



# HHS Public Access

Author manuscript

*Annu Rev Genomics Hum Genet.* Author manuscript; available in PMC 2025 June 04.

Published in final edited form as:

*Annu Rev Genomics Hum Genet.* 2024 August ; 25(1): 353–367. doi:10.1146/annurev-genom-021623-121812.

## RNA Sequencing in Disease Diagnosis

Craig Smail<sup>1</sup>, Stephen B. Montgomery<sup>2</sup>

<sup>1</sup>Genomic Medicine Center, Children's Mercy Research Institute, Children's Mercy Kansas City, Kansas City, Missouri, USA

<sup>2</sup>Department of Biomedical Data Science, Department of Genetics, and Department of Pathology, Stanford University School of Medicine, Stanford, California, USA

### Abstract

RNA sequencing (RNA-seq) enables the accurate measurement of multiple transcriptomic phenotypes for modeling the impacts of disease variants. Advances in technologies, experimental protocols, and analysis strategies are rapidly expanding the application of RNA-seq to identify disease biomarkers, tissue- and cell-type-specific impacts, and the spatial localization of disease-associated mechanisms. Ongoing international efforts to construct biobank-scale transcriptomic repositories with matched genomic data across diverse population groups are further increasing the utility of RNA-seq approaches by providing large-scale normative reference resources. The availability of these resources, combined with improved computational analysis pipelines, has enabled the detection of aberrant transcriptomic phenotypes underlying rare diseases. Further expansion of these resources, across both somatic and developmental tissues, is expected to soon provide unprecedented insights to resolve disease origin, mechanism of action, and causal gene contributions, suggesting the continued high utility of RNA-seq in disease diagnosis.

### Keywords

transcriptomics; RNA sequencing; genetic disease

## INTRODUCTION

RNA sequencing (RNA-seq) has continued to evolve as a molecular assay since its introduction in 2009 (116). RNA-seq characterizes and quantifies sequences of genetic code resulting from the process of transcription in a particular cellular context (77), providing unique insights over DNA sequencing in understanding the relative expression levels of gene products (32), transcript isoform usage (69), and alternative splicing (71) underlying disease phenotypes. Sequencing of mRNA—the code for amino acids, the building blocks of proteins—is the most common application of RNA-seq, achieved by applying a polyA

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See credit lines of images or other third-party material in this article for license information.

csmail@cmh.edu .

DISCLOSURE STATEMENT

S.B.M. is a consultant for BioMarin, MyOme, and Tenaya Therapeutics.

enrichment process to total RNA isolated from donor blood or tissue samples (125). However, abundant protocols are available for other RNA molecules of interest, such as RNAs involved in gene regulation (101).

In this review, we present an overview of established and emerging technologies, protocols, and analytical strategies for the generation and processing of RNA-seq data, with a focus on rare-disease applications (see Figure 1). We highlight state-of-the-art approaches using short-read sequencing of bulk RNA samples, as well as newer technologies that are rapidly increasing our ability to resolve the complexities of transcriptomic landscapes through the use of long-read, single-cell, and spatial transcriptomics innovations. We also discuss applications of RNA-seq to the diagnosis of rare diseases, as well as in identifying RNA biomarkers of disease progression and potential therapeutic targets. We further outline how the increasing availability of population-scale multiomic resources is providing better understanding of gene dosage effects and mechanisms of action of candidate disease variants. Overall, we outline the many impacts of RNA-seq since its introduction and show how this technology will continue to offer unprecedented insights into transcriptomic profiles of health and disease.

## ADVANCES IN RNA-SEQUENCING APPROACHES

Approaches for RNA-seq using next-generation sequencing technologies are now well established. Longer read lengths and increased depth of sequencing have steadily provided deeper resolution of transcript-level expression using Illumina platforms (7, 20, 23, 98). At present, paired-end RNA-seq using an Illumina MiSeq can generate 2×201-bp data, enabling contiguous sequencing across multiple internal exons of a gene. However, this approach is more cost-prohibitive, and most RNA-seq applications focus on 2×150-bp data provided by higher-throughput machines. Third-generation platforms from companies such as PacBio and Oxford Nanopore now also provide the opportunity to significantly extend sequencing of cDNA molecules with average read lengths ranging from 700 to 5,000 bp and beyond (38, 39, 58). Such platforms still present several biases, including in their coverage, sequence, and higher base error rates (2, 79). However, continued advances are increasingly allowing for direct RNA-seq that, unlike conventional RNA-seq, also supports the measurement of RNA modifications and potential analysis without requiring a genome reference (27, 50, 114).

Methods for processing RNA-seq libraries have also evolved to enable more routine depletion or enrichment of specific gene products. Routine use of polyA enrichment and ribosomal RNA depletion enables capturing full-length mRNAs and removing highly abundant species that can dominate RNA-seq libraries (83, 125). For specific tissues, additional highly abundant RNAs may be selectively depleted. For example, in blood RNA-seq, globin depletion further removes the large quantities of globin RNAs that can dominate a sequencing library (76). The majority of these approaches use either magnetic bead pull-down or degradation with RNase H (1). The Depletion of Abundant Sequences by Hybridization (DASH) protocol uses the Cas9 nuclease to cut specific sequences and limit their amplification to remove unwanted sequences (44). The benefit of these specialized RNA-seq approaches will be to provide new opportunities to focus on low-abundance and

potentially more tissue-restricted transcripts involved in disease. For example, the *DMD* gene involved in muscular dystrophy has been surveyed for splice defects in patient muscle (24); in Genotype–Tissue Expression (GTEx) Project data (43), *DMD* is expressed at a TPM (transcripts per million) of ~20 in muscle samples but is also expressed at a low level in blood, at a TPM of ~0.03. In addition, targeted RNA-seq shows comparable opportunities in rare-disease settings to evaluate expression of a disease-relevant gene panel (8). This suggests that canonically tissue-specific expression may be more observable with depletion or enrichment approaches for RNA-seq, thereby avoiding the need for invasive patient biopsies.

Complementing advances in RNA-seq, the rapid advances in single-cell and spatial RNA-seq have opened new opportunities for identifying molecular features in health and disease states (70, 90, 107, 118). Primarily, the advantages of single-cell approaches are their ability to quantify RNA in cells that are hard to detect due to low abundance or transience in bulk samples or, further, are misrepresented by averaging RNA quantification across cells. A representative example of the power of this approach has been the characterization of nearly 300,000 immune cells from 367 prenatal samples to detect microglia-like cells and their subtypes outside the central nervous system and their respective developmental dynamics (117). Comparable analyses using bulk RNA-seq would rely heavily on cell sorting and computational deconvolution, significantly increasing time and expense. Spatial transcriptomics takes this approach one step further to quantify cells in situ and their gene regulatory programs in the context of their local environment and interactions. For example, application of spatial transcriptomics in Alzheimer’s disease has helped to identify local gene expression programs proximal to amyloid plaques (18). Spatial transcriptomics has been further applied to map degenerative processes in spinal cords of amyotrophic lateral sclerosis (ALS) mouse models and humans (73). The application of these approaches is only expected to grow, given their potential to identify cellular transitions to disease states and more complex, regional patterns of cellular heterogeneity.

## USING RNA SEQUENCING FOR BIOMARKER DETECTION IN RARE DISEASES

Transcriptome profiling provides an opportunity to identify RNA biomarkers that inform disease progression, disease subtypes, and response to therapy (9). RNA biomarkers can further aid in repurposing novel drugs or therapies for a wide range of genetic diseases—several studies have screened for drugs that show evidence of reverting a disease expression profile to a healthy profile (30, 56, 119). In comparison with biomarkers for diseases like cancer, RNA biomarkers for rare diseases are scarce as genetic information is considered sufficiently diagnostic; however, in therapeutic development, genetic data alone are unable to monitor patient progression through any intervention.

We highlight specific rare diseases that provide representative examples of RNA biomarker discovery attempts. In Huntington’s disease (HD), early attempts at finding RNA biomarkers in humans leveraged shared changes in skeletal muscle of an HD animal model and human patients to identify a shared muscle gene expression signature (104). Notably, this

signature represented multiple factors, including endocrine-system changes, muscle-specific polyglutamine effects, and aberrant signaling from the central nervous system, highlighting the ability to identify path-way insights into preclinical therapeutic responses. However, obtaining RNA biomarkers from muscle has a higher patient burden than blood-based RNA assays, and several studies in HD have compared blood to brain RNA signatures to assess their commensurability for monitoring progression (11, 75, 94). Despite a larger number of candidate RNA biomarkers discovered in blood over the past decade, including *mHTT*, few have robustly replicated across studies, and none have been used as routine clinical tests for HD (74), suggesting that even in well-studied rare diseases with a known genetic etiology, identification of robust RNA biomarkers remains challenging.

RNA biomarker discovery in cystic fibrosis (CF) provides a separate representative example of the potential of RNA biomarkers to identify signatures of treatment response. One such effort by Jiang et al. (52) reported on the discovery of a six-gene expression signature in neutrophils from RNA-seq of 16 patients that separated patients between admission and after treatment. However, as with many biomarker discovery studies, the study had a reduced sample size, and its impact was mitigated by patient-level variance in disease-relevant responses. To reduce variance, Sun et al. (106) focused on a larger, genetically homogeneous population of CF patients with at least a G551D mutation in the *ABP2* gene during treatment with ivacaftor, a CF potentiator approved by the US Food and Drug Administration (FDA) for carriers of G551D. Notably, they identified a pretreatment signature that separated “good” from “moderate” response, highlighting the potential for an RNA biomarker to classify therapeutic response.

As of November 2022, several dozen studies are listed at [ClinicalTrials.gov](https://clinicaltrials.gov) for “RNA biomarker” or “gene expression biomarker.” However, despite advances in RNA-seq and single-cell RNA-seq, there remain multiple impediments to translating an RNA biomarker to the clinic (13). Two major roadblocks to more routine use of RNA-seq in clinical settings have the potential to be mitigated soon. First, multiple countries have now approved population-based RNA-seq through their health systems and biobanks. Future increased availability of population-scale data linking transcriptome data to clinical measurements will provide large-scale normative reference data and standards for RNA-seq that can accelerate robust discovery of RNA biomarkers. Second, RNA-based therapeutics for rare diseases such as antisense oligonucleotides (ASOs) in spinal muscular atrophy have recently provided new opportunities to treat rare diseases (91). These therapeutic approaches require an increased need to monitor the transcriptome for successful intervention, the cell-type specificity of their impact, and durability over time.

RNA biomarkers are also providing new opportunities for identifying or repurposing novel drugs. The National Institutes of Health Library of Integrated Network-Based Cellular Signatures (LINCS) consortium has provided drug-induced transcriptome signatures across a range of cell types (105). Such signatures, when intersected with robust RNA biomarkers, can nominate candidate therapeutics that revert from a disease to a normal expression profile (30, 56, 119). However, the major challenge with this approach has been that a disease expression profile may also include protective, homeostatic factors (88). A new method that investigates reversion from genetically predicted expression has the potential to

minimize these issues and has demonstrated potential for broad identification of repurposing candidates (119). As these data resources and methods continue to develop, RNA-based biomarkers, whether directly observed or genetically predicted, can be increasingly screened for candidates for drug repurposing in individuals with a broad range of genetic diseases.

## RNA SEQUENCING IN RARE-DISEASE DIAGNOSIS

RNA-seq has been increasingly applied to uncovering dysregulated molecular profiles underlying rare-disease patient cases. In one of the first large-scale applications of this approach, Cummings et al. (24) reported a 35% diagnosis rate using RNA-seq of affected tissue biopsies in a cohort of 50 patients with rare muscle disorders. Their approach further aided in interpreting a de novo intronic mutation in *COL6A1* that explained 25% of cases in an external collagen VI-related dystrophy patient cohort. This study also demonstrated the utility of using external matched tissue control cohorts, in this case skeletal muscle samples from the GTEx consortium (42).

Previous studies have also demonstrated how RNA-seq can aid in prioritizing candidate rare-disease variants from DNA sequencing. In a pediatric muscular dystrophy case, Gonorazky et al. (40) used RNA-seq to show how a deep intronic variant in *DMD* leads to a novel exon. Applying this approach more broadly, RNA-seq of blood, fibroblast, muscle, or bone marrow across a cohort of 113 rare-disease cases of diverse pathophysiologies increased diagnosis from 31% to 38% compared with DNA sequencing alone (60). A subset of cases diagnosed only through the addition of RNA-seq included neurology cases, diagnosed using patient blood samples. This showed that—at least for a subset of cases—blood can be used as a proxy for disease-affected tissues, which is important where affected tissue is not clinically accessible or biopsy would carry high risk. Frésard et al. (36) systematically assessed the diagnostic potential of RNA-seq of blood samples from 94 patients across 16 disease categories. Importantly, they found that 70.6% of OMIM rare-disease genes were expressed in blood, and that rate was similar in OMIM disease subcategories such as neurological, musculoskeletal, hematological, and ophthalmological diseases. Across the full cohort, a diagnosis was returned for 7.5% of cases, with a further 16.7% where RNA-seq improved candidate causal genes through the addition of transcriptomic dysregulation evidence.

Although blood has typically been the most used tissue to date in RNA-seq applications to rare-disease diagnosis, the expression of known rare-disease genes in this tissue captures only 60–80% of genes across OMIM disease categories (36). However, expanding to other clinically accessible tissues (defined as whole blood, Epstein-Barr virus-transformed lymphocytes, skeletal muscle, and skin-derived fibroblasts), studies using GTEx data show that more than 90% of OMIM disease genes are expressed in at least one clinically accessible tissue (121), increasing the potential for finding pathogenic variants with observable impacts on transcriptomic signatures. Demonstrating this approach, Kremer et al. (57) used RNA-seq of skin-derived fibroblasts in a mitochondriopathy patient cohort, reporting a 10% diagnosis rate from assessing aberrant gene expression, splicing events, and allele-specific expression. Furthermore, human induced pluripotent stem cells (hiPSCs)—most commonly generated from skin-derived fibroblasts—also express known rare-disease

genes (as listed in OMIM) not expressed in blood (3, 10, 15). Bonder et al. (10) used a population-scale cohort of hiPSCs to model the transcriptomic impacts of candidate rare disease variants in a hereditary cerebellar ataxia cohort proximal to *CACNA1A*, a gene broadly expressed primarily in the cerebellum and with absent or extremely minimal expression in nonbrain tissues. Furthermore, hiPSC programming protocols are available that promote lineage-specific differentiation, with application to the recapitulation of developmental trajectories in brain (4) and heart (21) tissues, among others.

Recent technological advances have enabled long-read RNA-seq approaches applied to characterizing full-length gene isoforms, improving the resolution possible with short-read technologies. Full-length transcript isoform sequencing resolves full mRNA molecules and as such provides higher resolution than short reads, which require a subsequent computational mapping step that can be error prone. Long-read approaches typically use PacBio or Oxford Nanopore platforms to measure full-length transcripts. For example, Stergachis et al. (103) utilized PacBio Iso-Seq to identify alternative splicing impacts in Charcot–Marie–Tooth disease. In a study by Farrow et al. (34) using the same technology, the discovery of a novel splice site solved a patient case of caseinolytic peptidase B (CLPB) deficiency where clinical testing using genome sequencing was negative. Similarly, using nanopore full-length RNA-seq, Helman et al. (45) identified a cryptic exon underlying a rare mitochondrial disorder.

Long-read approaches have further enabled improved reference genome assemblies for the analysis of previously unmappable RNA-seq reads. For example, one such assembly—Telomere-to-Telomere (T2T) Consortium CHM13—reports almost 2,000 novel genes, including 99 predicted as protein coding, compared with GRCh38 (86). Long-read assemblies, combined with RNA-seq, is a particularly useful strategy in resolving and modeling the impact of repeat expansions within genes (79) and longer-range segments overlapping multiple gene regions, such as the rare-disease gene *REXOIL* contained within the 1-Mb variable number tandem repeat (VNTR) region (46).

## COMPUTATIONAL APPROACHES FOR USE OF RNA SEQUENCING IN RARE-DISEASE CONTEXTS

Ongoing development of computational approaches has increasingly enabled use of RNA-seq approaches. Beyond gene expression levels, accurate quantification of splicing (66, 111, 113), alternative gene structure (64, 123, 126), RNA modifications (65), and allele-specific expression (6, 17, 59) all provide insight into normal and disease-relevant molecular activities. In the context of rare-disease diagnosis, the ability to measure normal and aberrant splicing has enabled the increased ability to classify pathogenic noncoding alleles (31) and identify potential therapeutic targets (16). These approaches are also rapidly being advanced with the availability of computational tools such as SpliceAI and Pangolin (49, 124) and via the reprocessing of hundreds of thousands of RNA-seq files to identify prior evidence of splice events (25). Computational approaches that measure allele-specific expression from RNA-seq data have further provided insight into the mechanisms of nonsense-mediated decay and provide candidate therapeutic targets (22, 84, 92, 109). Further single-cell

Author Manuscript

pipelines that enable improved detection of individual cell types have enabled profiling of rare cell types or differentiation trajectories that underlie rare-disease effects (47, 48, 82, 108). Single-cell approaches offer further unique opportunities to compare wild-type with disease cells in the same patient as a result of mosaic expression of the candidate disease allele (19).

Author Manuscript

Assuming a FASTQ file output from an RNA-seq run, several computational processing steps are necessary to prepare data for downstream analysis. Mapping aligns mRNA reads to a reference genome; multiple software packages are available, including Spliced Transcripts Alignment to a Reference (STAR) (28) and Minimap2 (63). Mapped reads can then be quantified at the gene level, or further at the level of transcript isoforms and individual exons (26, 62, 67, 89, 110). Additional molecular phenotypes obtainable from RNA-seq include known or novel splicing variation quantification (66), alternative polyadenylation, and—if genetic data are available—allele-specific expression (112).

Author Manuscript

Various sources of technical noise, combined with intrinsic transcriptomic heterogeneity across genes, can adversely affect downstream analyses that use RNA-seq data. To alleviate this issue, a common approach is to implement a noise correction algorithm on filtered, transformed, and normalized data. Filtering is performed to remove genes expressed at low levels across the study cohort and typically uses empirical thresholds (96), which controls for sparseness in individual observations and removes genes where low expression might bias downstream analyses. Log transformation and standardization for variance stabilization accounts for differences in the magnitude of expression across genes (127). Approaches implementing technical noise correction discover latent variables capturing shared patterns of expression variation across individual samples, contributed by experimental factors such as library preparation and sequencing batch (61, 102).

Author Manuscript

Author Manuscript

Comparing transcriptomic readouts within or between individual samples enables the identification of outlier molecular signatures. These outlier signatures potentially result from proximal large-effect variants impacting normal gene function. Outlier signatures at the level of total gene expression are typically ascertained through *Z*-score transformation (36, 41), enabling a fixed outlier threshold to be set across genes. As a further step, to separate genetic from technical/nongenetic outliers, one potential strategy is to summarize individual outlier counts across the full dataset—for example, by calculating the proportion of outlier genes per individual across all outlier calls (35). Individuals contributing a relatively higher proportion of outlier calls might be reflective of uncorrected technical noise and can be removed from further analysis. An alternative approach is Outlier in RNA-Seq Finder (OUTRIDER) (12), which identifies gene expression outliers after first accounting for confounding variation using expression correlation learned by an autoencoder. Relevant to additional molecular phenotypes obtainable from RNA-seq, Analysis of Expression Variation–Dosage Outlier Test (ANEVA-DOT) (81) can characterize allele-specific expression to measure the magnitude of deviation in equilibrium of gene expression between paternal and maternal haplotypes for a candidate pathogenic variant informed by a reference/control population. For identifying splicing events, LeafCutter (66) uses mapped split reads to identify intron removal and indicate splicing junctions. The Detection of RNA Outliers Pipeline (DROP) (122) provides a unified workflow to

detect aberrant expression, splicing, and monoallelic expression. Increasingly, the predicted molecular effects can be tested when suspect expression or splicing-modulating variation is identified through patient assays that compare isogenic cell lines through gene-editing-based correction of candidate disease variants (99).

There are multiple opportunities for future development of computational methods. Downstream of accurate quantification of molecular phenotypes from RNA-seq data, analysis strategies for comparison with reference/control resources are required to identify and interpret candidate disease variants linked to outlier molecular profiles. Trio RNA-seq of proband and parents has begun to aid in modeling transcriptomic impacts of candidate disease variants (60). However, tools that routinely integrate family-based transcriptomics in disease diagnosis are largely missing. Further, approaches like Watershed that integrate genome annotation and personal transcriptome profiles are providing new opportunities to prioritize specific rare variants (35). But unlike the multiple tools that predict variant effects from sequence alone, there are limited comparable approaches that integrate personal omics to identify signatures of pathogenic variants. As an example, dynamic, personal transcriptome signatures will need to be analyzed in future tools as repeated patient assays are used to assess temporal outlier effects such as those that occur across developmental time points in iPSC models (85) or in response to the environment.

## RNA SEQUENCING FOR EVALUATING RARE-DISEASE THERAPIES

We offer some perspectives on how RNA-seq can provide opportunities for evaluating rare-disease therapies. Currently, epidemiology and genetic data from health records and large-scale population biobanks are routinely used to identify the prevalence of a rare disease, which informs not only a potential market for a therapeutics company but also what types of approaches might be successfully employed. In many cases, accurate reporting of rare-disease prevalence can be a challenge from epidemiological and/or genetic data alone, given the different national reporting mechanisms, subclinical presentations, unidentified disease alleles, variable penetrance, and epistatic effects (120). As population-scale RNA-seq data continue to grow, direct observations of expression or splicing levels can potentially enhance our understanding of the prevalence of particular diseases, particularly as we further map the role dosage plays in rare-disease etiology. In this regard, efforts have been made to accurately estimate the sizes of genetic effects on gene expression as a means to subsequently identify physiologically relevant dosage effects (80). Further, recent efforts by Dong et al. (29) to map tissue-specific levels of dosage sensitivity using population-scale DNA sequencing and RNA-seq have identified constraints in genes based on their functional categories to subsequently better infer pathogenetic modes of inheritance and levels of gene expression.

While information about dosage informs disease etiology and prevalence, RNA-seq can also inform the targeting and efficacy of therapeutic strategies for rare diseases. Emerging RNA medicines can modulate dosage through small interfering RNAs or ASOs. For ASOs, steric blocking of splice modulators can selectively include or exclude exons to remove out-of-frame effects or induce specific isoform switching—here, RNA-seq data inform potential splice junctions to target. One example is ongoing clinical trials and

FDA approval of skipping ASOs in the *DMD* gene for Duchenne's muscular dystrophy to remove out-of-frame mutations in patients (93). Reference patient data can also confirm expected pathogenic splicing patterns in individuals with rare diseases and have helped to enable patient-customized ASO therapy in rare orphan diseases (55). But ASO strategies can also leverage reference, population-scale RNA-seq to identify cryptic or nonproductive splice products whose blocking may lead to isoform switching and increases in gene expression (68). In a study by Lim et al. (68), analysis of 78 RNA-seq samples from diverse tissues identified ~1,200 nonproductive splice events in disease-associated genes. Analysis of public data from hundreds of thousands of RNA-seq samples has since only increased the discovery of common, unannotated splicing events and potential targets (25, 97). The development of bioinformatic tools for mining both patient and population RNA-seq data for promising targets remains an ongoing challenge. Further, even when targets are identified, subsequent follow-up RNA-seq in treated cell lines has demonstrated its efficacy in identifying unwanted off-target effects (78).

Small-molecule therapies have also increasingly benefited from large-scale RNA-seq. The LINCS program has compiled more than 1.5 million gene expression signatures for more than 40,000 small-molecular perturbations (54), enabling rapid screening for gene expression signatures that reverse or mimic a disease state (33, 87, 95). However, challenges with these approaches are that they often have limitations on the range of cell types explored and that repositioning activities have focused on reversing signatures in the disease state where multiple homeostatic and protective factors could also be active. Studies that predict gene expression from genetic data provide another opportunity to identify gene expression signatures that preempt the disease state (37, 115). Typically, the prediction of gene expression is used to perform transcriptome-wide association studies (TWASs) to identify genes whose expression is correlated to disease state, in lieu of collecting gene expression data in these individuals. Song et al. (100) recently used this approach to assess computational drug repositioning for atopic dermatitis by evaluating enrichment of TWAS genes within drug signatures. However, as TWASs can generate risk and protective predicted risk profiles, they can also generate signatures that inform directions of risk. Recent approaches like PharmGWAS (53) provide multiple approaches for drug repositioning using TWAS data predicted from population-scale RNA-seq.

## FUTURE OPPORTUNITIES

A major challenge with RNA-seq compared with DNA sequencing in the context of rare-disease diagnosis is the unknown developmental and tissue specificity of disease effects. As yet, there are no convenient approaches to predict what developmental stages and cell types would be informative for a patient with an undiagnosed condition. Unlike complex diseases with multigene involvement, where heritability approaches can be a means to identify cell types and developmental time points, in rare diseases this often requires understanding the cell-type biology of a single gene. Further, detecting a disease-associated transcriptome signature may not always provide immediate insights into effective therapies as RNA signatures can have high interindividual variability and inaccuracy with respect to transcript isoform annotation, and may even be buffered at the protein-coding level (5). However, the increasing availability of long-read transcriptomic resources is expected to improve the

resolution of transcript isoforms, which may in part increase the resolution for modeling the impact of variants from mRNA abundance through protein expression (39). Further, the growing availability of reference multiomics provides new signatures beyond transcriptome features alone, and recent studies have successfully utilized both methylation and proteomics to provide diagnoses for individuals with rare diseases (72). In addition, the increasing availability of reference resources across the human life span will be useful in identifying molecular profiles of disease matched for age of onset. Efforts led by the Developmental Genotype–Tissue Expression (dGTEx) consortium aim to provide large-scale reference profiles of multiple tissues across early to late childhood. Complementary to this resource, transcriptomic characterization of embryonic and fetal tissues provides opportunities to understand transcriptomic profiles of disease at the point of first emergence (14).

There are several opportunities for future computational analysis of RNA-seq data for rare-disease applications. Tools for discovery of meaningful aberrant transcriptome signatures have emerged, such as DROP and LeafCutterMD (51, 122). However, there is a need for strategies that consider the network effects of outliers, integrate data across omics layers, and integrate genetic information and phenotype information. In the context of rare-disease therapy, there are numerous new opportunities to consider how integrating rare-disease and population-scale transcriptomics can help refine understanding of disease prevalence and nominate new therapeutic options.

Since the emergence of RNA-seq, there has been rapid progress in its use for rare diseases (116), and since 2017, this progress has only accelerated in diagnostic applications. Routine use of RNA-seq in clinics will soon be possible; however, this will first require standardized workflows that indicate when RNA-seq will be useful for specific conditions or in the presence of specific prior data. Efforts to nominate RNA-seq for specific biosamples informed by patient-specific, noncoding variants of unknown significance would be one practical means of promoting RNA-seq in a clinical setting. In addition, harmonization of control RNA-seq data to establish more confident measures of outlier status will also enable clinical geneticists and genetic counselors to know how truly aberrant an observed effect is (i.e., 1 in 100 or 1 in 10,000). Finally, as RNA medicines rapidly provide patient-specific therapeutic options, the validation of molecular effects and minimization of off targets will require RNA-seq as part of the design workflow. The progress and promise of RNA-seq for rare diseases have clearly begun.

## ACKNOWLEDGMENTS

S.B.M. is supported by National Institutes of Health/National Human Genome Research Institute grant U01HG011762. C.S. is supported by National Institutes of Health/National Institute of General Medical Sciences grant R35GM146966.

## LITERATURE CITED

1. Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, et al. 2013. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods* 10(7):623–29 [PubMed: 23685885]
2. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. 2020. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21:30 [PubMed: 32033565]

3. Anderson RH, Francis KR. 2018. Modeling rare diseases with induced pluripotent stem cell technology. *Mol. Cell. Probes* 40:52–59 [PubMed: 29307697]
4. Barak M, Fedorova V, Pospisilova V, Raska J, Vochyanova S, et al. 2022. Human iPSC-derived neural models for studying Alzheimer’s disease: from neural stem cells to cerebral organoids. *Stem Cell Rev. Rep.* 18(2):792–820 [PubMed: 35107767]
5. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, et al. 2015. Impact of regulatory variation from RNA to protein. *Science* 347(6222):664–67 [PubMed: 25657249]
6. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, et al. 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* 24(1):14–24 [PubMed: 24092820]
7. Bayega A, Fahiminiya S, Oikonomopoulos S, Ragoussis J. 2018. Current and future methods for mRNA analysis: a drive toward single molecule sequencing. *Methods Mol. Biol.* 1783:209–41 [PubMed: 29767365]
8. Berger K, Arafat D, Chandrakasan S, Snapper SB, Gibson G. 2022. Targeted RNAseq improves clinical diagnosis of very early-onset pediatric immune dysregulation. *J. Pers. Med.* 12(6):919 [PubMed: 35743704]
9. Bhattacharya S, Mariani TJ. 2009. Array of hope: expression profiling identifies disease biomarkers and mechanism. *Biochem. Soc. Trans.* 37(4):855–62 [PubMed: 19614607]
10. Bonder MJ, Smail C, Gloudemans MJ, Frésard L, Jakubosky D, et al. 2021. Identification of rare and common regulatory variants in pluripotent cells using population-scale transcriptomics. *Nat. Genet.* 53(3):313–21 [PubMed: 33664507]
11. Borovecki F, Lovrecic L, Zhou J, Jeong H, Then F, et al. 2005. Genome-wide expression profiling of human blood reveals biomarkers for Huntington’s disease. *PNAS* 102(31):11023–28 [PubMed: 16043692]
12. Brechtmann F, Mertes C, Matusevičiūtė A, Yépez VA, Avsec Ž, et al. 2018. OUTRIDER: a statistical method for detecting aberrantly expressed genes in RNA sequencing data. *Am. J. Hum. Genet.* 103(6):907–17 [PubMed: 30503520]
13. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. 2016. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat. Rev. Genet.* 17(5):257–71 [PubMed: 26996076]
14. Cao J, O’Day DR, Pliner HA, Kingsley PD, Deng M, et al. 2020. A human cell atlas of fetal gene expression. *Science* 370(6518):eaba7721 [PubMed: 33184181]
15. Carlessi L, Fusar Poli E, Bechi G, Mantegazza M, Pascucci B, et al. 2014. Functional and molecular defects of hiPSC-derived neurons from patients with ATM deficiency. *Cell Death Dis.* 5(7):e1342 [PubMed: 25032865]
16. Carvill GL, Mefford HC. 2020. Poison exons in neurodevelopment and disease. *Curr. Opin. Genet. Dev.* 65:98–102 [PubMed: 32615329]
17. Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. 2015. Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16:195 [PubMed: 26381377]
18. Chen W-T, Lu A, Craessaerts K, Pavie B, Sala Frigerio C, et al. 2020. Spatial transcriptomics and in situ sequencing to study Alzheimer’s disease. *Cell* 182(4):976–91.e19 [PubMed: 32702314]
19. Cheung AYL, Horvath LM, Grafodatskaya D, Pasceri P, Weksberg R, et al. 2011. Isolation of *MECP2*-null Rett Syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation. *Hum. Mol. Genet.* 20(11):2103–15 [PubMed: 21372149]
20. Cho H, Davis J, Li X, Smith KS, Battle A, Montgomery SB. 2014. High-resolution transcriptome analysis with long-read RNA sequencing. *PLOS ONE* 9(9):e108095 [PubMed: 25251678]
21. Cho S, Lee C, Skylar-Scott MA, Heilshorn SC, Wu JC. 2021. Reconstructing the heart using iPSCs: engineering strategies and applications. *J. Mol. Cell. Cardiol.* 157:56–65 [PubMed: 33895197]
22. Coban-Akdemir Z, White JJ, Song X, Jhangiani SN, Fatih JM, et al. 2018. Identifying genes whose mutant transcripts cause dominant disease traits by potential gain-of-function alleles. *Am. J. Hum. Genet.* 103(2):171–87 [PubMed: 30032986]

23. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, et al. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17:13 [PubMed: 26813401]
24. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, et al. 2017. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* 9(386):eaal5209 [PubMed: 28424332]
25. Dawes R, Bournazos AM, Bryen SJ, Bommireddipalli S, Marchant RG, et al. 2023. SpliceVault predicts the precise nature of variant-associated mis-splicing. *Nat. Genet.* 55(2):324–32 [PubMed: 36747048]
26. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire M-D, et al. 2012. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28(11):1530–32 [PubMed: 22539670]
27. Depledge DP, Srinivas KP, Sadaoka T, Bready D, Mori Y, et al. 2019. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat. Commun.* 10:754 [PubMed: 30765700]
28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21 [PubMed: 23104886]
29. Dong D, Shen H, Wang Z, Liu J, Li Z, Li X. 2023. An RNA-informed dosage sensitivity map reflects the intrinsic functional nature of genes. *Am. J. Hum. Genet.* 110(9):1509–21 [PubMed: 37619562]
30. Dudley JT, Sirota M, Shenoy M, Pai RK, Roedder S, et al. 2011. Computational repositioning of the anticonvulsant topiramate for inflammatory bowel disease. *Sci. Transl. Med.* 3(96):96ra76
31. Ellingford JM, Ahn JW, Bagnall RD, Baralle D, Barton S, et al. 2022. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Med.* 14:73 [PubMed: 35850704]
32. Ergin S, Kherad N, Alagoz M. 2022. RNA sequencing and its applications in cancer and rare diseases. *Mol. Biol. Rep.* 49(3):2325–33 [PubMed: 34988891]
33. Evangelista JE, Clarke DJB, Xie Z, Lachmann A, Jeon M, et al. 2022. SigCom LINCS: data and metadata search engine for a million gene expression signatures. *Nucleic Acids Res.* 50(W1):W697–709 [PubMed: 35524556]
34. Farrow E, Jay A, Means J, Younger S, Biswell R, et al. 2023. Case of CLPB deficiency solved by HiFi long read genome sequencing and RNAseq. *Am. J. Med. Genet. A* 191(12):2908–912 [PubMed: 37548286]
35. Ferraro NM, Strober BJ, Einson J, Abell NS, Aguet F, et al. 2020. Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science* 369(6509):eaaz5900 [PubMed: 32913073]
36. Frésard L, Smail C, Ferraro NM, Teran NA, Li X, et al. 2019. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat. Med.* 25(6):911–19 [PubMed: 31160820]
37. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, et al. 2015. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* 47(9):1091–98 [PubMed: 26258848]
38. Garalde DR, Snell EA, Jachimowicz D, Sipos B, Lloyd JH, et al. 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nat. Methods* 15(3):201–6 [PubMed: 29334379]
39. Glinos DA, Garborcauskas G, Hoffman P, Ehsan N, Jiang L, et al. 2022. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature* 608(7922):353–59 [PubMed: 35922509]
40. Gonorazky HD, Liang M, Cummings B, Lek M, Micallef J, et al. 2016. RNAseq analysis for the diagnosis of muscular dystrophy. *Ann. Clin. Transl. Neurol.* 3(1):55–60 [PubMed: 26783550]
41. Gonorazky HD, Naumenko S, Ramani AK, Nelakuditi V, Mashouri P, et al. 2019. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *Am. J. Hum. Genet.* 104(3):466–83 [PubMed: 30827497]
42. Consort GTEx. 2015. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348(6235):648–60 [PubMed: 25954001]
43. Consort GTEx. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369(6509):1318–30 [PubMed: 32913098]

44. Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, et al. 2016. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol.* 17:41 [PubMed: 26944702]
45. Helman G, Compton AG, Hock DH, Walkiewicz M, Brett GR, et al. 2021. Multiomic analysis elucidates Complex I deficiency caused by a deep intronic variant in *NDUFB10*. *Hum. Mutat.* 42(1):19–24 [PubMed: 33169436]
46. Hoyt SJ, Storer JM, Hartley GA, Grady PGS, Gershman A, et al. 2022. From telomere to telomere: the transcriptional and epigenetic state of human repeat elements. *Science* 376(6588):eabk3112 [PubMed: 35357925]
47. Hua Y, Wu N, Miao J, Shen M. 2023. Single-cell transcriptomic analysis in two patients with rare systemic autoinflammatory diseases treated with anti-TNF therapy. *Front. Immunol.* 14:1091336 [PubMed: 36911721]
48. Jagadeesh KA, Dey KK, Montoro DT, Mohan R, Gazal S, et al. 2022. Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat. Genet.* 54(10):1479–92 [PubMed: 36175791]
49. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, et al. 2019. Predicting splicing from primary sequence with deep learning. *Cell* 176(3):535–48.e24 [PubMed: 30661751]
50. Jain M, Abu-Shumays R, Olsen HE, Akeson M. 2022. Advances in nanopore direct RNA sequencing. *Nat. Methods* 19(10):1160–64 [PubMed: 36203024]
51. Jenkinson G, Li YI, Basu S, Cousin MA, Oliver GR, Klee EW. 2020. LeafCutterMD: an algorithm for outlier splicing detection in rare diseases. *Bioinformatics* 36(17):4609–15 [PubMed: 32315392]
52. Jiang K, Poppenberg KE, Wong L, Chen Y, Borowitz D, et al. 2019. RNA sequencing data from neutrophils of patients with cystic fibrosis reveals potential for developing biomarkers for pulmonary exacerbations. *J. Cyst. Fibros.* 18(2):194–202 [PubMed: 29941318]
53. Kang H, Pan S, Lin S, Wang Y-Y, Yuan N, Jia P. 2024. PharmGWAS: a GWAS-based knowledgebase for drug repurposing. *Nucleic Acids Res.* 52(D1):D972–79 [PubMed: 37831083]
54. Keenan AB, Jenkins SL, Jagodnik KM, Koplev S, He E, et al. 2018. The Library of Integrated Network-Based Cellular Signatures NIH program: system-level cataloging of human cells response to perturbations. *Cell Syst.* 6(1):13–24 [PubMed: 29199020]
55. Kim J, Hu C, Moufawad El Achkar C, Black LE, Douville J, et al. 2019. Patient-customized oligonucleotide therapy for a rare genetic disease. *N. Engl. J. Med.* 381(17):1644–52 [PubMed: 31597037]
56. Koudijs KKM, Terwisscha van Scheltinga AGT, Böhringer S, Schimmel KJM, Guchelaar H-J. 2019. Transcriptome signature reversion as a method to reposition drugs against cancer for precision oncology. *Cancer J.* 25(2):116–20 [PubMed: 30896533]
57. Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, et al. 2017. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* 8:15824 [PubMed: 28604674]
58. Kuo RI, Cheng Y, Zhang R, Brown JWS, Smith J, et al. 2020. Illuminating the dark side of the human transcriptome with long read transcript sequencing. *BMC Genom.* 21:751
59. Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501(7468):506–11 [PubMed: 24037378]
60. Lee H, Huang AY, Wang L-K, Yoon AJ, Renteria G, et al. 2020. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet. Med.* 22(3):490–99 [PubMed: 31607746]
61. Leek JT. 2014. svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res.* 42(21):e161 [PubMed: 25294822]
62. Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12:323
63. Li H 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34(18):3094–100 [PubMed: 29750242]

64. Li L, Huang K-L, Gao Y, Cui Y, Wang G, et al. 2021. An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.* 53(7):994–1005 [PubMed: 33986536]
65. Li Q, Gloudemans MJ, Geisinger JM, Fan B, Aguet F, et al. 2022. RNA editing underlies genetic risk of common inflammatory diseases. *Nature* 608(7923):569–77 [PubMed: 35922514]
66. Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, et al. 2018. Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* 50(1):151–58 [PubMed: 29229983]
67. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923–30 [PubMed: 24227677]
68. Lim KH, Han Z, Jeon HY, Kach J, Jing E, et al. 2020. Antisense oligonucleotide modulation of nonproductive alternative splicing upregulates gene expression. *Nat. Commun.* 11:3501 [PubMed: 32647108]
69. Liu Q, Fang L, Wu C. 2022. Alternative splicing and isoforms: from mechanisms to diseases. *Genes* 13(3):401 [PubMed: 35327956]
70. Lohoff T, Ghazanfar S, Missarova A, Kouloua N, Pierson N, et al. 2022. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat. Biotechnol.* 40(1):74–85 [PubMed: 34489600]
71. Lord J, Baralle D. 2021. Splicing in the diagnosis of rare disease: advances and challenges. *Front. Genet.* 12:689892 [PubMed: 34276790]
72. Lunke S, Bouffler SE, Patel CV, Sandaradura SA, Wilson M, et al. 2023. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat. Med.* 29(7):1681–91 [PubMed: 37291213]
73. Maniatis S, Åijö T, Vickovic S, Braine C, Kang K, et al. 2019. Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science* 364(6435):89–93 [PubMed: 30948552]
74. Martí-Martínez S, Valor LM. 2022. A glimpse of molecular biomarkers in Huntington’s disease. *Int. J. Mol. Sci.* 23(10):5411 [PubMed: 35628221]
75. Mastrokolas A, Ariyurek Y, Goeman JJ, van Duijn E, Roos RAC, et al. 2015. Huntington’s disease biomarker progression profile identified by transcriptome sequencing in peripheral blood. *Eur. J. Hum. Genet.* 23(10):1349–56 [PubMed: 25626709]
76. Mastrokolas A, den Dunnen JT, van Ommen GB, ’t Hoen PAC, van Roon-Mom WMC. 2012. Increased sensitivity of next generation sequencing-based expression profiling after globin reduction in human blood RNA. *BMC Genom.* 13:28
77. McGettigan PA. 2013. Transcriptomics in the RNA-seq era. *Curr. Opin. Chem. Biol.* 17(1):4–11 [PubMed: 23290152]
78. Michel S, Schirduan K, Shen Y, Klar R, Tost J, Jaschinski F. 2021. Using RNA-seq to assess off-target effects of antisense oligonucleotides in human cell lines. *Mol. Diagn. Ther.* 25(1):77–85 [PubMed: 33314011]
79. Mikheenko A, Prjibelski AD, Joglekar A, Tilgner HU. 2022. Sequencing of individual barcoded cDNAs using Pacific Biosciences and Oxford Nanopore Technologies reveals platform-specific error patterns. *Genome Res.* 32(4):726–37 [PubMed: 35301264]
80. Mohammadi P, Castel SE, Brown AA, Lappalainen T. 2017. Quantifying the regulatory effect size of *cis*-acting genetic variation using allelic fold change. *Genome Res.* 27(11):1872–84 [PubMed: 29021289]
81. Mohammadi P, Castel SE, Cummings BB, Einson J, Sousa C, et al. 2019. Genetic regulatory variation in populations informs transcriptome analysis in rare disease. *Science* 366(6463):351–56 [PubMed: 31601707]
82. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, et al. 2018. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* 560(7718):319–24 [PubMed: 30069044]
83. Newton Y, Sedgewick AJ, Cisneros L, Golovato J, Johnson M, et al. 2020. Large scale, robust, and accurate whole transcriptome profiling from clinical formalin-fixed paraffin-embedded samples. *Sci. Rep.* 10:17597 [PubMed: 33077815]
84. Nomakuchi TT, Rigo F, Aznarez I, Krainer AR. 2016. Antisense oligonucleotide-directed inhibition of nonsense-mediated mRNA decay. *Nat. Biotechnol.* 34(2):164–66 [PubMed: 26655495]

85. Novak G, Kyriakis D, Grzyb K, Bernini M, Rodius S, et al. 2022. Single-cell transcriptomics of human iPSC differentiation dynamics reveal a core molecular network of Parkinson's disease. *Commun. Biol.* 5:49 [PubMed: 35027645]
86. Nurk S, Koren S, Rhie A, Rautiainen M, Bizakdzic AV, et al. 2022. The complete sequence of a human genome. *Science* 376(6588):44–53 [PubMed: 35357919]
87. Pilarczyk M, Fazel-Najafabadi M, Kouril M, Shamsaei B, Vasiliauskas J, et al. 2022. Connecting omics signatures and revealing biological mechanisms with iLINCS. *Nat. Commun.* 13:4678 [PubMed: 35945222]
88. Porcu E, Sadler MC, Lepik K, Auwerx C, Wood AR, et al. 2021. Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. *Nat. Commun.* 12:5647 [PubMed: 34561431]
89. Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. 2022. Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* 38(10):2943–45 [PubMed: 35561197]
90. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, et al. 2017. The Human Cell Atlas. *eLife* 6:e27041 [PubMed: 29206104]
91. Rigo F, Hua Y, Krainer AR, Bennett CF. 2012. Antisense-based therapy for the treatment of spinal muscular atrophy. *J. Cell Biol.* 199(1):21–25 [PubMed: 23027901]
92. Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, et al. 2015. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* 348(6235):666–69 [PubMed: 25954003]
93. Roberts TC, Wood MJA, Davies KE. 2023. Therapeutic approaches for Duchenne muscular dystrophy. *Nat. Rev. Drug Discov.* 22(11):917–34 [PubMed: 37652974]
94. Runne H, Kuhn A, Wild EJ, Pratyaksha W, Kristiansen M, et al. 2007. Analysis of potential transcriptomic biomarkers for Huntington's disease in peripheral blood. *PNAS* 104(36):14424–29 [PubMed: 17724341]
95. Setoain J, Franch M, Martínez M, Tabas-Madrid D, Sorzano COS, et al. 2015. NFFinder: an online bioinformatics tool for searching similar transcriptomics experiments in the context of drug repositioning. *Nucleic Acids Res.* 43(W1):W193–99 [PubMed: 25940629]
96. Sha Y, Phan JH, Wang MD. 2015. Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data. In 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6461–64. Piscataway, NJ: IEEE
97. Shiraishi Y, Okada A, Chiba K, Kawachi A, Omori I, et al. 2022. Systematic identification of intron retention associated variants from massive publicly available transcriptome sequencing data. *Nat. Commun.* 13:5357 [PubMed: 36175409]
98. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* 15(2):121–32 [PubMed: 24434847]
99. Sladen PE, Perdigão PRL, Salisbury G, Novoselova T, van der Spuy J, et al. 2021. CRISPR-Cas9 correction of *OPA1* c.1334G>A: p.R445H restores mitochondrial homeostasis in dominant optic atrophy patient-derived iPSCs. *Mol. Ther. Nucleic Acids* 26:432–43 [PubMed: 34589289]
100. Song J, Kim D, Lee S, Jung J, Joo JWJ, Jang W. 2022. Integrative transcriptome-wide analysis of atopic dermatitis for drug repositioning. *Commun. Biol.* 5:615 [PubMed: 35729261]
101. Statello L, Guo C-J, Chen L-L, Huarte M. 2021. Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* 22(2):96–118. Correction. 2021. *Nat. Rev. Mol. Cell Biol.* 22(2):159 [PubMed: 33353982]
102. Stegle O, Parts L, Piipari M, Winn J, Durbin R. 2012. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* 7(3):500–7 [PubMed: 22343431]
103. Stergachis AB, Blue EE, Gillentine MA, Wang L-K, Schwarze U, et al. 2023. Full-length isoform sequencing for resolving the molecular basis of Charcot-Marie-Tooth 2A. *Neurol. Genet.* 9(5):e200090 [PubMed: 37560121]
104. Strand AD, Aragaki AK, Shaw D, Bird T, Holton J, et al. 2005. Gene expression in Huntington's disease skeletal muscle: a potential biomarker. *Hum. Mol. Genet.* 14(13):1863–76 [PubMed: 15888475]

105. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, et al. 2017. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* 171(6):1437–52.e17 [PubMed: 29195078]
106. Sun T, Sun Z, Jiang Y, Ferguson AA, Pilewski JM, et al. 2019. Transcriptomic responses to ivacaftor and prediction of ivacaftor clinical responsiveness. *Am. J. Respir. Cell Mol. Biol.* 61(5):643–52 [PubMed: 30995102]
107. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, et al. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6(5):377–82 [PubMed: 19349980]
108. Tang Q, Li W, Huang J, Wu Y, Ma C, et al. 2023. Single-cell sequencing analysis of peripheral blood in patients with moyamoya disease. *Orphanet J. Rare Dis.* 18:174 [PubMed: 37400835]
109. Teran NA, Nachun DC, Eulalio T, Ferraro NM, Smail C, et al. 2021. Nonsense-mediated decay is highly stable across individuals and tissues. *Am. J. Hum. Genet.* 108(8):1401–8 [PubMed: 34216550]
110. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, et al. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28(5):511–15 [PubMed: 20436464]
111. van den Brandt PA, Goldbohm RA, van 't Veer P, Hermus RJ, Sturmans F. 1988. Dietary habits and the aetiology of cancer. *Int. J. Epidemiol.* 17(2):472 [PubMed: 3403144]
112. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* 43(1110):11.10.1–33
113. Vaquero-Garcia J, Barrera A, Gazzara MR, González-Vallinas J, Lahens NF, et al. 2016. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *eLife* 5:e11752 [PubMed: 26829591]
114. Volden R, Palmer T, Byrne A, Cole C, Schmitz RJ, et al. 2018. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *PNAS* 115(39):9726–31 [PubMed: 30201725]
115. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, et al. 2019. Opportunities and challenges for transcriptome-wide association studies. *Nat. Genet.* 51(4):592–99 [PubMed: 30926968]
116. Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10(1):57–63 [PubMed: 19015660]
117. Wang Z, Wu Z, Wang H, Feng R, Wang G, et al. 2023. An immune cell atlas reveals the dynamics of human macrophage specification during prenatal development. *Cell* 186(20):4454–71.e19 [PubMed: 37703875]
118. Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A. 2022. An introduction to spatial transcriptomics for biomedical research. *Genome Med.* 14:68 [PubMed: 35761361]
119. Wu P, Feng Q, Kerchberger VE, Nelson SD, Chen Q, et al. 2022. Integrating gene expression and clinical data to identify drug repurposing candidates for hyperlipidemia and hypertension. *Nat. Commun.* 13:46 [PubMed: 35013250]
120. Xiao Q, Lauschke VM. 2021. The prevalence, genetic complexity and population-specific founder effects of human autosomal recessive disorders. *npj Genom. Med.* 6:41 [PubMed: 34078906]
121. Yépez VA, Gusic M, Kopajtich R, Mertes C, Smith NH, et al. 2022. Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med.* 14:38 [PubMed: 35379322]
122. Yépez VA, Mertes C, Müller MF, Klapproth-Andrade D, Wachutka L, et al. 2021. Detection of aberrant gene expression events in RNA sequencing data. *Nat. Protoc.* 16(2):1276–96 [PubMed: 33462443]
123. Yoon OK, Hsu TY, Im JH, Brem RB. 2012. Genetics and regulatory impact of alternative polyadenylation in human B-lymphoblastoid cells. *PLOS Genet.* 8(8):e1002882 [PubMed: 22916029]
124. Zeng T, Li YI. 2022. Predicting RNA splicing from DNA sequence using Pangolin. *Genome Biol.* 23:103 [PubMed: 35449021]

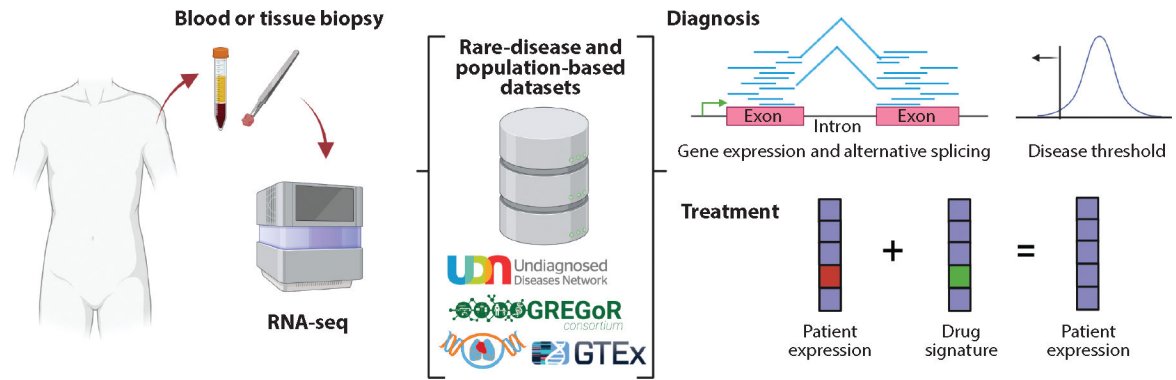
125. Zhao S, Zhang Y, Gamini R, Zhang B, von Schack D. 2018. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. *Sci. Rep.* 8:4781 [PubMed: 29556074]
126. Zhernakova DV, de Klerk E, Westra H-J, Mastrokolias A, Amini S, et al. 2013. DeepSAGE reveals genetic variants associated with alternative polyadenylation and expression of coding and non-coding transcripts. *PLOS Genet.* 9(6):e1003594 [PubMed: 23818875]
127. Zwiener I, Frisch B, Binder H. 2014. Transforming RNA-Seq data to improve the performance of prognostic gene signatures. *PLOS ONE* 9(1):e85150 [PubMed: 24416353]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 1.**

RNA-seq in disease diagnosis and treatment. A patient's biospecimen is subject to RNA-seq and subsequently deposited in or compared with rare-disease and population-based RNA-seq datasets. Outlier gene expression or alternative splicing can inform an underlying pathogenic effect to assist with diagnosis, and comparison with drug signatures that may restore normal molecular function provides an opportunity to identify treatment strategies. Abbreviations: GREGoR, Genomics Research to Elucidate the Genetics of Rare Diseases; GTEx, Genotype–Tissue Expression; RNA-seq, RNA sequencing. Figure adapted from images created with [BioRender.com](https://www.biorender.com).