2] Given

that WES profiles only ~1% of the human genome and has low performance in detecting structural variation, it was expected that WGS would close the gap in diagnostic yield. However, WGS provides only a modest 5%–10% increase in diagnostic performance,[2] suggesting that advances in DNA sequencing alone are necessary but mostly insufficient for molecular diagnosis. With novel sequencing approaches there is almost no lack of variant identification, rare variants can be detected in almost every disease-associated gene. There is a lack of knowledge of the functional consequence and clinical relevance of these variants. Thus, there is a great need in clinical genetics for robust high-throughput approaches for variant prioritisation and interpretation.

To address the diagnostic gap, researchers began almost a decade ago to apply a variety of omics technologies to complement clinical characterisation and DNA sequencing approaches. Omics methods provide high-throughput and comprehensive profiling of various clinical and molecular phenotypes, and some of them are now being implemented in advanced diagnostic centres. The systematic collection of patient signs and symptoms or imaging data constitutes phenomics. It helps to understand how specific a clinical presentation is for predicting a particular disease gene or how closely the patient's phenotype matches the predicted phenotype of a genetic finding. Molecular profiling of gene products using transcriptomics or proteomics allows the reclassification of potential splice variants or ambiguous missense variants and the prioritisation of causal genes based on aberrantly expressed gene products. Metabolomics and DNA methylation profiling are used to quantify thousands of molecular biomarkers with applications in cohort stratification, disease classification or disease gene prediction. Although a number of other omics approaches such as glycomics, fluxomics, metagenomics and others are being used extensively to elucidate the pathomechanisms of rare genetic diseases, this review focuses on the clinical application of transcriptomics, proteomics, metabolomics and methylation profiling. This review identifies diagnostic challenges and discusses omics-based strategies currently in use to potentially overcome them.

2 | KEY OMICS DATA ANALYSIS CONCEPTS

Although each omic method profiles a different molecular or physiological phenotype, the approaches to data analysis are quite similar. The basis is the collection and integration of large data sets to gain statistical power for analysis. Two different global concepts of omics data analysis can then be outlined: outlier detection and differential analysis (Figure 1). The first, outlier detection, has been successfully applied to transcriptomic and proteomic data. This approach is suitable for identifying rare events with a large effect size. In clinical practice, the most typical application of outlier analysis is the interpretation of chemical test results to support a clinical diagnosis. In this case, a standardised biochemical assay has been applied to thousands of individuals to establish a physiological (normal) range (distribution), and the abnormal measurements for the affected individual would appear as an outlier in this distribution. If the assay targets a metabolite, the specific assay can be replaced by metabolomics. Similarly, pathogenic variants may affect gene expression or protein stability, resulting in a significant reduction in the amount of gene product that would appear as an outlier in transcript and/or protein expression values. Here, transcriptomics studies replace individual gene expression assays and proteomics replaces Western blotting. A substantial sample size is required to detect outliers as statistical distribution must be robustly fitted to estimate the physiological range. Physiological parameters can differ drastically in healthy individuals and only extreme changes indicate pathological situations. For example, gigantism is defined as three standard deviations above the mean of the normal distribution, which has been determined by measuring the height of millions of individuals worldwide.[1] Omics studies have the advantage that thousands of analytes are quantified in each experiment, that is, data for the normal distribution are generated for all analytes. The variance can be biological or technical. However, the cost of omics analyses and the availability of biomaterials may limit the sample size and lead to a false estimate of the physiological ranges of analytes. The minimum sample size depends on the type of omics and the statistical test. However, an empirical 'rule of thumb' is that 50 samples are required for a robust analysis. Ideally, the samples should be healthy controls or clinically and genetically highly heterogeneous. Cases with the same defect will lead to an incorrect estimate of the physiological range, which would be shifted towards abnormal values.

The second approach of differential analysis, such as differential expression or signature detection, is more conventional for modern bioinformatics. Its major purpose is to identify features that are significantly different between two or more groups (e.g., of patients). In the context of rare genetic diseases, it relies on the availability of retrospectively collected knowledge, such as gene-phenotype associations, or pre-trained machine-learning classifiers that could differentiate individuals with and without a certain feature (e.g., with a defect in a certain gene). In order to develop such a signature for molecular diagnostics, the main requirement is to have a substantial

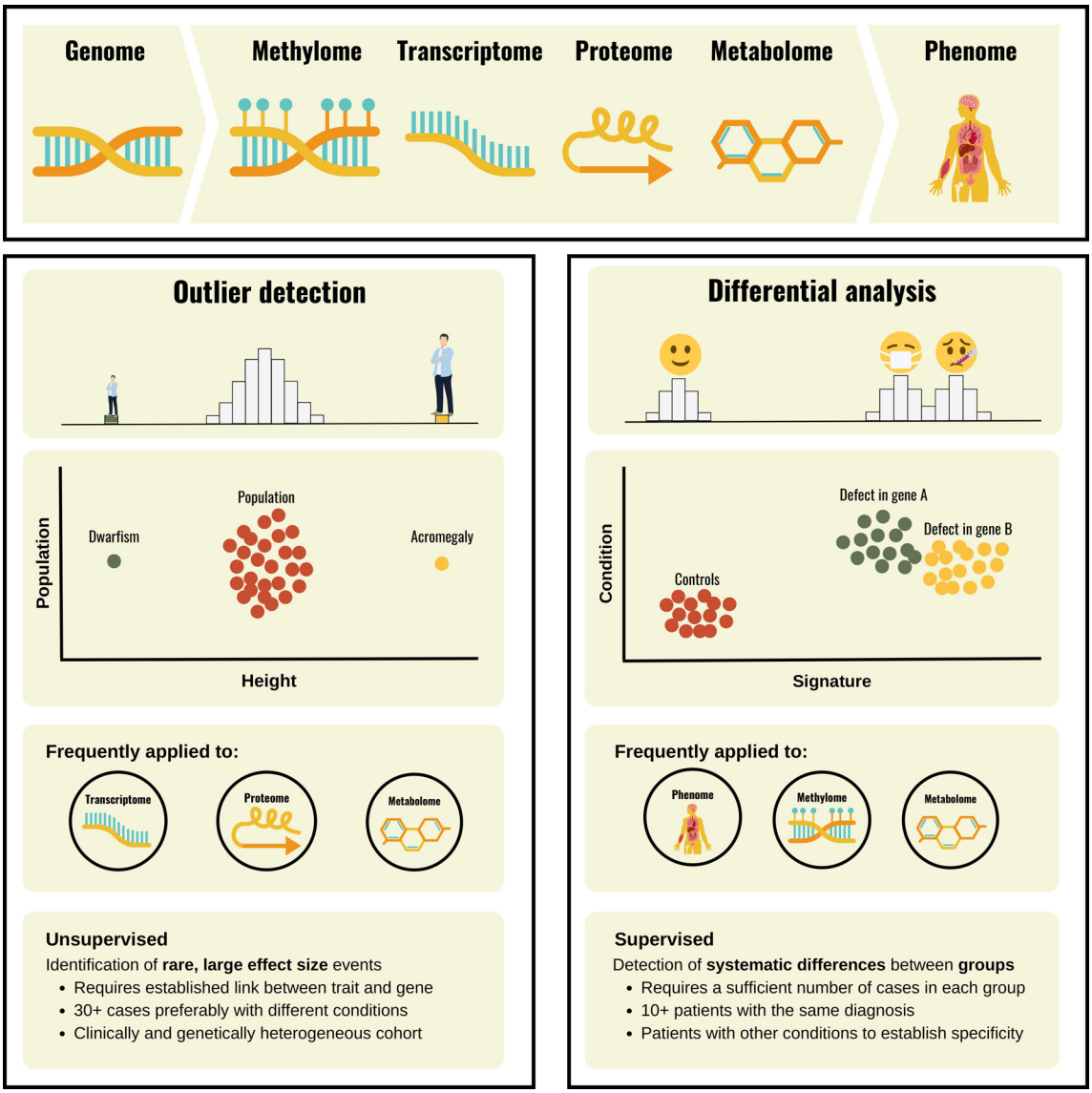

FIGURE 1 Overview of omics techniques and major approaches to their data analysis.

number of individuals with a defect in the same gene, healthy controls and a heterogeneous group of patients affected by other diseases to ensure specificity. As this method is supervised, it could be applied to any omics type, however, it is limited to a prediction of a certain known feature. Most of the signatures lack specificity, for example, there are just a few conditions where a genetic cause could be predicted purely from phenotype. Nevertheless, recent advances in omics data analysis allowed the development of signatures specific to defects in individual genes.

3 | PHENOMICS

Fundamentally phenomics is based on the systematic collection and interpretation of high-dimensional phenotype data, including physiological and morphological traits, imaging data and some biochemical and biomarker evidence. A major influence has been produced by the development of a systematic and hierarchical system for capturing patients' phenotypes—the human phenotype ontology (HPO).[2,3] HPO is a comprehensive, standardised vocabulary of phenotypic abnormalities that can be used to describe human diseases and their corresponding signs and symptoms. The popularity of HPO facilitated the development of a number of gene prioritisation algorithms based on the similarity analysis of patient phenotypes to clinical symptoms described for known disease genes[4–7] (Table S1). These methods provide a framework for systematic reanalysis of genetic data.[8–11] Because of updates to variant pathogenicity records in databases such as ClinVar, novel genotype–phenotype associations and disease-genes discoveries, periodic reanalysis over

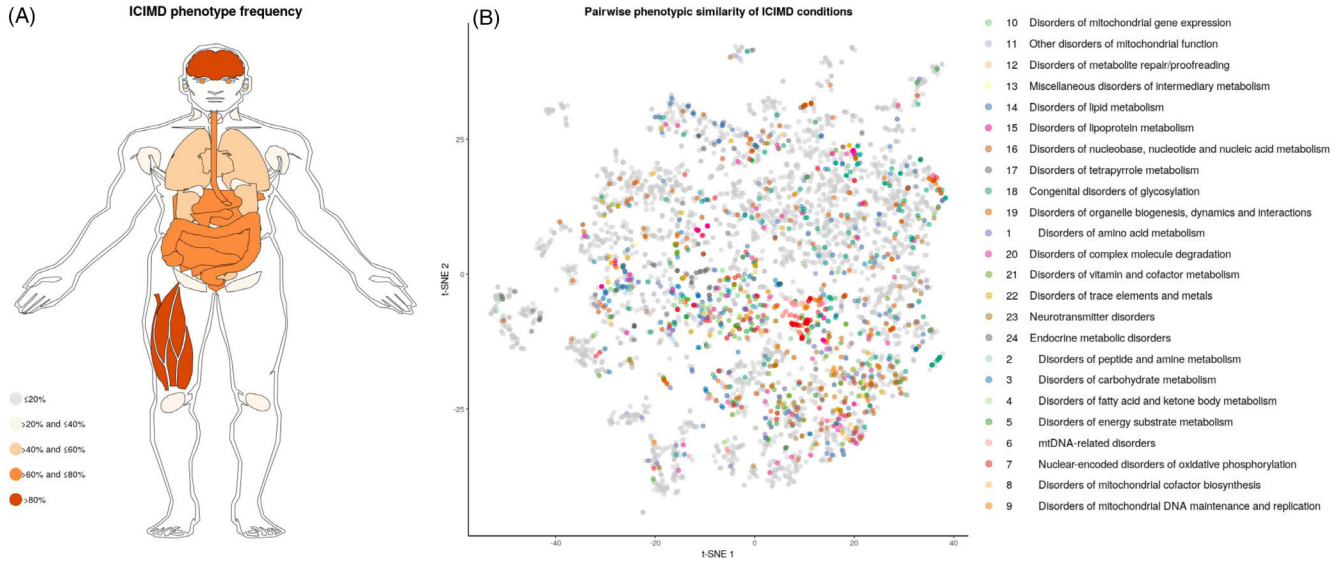

FIGURE 2 Phenotypic properties of inborn errors of metabolism. (A) Frequency of organ system involvement for conditions included in the international classification of inborn metabolic disorders (ICIMD). Percentages were calculated by counting ICIMD conditions with affected organ systems as described in level 3 of human phenotype ontology (HPO). (B) t-distributed stochastic neighbour embedding (t-SNE) of pairwise phenotypic similarity between OMIM disease genes, as encoded according to their HPO terms. For each disease gene, colour coding indicates classification according to ICIMD, and grey dots represent other Mendelian disease conditions listed in OMIM.

some years can yield a diagnostic improvement of approximately 14% ± 7%.[12–15]

Phenomics studies are particularly important for inborn errors of metabolism (IEMs) to understand the natural history of these diseases. IEMs represent a heterogeneous group of more than 1450 disorders displaying a broad phenotypic spectrum.[16] Almost every organ system could be affected with neurological and muscular phenotypes being the most common (Figure 2A). The phenotypic heterogeneity of IEMs, coupled with overlapping clinical presentations, has traditionally complicated the diagnostic process, leading to diagnostic delays and uncertainty.[15,17] This led to the development of a number of resources focused on the curation of clinical, biochemical and genetic properties of IEMs.[18–22,15] Phenotypic complexity becomes especially pronounced when looking at phenotypic similarity between conditions listed in the international classification of inborn metabolic disorders (ICIMD) (Figure 2B, Figure S1). The phenotypic spectrum of diseases within a biochemically defined class is broad and the biochemically clustered group rarely results in strong phenotypic similarity. Only some clustering can be seen for a subset of mitochondrial diseases or disorders of nucleic acid metabolism, indicating that the phenotype in itself is insufficient to make a specific diagnosis. Despite the success of a combination of conventional diagnostic tools such as biochemical analyses and exome or genome sequencing,[23] many challenges persist in diagnosing and managing IEMs, necessitating ongoing efforts in technological and methodological advancements.

4 | TRANSCRIPTOMICS

The transcriptomic approach is potentially one of the most rapidly developing omics technologies for rare diseases in recent years. It was successfully applied as a complementary approach to WES and WGS in more than 18 studies (Table S2). There are multiple case reports in which RNA-sequencing (RNA-seq) provided the necessary functional evidence to support a molecular diagnosis.[24–39] Transcriptomics can be used for VUS reclassification and for prioritisation of overlooked likely pathogenic variants. Due to the high impact of pathogenic variation on gene expression, it also supports disease gene identification by focusing on transcripts with extreme changes in expression and functional relation to disease pathomechanisms. Altogether, RNA-seq is leading to an increase in the diagnostic rate over WES/WGS alone of 16% ± 3% (Table S2). A meta-analysis of diagnosed cases from eight studies showed that the highest value of RNA-seq is to provide clinical interpretation of non-coding variants, a long-standing challenge in clinical genomics.[40] Currently, three RNA phenotypes are routinely analysed: aberrant RNA expression, aberrant splicing and allele-specific expression (ASE) with its extreme case of monoallelic expression (MAE; Figure 3). However, the power of RNA-seq is not limited to the detection of these three phenotypes, as it can also be used for variant calling, haplotype phasing, detection of gene fusions or repeat expansions. Developments in long-read RNA-seq approaches provided a novel phenotype called

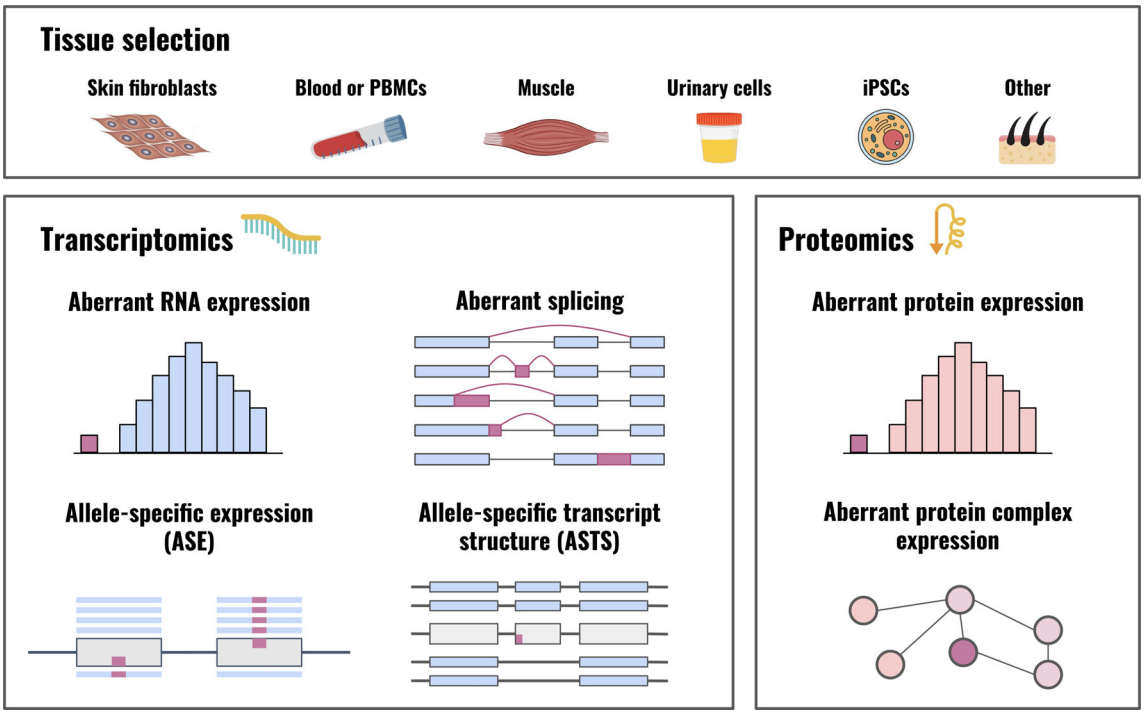

FIGURE 3 Overview of outlier-based approaches applied to transcriptomics and proteomics for rare genetic disease diagnostics.

allele-specific transcript structure (ASTS) with potential applications to variant classification tasks. In the following chapter, we will discuss diagnostic evidence provided by clinical transcriptomics as well as recommendations for the selection of an appropriate biomaterial for the analysis.

4.1 | Clinically relevant molecular events detected by transcriptomics

4.1.1 | Aberrant expression

Aberrant expression, or transcript expression outside of the physiological range, can be caused by a wide range of deleterious variants in coding and non-coding regions. Therefore, the detection of expression outliers became a useful diagnostic tool. In the meta-analysis of 120 cases diagnosed with RNA-seq collected across eight studies, aberrant expression identified the disease-causing genes in 64% of cases, while aberrant splicing was helpful in 62% and ASE in 27%.[40] Within the ACMG/AMP framework for the interpretation of potential disease-causing variants, the detection of underexpression outliers can be considered functional evidence of pathogenicity (PS3) and can reach strong evidence strength.[40,41] Detection of aberrant expression allows the prioritisation of candidate genes for further investigation for deleterious variants. To ensure robust identification of clinically relevant outliers and consideration of multiple technical and biological factors affecting gene expression readout, several specialised tools were developed and are summarised in Table S2.[42–45]

4.1.2 | Allele-specific expression

Another expression phenotype detected in RNA-seq data is ASE. In the extreme case, when one allele is completely silenced monoallelic expression can be observed. If only one allele is expressed, for example, due to the imprinting of another allele, a heterozygous variant can determine the phenotype even in recessive diseases. Some variants are not detected by WES such as larger deletions or epigenetic modifications; however, they can result in a significant imbalance between the amounts of RNA transcribed from each allele. To detect allele-specific expression both RNA-seq and DNA sequencing data are needed to compare expression levels from reference and alternative alleles with statistical tests, like the negative binomial test[24] or specialised tools such as ANEVA-DOT.[46] For autosomal dominant diseases, the detection of significant allelic imbalance could be considered functional evidence of pathogenicity for the candidate variant (ACMG/AMP framework: PS3).[40] In recessive disorders, ASE supports the prioritisation of rare monoallelic variants and serves as an indication of a deleterious variant on another allele that was not detected. In addition, ASE provides supporting allelic evidence of pathogenicity for expressed VUS (ACMG/AMP framework: PM3).[40]

4.1.3 | Aberrant splicing

Alternative splicing is a mechanism by which a single gene can encode multiple transcript isoforms and consequently several protein isoforms by combining different exons at the post-transcriptional stage. Most human genes (~95%) exhibit alternative splicing, frequently in a tissue-specific manner.[47] Aberrant splicing refers to changes in the normal process of alternative splicing that lead to the production of abnormal, non-functional or disease-causing proteins. These changes can be caused by pathogenic variation and a recent study by Jaganathan et al. estimated that about 10% of pathogenic variants in Mendelian diseases affect splicing.[48] There are several types of aberrant splicing events: exon skipping, exon extension, exon truncation, exon creation and intron retention. In RNA-seq data, they could be identified by the detection of splicing outliers and several tools were developed for aberrant splicing detection (Table S1).[37,49–55]

The ability of RNA-seq to robustly detect splice defects has opened new possibilities for the development of therapies that target these defects. The study by Hong et al. demonstrated this by using RNA-seq data to facilitate the development of antisense oligonucleotides (ASOs) as a personalised treatment approach for two patients.[35] By providing a comprehensive view of the transcriptome, routine implementation of RNA-seq in clinical practice is an important step towards a more personalised approach to the diagnosis and treatment of rare genetic diseases.

4.1.4 | Gene fusion detection

Structural variants such as inversions, deletions, duplications and translocations can result in gene fusions, a molecular event when two genes form a single hybrid transcript.[56] The majority of pathogenic gene fusion events were described for cancer, several studies reported this event in Mendelian diseases.[57–63] Therefore, several groups developed specific tools for the discovery of gene fusion transcripts in RNA-seq data.[64–66]

4.1.5 | Repeat expansions detection

Another molecular event that could be detected in the RNA-seq data is short tandem repeat (STRs) expansions, also known as microsatellite repeat expansions. They represent one to six nucleotide-long motifs that could be repeated hundreds of times and are involved in multiple hereditary diseases.[67–70] Due to their homopolymeric structure, they are usually difficult to analyse with short-read sequencing; however, Fearnley et al. recently developed superSTR, a method for repeat expansion detection in NGS data.[71] They applied superSTR to two independent patients' RNA-seq datasets and obtained high-quality results. The discovery of repeat expansions in transcriptomic data holds a promise to additionally increase diagnostic rates; however, the performance should be additionally compared to the analysis in WGS data or with southern blotting.

4.1.6 | RNA-seq variant calling

Being a sequencing technology, by its nature, RNA-seq data allow variant calling and thereby can complement WES data by the detection of non-coding variants. RNA-seq variant calling yields around ~45 000 variants per sample, in comparison to ~64 000 variants from WES. This includes around 19 000 variants not detected in WES, making it a useful complementary technology specifically to call deep intronic cryptic splice variants.[31,35]

4.1.7 | Haplotype phasing

In the absence of parental DNA samples and the detection of two variants within the same gene, the question of mono- or biallelic variants becomes challenging.[72] An important feature that is carried out by short-read RNA-seq is partial haplotype phasing. Castel et al. demonstrated that by combining evidence from WES and RNA-seq data phasing can be performed for variants in the same gene up to hundreds of kilobases away.[73] Their approach was able to provide phasing to almost 30% of rare coding variants and ~6% of all rare variants, which accounts for a ~2-fold increase compared to WES alone.

4.1.8 | Allele-specific transcript structure and long-read RNA-seq

Still, short-read sequencing has clear limitations in phasing, gene-fusion calling and aberrant splicing analysis which can be better addressed by long-read RNA-seq.[74] Long-read RNA-seq allows quantification of entire isoforms, providing a better overview of physiological splicing.[75–81] Glinos et al. applied long-read nanopore sequencing to 88 samples from 14 tissues from GTEx[82,83] and described 'allele-specific transcript structure (ASTS)' as a novel RNA phenotype.[83] Similarly to how ASE analysis examines differences in expression between maternal and paternal alleles, ASTS detects changes in splicing patterns by allele-specific isoform analysis. Long-read RNA-seq also allows the detection of epigenetic

modifications of transcripts that may serve as useful diagnostic phenotypes in the future.[84] Altogether, long-read transcriptomics provides more information for future applications in rare genetic disease diagnostics; however, due to the high price its routine clinical application is currently limited.

4.2 | Selection of the source material for clinical transcriptome

Though RNA-seq is an extremely versatile tool, the tissue-specificity of the gene expression provides a limitation for the quantification of the full spectrum of transcripts related to the disease. While the affected tissue most adequately represents pathogenic variant effects and splicing patterns, the affected tissue could show secondary downstream effects that overlay the primary cause. In clinical practice, diseased tissue specimens are rarely available from patients and controls as they typically require an invasive biopsy. For these reasons, most studies focussed on sequencing of clinically available tissues (CATs). Fibroblast cell lines and blood are currently the best resources for clinical RNA-seq and therefore most frequently used for diagnostic purposes (Table S2). However, several studies demonstrated that clinically relevant genes had higher and more consistent expression in fibroblasts (90% of OMIM genes and 85% of ICIMD genes Table S3) than in blood (<50%).[27,28,30] If the gene of interest is not expressed in available tissue, transdifferentiation into induced pluripotent stem cells (iPSCs) or affected-tissue-related cell lines, such as T-myotubes, is one way to overcome this limitation.[26,33,82,85] If genes of interest demonstrate low expression in available biosamples, cycloheximide treatment can be applied in cell culture to increase the sensitivity for splicing defects which cause nonsense-mediated RNA decay.[37] In addition, CRISPR/Cas9 technology could be used to increase coverage of low-expressed genes by depletion of highly abundant transcripts.[86]

Though blood samples and skin fibroblasts are routinely available in clinical practice, they still require a minimally invasive procedure for sample collection. Therefore, non-invasive biospecimens for transcriptomic approach recently got more attention. Lee et al. examined the potential of RNA-seq from amniotic fluid cells for prenatal diagnostics.[29] They identified a high similarity of gene expression profiles to fibroblasts and provided a diagnosis to 4 patients (8%). Martorella and colleagues investigated the performance of RNA-seq in four fully non-invasive CATs: buccal swabs, hair follicles, saliva and urine cell pellets.[87] Hair samples provided the highest detection rate of Mendelian disease genes (63%), followed by saliva (53%), urine (43%) and buccal swabs (33%). Analysis of non-invasive CATs holds great promise; however, a more in-depth assessment of the clinical utility of transcriptomics in these tissues in larger patient cohorts is still required.

To reach a sufficient sample size (>50) for outlier detection, internally sequenced data could be combined with public datasets.[27,43,49,88] Sharing of count data in public repositories and raw RNA-seq data in restricted access repositories is highly encouraged.[82,89,90] Due to differences in expression across tissues, it is not recommended to combine data from different biological sources.[88]

Given the above-listed factors affecting the performance of clinical RNA-seq, multiple tools were developed to support tissue selection (Table S1).[26,82,91–94] To assist the interpretation of RNA-seq data we provide detection rates for ICIMD genes across 25 tissues using RNA-seq and proteomics data from Kopajtich et al. and Jiang et al. studies in Table S3.[32,95]

5 | PROTEOMICS

The power of RNA-seq to increase diagnostic rates of rare genetic diseases has been proven by multiple studies; however, transcript expression levels represent only a proxy of protein abundance. Often low RNA expression levels can be compensated at the protein level. This is shown by the modest correlation of RNA and protein levels (Spearman Rho = ~0.4) in different tissues.[95,96] Therefore, proteomics, which provides direct quantification of protein levels, is a good complementary assay to RNA-seq and DNA analysis. Moreover, pathogenic missense mutations rarely result in aberrant RNA phenotypes but often affect protein stability. Strongly reduced protein levels provide a valuable readout of functional consequences of missense and regulatory variants.[32,97] Proteomics provides an additional level of functional evidence that validates and extends transcriptomic data and so opens the perspective to increase diagnostic rates.

Currently, mass-spectrometry-based proteomics quantifies fewer gene products compared to RNA-seq. In the first study, integrating proteomics and RNA-seq in diagnostics using fibroblast cell lines Kopajtich et al. detected ~12 000 transcripts and ~8000 proteins per sample and a median of 85% and 76% of ICIMD disease genes, respectively (Table S3). In the GTEx study of 32 different tissues, authors identified ~7500 proteins per tissue with 85% of them present in all tissues.[95] A limiting factor in mass spectrometry analysis is a relatively expensive and long machine running time. Multiplexing approaches such as quantitative tandem mass tag (TMT) labelling or isobaric tagging for relative and absolute quantification

(iTRAQ) allow the processing of up to 27 samples in parallel.[98] Multiplexing proteomics approaches provide homogeneous detection of peptides and proteins within the batch, which is especially important for low-expressed genes which suffer usually from widespread missing values.

5.1 | Aberrant protein expression

Proteomics provides functional evidence through the detection of protein expression outliers (Figure 3). Several studies applied proteomics for the investigation of the mechanism of disease, validation of findings in RNA-seq data, reclassification of VUS and causal gene prioritisation.[24,32,38,99,100] Dedicated methods for aberrant protein expression detection were developed (Table S2).[32,43] Forny et al. used proteomics in the cohort of 210 patients diagnosed with Methylmalonic aciduria and 20 healthy controls.[38] They performed methylmalonyl-CoA mutase (MMUT) enzyme activity assay and identified a significant reduction of MMUT protein expression levels in 150 enzyme-deficient samples. They were able to identify disease-causing variants in 148 samples, the majority of which were missense (56%), and indicating that proteomics can replace enzyme activity assays for some conditions. Kopajtich et al. applied quantitative proteomics in fibroblast cell lines from a cohort of 143 patients with suspected Mendelian disease to investigate the diagnostic potential of aberrant protein expression detection.[32] Via detection of protein expression outliers and investigation of rare variants in corresponding genes, the proteomic approach reached a genetic resolution of 21% and allowed the discovery of two novel disease genes. They also performed RNA-seq analysis in the same cohort that reached a diagnostic rate only of 11%. The major advantage of proteomics is to detect functional consequences of in-frame indel and missense variants, the most frequent types of VUS in ClinVar.[101] This makes proteomics a valuable complementary approach to DNA and RNA sequencing.

5.2 | Aberrant protein complex expression and complexomics

Proteins often function in large and dynamic protein complexes. The function and stability of these complexes rely on the presence of all their components. Therefore detection of aberrant protein complex expression can provide additional evidence for VUS reclassification and give insights into the mechanism of the disease.[100,102] Lake et al. demonstrated that defect in the small mitoribosomal subunit MRPS34 caused a reduction in the expression of all small subunits, while the expression of the large ribosomal subunits was not affected.[100] In addition, this defect consequently led to the downregulation of respiratory chain complexes I, III and IV, and their enzymatic activity correlated with protein complex levels, calculated as the mean intensity of all subunits. Kopajtich et al. detected aberrant protein complex expression in 39% of all diagnosed cases. In this way, proteomics detects not only direct but also additional downstream functional consequences of pathogenic variants increasing evidence of measured proteins but also providing indirect evidence for protein defects which have not been detected or for which the interpretation is difficult.

While conventional proteomics can only provide estimates of the abundance of protein complexes, a dedicated complexome profiling technology has been developed to study the stability, size and composition of protein complexes and their subunits.[103] Complexome profiling has become particularly popular in mitochondrial research, allowing the identification of novel protein interactions and the study of subunit assembly in large structures such as mitochondrial respiratory chain complexes and ribosomes.[104–107] Although several techniques have been developed for complexome analysis, all methods share a common concept and start with size separation of native protein extracts, for example by blue native polyacrylamide gel electrophoresis (BN-PAGE), followed by MS identification.[108,109] The overall composition and stability of a protein complex depends on the availability of all its components. Therefore, studying the assembly and functionality of these complexes provides a valuable tool for VUS interpretation and novel disease gene characterisation.[103,110]

6 | METABOLOMICS AND LIPIDOMICS

Metabolite measurements have a long history of use in the diagnosis of IMD and other genetic disorders and are essential in routine clinical practice.[111–113] Metabolomics and lipidomics can provide a comprehensive and holistic view of cellular metabolism and are able to detect the metabolic perturbations caused by pathogenic variants. By comparing the levels of these molecular species with those of healthy individuals or the general population, it is possible to identify characteristic metabolic biomarkers and signatures associated with specific genetic conditions. Metabolic biomarkers and signatures can directly inform clinical diagnosis, guide patient treatment and evaluate genetic findings.[111,114] While this review focuses on omics approaches to prioritise causal genes and interpret candidate variants, several studies have comprehensively reviewed advances in targeted and

untargeted metabolomics and lipidomics approaches to provide clinical diagnosis, identify appropriate treatments and monitor disease progression.[111–113,115–120]

Traditionally, targeted metabolomics approaches and enzyme assays have been used in the diagnostic setting. They provide absolute quantification and allow comparison across laboratories. Untargeted metabolomics and lipidomics are emerging fields in the diagnosis of rare genetic diseases and have shown promising results. Untargeted methods can quantify tens of thousands of features, providing an unbiased view of the human metabolome and can be used for screening for novel biomarkers. Several studies have used untargeted metabolomics to provide additional evidence to confirm the pathogenicity of variants.[17,24,38,114,121,122] The application of metabolite set enrichment analysis on untargeted metabolomics data has also been shown to prioritise relevant pathways for inherited metabolic disorders.[120,123] In addition, abnormal metabolite levels and biochemical measurements could be encoded as HPO terms and, therefore, integrated with phenotype-based prioritisation algorithms.[23] Already in 2015, Guo et al. demonstrated the utility of integrated WES and targeted metabolic profiling to assess variant pathogenicity using an outlier detection approach in a cohort of 80 healthy individuals.[124] In a systematic analysis of WES and metabolomics data from 170 individuals with predominantly neurological symptoms, Alaimo and colleagues were able to diagnose 21 patients, resulting in a diagnostic rate of 12%.[125] Webb-Robertson et al. further demonstrated how metabolite outlier detection can facilitate diagnosis by providing functional evidence necessary for VUS reclassification.[126,127] This validation was made possible by a previously established functional association between the candidate gene *NADK2*, 2,4-dienoyl-coenzyme A reductase deficiency and elevated urinary lysine.[128] Pathway and network-based approaches can facilitate causal gene identification using metabolomics data. Kerkhofs et al. recently developed a cross-omics pipeline that detects metabolic outliers and assesses their relevance to gene defects using pathway information.[129] Environmental factors such as dietary preferences, medication use and exposure to toxins have a significant impact on the metabolome of patients. Therefore, advanced analytical methods and improved pipelines that account for these and other known and unknown confounders are expected to further improve the accuracy of metabolomics and lipidomics in the clinical setting.

7 | DNA METHYLATION EPISIGNATURES AND EPIVARIATIONS

Epigenetic modifications, such as DNA methylation and histone modifications, have been extensively studied and have been shown to play a critical role in the development and manifestation of several rare diseases.[130] In recent years, the detection of DNA methylation episignatures has gained the most popularity among epigenetic approaches for rare disease diagnosis. Several studies have shown that pathogenic variants in disease-causing genes can induce stable changes in DNA methylation patterns at multiple positions across the genome, referred to as episignatures.[131–141] They are typically detected by conducting epigenome-wide association studies and then training binary or multi-class machine learning classifiers.[142] To date, episignatures have been described for more than 65 Mendelian neurodevelopmental disorders. Most of the affected genes encode proteins of the epigenetic machinery (Table S4). Described signatures have shown high specificity for clinical syndromes, protein complexes, specific genes, protein domains and even single nucleotide pathogenic variants.[133] The main application of episignatures is the functional evaluation of candidate variants and VUS reclassification. However, episignatures can also support the identification of copy number variants and the interpretation of mosaic variants.[143,144] In addition, the degree of hypermethylation of CpG sites underlying the KMT2B episignature has been shown to correlate with disease onset and severity.[138] In the context of causal gene prioritisation, episignatures provide a diagnostic uplift of approximately 10% in cases of neurodevelopmental disorders.[39,145] When used for VUS reclassification in a selected cohort of patients, the validation rate can be as high as 35%.[145] Although the number of episignatures detected is substantial and still growing, the majority of Mendelian disease genes (~85%) have not yet been evaluated or do not show a stable change in DNA methylation in blood. Therefore, the specificity of the available signatures is still incompletely defined. The results of epigenetic analysis, as with other omics, should be evaluated in the context of all available evidence, for example using the ACMG/AMP framework.[41] As the detection of robust and generalisable signatures depends on a large number of observations, data sharing in public or controlled access repositories and the implementation of federated learning approaches are essential to increase sensitivity, specificity and the number of episignatures.

Methylation profiling in large cohorts allows the detection of another clinically relevant event—epivariations. Similar to genetic variants, rare changes in DNA methylation such as promoter hypermethylation may be associated with Mendelian diseases.[146] Epimutations are detected as regions with outlier changes in DNA methylation and can have an effect on gene expression comparable to loss-of-function mutations.[97,146,147] A systematic analysis of DNA methylation data from 23 116 individuals identified 4452 autosomal epimutations.[147] 70% of epivariations are segregated on specific haplotypes indicating

potential genetic origin and more than ~400 of them are likely caused by rare variants.

Currently, episignatures and epivariations are typically analysed using DNA methylation microarrays. Recent advances in long-read genome sequencing allow the detection of DNA methylation. Wide application of these technologies will greatly facilitate the discovery of episignatures and clinical interpretation of epivariants, as both genetic and epigenetic analysis will be performed from a single sample, dramatically increasing the knowledge of pathomechanisms involving epigenetic changes.

8 | MULTI-OMICS INTEGRATION

With the democratisation of high-throughput technologies, the exploration of multiple 'omics' layers has become more feasible and is increasingly used in rare disease diagnostics. The integration of multi-omics data provides a more comprehensive understanding of the molecular landscape of disease. It strengthens the identification of disease-causing variants through the accumulation of evidence.

A conventional strategy for multi-omic integration is the filtering approach, where a set of criteria is applied to each individual omic dataset to narrow down the list of candidate genes. Frésard et al. used this approach to systematically evaluate genetic, phenotypic and transcriptomic data. It is also possible to integrate multi-omics data using the ACMG/AMP framework and guidelines for the clinical interpretation of RNA phenotypes have recently been proposed.[27,37,40,88] Kopajtich et al. (2021) also integrated proteomics with genetic, phenotypic and RNA-seq data and proposed a visualisation technique to facilitate clinical interpretation.

Multi-omics integration could also be performed using computational models. Çelik et al. showed that the integration of RNA-seq and genetic data in a model provides a 5-fold increase in the accuracy of aberrant splicing prediction compared to state-of-the-art computational tools (SpliceAI and MMsplice).[148] Unsupervised approaches, such as multi-omics factor analysis, have been shown to be powerful methods for increasing sensitivity.[149] For example, Forny et al. were able to identify mitochondrial pathways enriched in MMUT-deficient samples only after integrating multiple omic layers.[38]

Another powerful approach to multi-omics data integration is the use of knowledge graphs, such as the Clinical Knowledge Graph. In this data-driven model, nodes represent entities (e.g., genes, proteins, diseases), and edges represent the relationships between these entities. By integrating genomics, proteomics and metabolomics data within the framework of knowledge graphs, researchers can pinpoint affected metabolic pathways, identify potential biomarkers and even suggest potential therapeutic strategies.[150]

9 | CONCLUSION

The application of complementary omics technologies for the diagnostics of rare genetic disorders is a rapidly growing field. Transcriptomics, proteomics, phenomics, metabolomics, lipidomics and methylomics are already routinely implemented as a part of the molecular diagnostic workflow in multiple centres worldwide. Other technologies and analyses such as epi-transcriptomics, metagenomics, fluxomics, glycomics and other omics are yet to be evaluated for the purposes of disease-gene prioritisation.[151–155] Currently, the main limitation of the wide application of these approaches is the cost. However, falling costs, increasing democratisation and constantly growing evidence of the diagnostic power of multi-omics profiling pave the way for routine clinical implementation of these technologies.

All the abovementioned omics analyses rely on the systematic collection of well-annotated large-scale datasets. Therefore, sharing patient-level data in public and controlled access repositories is highly relevant and encouraged. However, an increasing number of analysis pipelines results in frequent reanalysis of data with usually arbitrary changes in performance and producing serious ecological burden due to the high computing times. To overcome this problem, the establishment of reference, publicly available benchmark datasets for each omics data type is essential. This will also provide harmonisation of the data analysis steps and may allow routine implementation of federated learning approaches. Federated learning provides the framework for the collaborative training of complex algorithms on extremely large datasets while preserving data security and privacy. For example, the federation of phenotype-based prioritisation algorithms running in parallel in a number of large diagnostic centres can potentially prioritise novel disease variants overlooked by multiple groups, which could be further connected by matchmaking systems, and thereby enhance diagnostics.

We believe that increasing democratisation and application of omics technologies together with continuous improvements in analysis pipelines and algorithms will be able to bridge the diagnostic gap and also facilitate the discovery of personalised treatment options.

AUTHOR CONTRIBUTIONS

This review was written by Dmitrii Smirnov under the guidance of Holger Prokisch. Dmitrii Smirnov and Nikita

Konstantinovskiy conceptualised figures. All authors have reviewed and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We would like to thank the Prokisch lab members for the constructive discussions on the future applications of multi-omics in molecular diagnostics. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

This study was supported by the BMBF (German Federal Ministry of Education and Research) through the mitoNET German Network for Mitochondrial Diseases (grant number 01GM1906B), PerMiM Personalized Mitochondrial Medicine (grant number 01KU2016A) and E-Rare project GENOMIT (grant number 01GM1207). The Bavarian State Ministry of Health and Care funded this study within its framework of DigiMed Bayern (grant number DMB-1805-0002) and the European Union's Horizon 2020 research and innovation program project Recon4IMD (grant number 101080997).

CONFLICT OF INTEREST STATEMENT

Authors declare no competing interests associated with the manuscript.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analysed in this study.

ETHICS STATEMENT

This article does not contain any studies with human or animal subjects performed by any authors.

ORCID

Dmitrii Smirnov https://orcid.org/0000-0002-5802-844X
Nikita Konstantinovskiy https://orcid.org/0000-0002-8802-8125
Holger Prokisch https://orcid.org/0000-0003-2379-6286

REFERENCES

1. Bello MO, Garla VV. Gigantism and acromegaly. *StatPearls [Internet]*. StatPearls Publishing; 2023. http://www.ncbi.nlm.nih.gov/books/NBK538261/
2. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. 2008;83(5):610-615.
3. Köhler S, Vasilevsky NA, Engelstad M, et al. The human phenotype ontology in 2017. *Nucleic Acids Res*. 2017;45(D1):D865-D876.
4. Köhler S, Schulz MH, Krawitz P, et al. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Hum Genet*. 2009;85(4):457-464.
5. Greene D, Richardson S, Turro E. ontologyX: a suite of R packages for working with ontological data. *Bioinformatics*. 2017;33(7):1104-1106.
6. Deng Y, Gao L, Wang B, Guo X. HPOSim: an R package for phenotypic similarity measure and enrichment analysis based on the human phenotype ontology. *PLoS One*. 2015;10(2):e0115692.
7. Zhai W, Huang X, Shen N, Zhu S. Phen2Disease: a phenotype-driven semantic similarity-based integrated model for disease and gene prioritization [Internet]. *Brief Bioinform*. 2023;24(4):bbad172. doi:10.1093/bib/bbad172/
8. Birgmeier J, Haeussler M, Deisseroth CA, et al. AMELIE speeds Mendelian diagnosis by matching patient phenotype and genotype to primary literature. *Sci Transl Med*. 2020;12(544):eaau9113.
9. Smedley D, Jacobsen JOB, Jäger M, et al. Next-generation diagnostics and disease-gene discovery with the exomiser. *Nat Protoc*. 2015;10(12):2004-2015.
10. Sifrim A, Popovic D, Tranchevent LC, et al. eXtasy: variant prioritization by genomic data fusion. *Nat Methods*. 2013;10(11):1083-1084.
11. Li Q, Zhao K, Bustamante CD, Ma X, Wong WH. Xrare: a machine learning method jointly modeling phenotypes and genetic evidence for rare disease diagnosis. *Genet Med*. 2019;21(9):2126-2134.
12. Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet Med*. 2017;19(2):209-214.
13. Eldomery MK, Coban-Akdemir Z, Harel T, et al. Lessons learned from additional research analyses of unsolved clinical exome cases. *Genome Med*. 2017;9(1):26.
14. Bruel AL, Nambot S, Quéré V, et al. Increased diagnostic and new genes identification outcome using research reanalysis of singleton exome sequencing. *Eur J Hum Genet*. 2019;27(10):1519-1531.
15. Stenton SL, Shimura M, Piekutowska-Abramczuk D, et al. Diagnosing pediatric mitochondrial disease: lessons from 2,000 exomes [Internet]. *medRxiv*. 2021. doi:10.1101/2021.06.21.21259171v1
16. Ferreira CR, Rahman S, Keller M, Zschocke J, Group IA. An international classification of inherited metabolic disorders (ICIMD). *J Inherit Metab Dis*. 2021;44(1):164-177.
17. Tarailo-Graovac M, Shyr C, Ross CJ, et al. Exome sequencing and the management of neurometabolic disorders. *N Engl J Med*. 2016;374(23):2246-2255.
18. Falk MJ, Shen L, Gonzalez M, et al. Mitochondrial disease sequence data resource (MSeqDR): a global grass-roots consortium to facilitate deposition, curation, annotation, and integrated analysis of genomic data for the mitochondrial disease clinical and research communities. *Mol Genet Metab*. 2015;114(3):388-396.
19. Lee JJY, Wasserman WW, Hoffmann GF, van Karnebeek CDM, Blau N. Knowledge base and mini-expert platform for the diagnosis of inborn errors of metabolism. *Genet Med*. 2018;20(1):151-158.
20. Zastrow DB, Baudet H, Shen W, et al. Unique aspects of sequence variant interpretation for inborn errors of metabolism (IEM): the ClinGen IEM working Group and the phenylalanine hydroxylase gene. *Hum Mutat*. 2018;39(11):1569-1580.

21. Rahman J, Noronha A, Thiele I, Rahman S. Leigh map: a novel computational diagnostic resource for mitochondrial disease. *Ann Neurol*. 2017;81(1):9-16.
22. Rahman J, Rahman S. The utility of phenomics in diagnosis of inherited metabolic disorders. *Clin Med*. 2019;19(1):30-36.
23. Lee JJY, Gottlieb MM, Lever J, et al. Text-based phenotypic profiles incorporating biochemical phenotypes of inborn errors of metabolism improve phenomics-based diagnosis. *J Inherit Metab Dis*. 2018;41(3):555-562.
24. Kremer LS, Bader DM, Mertes C, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun*. 2017;8(1):15824.
25. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med*. 2017;9(386):eaal5209.
26. Gonorazky HD, Naumenko S, Ramani AK, et al. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *Am J Hum Genet*. 2019;104(3):466-483.
27. Frésard L, Smail C, Ferraro NM, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med*. 2019;25(6):911-919.
28. Lee H, Huang AY, Kai WL, et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med*. 2020;22(3):490-499.
29. Lee M, Kwong AKY, Chui MMC, et al. Diagnostic potential of the amniotic fluid cells transcriptome in deciphering mendelian disease: a proof-of-concept. *Npj Genomic Med*. 2022;7(1):1-10.
30. Murdock DR, Dai H, Burrage LC, et al. Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J Clin Invest*. 2021;131(1):e141500. https://www.jci.org/articles/view/141500
31. Yépez VA, Gusic M, Kopajtich R, et al. Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med*. 2022;14(1):38.
32. Kopajtich R, Smirnov D, Stenton SL, et al. Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders [Internet]. *medRxiv*. 2021;21253187. doi:10.1101/2021.03.09.21253187v2
33. Rentas S, Rathi KS, Kaur M, et al. Diagnosing Cornelia de Lange syndrome and related neurodevelopmental disorders using RNA sequencing. *Genet Med*. 2020;22(5):927-936.
34. von Elsner L, Chai G, Schneeberger PE, et al. Biallelic FRA10AC1 variants cause a neurodevelopmental disorder with growth retardation. *Brain*. 2022;145(4):1551-1563.
35. Hong SE, Kneissl J, Cho A, et al. Transcriptome-based variant calling and aberrant mRNA discovery enhance diagnostic efficiency for neuromuscular diseases. *J Med Genet*. 2022;59(11):1075-1081.
36. Hiraide T, Shimizu K, Miyamoto S, et al. Genome sequencing and RNA sequencing of urinary cells reveal an intronic FBN1 variant causing aberrant splicing. *J Hum Genet*. 2022;67(7):387-392.
37. Dekker J, Schot R, Bongaerts M, et al. Web-accessible application for identifying pathogenic transcripts with RNA-seq: increased sensitivity in diagnosis of neurodevelopmental disorders. *Am J Hum Genet*. 2023;110:251-272. https://www.sciencedirect.com/science/article/pii/S0002929722005493
38. Forny P, Bonilla X, Lamparter D, et al. Integrated multi-omics reveals anaplerotic rewiring in methylmalonyl-CoA mutase deficiency. *Nat Metab*. 2023;5(1):80-95.
39. Colin E, Duffourd Y, Tisserant E, et al. OMIXCARE: OMICS technologies solved about 33% of the patients with heterogeneous rare neuro-developmental disorders and negative exome sequencing results and identified 13% additional candidate variants. *Front Cell Dev Biol*. 2022;10:10. doi:10.3389/fcell.2022.1021785
40. Smirnov D, Schlieben LD, Peymani F, Berutti R, Prokisch H. Guidelines for clinical interpretation of variant pathogenicity using RNA phenotypes. *Hum Mutat*. 2022;43(8):1056-1070.
41. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424.
42. Richter F, Hoffman GE, Manheimer KB, et al. ORE identifies extreme expression effects enriched for rare variants. *Bioinformatics*. 2019;35(20):3906-3912.
43. Brechtmann F, Mertes C, Matusevičiūtė A, et al. OUTRIDER: a statistical method for detecting aberrantly expressed genes in RNA sequencing data. *Am J Hum Genet*. 2018;103(6):907-917.
44. Salkovic E, Bensmail H. A novel Bayesian outlier score based on the negative binomial distribution for detecting aberrantly expressed genes in RNA-Seq gene expression count data. *IEEE Access*. 2021;9:75789-75800.
45. Labory J, Le Bideau G, Pratella D, et al. ABEILLE: a novel method for ABerrant expression identification empLoying machine LEarning from RNA-sequencing data. *Bioinformatics*. 2022;38(20):4754-4761.
46. Mohammadi P, Castel SE, Cummings BB, et al. Genetic regulatory variation in populations informs transcriptome analysis in rare disease. *Science*. 2019;366(6463):351-356.
47. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 2008;40(12):1413-1415.
48. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176(3):535-548.e24.
49. Mertes C, Scheller IF, Yépez VA, et al. Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun*. 2021;12(1):529.
50. Kamps-Hughes N, Carlton VEH, Fresard L, et al. A systematic method for detecting abnormal mRNA splicing and assessing its clinical impact in individuals undergoing genetic testing for hereditary cancer syndromes. *J Mol Diagn*. 2023;25(3):156-167. https://www.jmdjournal.org/article/S1525-1578(22)00348-8/fulltext#relatedArticles
51. Halperin RF, Hegde A, Lang JD, et al. Improved methods for RNAseq-based alternative splicing analysis. *Sci Rep*. 2021;11(1):10740.
52. Jenkinson G, Li YI, Basu S, Cousin MA, Oliver GR, Klee EW. LeafCutterMD: an algorithm for outlier splicing detection in rare diseases. *Bioinformatics*. 2020;36(17):4609-4615.
53. Ferraro NM, Strober BJ, Einson J, et al. Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science*. 2020;369(6509):eaaz5900.
54. Zhang D, Reynolds RH, Garcia-Ruiz S, et al. Detection of pathogenic splicing events from RNA-sequencing data using

dasper [Internet]. *bioRxiv*. 2021;437534. doi:10.1101/2021.03.29.437534v1
55. Li YI, Knowles DA, Humphrey J, et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet*. 2018;50(1):151-158.
56. Dai X, Theobard R, Cheng H, Xing M, Zhang J. Fusion genes: a promising tool combating against cancer. *Biochim Biophys Acta BBA—Rev Cancer*. 2018;1869(2):149-160.
57. van Heesch S, Simonis M, van Roosmalen MJ, et al. Genomic and functional overlap between somatic and germline chromosomal rearrangements. *Cell Rep*. 2014;9(6):2001-2010.
58. Nothwang HG, Kim HG, Aoki J, et al. Functional hemizygosity of PAFAH1B3 due to a PAFAH1B3–CLK2 fusion gene in a female with mental retardation, ataxia and atrophy of the brain. *Hum Mol Genet*. 2001;10(8):797-806.
59. Mayo S, Monfort S, Roselló M, et al. Chimeric genes in deletions and duplications associated with intellectual disability. *Int J Genomics*. 2017;24(2017):e4798474.
60. Rippey C, Walsh T, Gulsuner S, et al. Formation of chimeric genes by copy-number variation as a mutational mechanism in schizophrenia. *Am J Hum Genet*. 2013;93(4):697-710.
61. Boone PM, Yuan B, Campbell IM, et al. The Alu-rich genomic architecture of SPAST predisposes to diverse and functionally distinct disease-associated CNV alleles. *Am J Hum Genet*. 2014;95(2):143-161.
62. Ceroni F, Sagar A, Simpson NH, et al. A deletion involving CD38 and BST1 results in a fusion transcript in a patient with autism and asthma. *Autism Res*. 2014;7(2):254-263.
63. Bertelsen B, Melchior L, Jensen LR, et al. A t(3;9)(q25.1;q34.3) translocation leading to OLFM1 fusion transcripts in Gilles de la Tourette syndrome, OCD and ADHD. *Psychiatry Res*. 2015;225(3):268-275.
64. Choi HY, Jo H, Zhao X, et al. SCISSOR: a framework for identifying structural changes in RNA transcripts. *Nat Commun*. 2021;12(1):286.
65. Kim D, Salzberg SL. TopHat-fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol*. 2011;12(8):R72.
66. Oliver GR, Tang X, Schultz-Rogers LE, et al. A tailored approach to fusion transcript identification increases diagnosis of rare inherited disease. *PLoS One*. 2019;14(10):e0223337.
67. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. *Nat Rev Genet*. 2018;19(5):286-298.
68. Bahlo M, Bennett MF, Degorski P, Tankard RM, Delatycki MB. Lockhart PJ. Recent advances in the detection of repeat expansions with short-read next-generation sequencing [Internet]. *F1000Research*. 2018;7:F1000 Faculty Rev-736. https://f1000research.com/articles/7-736
69. La Spada AR, Taylor JP. Repeat expansion disease: progress and puzzles in disease pathogenesis. *Nat Rev Genet*. 2010;11(4):247-258.
70. Khristich AN, Mirkin SM. On the wrong DNA track: molecular mechanisms of repeat-mediated genome instability. *J Biol Chem*. 2020;295(13):4134-4170.
71. Fearnley LG, Bennett MF, Bahlo M. Detection of repeat expansions in large next generation DNA and RNA sequencing data without alignment. *Sci Rep*. 2022;12(1):13124.
72. Tewhey R, Bansal V, Torkamani A, Topol EJ, Schork NJ. The importance of phase information for human genomics. *Nat Rev Genet*. 2011;12(3):215-223.
73. Castel SE, Mohammadi P, Chung WK, Shen Y, Lappalainen T. Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nat Commun*. 2016;7(1):12817.
74. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Rev Genet*. 2018;19(6):329-346.
75. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol*. 2009;4(4):265-270.
76. Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323(5910):133-138.
77. Tilgner H, Jahanbani F, Blauwkamp T, et al. Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nat Biotechnol*. 2015;33(7):736-742.
78. Anvar SY, Allard G, Tseng E, et al. Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biol*. 2018;19(1):46.
79. Tardaguila M, de la Fuente L, Marti C, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res*. 2018;28(3):396-411.
80. Workman RE, Tang AD, Tang PS, et al. Nanopore native RNA sequencing of a human poly(a) transcriptome. *Nat Methods*. 2019;16(12):1297-1305.
81. Tilgner H, Grubert F, Sharon D, Snyder MP. Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc Natl Acad Sci*. 2014;111(27):9869-9874.
82. Aguet F, Brown AA, Castel SE, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213.
83. Glinos DA, Garborcauskas G, Hoffman P, et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature*. 2022;608(7922):353-359.
84. Lucas MC, Novoa EM. Long-read sequencing in the era of epigenomics and epitranscriptomics. *Nat Methods*. 2023;20(1):25-29.
85. Bonder MJ, Smail C, Gloudemans MJ, et al. Identification of rare and common regulatory variants in pluripotent cells using population-scale transcriptomics. *Nat Genet*. 2021;53(3):313-321.
86. Gu W, Crawford ED, O'Donovan BD, et al. Depletion of abundant sequences by hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol*. 2016;17(1):41.
87. Martorella M, Kasela S, Garcia-Flores R, Gokden A, Castel SE, Lappalainen T. Noninvasive, low-cost RNA-sequencing enhances discovery potential of transcriptome studies [Internet]. *bioRxiv*. 2022;506813. doi:10.1101/2022.09.06.506813v1
88. Yépez VA, Mertes C, Müller MF, et al. Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc*. 2021;16(2):1276-1296.
89. Wilks C, Zheng SC, Chen FY, et al. recount3: summaries and queries for large-scale RNA-seq expression and splicing. *Genome Biol*. 2021;22(1):323.

90. Lachmann A, Torre D, Keenan AB, et al. Massive mining of publicly available RNA-seq data from human and mouse. *Nat Commun*. 2018;9(1):1366.
91. Velluva A, Radtke M, Horn S, et al. Phenotype-tissue expression and exploration (PTEE) resource facilitates the choice of tissue for RNA-seq-based clinical genetics studies. *BMC Genomics*. 2021;22(1):802.
92. Rowlands CF, Taylor A, Rice G, et al. MRSD: a quantitative approach for assessing suitability of RNA-seq in the investigation of mis-splicing in Mendelian disease. *Am J Hum Genet*. 2022;109(2):210-222.
93. Aicher JK, Jewell P, Vaquero-Garcia J, Barash Y, Bhoj EJ. Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genet Med*. 2020;22(7):1181-1190.
94. Cummings BB, Karczewski KJ, Kosmicki JA, et al. Transcript expression-aware annotation improves rare variant interpretation. *Nature*. 2020;581(7809):452-458.
95. Jiang L, Wang M, Lin S, et al. A quantitative proteome map of the human body. *Cell*. 2020;183(1):269-283.e19.
96. Wang D, Eraslan B, Wieland T, et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol Syst Biol*. 2019;15(2):e8503.
97. Li T, Ferraro N, Strober BJ, et al. The functional impact of rare variation across the regulatory cascade [Internet]. *bioRxiv*. 2022;507008. doi:10.1101/2022.09.07.507008v1
98. Wang Z, Yu K, Tan H, et al. 27-Plex tandem mass tag mass spectrometry for profiling brain proteome in Alzheimer's disease. *Anal Chem*. 2020;92(10):7162-7170.
99. Grabowski P, Hesse S, Hollizeck S, et al. Proteome analysis of human neutrophil granulocytes from patients with monogenic disease using data-independent acquisition*[S]. *Mol Cell Proteomics*. 2019;18(4):760-772.
100. Lake NJ, Webb BD, Stroud DA, et al. Biallelic mutations in MRPS34 Lead to instability of the small Mitoribosomal subunit and Leigh syndrome. *Am J Hum Genet*. 2017;101(2):239-254.
101. Pérez-Palma E, Gramm M, Nürnberg P, May P, Lal D. Simple ClinVar: an interactive web server to explore and retrieve gene and disease variants aggregated in ClinVar database. *Nucleic Acids Res*. 2019;47(W1):W99-W105.
102. Stroud DA, Surgenor EE, Formosa LE, et al. Accessory subunits are integral for assembly and function of human mitochondrial complex I. *Nature*. 2016;538(7623):123-126.
103. Heide H, Bleier L, Steger M, et al. Complexome profiling identifies TMEM126B as a component of the mitochondrial complex I assembly complex. *Cell Metab*. 2012;16(4):538-549.
104. Kiirika L, Behrens C, Braun HP, Colditz F. The mitochondrial Complexome of Medicago truncatula. *Front Plant Sci*. 2013;4:4. doi:10.3389/fpls.2013.00084
105. Guerrero-Castillo S, van Strien J, Brandt U, Arnold S. Ablation of mitochondrial DNA results in widespread remodeling of the mitochondrial complexome. *EMBO J*. 2021;40(21):e108648.
106. Senkler J, Senkler M, Eubel H, et al. The mitochondrial complexome of Arabidopsis thaliana. *Plant J*. 2017;89(6):1079-1092.
107. Wittig I, Karas M, Schägger H. High resolution clear native electrophoresis for In-gel functional assays and fluorescence studies of membrane protein complexes*. *Mol Cell Proteomics*. 2007;6(7):1215-1225.
108. Wittig I, Malacarne FP. Complexome profiling: assembly and remodeling of protein complexes. *Int J Mol Sci*. 2021;22(15):7809. https://www.mdpi.com/1422-0067/22/15/7809
109. Cabrera-Orefice A, Potter A, Evers F, Hevler JF, Guerrero-Castillo S. Complexome profiling—exploring mitochondrial protein complexes in health and disease. *Front Cell Dev Biol*. 2022;9:796128. doi:10.3389/fcell.2021.796128
110. Stenton SL, Sheremet NL, Catarino CB, et al. Impaired complex I repair causes recessive Leber's hereditary optic neuropathy. *J Clin Invest*. 2021;131(6):e138267. https://www.jci.org/articles/view/138267
111. Thistlethwaite LR, Li X, Burrage LC, et al. Clinical diagnosis of metabolic disorders using untargeted metabolomic profiling and disease-specific networks learned from profiling data. *Sci Rep*. 2022;12(1):6556.
112. Graham E, Lee J, Price M, et al. Integration of genomics and metabolomics for prioritization of rare disease variants: a 2018 literature review. *J Inherit Metab Dis*. 2018;41(3):435-445.
113. Lydic TA, Goo YH. Lipidomics unveils the complexity of the lipidome in metabolic diseases. *Clin Transl Med*. 2018;26(7):4.
114. Almeida LS, Pereira C, Aanicai R, et al. An integrated multiomic approach as an excellent tool for the diagnosis of metabolic diseases: our first 3720 patients. *Eur J Hum Genet*. 2022;30(9):1029-1035.
115. Mussap M, Zaffanello M, Fanos V. Metabolomics: a challenge for detecting and monitoring inborn errors of metabolism. *Ann Transl Med*. 2018;6(17):338.
116. DeBerardinis RJ, Keshari KR. Metabolic analysis as a driver for discovery, diagnosis, and therapy. *Cell*. 2022;185(15):2678-2689.
117. Coene KLM, Kluijtmans LAJ, van der Heeft E, et al. Next-generation metabolic screening: targeted and untargeted metabolomics for the diagnosis of inborn errors of metabolism in individual patients. *J Inherit Metab Dis*. 2018;41(3):337-353.
118. van Karnebeek CDM, Wortmann SB, Tarailo-Graovac M, et al. The role of the clinician in the multi-omics era: are you ready? *J Inherit Metab Dis*. 2018;41(3):571-582.
119. Wortmann SB, Oud MM, Alders M, et al. How to proceed after "negative" exome: a review on genetic diagnostics, limitations, challenges, and emerging new multiomics techniques. *J Inherit Metab Dis*. 2022;45(4):663-681.
120. Liu N, Xiao J, Gijavanekar C, et al. Comparison of untargeted metabolomic profiling vs traditional metabolic screening to identify inborn errors of metabolism. *JAMA Netw Open*. 2021;4(7):e2114155.
121. Ching-López A, Martinez-Gonzalez LJ, Arrabal L, et al. Combined genome, transcriptome and metabolome analysis in the diagnosis of childhood cerebellar ataxia. *Int J Mol Sci*. 2021;22(6):2990.
122. Graham Linck EJ, Richmond PA, Tarailo-Graovac M, et al. metPropagate: network-guided propagation of metabolomic information for prioritization of metabolic disease genes. *Npj Genomic Med*. 2020;5(1):1-11.
123. Hoegen B, Hampstead JE, Engelke UFH, et al. Application of metabolite set enrichment analysis on untargeted metabolomics data prioritises relevant pathways and detects novel biomarkers for inherited metabolic disorders. *J Inherit Metab Dis*. 2022;45(4):682-695.

124. Guo L, Milburn MV, Ryals JA, et al. Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. *Proc Natl Acad Sci*. 2015;112(35):E4901-E4910.
125. Alaimo JT, Glinton KE, Liu N, et al. Integrated analysis of metabolomic profiling and exome data supplements sequence variant interpretation, classification, and diagnosis. *Genet Med*. 2020;22(9):1560-1566.
126. Webb-Robertson BJM, Stratton KG, Kyle JE, et al. Statistically driven metabolite and lipid profiling of patients from the undiagnosed diseases network. *Anal Chem*. 2020;92(2):1796-1803.
127. Kyle JE, Stratton KG, Zink EM, et al. A resource of lipidomics and metabolomics data from individuals with undiagnosed diseases. *Sci Data*. 2021;8(1):114.
128. Houten SM, Denis S, Te Brinke H, et al. Mitochondrial NADP(H) deficiency due to a mutation in NADK2 causes dienoyl-CoA reductase deficiency with hyperlysinemia. *Hum Mol Genet*. 2014;23(18):5009-5016.
129. Kerkhofs MHPM, Haijes HA, Willemsen AM, et al. Cross-omics: integrating genomics with metabolomics in clinical diagnostics. *Metabolites*. 2020;10(5):206.
130. Martinez-Delgado B, Barrero MJ. Epigenomic approaches for the diagnosis of rare diseases. *Epigenomes*. 2022;6(3):21.
131. Velasco G, Francastel C. Genetics meets DNA methylation in rare diseases. *Clin Genet*. 2019;95(2):210-220.
132. Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA methylation Episignatures for diagnosis and phenotype correlations in 42 Mendelian neurodevelopmental disorders. *Am J Hum Genet*. 2020;106(3):356-370.
133. Levy MA, McConkey H, Kerkhof J, et al. Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. *Hum Genet Genomics Adv*. 2022;3(1):100075.
134. Haghshenas S, Levy MA, Kerkhof J, et al. Detection of a DNA methylation signature for the intellectual developmental disorder, X-linked, syndromic, Armfield type. *Int J Mol Sci*. 2021;22(3):1111.
135. Verberne EA, van der Laan L, Haghshenas S, et al. DNA methylation signature for JARID2-neurodevelopmental syndrome. *Int J Mol Sci*. 2022;23(14):8001.
136. Ciolfi A, Aref-Eshghi E, Pizzi S, et al. Frameshift mutations at the C-terminus of HIST1H1E result in a specific DNA hypomethylation signature. *Clin Epigenetics*. 2020;12(1):7.
137. Choufani S, McNiven V, Cytrynbaum C, et al. An HNRNPK-specific DNA methylation signature makes sense of missense variants and expands the phenotypic spectrum of Au-Kline syndrome. *Am J Hum Genet*. 2022;109(10):1867-1884.
138. Mirza-Schreiber N, Zech M, Wilson R, et al. Blood DNA methylation provides an accurate biomarker of KMT2B-related dystonia and predicts onset. *Brain*. 2022;145(2):644-654.
139. Schenkel LC, Aref-Eshghi E, Skinner C, et al. Peripheral blood epi-signature of Claes-Jensen syndrome enables sensitive and specific identification of patients and healthy carriers with pathogenic mutations in KDM5C. *Clin Epigenetics*. 2018;10(1):21.
140. Hood RL, Schenkel LC, Nikkel SM, et al. The defining DNA methylation signature of Floating-Harbor syndrome. *Sci Rep*. 2016;6(1):38803.
141. Radio FC, Pang K, Ciolfi A, et al. SPEN haploinsufficiency causes a neurodevelopmental disorder overlapping proximal 1p36 deletion syndrome with an episignature of X chromosomes in females. *Am J Hum Genet*. 2021;108(3):502-516.
142. Turinsky AL, Choufani S, Lu K, et al. EpigenCentral: portal for DNA methylation data analysis and classification in rare diseases. *Hum Mutat*. 2020;41(10):1722-1733.
143. Van Der Laan L, Rooney K, Trooster TM, Mannens MM, Sadikovic B, Henneman P. DNA methylation episignatures: insight into copy number variation. *Epigenomics*. 2022;14(21):1373-1388.
144. Episignature analysis of moderate effects and mosaics [Internet]. 2023 https://www.researchsquare.com
145. Sadikovic B, Levy MA, Kerkhof J, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genet Med*. 2021;23(6):1065-1074.
146. Barbosa M, Joshi RS, Garg P, et al. Identification of rare de novo epigenetic variations in congenital disorders. *Nat Commun*. 2018;9(1):2064.
147. Garg P, Jadhav B, Rodriguez OL, et al. A survey of rare epigenetic variation in 23,116 human genomes identifies disease-relevant Epivariations and CGG expansions. *Am J Hum Genet*. 2020;107(4):654-669.
148. Çelik MH, Wagner N, Hölzlwimmer FR, et al. Aberrant splicing prediction across human tissues [Internet]. *bioRxiv*. 2022;495326. doi:10.1101/2022.06.13.495326v1
149. Argelaguet R, Velten B, Arnol D, et al. Multi-omics factor analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*. 2018;14(6):e8124.
150. Santos A, Colaço AR, Nielsen AB, et al. A knowledge graph to interpret clinical proteomics data. *Nat Biotechnol*. 2022;40(5):692-702.
151. Abu Bakar N, Ashikov A, Brum JM, et al. Synergistic use of glycomics and single-molecule molecular inversion probes for identification of congenital disorders of glycosylation type-1. *J Inherit Metab Dis*. 2022;45(4):769-781.
152. Ashikov A, Abu Bakar N, Wen XY, et al. Integrating glycomics and genomics uncovers SLC10A7 as essential factor for bone mineralization by regulating post-Golgi protein transport and glycosylation. *Hum Mol Genet*. 2018;27(17):3029-3045.
153. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet*. 2019;20(6):341-355.
154. Reijngoud DJ. Flux analysis of inborn errors of metabolism. *J Inherit Metab Dis*. 2018;41(3):309-328.
155. van Harskamp D, Garrelfs SF, Oosterveld MJS, Groothoff JW, van Goudoever JB, Schierbeek H. Development and validation of a new gas chromatography–tandem mass spectrometry method for the measurement of enrichment of Glyoxylate metabolism Analytes in Hyperoxaluria patients using a stable isotope procedure. *Anal Chem*. 2020;92(2):1826-1832.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Smirnov D, Konstantinovskiy N, Prokisch H. Integrative omics approaches to advance rare disease diagnostics. *J Inherit Metab Dis*. 2023;46(5):824-838. doi:10.1002/jimd.12663