Received: 22 December 2021 | Revised: 16 March 2022 | Accepted: 26 April 2022
DOI: 10.1002/humu.24394

REVIEW

Transcriptome analysis provides critical answers to the "variants of uncertain significance" conundrum

Mackenzie D. Postel[1,2] | Julie O. Culver[2] | Charité Ricker[2] | David W. Craig[1,2]

[1]Department of Translational Genomics, University of Southern California, Los Angeles, California, USA

[2]Keck School of Medicine of USC, University of Southern California, Los Angeles, California, USA

Correspondence
David W. Craig, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA 90033, USA.
Email: davidwcr@usc.edu

Funding information
National Cancer Institute, Grant/Award Number: 1U2CCA252971-01A1

Abstract

While whole-genome and exome sequencing have transformed our collective understanding of genetics' role in disease pathogenesis, there are certain conditions and populations for whom DNA-level data fails to identify the underlying genetic etiology. Specifically, patients of non-White race and non-European ancestry are disproportionately affected by "variants of unknown/uncertain significance" (VUS), limiting the scope of precision medicine for minority patients and perpetuating health disparities. VUS often include deep intronic and splicing variants which are difficult to interpret from DNA data alone. RNA analysis can illuminate the consequences of VUS, thereby allowing for their reclassification as pathogenic versus benign. Here we review the critical role transcriptome analysis plays in clarifying VUS in both neoplastic and non-neoplastic diseases.

KEYWORDS

deep intronic variants, genetic ancestry, splicing variants, variants of uncertain significance, variants of unknown significance

1 | INTRODUCTION

Next-generation sequencing (NGS) technology has revolutionized and facilitated the genetic diagnosis of disease. Clinicians and researchers can now routinely sequence at a genome-wide level rather than performing the tedious process of sequencing and interrogating one gene at a time. However, with this scientific windfall a new problem emerges: the challenge of identifying pathogenic variants from millions of benign variants. Some variants can be excluded with confidence upon the basis of allele frequencies or gene pathways. However, there are many variants of unknown/uncertain significance (VUS) for which it is not as evident whether a variant is benign or pathogenic. VUS can take a significant psychological and emotional toll on patients and their families and may lead to clinical mismanagement. Importantly, VUS affect minority populations disproportionately (Culver et al., 2013; Gelfman et al., 2017; Maurano et al., 2012; Park et al., 2018; Qian et al., 2021; Vaz-Drago et al., 2017).

VUS have served as the subject of entire special issues of this journal, *Human Mutation* (e.g., 2008) (Tavtigian et al., 2008). The precise meaning of VUS depends upon context and connotation. The American College of Medical Genetics (ACMG) (Richards et al., 2018), the Association for Molecular Pathology (AMP) (M. M. Li, et al., 2017), and the International Agency for Research on Cancer (IARC) (Plon et al., 2008) have delineated definitions and guidelines for the interpretation of VUS. At a simple level, the sheer numbers of intronic, synonymous, and other "deep" variants contribute to their overall predictive value as a variant class. However, in clinical testing,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2022 The Authors. *Human Mutation* published by Wiley Periodicals LLC.

entire variant classes (such as synonymous variants) are presumed to be likely benign from the outset, ensuring that VUS are less frequently reported. This is despite the fact that important disease-causing variants can fall into these neglected, "opaque" areas of variant interpretation.

There is an overwhelming need for better tools to elucidate the functional consequences of rare molecular events and, thereby, illuminate how genetic variation impacts biology and disease pathogenesis. Analysis of DNA-level information alone in the context of clinical phenotypes is unlikely to provide clarity given rare events' limited sample size and lack of statistical power. Compared to genomic analysis—which outlines the repertoire of functions encoded in DNA that a cell *could* perform—transcriptomic analysis enables a deeper understanding of how a cell is *actually* functioning. In other words, RNA expression profiling provides critical insight into how genotype translates into phenotype (Mantione et al., 2014).

Here, we will highlight how gene expression profiling provides functional insight into variants beyond the initial 2 base pairs (bps) of a splice junction. We will review the emerging role of RNA sequencing as a tool to differentiate VUS as pathogenic versus benign, especially in populations of diverse races, ethnicities, and genetic ancestries, as individuals of non-European ancestry have significantly higher rates of VUS (Petrovski & Goldstein, 2016). Finally, we will explore how transcriptomic analysis can be translated into clinical diagnostics for cancer (i.e., detection of causative variants, both germline and somatic) and how it can be applied to address cancer disparities.

1.1 | Leveraging the transcriptome for molecular diagnoses

The term "RNA-seq" was coined in 2008. Around the same time, several landmark studies promoted the potential diagnostic utility of RNA sequencing data using short-read NGS. One major RNA-focused effort was the ENCODE Project Consortium, which stated: "*to understand the human genome… and the ways in which its defects can give rise to disease, we need a more transparent view of the information it* ***encodes***" (Birney et al., 2007). An understanding of human disease necessitates an understanding of the link between genomic variation and transcription; this was also the goal of the Genotype-Tissue Expression (*GTEx*) project started in 2010 (GTex Consortium, 2013). Analogous to the 1000 Genomes Project, which aimed to provide a catalog of common variation, *GTEx* aimed to catalog genotype and gene-expression correlations across dozens of tissues and thousands of individuals.

Addressing VUS necessitates an improved ability to predict the consequences of genetic variation on transcription. RNA analysis allows for the appreciation of phenomena like differential expression, allele-specific expression (ASE), alternative splicing, isoform switching, and many other events that would not be readily apparent in DNA data. Current approaches are largely limited to the annotation of variant classes that are both rare and predictable for transcript disruption, such as nonsense mutations that lead to loss of transcript expression. Compared to rare variants, more common variant classes have higher prevalence but lower accuracy when it comes to predicting gene expression consequences.

Germline variants and somatic mutations can impact RNA species in a variety of ways. They are often grouped into categories according to functional impact (i.e., High, Moderate, Low/Modifier) using annotation ontology software such as *ANNOVAR* (Wang et al., 2010; Yang & Wang, 2015) and *SnpEff* (Cingolani et al., 2012) (Table 1). Variants falling into the *High* functional impact category include nonsense variants that introduce premature stop-codons, start- and stop-loss mutations, and 5-prime (5′) untranslated region (UTR) premature start codons. Many of these lead to lowered transcription levels or full loss of transcripts through nonsense-mediated decay (NMD) (Chakravorty & Hegde, 2018). Another relatively rare variant class includes splice-site disrupting variants within the exon-intron boundaries' +/−2bp region. These variant classes can impact splicing and result in exon skipping, exon usage, mutually exclusive/inclusive exons, or intron retention (Woolfe et al., 2010). Variants within the *Moderate* category are typically those predicted to impact amino acid composition (missense), including non-synonymous variants and in-frame insertions/deletions.

The remainder of variants are frequently found on a per-genome basis and are often pragmatically filtered out due to their low predictive value. These include synonymous variants, regulatory region variants, 3-prime (3′) UTR variants, 5′ UTR variants, and intronic variants (e.g., splice region variants proximal to exon boundaries). Importantly, as shown in Figure 1A, events beyond the 2bp exon donor/acceptor window can still impact function. Variants annotated as missense and synonymous can alter splicing, exposing so-called cryptic splice-sites and splice-enhancing or silencing elements (Woolfe et al., 2010). Variants deep within the intron—as far out as 18bp—can impact spliceosome binding at a key motif referred to as the "lariat junction."

There are numerous examples of disease driven by these "opaquer" variants. These include cryptic splice-sites (e.g., NM_000518.5:c.316-146T>G in the hemoglobin subunit beta gene, *HBB*, causing β-thalassemia), UTR variants (e.g., NM_000166.6:c.−103C>T in the gap junction protein beta 1 gene, *GJB1*, causing Charcot-Marie-Tooth neuropathy), and other deep intronic events (e.g., NM_033380.3:c.385-719G>A in collagen type IV alpha 5 chain, *COL4A5*, causing Alport Syndrome) (Dobkin et al., 1983; King et al., 2002). The mechanism of impact is frequently evident in transcriptome data.

Novel RNA functions/phenomena have been elucidated extending far beyond the traditional protein-coding role described by the "central dogma." We are also beginning to better understand how messenger RNA (mRNA) can be processed—chiefly by alternative splicing, RNA editing, and crosstalk between the two processes – to create the great variety that characterizes the human transcriptome (Tang et al., 2020). Examples of unique phenomena into which gene expression data provides insight, and which have significant disease implications, include splicing mutations, deep intronic truncating mutations leading to NMD and loss of mRNA, and ASE.

TABLE 1 Selected RNA-seq software and databases.

Selected tools Software	Purpose	Source
ANNOVAR	Variant annotation and functional effect prediction	https://github.com/WGLab/doc-ANNOVAR/
SnpEff	Variant annotation and functional effect prediction	http://pcingola.github.io/SnpEff/
DeepSplice	Deep learning-based splice junction sequence classifier	https://github.com/zhangyimc/DeepSplice/
SQUIRLS	Interpretation of splicing variants outside of the canonical splice sites	https://github.com/TheJacksonLaboratory/Squirls/
SpliceAI	Deep learning-based splice variant identification	https://github.com/Illumina/SpliceAI/
LeafCutterMD	Outlier splicing detection	https://davidaknowles.github.io/leafcutter/
FRASER	Detection of rare splicing events	https://github.com/c-mertes/FRASER/
SpliceSeq	Analysis/visualization of alternative splicing and functional impacts	https://github.com/MD-Anderson-Bioinformatics/SpliceSeq/
SpliceV	Analysis/visualization of linear and circRNA splicing, expression, and regulation	https://github.com/flemingtonlab/SpliceV/
STAR	RNA-seq aligner with support for splice-junction detection	https://github.com/alexdobin/STAR/
Cufflinks/cuffdiff2	Transcriptome assembly and differential expression analysis for RNA-seq data	http://cole-trapnell-lab.github.io/cufflinks/
DiffSplice	Detection of differential splicing events from RNA-seq data	http://www.netlab.uky.edu/p/bioinfo/DiffSplice/
DEXSeq	Inference of differential exon usage from RNA-seq data	https://bioconductor.org/packages/release/bioc/html/DEXSeq.html/
edgeR	Differential expression analysis of RNA-seq expression profiles	https://bioconductor.org/packages/release/bioc/html/edgeR.html/
JunctionSeq	Detection/visualization of differential exon and splice juntion usage from RNA-seq data	https://github.com/hartleys/JunctionSeq/
limma	Differential expression analysis of microarray data	https://bioconductor.org/packages/release/bioc/html/limma.html/
dSpliceType	Detection of differential splicing and expression events from RNA-seq data	http://dsplicetype.sourceforge.net/
MAJIQ/Voila	Detection, quantification, and visualization of local splicing variations from RNA-seq data	https://majiq.biociphers.org/
rMATS	Detection of differential alternative splicing events from RNA-seq data	http://rnaseq-mats.sourceforge.net/
MISO	Quantification of expression of alternatively spliced genes and identification of differentially regulated isoforms or exons from RNA-seq data	http://hollywood.mit.edu/burgelab/miso/
SUPPA/SUPPA2	Differential splicing analysis	https://github.com/comprna/SUPPA/
Salmon	RNA-seq expression quantification	https://combine-lab.github.io/salmon/
DESeq2	RNA-seq differential gene expression analysis based on the negative binomial distribution	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Kallisto	RNA-seq expression quantification based on pseudoalignment	https://pachterlab.github.io/kallisto/about/
UclncR	Detection of lncRNAs from RNA-seq data	https://bioinformaticstools.mayo.edu/research/uclncr-pipeline/
miRDeep2	Identification of novel and known miRNAs in deep sequencing data	https://github.com/rajewsky-lab/mirdeep2/
CAP-miRSeq	Analysis of miRNA-seq data	https://bioinformaticstools.mayo.edu/research/cap-mirseq/
iMir	Analysis of small ncRNA data	https://tools4mirs.org/software/isomirs_identification/imir/

TABLE 1 (Continued)

Selected tools Software	Purpose	Source
piPipes	Analysis of piRNA and transposons via small RNA-seq, RNA-seq, degradome- and CAGE-seq, ChIP-seq, and genomic DNA sequencing	https://github.com/bowhan/piPipes/
UEA sRNA workbench	Analysis of smallRNA data	https://github.com/sRNAworkbenchuea/UEA_sRNA_Workbench/
omiRas	Differential expression analysis of miRNAs derived from small RNA-seq data	http://tools.genxpro.net/omiras/
sRNAtoolbox	Analysis of smallRNA data	https://arn.ugr.es/srnatoolbox/
FlaiMapper	Annotation of small ncRNA-derived fragments	https://github.com/yhoogstrate/flaimapper
tDRMapper	Mapping, naming, and quantification of tRNA-derived RNAs	https://github.com/sararselitsky/tDRmapper
UROBORUS	Identification of circRNA from total RNA-seq data	https://github.com/WGLab/UROBORUS
PredCircRNA	Classification of circular RNA from other long noncoding RNA	https://github.com/xypan1232/PredcircRNA
find_circ	Analysis of circular RNAs	https://github.com/marvin-jens/find_circ
CircExplorer	Annotation and analysis of circRNA	https://github.com/YangLab/CIRCexplorer
CIRI	*De novo* circular RNA identification	https://bio.tools/CIRI-full
iSeeRNA	Identification of lincRNA transcripts from RNA-seq data	https://sunlab.cpy.cuhk.edu.hk/iSeeRNA/webserver.html
Sebnif	Identification of novel lincRNAs	https://sunlab.cpy.cuhk.edu.hk/sebnif/webserver/
LncRNA2Function	Functional investigation of lncRNAs from RNA-seq data	http://mlg.hit.edu.cn/lncrna2function
Database		
dbscSNV	Database of SNVs within splicing consensus regions and their functional annotations	http://www.liulab.science/dbscsnv.html
ValidSpliceMut	Database of validated splicing mutations allowing for prediction/exploration of user data	https://validsplicemut.cytognomix.com/
ClinVar	Database of human variation and associated phenotypes	https://www.ncbi.nlm.nih.gov/clinvar/
gnomAD	Database of aggregated/harmonized, summary-level exome and genome sequencing data	https://gnomad.broadinstitute.org/
RNAcentral	Database of noncoding RNA sequences	https://rnacentral.org/
SpliceDB	Database of canonical and noncanonical mammalian splice sites	http://www.softberry.com/spldb/SpliceDB.html
ORNA-seq	Ontology for annotation of RNA-seq data	https://github.com/safisher/oRNA-seq
GTEX	Database of tissue-specific gene expression and regulation	https://gtexportal.org/home/
miREV	Database of small RNA-seq data from extracellular vesicle enriched samples	https://www.physio.wzw.tum.de/mirev/

1.2 | Aberrant splicing mutations beyond the canonical donor/acceptor window

Functionally, splicing occurs via a conglomerate of proteins and small nuclear RNAs called the spliceosome, forming complementary RNA-RNA complexes with target RNAs (Anna & Monika, 2018). The spliceosome catalyzes the splicing of pre-mRNA into mRNA and circular lariat RNAs (Talhouarne & Gall, 2018). Variants disrupting RNA splicing can occur throughout a given gene and, without RNA-based functional evidence, can be annotated into less-obvious categories including synonymous, intronic, and UTR variants. The spliceosome catalyzes the splicing of pre-mRNA into mRNA and circular lariat RNAs, the latter of which are destroyed (Talhouarne & Gall, 2018). Variants at the canonical boundary motifs between exons and introns—that is, *GT* (at the 5′ end) and *AG* (at the 3′ end)—are generally prioritized but, as shown in Figure 1a, variants upstream and downstream can disrupt function at a lower certainty/frequency (Anna & Monika, 2018). Adding to this mechanistic complexity is the existence of a variety of spliceosomes

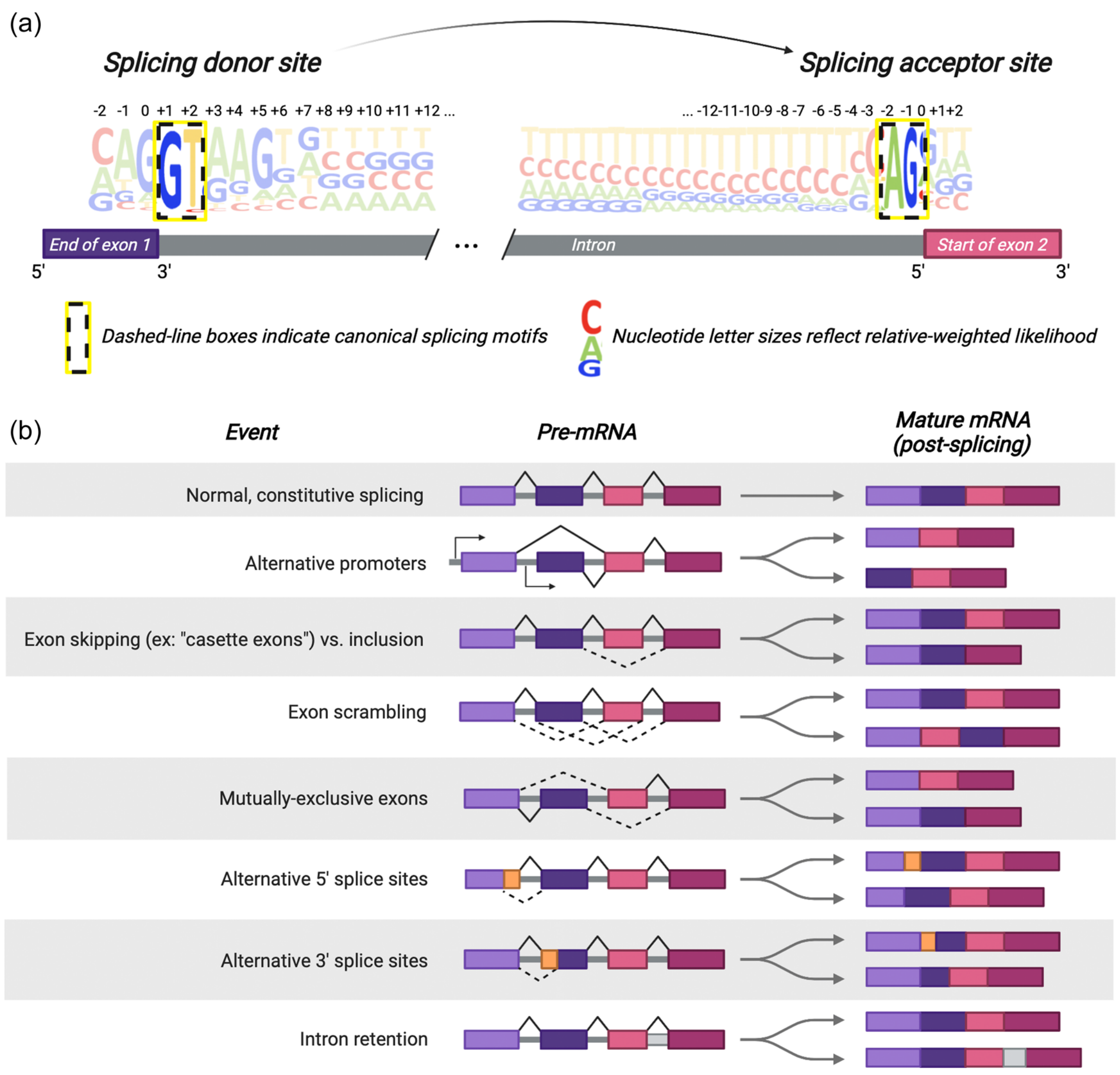

FIGURE 1 Splice site motifs and variants. (a) A schematic of two exons separated by an intron in pre-mRNA. Shown above the pre-mRNA are the corresponding, relative-weighted DNA nucleotide motifs at splicing donor and acceptor sites. Though the canonical/consensus boundary motifs are GT (or GU, in RNA, after the 3′ end of exon #1) and AG (before the 5′ end of exon #2), these are probabilistic. Other boundary motifs exist and positions up/downstream of the motifs are also important for splicing. RNA-seq data can help isolate functionally relevant, potentially pathogenic splicing variants. Canonical boundary motifs are demarcated with dashed boxes. (b) Depicted are schematics of seven representative types of splicing mutations, their mechanisms in pre-mRNA, and their potential consequences in mature mRNA. mRNA, messenger RNA. *Figure created with BioRender.com*

("major" and "minor") as well as noncanonical—aka "cryptic" or "pseudo"—splice site sequences (e.g., *GC-AG* and *AT-AC*). The presence of cis-acting regulatory elements (i.e., splicing enhancers vs. suppressors) and the physical structures of a branch site and polypyrimidine tract (sequences that bind spliceosome proteins) influence which splice site ends up being used (Anna & Monika, 2018).

To minimize false-positive rates, candidate variant lists derived from DNA sequencing typically only take into consideration *canonical* changes to boundary motifs (as mentioned earlier: 5′-*GT* and 3′-*AG*) or variants that lie within 100 base pairs of canonical splice sites (i.e., *not* deep intronic variants). This approach is pragmatic, as these variants are both rare and highly-penetrant, allowing for overall high precision (or positive predictive value). Without additional functional data, the large number of variants existing further down- or upstream would lead to low overall precision. However, evidence is mounting in support of the fact that deep-intronic, noncanonical splicing variants can drive disease pathology (Blakes et al., 2022; Koster et al., 2021).

RNA-seq data can provide a direct functional measure of splicing, allowing for fewer variants (with higher inherent accuracy and precision) to be considered. Furthermore, it provides a direct measure of spliced events. As shown in Figure 1b, there are *at least* 7 classes of splicing events including: (1) exon skipping, (2) mutually-exclusive exons (i.e., a coordinated set of splicing events where, as a result, only one of two exons is retained), (3) exon scrambling, (4) intron retention, (5) alternative promoters and terminators, and (6, 7) alternative donor and acceptor sites (aka alternative 5′ and 3′ splice

sites). We have used the phrasing "at least" here because there is no true consensus as to the number of splicing events possible and most bioinformatic tools do not define/detect all seven of these events (Halperin et al., 2021; Shen et al., 2014; Wang et al., 2015). Transcriptome data also facilitates consideration of other, typically filtered-out variants including synonymous and UTR variants (Chen & Weiss, 2015; Pohl et al., 2013; Shi et al., 2018).

The emergence of RNA-seq in conjunction with germline DNA sequencing spurred the optimization, development, and training of new bioinformatic tools for splice-site prediction and detection. Splice-site prediction focuses on using advanced algorithms and artificial intelligence (AI) methods to predict whether a variant will impact splicing from DNA data alone. Pragmatically, these approaches have higher utility when RNA-seq data is also available, as the combined analysis reduces the overall search space. Splice variant prediction tools typically employ one of two strategies: adaptive models or random forest analysis. The essential difference between these two methods is whether one uses large databases from which nonpathogenic versus pathogenic splice variants can be trained. Methods like those employed by Liu et al. (2016) utilize random forests to generate probability scores that can be used to estimate the potential of a 3–12bp window within the lariat RNA junction. Super Quick Information-content Random-forest Learning of Splice variants (*SQUIRLS*) is a recently-published algorithm specializing in the interpretation of splicing mutations outside of canonical splice sites (Danis et al., 2021) (Table 1). *DeepSplice* is an example of a tool that utilizes deep convolutional neural networks, as well as paired events to reduce false positives (Zhang et al., 2018) (Table 1). *SpliceAI* uses a deep neural network to model mRNA splicing from noncoding sites, yielding a 10% rate of pathogenic variant discovery in neurodevelopmental disorders (Jaganathan et al., 2019; Sanders et al., 2020) (Table 1). These are just a few emerging algorithms and, in practice, databases serve as vehicles for querying calculated predictive scores, such as with *dbscSNV* (a database of SNVs within splicing consensus regions and their functional annotations) (Jian et al., 2014) (Table 1). *ValidSpliceMut* is a database of validated splicing mutations that allows users to predict and explore splicing variants (Mucaki et al., 2020) (Table 1).

Other informatic algorithms directly utilize RNA-seq data in their analysis, such as *LeafCutterMD* (Jenkinson et al., 2020) and *FRASER* (Find RAre Splicing Events in RNA-seq) (Mertes et al., 2021) (Table 1). *SpliceSeq* is an example of a tool that utilizes "splice graph" models (Ryan et al., 2012) (Table 1). Tools like *SpliceV* facilitate discovery and visualization of splicing events (Ungerleider & Flemington, 2019) (Table 1). Underlying these approaches lies layers of additional nuance with variation in alignment strategies (such as variant-aware aligners like *STAR*) and assembly (Hong et al., 2018). Other RNA-seq bioinformatic tools that can aid RNA splicing analysis include *Cufflinks/cuffdiff2*, *DiffSplice*, *DEXSeq*, *edgeR*, *JunctionSeq*, *limma*, *dSpliceType*, *MAJIQ/Voila*, *rMATS*, *MISO*, *SUPPA/SUPPA2*, *Salmon*, *Kallisto*, and *DESeq2* (Halperin et al., 2021; Mehmood et al., 2020; Muller et al., 2021) (Table 1).

As methods rapidly evolve, implemented approaches vary substantially across and within labs, adversely impacting database resources. One useful metric when it comes to assessing alternative splicing events and harmonizing data is "PSI" or percent spliced in; also known as the exon-inclusion ratio, PSI indicates how often a given exon occurs in all isoforms of the gene that contains said exon (Tanner et al., 2021). One recently published study found that a standardized RNA diagnostic protocol was capable of reclassifying 75% of putative splicing variants (Bournazos et al., 2022).

It is worth noting that such splicing analysis has both diagnostic and therapeutic utility. For instance, tumors with identified splicing variants could be targeted with therapies that inhibit the spliceosome, splicing regulatory proteins, or aberrant splicing products (Lee & Abdel-Wahab, 2016; Scotti & Swanson, 2016).

1.3 | Indirect insights from allele-specific expression and nonsense-mediated decay

Another indirect phenomenon into which RNA-seq can provide insight (and which DNA-only analysis cannot appreciate) is ASE. The initial definition of ASE was a purely germline one: preferential expression of one parental allele in a heterozygous individual (Shao et al., 2019). The definition of ASE has since expanded to encompass somatic events. Fundamentally with RNA-seq, ASE requires heterozygous proxy variants, often single-nucleotide polymorphism (SNPs), that facilitate phasing (or at least quantification) of the relative expression between maternally and paternally inherited chromosomes (Figure 2). These proxy-SNPs include common and rare synonymous, UTR, and missense variants. With ASE, one often infers the loss of a pathogenic allele by monogenic expression of the wild-type allele. Processes such as NMD will actively remove transcripts containing pathogenic mutations. NMD is the process whereby a trimeric complex of proteins degrades mRNA that would otherwise result in potentially-pathogenic truncated proteins; it has been posited that NMD also plays a role in the regulation of *normal* mRNA expression, as well as regulation of alternative splicing via degradation of splice variants with premature termination codons (Brogna & Wen, 2009).

Detection of ASE can lead to discoveries of VUS and cancer-related genes. One common mechanism in the pathogenesis of cancer is a nonsense, somatic mutation in a tumor suppressor that leads to NMD and, by means of haploinsufficiency or dominant-negative effects, loss of function. A study in individuals with hereditary pancreatic cancer identified a heterozygous SNP (rs144848) in the gene BRCA2 DNA repair associated (*BRCA2*) that displayed ASE in RNA. The mechanism underlying this VUS was determined to be a truncating mutation leading to NMD (Tan et al., 2008).

One key historical example of ASE involves the measure of X-chromosome inactivation or "skewing" (a physiological process normal in females) and genomic imprinting to gain insight into the preferential selection of variants along the X-chromosome (Shao et al., 2019). In genetic females, ASE also reflects the loss of transcript expression from one allele due to X-inactivation, an important methylation-driven "dosage

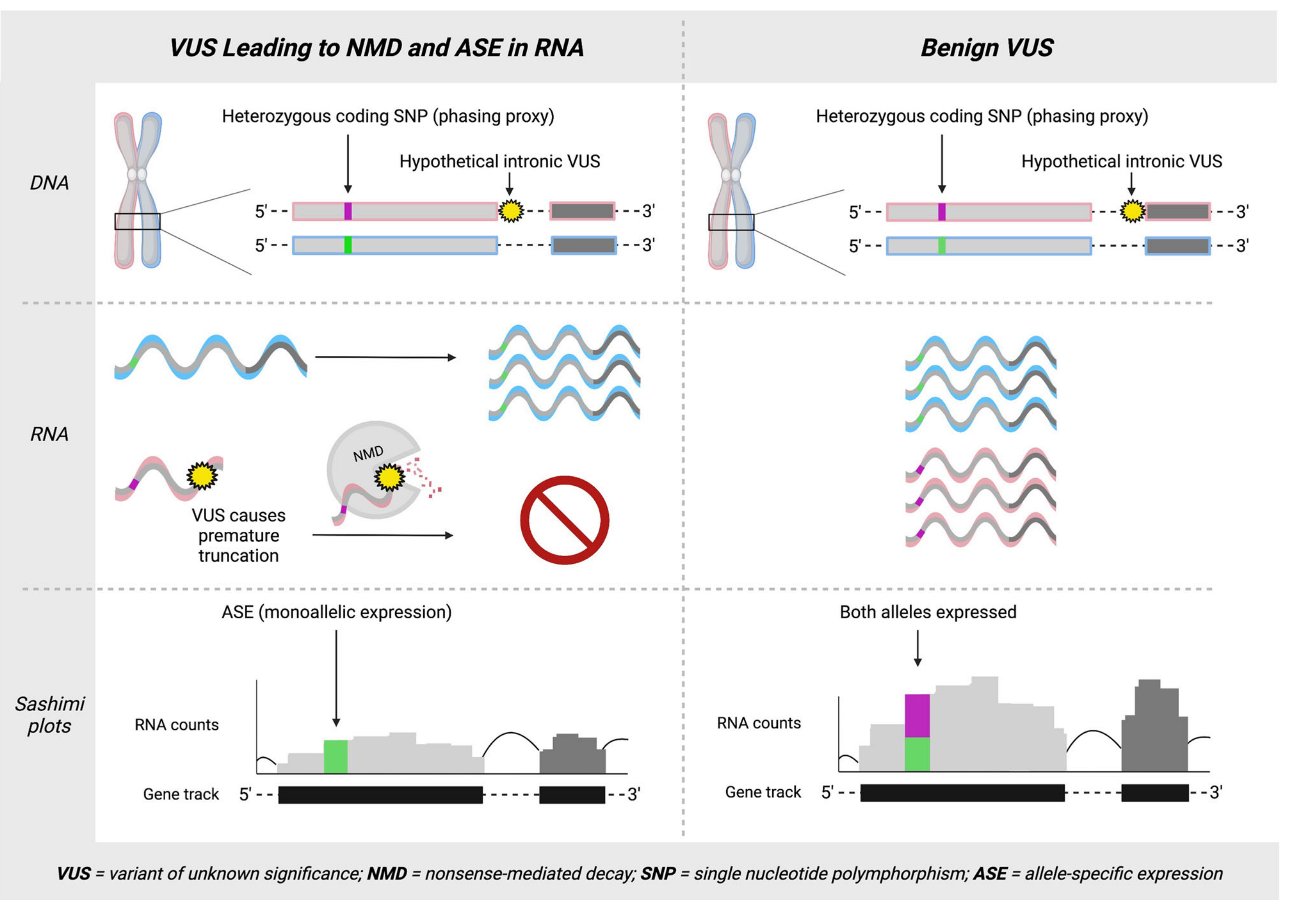

FIGURE 2 A hypothetical variant of unknown significance (VUS) leads to nonsense-mediated decay (NMD), monoallelic allele-specific expression (ASE), and loss of messenger RNA (mRNA) transcripts. Phasing of alleles (maternal vs. paternal) is indicated by pink (maternal) and blue (paternal) highlights. In the left-hand panels we depict the hypothetical case of a loss-of-function intronic VUS (depicted as a yellow sunburst symbol) on the DNA-level. Also shown schematically is a coding heterozygous single-nucleotide polymorphism (SNP) upstream (shown as green and magenta color-coded alleles). The VUS in this case causes intron fragment inclusion and when a premature stop codon is reached, premature truncation. The truncated mRNA is degraded via NMD, and complete loss of associated mRNA transcripts is observed. This is evidenced by the fact that in the bottom left sashimi plot, (1) we do not see expression of the magenta allele (a phasing proxy for the VUS), and (2) we do see monoallelic ASE of the green allele. ASE is the process whereby only one allele is expressed in RNA despite the fact that an individual is heterozygous at that position in their DNA. Key to recognizing whether an individual is displaying ASE is the identification of a coding heterozygous SNP up/downstream that can act as a phasing proxy. In contrast, in the right-hand panels, we depict the scenario of a benign VUS whereby we do not see a loss of function in RNA; instead, in the sashimi plot, we see a roughly equal expression of both alleles (magenta and green) at the upstream heterozygous locus, as well as the correct "dosage" of expression. *Figure created with BioRender.com.*

compensation mechanism" (Shvetsova et al., 2019). Historically, X-skewing has been detected using the human androgen receptor (*HUMARA)* assay, which takes advantage of differential methylation at the androgen receptor locus. However, RNA-seq along expressed X-chromosome genes provides additional insights. One can better observe skewing directionality (e.g., 95% skewing in X and preferential expression of wild-type alleles) (Szelinger et al., 2014).

A specific, translational example of how consideration of ASE in RNA-seq data can lead to a molecular diagnosis is that of a candidate variant within the X-linked gene TATA-box binding protein associated factor 1 gene (*TAF1*) in a male pediatric patient. Our group leveraged RNA-seq data for the boy's healthy mother, which showed extreme X-skewing towards the protective wild-type allele and away from our candidate VUS (Hurst et al., 2018). In other words, RNA-seq was used to evaluate a model whereby X-skewing protected the mother. Further sequencing showed that the de novo event was founded within the mother and was likely selected for, given the boy's parents' history of unsuccessful pregnancies. This example has generalizability as it is relevant to other X-linked disorders, including Alport, Charcot-Marie-Tooth, Fabry, Fanconi, Fragile X, Hunter, Rett, and Wiskott-Aldrich syndromes (Migeon, 2020).

1.4 | Outlier analysis as a solution to power problems

Case–control studies are feasible for *common* variants but are far more difficult for *rare* VUS where the controls overwhelmingly

outnumber the cases. Various solutions have been put forth to solve this problem. These include likelihood ratio tests, burden tests with genetic scores, adaptive burden tests with data-adaptive weights/thresholds, variance-component tests, exponential combination tests, normal transformation with trait "winsorization" (to find a balance between Type I error and statistical power), or a combination thereof (Auer et al., 2016; Li et al., 2021).

Alternatively, one key solution to address the inherent lack of statistical power when analyzing rare variants is to use *gene expression outlier* analysis. Rare variants are often associated with extremes of expression, whether over- or underexpression (X. Li, Kim, et al., 2017). Variants can be interpreted through careful integration of DNA data with RNA-seq data from other patients or from public resources. Cummings et al. (2017) were one of the first groups to demonstrate the fact that consideration of such "outlier" variants in RNA-seq data results in an improved diagnostic rate (35% in that study, specifically) (Cummings et al., 2017). Gene expression outlier analysis has often been employed in cancer genomics to identify cancer drivers for a specific subset of cancer types or cancer outliers (Alshalalfa et al., 2012; Mori et al., 2013). Outlier analysis does not actually identify a mechanism but rather a gene that is different from others expressed in a cohort. As will be discussed later, there is a fundamental presumption that the comparison cohort is relevant.

2 | UTILITY OF RNA FOR THE DIAGNOSIS OF NEOPLASTIC AND NON-NEOPLASTIC DISEASE

While *analytical* validity refers to the sensitivity, specificity, and accuracy of a diagnostic test in terms of its ability to measure a biomarker in a lab setting, *clinical* validity refers to that test's accuracy and predictive value when it comes to predicting clinical diagnosis. Both terms are distinct from clinical *utility*, which refers to a test's ability to make a difference—that is, its potential to impact patient quality of care/life by guiding clinical decision-making (Byron et al., 2016). Here we will outline the clinical utility of transcriptome analysis for diagnosing both neoplastic and non-neoplastic diseases.

2.1 | RNA diagnostics for Mendelian disorders

Transcriptome analysis is a boon to the diagnosis of rare Mendelian diseases. Historically, genetic counseling has relied upon whole-exome sequencing to identify causative disease variants; however, this DNA-only approach has left up to 75% of patients without genetic diagnoses (Abou Tayoun et al., 2016; Aggarwal, 2021; Ellingford et al., 2016; Kopajtich et al., 2021; Lee et al., 2014; Liu et al., 2021; Rajagopalan et al., 2021; Stenton & Prokisch, 2020; Volodarsky et al., 2021; Yang et al., 2014). When integrated with genome sequencing—and, especially, in situations when said genome sequencing encompasses both exons and introns—gene expression profiling has been shown to significantly boost molecular diagnostic rates; yields have been shown to increase by 10%–35% (Lee et al., 2020; Maddirevula et al., 2020; Murdock et al., 2021; Stenton & Prokisch, 2020; Yépez et al., 2021). This is because RNA data both: (1) puts any variants identified in DNA into context by revealing their transcript-level consequences (e.g., ASE due to NMD, imprinting, and/or expression of splice variants), and (2) illuminates phenomena (like gene expression outliers) that may not pass the threshold of detection in DNA data alone (but that are crucial to the pathogenesis of a given disease) (Lee et al., 2020).

Gene expression profiling has also improved clinicians' ability to diagnose, stratify, and subtype autoimmune diseases, like systemic lupus erythematosus, as well as degenerative diseases like Age-Related Macular Degeneration (Alarcón-Riquelme, 2019). Additionally, transcriptome analysis has shed light on the fact that many of these diseases are heterogenous with a spectrum of causative molecular events (Morello et al., 2019). RNA-seq is also capable of overcoming the "bottleneck of variant interpretation" in patients with inborn errors of metabolism, mitochondriopathies, and/or unsolved muscle disorders, leading to significantly increased diagnostic yields (Kremer et al., 2018; Thompson et al., 2020).

It is important to note that recent studies have shown, particularly for monogenetic neuromuscular disorders, that blood-based RNA-seq is not sufficient for diagnosis. However, RNA-seq performed on myotubes generated by trans-differentiation of patient fibroblasts could identify a molecular culprit (predominantly splicing variants) in 36% of patients for whom DNA-only analysis had failed to do so (Gonorazky et al., 2019). These findings highlight the fact that several methodological improvements must be made to hasten the progress of translating transcriptome analysis from the benchtop to the bedside and to enhance diagnostic sensitivity. These include refinement of *ex vivo* trans-differentiation of accessible cells to more disease-relevant cell types (Lee et al., 2020).

2.2 | RNA-seq facilitates molecular diagnoses for hereditary cancers

Cancer genomic analysis involves the identification of inherited ("germline") risk variants and acquired ("somatic") mutations in DNA and RNA (Koeppel et al., 2018). Transcriptome analysis has been shown to be capable of identifying rare, causative variants by revealing changes in splicing and gene expression that were undetected by DNA sequencing (Yuan et al., 2020). Since examples of RNA-seq analysis in the conjunction of cancer risk prediction is more recent, we will dissect these papers in greater detail. Before we do, two key distinctions should be made regarding hereditary cancer studies. First, there is a general bias towards using RNA-sequencing in conjunction with panel-based clinical sequencing to reduce the genomic search space. If one focuses on *all* oncogenic or tumor-suppressor genes, the prevalence of background events changes and, with it, the precision and diagnostic yield. Second, the context of variant reporting is distinct from that for studies of Mendelian disease.

A series of papers from 2019 2021 illustrate and give further insight into these distinctions. First, Conner et al. (2019) found that by supplementing DNA genetic testing with RNA, heterozygous duplication events in mutS homolog 2 (*MSH2*), which were previously classified as VUS in five individuals with Lynch Syndrome, were able to be reclassified as pathogenic or likely pathogenic (Conner et al., 2019). Similarly, Karam et al. (2019) showed that, by supplementing DNA with RNA genetic testing in cases suspicious for hereditary cancer in which the variant in question involved a potential splice site alteration, (1) inconclusive DNA-based results were resolved in 49 of 56 inconclusive cases (88%) studied, with 26 (47%) being reclassified as clinically actionable and 23 (41%) being clarified as benign; and (2) approximately 2% of patients receiving paired DNA/RNA testing would benefit by the addition of RNA by further characterization of splice-site VUS (Karam et al., 2019). Two other studies found that the addition of transcriptomic analysis to hereditary cancer testing enabled 60% and 20%, respectively, of splicing VUS to be reclassified as (likely) pathogenic (Agiannitopoulos et al., 2021; Rofes et al., 2020). Landrith et al. (2020) performed germline RNA-seq to profile 18 genes—that is, APC regulator of WNT signaling pathway (*APC*); ATM serine/threonine kinase (*ATM*); BRCA1 and BRCA2 DNA repair associated (*BRCA1*, *BRCA2*); BRCA1 interacting helicase 1 (*1BRIP1*); cadherin 1 (*CDH1*); checkpoint kinase 2 (*CHEK2*); mutL homologs 1, 2, and 6 (*MLH1*, *MSH2*, and *MSH6*); mutY DNA glycosylase (*MUTYH*); neurofibromin 1 (*NF1*); partner and localizer of BRCA2 (*PALB2*); PMS1 homolog 2 (*PMS2*); phosphatase and tensin homolog (*PTEN*); tumor protein p53 (*TP53*); and RAD51 paralogs C and D (*RAD51C* and *RAD51D*)—in patients with suspected hereditary cancer syndromes. The investigators demonstrated a 9.1% relative increase in the detection of pathogenic variants afforded by augmenting DNA data with RNA analysis (Landrith et al., 2020). Deep intronic variants have also been identified in *BRCA1/2*, by virtue of RNA analysis, in patients with familial breast and ovarian cancers (Anczuków et al., 2012; Montalban et al., 2019).

As is evident from the studies mentioned above, deep intronic mutations and splicing aberrations are unique mechanisms of carcinogenesis which, based upon DNA data alone, are still often classified as VUS (Urbanski et al., 2018). Splicing mutations (which can be present in both pre-mRNA exons and introns, the latter of which has historically been harder to detect using traditional DNA analyses) lead to abnormal mRNA phenomena (e.g., exon skipping, intron inclusion, and cryptic splice site activation) and the production of abnormal proteins with diagnostic value (Shi et al., 2018). Expression changes in splicing regulators can be used as biomarkers for cancer diagnosis (e.g., heterogeneous nuclear ribonucleoprotein A2/B1, *hnRNPA2/B1*, an RNA-binding protein involved in mRNA splicing, is a sensitive and specific early-diagnostic marker of lung neoplasms) (Zhang et al., 2021). RNA-seq has shown utility for the diagnosis of germline splicing variants in hereditary cancer genes that were not evident in DNA analysis (Urbanski et al., 2018). While splicing variants makeup 11% of hereditary cancer gene VUS, they make up 55% of those VUS that are “likely pathogenic” (Parsons et al., 2019).

Larger-scale reports have been published by clinical genetic companies where RNA-seq was used in conjunction with panel-based studies across thousands of individuals. Ambry recently released a series of “RNA Case Studies” that demonstrate the clinical diagnostic utility of transcriptomic data, particularly for identifying intronic variants (AmbryGenetics, 2019). One such scenario was the case of a 33-year-old male, with a personal and family history of colon polyps, for whom no clinically significant variants could be detected via DNA-only analysis. When the genetic analysis was supplemented with transcriptomic analysis (i.e., Ambry's *+RNAinsight®* panel), abnormal *APC* transcripts were detected, prompting further investigation via targeted Sanger DNA sequencing. This resulted in the confirmation of a deep intronic, likely pathogenic variant. Transcriptomic data enabled the patient's provider to make a genetic diagnosis of Familial denomatous polyposis (AmbryGenetics, 2019). Other examples include a likely pathogenic intronic variant identified outside of DNA analytical range in the gene *ATM* (NM_000051.4:c.497-2661A>G), and exon skipping variants in *MSH6* leading to Lynch Syndrome. Ambry's *+RNAinsight®* panel, mentioned in the cases above, analyzes 91 cancer driver genes, and can be paired with most DNA panels; it has shown to be capable of reclassifying >70% of VUS (AmbryGenetics, 2021).

Similarly, a recent study by Invitae aimed to exemplify the utility of RNA analysis for reclassifying splicing VUS (Truty et al., 2021). The investigators analyzed a significantly large sample consisting of nearly 700,000 patients from a clinical cohort plus individuals from two large public datasets (i.e., *ClinVar* and Genome Aggregation Database/*gnomAD*) (Truty et al., 2021) (Table 1). In their clinical cohort, Invitae found that 5.4% of individuals had at least one splicing VUS (most of which were identified outside of essential splice sites), and that splicing variants represented 13% of all variants classified as (likely) pathogenic or VUS. They estimated that, in the clinical cohort, RNA analysis would be capable of clarifying/reclassifying splicing VUS in 1.7% of cases. In comparison to the clinical cohort, in *ClinVar* and *gnomAD*, Invitae observed that splicing VUS comprised nearly 5% and 9% of reported variants, respectively. Invitae concluded that, in all three cohorts, individuals would have a tangible, clinical-diagnostic benefit from RNA testing (Truty et al., 2021).

Not only can transcriptome characterization classify VUS as (likely) pathogenic, but it can also clarify variants as *benign*. For example, RNA data supported a variant *downgrade* of a likely pathogenic splice site variant at a canonical splice site (Shamseldin et al., 2021). In the case of *CDH1* NM_004360.5:c.387+1G>A, various clinical laboratories initially reported the variant in multiple Hispanic/Latino patients as “likely pathogenic” on the basis of the “+1” position of the variant. This led to the diagnosis of hereditary diffuse gastric cancer syndrome, a condition requiring complex management because of its association with a very high risk of early onset gastric cancer and lobular breast cancer. However, the variant was studied in more detail because the patients with this variant lacked the associated phenotype of the condition. The variant was experimentally demonstrated to result in the activation of a cryptic in-frame donor splice site, leading to the recommendation by ACMG

and AMP that variants at this position *not* be considered as likely pathogenic (Maoz & Culver, 2016).

Due to space and scope, we have limited this review to germline-inherited variation. However, RNA-sequencing has utility in the context of *somatic* variation and can guide treatment decisions. The histological subtypes of certain cancers are more strongly associated with transcriptomic signatures than genomic ones (Ghatak et al., 2022; Tang et al., 2021). Indeed, since the discovery of the *BCR::ABL* fusion gene in chronic myeloid leukemia, paired DNA/RNA-seq has allowed for groundbreaking discoveries of targetable fusion genes in both hematological and solid tumors (Tsang et al., 2021). Multiple, patented cancer diagnostic panels function by detecting fusions in RNA, such as the Fusion-STAMP targeted RNA-seq panel and the *Oncomine®Focus RNA Fusion* assay (Nohr et al., 2019; Williams et al., 2018). It is worth highlighting that a 2021 study in *Oncogene* examined somatic variation across over 1000 pan-cancer, paired whole genomes and transcriptomes to understand the role of splicing mutations in tumorigenesis (Jung et al., 2021). The investigators identified about 700 somatic intronic mutations; nearly half were within deep intronic regions and, of those, 38% activated cryptic splice sites. A subset of the deep intronic mutations resulted in splicing enhancer/silencer alterations.

3 | LIMITATIONS AND FUTURE DIRECTIONS

The progress of RNA-based diagnostics is encouraging, especially as new and translational transcriptomic techniques emerge (Wang et al., 2020). Gene expression profiling allows for the identification of fusion transcripts and the detection of phenomena like differential expression, ASE, alternative splicing, and the presence of noncoding RNAs (Conner et al., 2019). Both targeted RNA microarrays and RNA-seq have shown analytical validity in diagnostics for pediatric, adolescent/young adult, and adult patients (Vaske et al., 2019).

3.1 | Conflicting lines of evidence

One fallacy of reasoning—commonly and erroneously applied to the analysis of variant lists such as variant call format (VCF) files—is the assumption that the *absence* of a transcript variant means that the variant is absent from the specimen. This common misconception led to the development of genomic VCFs (gVCFs) which call every position—both variant and wild type/reference.

The only way to move forward with statistical power and confidence is through collaborative efforts and the creation of diverse and devoted databases. *ClinVar* (Rehm et al., 2017) and *gnomAD* (Karczewski et al., 2020) are under-appreciated summary-level datasets (Table 1). *gnomAD*'s focus on categorizing rare events was foundational. At the RNA-level, this approach has not yet been adopted outside of isolated cases; burgeoning examples are *RNAcentral* (a database of noncoding RNAs) (Petrov et al., 2015) and *SpliceDB* (a database of canonical and noncanonical mammalian splice sites) (Burset et al., 2001) (Table 1).

With the clinical implementation of any new "translational" technology, one must approach variant curation and interpretation of functional evidence with caution. Interpretation can be more complex than anticipated; there are many potential pitfalls. For example, Nix et al. once posited that a partial exon-skipping mutation identified in *BRCA2* was pathogenic; it was later found to occur in many healthy controls (Mundt et al., 2017).

3.2 | Differences in RNA-seq library preparation and analysis methods

Unlike genomic sequencing of DNA, differences in collection methods, library preparation, tissue sources, etc. fundamentally impact RNA-seq analysis and interpretation. The first and most apparent variable is RNA tissue source and its relevance to a given disease or phenotype. For example, how well can RNA from whole blood provide insight into neurological disorders? *GTEx* provides an initial framework to evaluate this question, showing that typically >40% of neurological genes are expressed at reasonably high levels in blood (GTEx Consortium, 2013). Still, investigators must carefully consider the tissue from which they are isolating RNA given that expression patterns differ across tissues (and, on the circadian level, RNA expression can even differ in the same tissue at different time points) (Maddirevula et al., 2020). Customized assays leveraging enrichment may increase the dynamic range of RNA species. Nevertheless, many of the studies highlighted showed >10% improvement in diagnostic yield despite such limitations.

The ability to probe DNA variation across thousands of individuals (using resources like *gnomAD*) has profoundly influenced how we interpret genomic variants. Normalization/harmonization of RNA-seq techniques is a necessary next step (and an active area of research beyond the scope of this review).

When examining consortiums such as *PsychENCODE* (Psych et al., 2015) and *AMP-AD* (Hodes & Buckholtz, 2016), it becomes clear that elimination of technical variation from RNA-seq data is challenging, especially if one is interested in rare events. To illustrate this point, we consider the recent release of 4,871 longitudinally collected samples from 1,570 clinically phenotyped individuals from the Parkinson's Progression Marker Initiative (*PPMI*), conducted using random priming for PaxGene-collected whole-blood with paired whole-genome sequencing (Craig et al., 2021). Forthcoming efforts from *TopMED* will utilize the same PaxGene whole-blood protocols but will differ in the use of mRNA-seq from poly-A priming. These two methods lead to different species with random priming, showing pre-spliced RNA and non-poly-A-tailed transcripts. Algorithms trained on these methods will fundamentally differ in their core measures, such as PSI. We have observed significant differences in gene/exon usage even within the same data set, depending on read lengths of paired 100 bp versus a 125 bp subset. While daunting, solutions are emerging for aggregating RNA, such as through the

ARCHS[4] aggregation across mouse and human RNA-seq studies (Lachmann et al., 2018).

3.3 | Fragmentation of RNA-seq databases and standards

Though the RNA-based diagnostics described here have potential, obstacles must be overcome before they become routine clinical practice. These challenges include the need for scientific rigor, reproducibility, accuracy, precision, clinical validity, and clinical utility. Standards must be created for test thresholds and normalized reporting, and databases must be established (Tahiliani et al., 2020; Wang et al., 2020). These databases must be designed to not fall prey to any logical fallacies (e.g. the "marker-positive fallacy").

Issues of database size, diversity, and representation (both in the sense of race/ethnicity and cases/controls), population structure, and cryptic relatedness must be considered (National Research Council (US) Committee, 1996). Fisher et al. created Ontology of RNA Sequencing (*ORNASEQ*) to capture, annotate, and manage provenance stores from RNA-seq studies (Fisher & Kim, 2018) (Table 1). We must also acknowledge, and attempt to address, limitations (e.g., the half-life/stability of RNA) and potential confounders (e.g., temporal changes in RNA expression, differences in RNA capture from fresh frozen vs. formalin-fixed paraffin-embedded samples, and phenomena like clonal hematopoiesis of indeterminate potential in liquid biopsies) (Wang et al., 2020).

It is important to balance preference for minimally invasive techniques with considerations of differential tissue expression. One recent study found that when comparing brain versus blood versus human B-lymphoblastoid cell lines (LCL), LCLs possessed isoform diversity for neurodevelopmental genes similar to that of brain tissue; LCLs also expressed these genes more highly compared to blood (Rentas et al., 2020). The authors of this article described an RNA-seq pipeline with 90% sensitivity and claimed that findings in LCLs outperformed those in blood and had implications for the molecular diagnosis of >1,000 genetic syndromes (Rentas et al., 2020). Rowlands et al. recently championed the use of a gene- and tissue-specific metric of their design—i.e., minimum required sequencing depth (MRSD)—for standardized prediction of RNA-seq utility, specifically for the diagnosis of splicing mutations in Mendelian diseases (Rowlands et al., 2022).

Another limitation is the fact that expression quantitative trait loci (eQTL) databases—like *GTEx* Portal—are limited to common variants (i.e., variants with a minor allele frequency >1%) (Table 1). This means that such datasets are not applicable toward understanding VUS which, although rare in the general/overall population, disproportionately impact non-White/European groups. RNA analysis is also limited by the fact that most tools utilize transcripts defined by a Gene Transfer Format (GTF) file and find it difficult to annotate the 3′ UTR (Shenker et al., 2015). Therefore, there exists a critical need for more rigorous, reproducible, and representative RNA databases and tools.

3.4 | Future direction: Investigating the "other" RNAs

A burgeoning area of transcriptomic research is the detection of noncoding RNAs (ncRNAs), so named because, unlike mRNAs, they are not translated into protein. There is hope that ncRNAs will have diagnostic utility (Ellingford et al., 2021; McQuerry et al., 2021), particularly for cancer diagnosis (Coley et al., 2021; Distefano et al., 2022; Lu et al., 2021; Osan et al., 2021; Ren et al., 2021; Sun et al., 2021; Tabury et al., 2022; Zhu et al., 2021). Examples of ncRNAs include miRNAs, small interfering RNA (siRNA), lncRNAs including circular RNAs (circRNAs, which are produced by non-canonical "back-splicing"), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), and PIWI-interacting RNA (piRNA) (Hasegawa et al., 2021; Junqueira-Neto et al., 2019; Mussack et al., 2020; Solé et al., 2021; Wen et al., 2021). Particular effort has been devoted to assessing the efficacy of minimally invasive "liquid biopsies" (samples of bodily fluids like saliva, serum, and urine) for the detection of cancer-related ncRNAs; these methods are currently being rigorously validated and must be standardized before they become universally adopted as best clinical practice (Zeuschner et al., 2020).

One of the benefits of using noncoding RNAs versus protein-coding RNAs is that ncRNAs have been shown to display more tissue-specific gene expression (Iaccarino & Klapper, 2021); this renders ncRNA a potential tumor-associated *and* tumor-*specific* marker. Both gene expression microarray and RNA-seq-based techniques can identify lncRNA biomarkers; RNA-seq allows for a discovery-type approach, whereas microarrays necessitate a targeted approach (Sun, 2015). The most comprehensive study, to date, analyzed >7,000 RNA-seq libraries and identified >7,000 cancer-associated lncRNAs (Sun, 2015). Several promising lncRNA candidates include prostate cancer-associated 3 (*PCA3*) and SWI/SNF complex antagonist associated with prostate cancer 1 (*SChLAP1*) (specific markers for prostate cancer which can be identified in urine); hepatocellular carcinoma upregulated long noncoding RNA (*HULC*) in pancreatic cancer; lung cancer-associated lncRNA 1 (*LCAL1*) in lung carcinomas; H19 imprinted maternally expressed transcript (*H19*), hepatocellular carcinoma upregulated EZH2-associated long noncoding RNA (*HEIH*), HOX transcript antisense RNA (*HOTAIR*), and HOXA distal transcript antisense RNA (*HOTTIP*) in hepatocellular carcinoma; long intergenic non-protein coding RNA 261 (*LINC00261*) and *PCA3* in choriocarcinoma; and ADAMTS9 antisense RNA 2 (*ADAMTS9-AS2*), HOXA11 antisense RNA (*HOXA11-AS*), and cancer susceptibility 2 (*CASC2*) in glioma (Di Fiore et al., 2021; Ebrahimi et al., 2021; Huang & Tang, 2021; Lanzafame et al., 2018; Sun, 2015; Xi et al., 2017). Recent research has shown that circRNAs—which (1) are dysregulated in a tumor-specific manner, (2) display stage-specific expression patterns, and (3) are more stable than linear RNAs due to their circular conformation—could be used as diagnostic markers for lung, liver, colorectal, gastric, and bladder cancers (Sawaki et al., 2018; Solé et al., 2021).

Before lncRNA-based diagnostics become commonplace in the clinical setting, investigators still must determine whether: (1)

tumor-associated lncRNAs are being expressed by the tumor cells themselves versus other cells in the tumor microenvironment; (2) whether lncRNA expression is confounded by cancer subtype and/or patient demographics (Iaccarino & Klapper, 2021); and (3) whether lncRNA structure reflects function (this could be probed using Clustered Regulatory Interspaced Short Palindromic Repeat, *CRISPR*, screening) (Wong & Wong, 2021). One example of a bioinformatic tool for RNA-seq-based detection of lncRNAs is *UclncR* (Sun et al., 2017) (Table 1).

Two other important species of noncoding RNA, which both play roles in gene regulation, are siRNAs (which target specific mRNA) and miRNAs (which have multiple mRNA targets) (Lam et al., 2015). Hundreds of clinical trials are currently evaluating the application of miRNAs to human diseases (Krishnan & Damaraju, 2018). Studies have suggested that miRNAs could be used for the molecular diagnosis of gastric cancer and high-grade cervical intraepithelial neoplasia (Causin et al., 2021; Sawaki et al., 2018). Two promising miRNAs are *miR-20a-5p* for the detection of renal cell carcinoma and *miR-21* for pancreatic cancer (Oto et al., 2021; Xi et al., 2017). miRNAs are also proving to be capable of serving as early diagnostic biomarkers for diseases lacking clear modes of inheritance and biomarkers (Tan et al., 2021).

Techniques for the isolation of miRNA must be optimized, and reference databases must be established. One group has developed an ultrasensitive, single-molecule, amplification-free, multiplexed assay for the detection of miRNAs directly from small samples of human serum (Cai et al., 2021). The database *miREV* contains >400 miRNA sequencing data sets; such datasets may prove to be critical resources going forward (Hildebrandt et al., 2021) (Table 1).

The application of ncRNA species other than those outlined above is still in its infancy (Byron et al., 2016). Promising candidates include the piRNA *piR-651* for the diagnosis of lymphoma; the snoRNAs *SNORD33*, *SNORD66, and SNORD76* for the diagnosis of non-small cell lung cancer; and the circRNA *Hsa_circ_002059* for the diagnosis of gastric cancer (Xi et al., 2017).

A recent paper published in *Human Mutation* enumerated the most used bioinformatics tools for the analysis of ncRNA data (Veneziano et al., 2016). These include *miRDeep2*, *CAP-miRSeq*, *iMir*, *piPipes*, *UEA sRNA workbench*, *omiRas*, *sRNAtoolbox*, *FlaiMapper*, and *tDRMapper* for analysis of small ncRNAs; *UROBORUS*, *PredCircRNA*, *find*_circ, *CircExplorer*, and *CIRI* for analysis of circular ncRNAs; and *iSeeRNA*, *Sebnif*, and *LncRNA2Function* for analysis of lncRNAs (Veneziano et al., 2016) (Table 1).

3.5 | Future direction: Targeting VUS to alleviate cancer disparities

One anecdotal trend that we have noticed within our group and across collaborative efforts is that RNA data identifies previously missed variation, particularly in individuals of non-European ancestry. In *Human Mutation* we reported a variant within 3bp of the exon boundary using an outlier approach in individuals of African ancestry. The molecular consequences of this variant included exon skipping, altered isoform usage, and loss of canonical isoform expression—events not evident in DNA data alone (McCullough et al., 2020). Patients who self-identify as Hispanic/Latino, Black/African, and Asian/Pacific Islander experience more advanced stage disease at the time of screening, significantly lower diagnostic yields, and higher rates of VUS and variant reclassification compared to their European/Caucasian counterparts (Dutil et al., 2019; Kinney et al., 2018; Kowalski et al., 2019; Marco-Puche et al., 2019; Ndugga-Kabuye & Issaka, 2019; Roberts et al., 2020; Slavin et al., 2018; Urbina-Jara et al., 2019). Individuals from non-European populations will have more "private" variation for one of three reasons: (1) they are poorly represented in reference data sets, (2) they have greater African ancestry, or (3) they come from a population that has undergone recent expansions (e.g., Bangladesh) (Halperin et al., 2017).

A recent study reported by Ambry Genetics found that their *BRCAplus*, *BreastNext*, and *CancerNext* panels yielded ≈2–3x fewer VUS for Non-Hispanic Whites than for minority populations (AmbryGenetics, 2017). Another study reports VUS frequencies in the tumor suppressor genes *BRCA1/2* to be 4.4% in Caucasians, 8.9% in African Americans, and 8.0% in Hispanic/Latinos; for larger hereditary cancer panels, this study reported VUS frequencies of 22.1% in Caucasians, 30.3% in African Americans, and 24.9% in Hispanic/Latinos (Appelbaum et al., 2020).

One important distinction to make here is the difference between race/ethnicity and genetic ancestry. While race and ethnicity are social constructs, ancestry is a biological/genetic construct resulting from human migrations throughout history resulting in biogeographical genetic variation (Batai et al., 2021). An example of how genetic ancestry can further clarify race/ethnicity-based disparities is the fact that higher African ancestry in Hispanic/Latinos (who are typically "admixed" with genetic contributions from African, European, and American Indian aka Native/Indigenous American ancestries) is associated with more aggressive breast cancer subtypes and a greater likelihood of receiving inconclusive VUS during genetic testing (Chapman-Davis et al., 2021; Dutil et al., 2019; Kinney et al., 2018; Kowalski et al., 2019; Marco-Puche et al., 2019; Ndugga-Kabuye & Issaka, 2019; Roberts et al., 2020; Slavin et al., 2018; Urbina-Jara et al., 2019; Virlogeux et al., 2015). Interestingly, in turn, genetic ancestry can clarify the interpretation of VUS by providing information about the local and global "genomic context" of a variant (Wang et al., 2021). Thus, gene expression profiling may be able to help shed light on and alleviate cancer disparities (Frésard et al., 2019; Wai et al., 2020).

4 | CONCLUSIONS

VUS cause significant psychological distress to patients and disproportionately limit the promise of precision medicine for minority patients (Landry et al., 2018). RNA data provide critical answers to the question of VUS, particularly in terms of clarifying deep intronic and splicing variants as pathogenic versus benign. This

necessitates the development of more rigorous, reproducible, and representative RNA databases and analytical tools.

ACKNOWLEDGMENTS

The authors would like to acknowledge support by NCI through 1U2CCA252971-01A1.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

ORCID

Mackenzie D. Postel http://orcid.org/0000-0002-9438-1283
Julie O. Culver http://orcid.org/0000-0001-8658-3507
Charité Ricker http://orcid.org/0000-0002-0620-9345
David W. Craig http://orcid.org/0000-0003-2040-1955

REFERENCES

Abou Tayoun, A. N., Krock, B., & Spinner, N. B. (2016). Sequencing-based diagnostics for pediatric genetic diseases: Progress and potential. *Expert Review of Molecular Diagnostics*, *16*(9), 987–999. https://doi.org/10.1080/14737159.2016.1209411

Aggarwal, S. (2021). Role of whole exome sequencing for unidentified genetic syndromes. *Current Opinion in Obstetrics and Gynecology*, *33*(2), 112–122. https://doi.org/10.1097/GCO.0000000000000688

Agiannitopoulos, K., Pepe, G., Papadopoulou, E., Tsaousis, G. N., Kampouri, S., Maravelaki, S., & Nasioulas, G. (2021). Clinical utility of functional RNA analysis for the reclassification of splicing gene variants in hereditary cancer. *Cancer Genomics & Proteomics*, *18*(3), 285–294. https://doi.org/10.21873/cgp.20259

Alarcón-Riquelme, M. E. (2019). New attempts to define and clarify lupus. *Current Rheumatology Reports*, *21*(4), 11. https://doi.org/10.1007/s11926-019-0810-4

Alshalalfa, M., Bismar, T. A., & Alhajj, R. (2012). Detecting cancer outlier genes with potential rearrangement using gene expression data and biological networks. *Advances in Bioinformatics*, *2012*, 373506. https://doi.org/10.1155/2012/373506

AmbryGenetics. (2017). *Improvements to patient care through hereditary cancer panels*. [online] Available at: ambrygen.com/file/material/view/1235/GeneralCancer_Specific_1117_final.pdf

AmbryGenetics. (2019). *RNA case study booklet*. [online]. Available at: ambrygen.com/file/material/view/1304/RNA%20Case%20Study%20Booklet.pdf

AmbryGenetics. (2021). *+RNAinsight® flyer*. [online]. Available at: ambrygen.com/file/material/view/1663/RNA_Flyer_FNL%20091521.pdf

Anczuków, O., Buisson, M., Léoné, M., Coutanson, C., Lasset, C., Calender, A., & Mazoyer, S. (2012). BRCA2 deep intronic mutation causing activation of a cryptic exon: Opening toward a new preventive therapeutic strategy. *Clinical Cancer Research*, *18*(18), 4903–4909. https://doi.org/10.1158/1078-0432.CCR-12-1100

Anna, A., & Monika, G. (2018). Splicing mutations in human genetic disorders: Examples, detection, and confirmation. *Journal of Applied Genetics*, *59*(3), 253–268. https://doi.org/10.1007/s13353-018-0444-7

Appelbaum, P. S., Parens, E., Berger, S. M., Chung, W. K., & Burke, W. (2020). Is there a duty to reinterpret genetic data? The ethical dimensions. *Genetics in Medicine*, *22*(3), 633–639. https://doi.org/10.1038/s41436-019-0679-7

Auer, P. L., Reiner, A. P., & Leal, S. M. (2016). The effect of phenotypic outliers and non-normality on rare-variant association testing. *European Journal of Human Genetics*, *24*(8), 1188–1194. https://doi.org/10.1038/ejhg.2015.270

Batai, K., Hooker, S., & Kittles, R. A. (2021). Leveraging genetic ancestry to study health disparities. *American Journal of Physical Anthropology*, *175*(2), 363–375. https://doi.org/10.1002/ajpa.24144

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., & Margulies, E. H., Institute, C. s. H. O. R. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799–816. https://doi.org/10.1038/nature05874

Blakes, A. J. M., Wai, H., Davies, I., Moledian, H. E., Ruiz, A., Thomas, T., & Lord, J. (2022). A systematic analysis of splicing variants identifies new diagnoses in the 100,000 Genomes Project. *medRxiv*. 22270002. https://doi.org/10.1101/2022.01.28.22270002

Bournazos, A. M., Riley, L. G., Bommireddipalli, S., Ades, L., Akesson, L. S., & Al-Shinnag, M., Diagnostics, A. C. f. R. (2022). Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants. *Genetics in Medicine*, *24*(1), 130–145. https://doi.org/10.1016/j.gim.2021.09.001

Brogna, S., & Wen, J. (2009). Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural & Molecular Biology*, *16*(2), 107–113. https://doi.org/10.1038/nsmb.1550

Burset, M., Seledtsov, I. A., & Solovyev, V. V. (2001). SpliceDB: Database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Research*, *29*(1), 255–259. https://doi.org/10.1093/nar/29.1.255

Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D., & Craig, D. W. (2016). Translating RNA sequencing into clinical diagnostics: Opportunities and challenges. *Nature Reviews Genetics*, *17*(5), 257–271. https://doi.org/10.1038/nrg.2016.10

Cai, S., Pataillot-Meakin, T., Shibakawa, A., Ren, R., Bevan, C. L., Ladame, S., & Edel, J. B. (2021). Single-molecule amplification-free multiplexed detection of circulating microRNA cancer biomarkers from serum. *Nature Communications*, *12*(1), 3515. https://doi.org/10.1038/s41467-021-23497-y

Causin, R. L., da Silva, L. S., Evangelista, A. F., Leal, L. F., Souza, K. C. B., Pessôa-Pereira, D., & Marques, M. M. C. (2021). MicroRNA biomarkers of high-grade cervical intraepithelial neoplasia in liquid biopsy. *BioMed Research International*, *2021*, 6650966. https://doi.org/10.1155/2021/6650966

Chakravorty, S., & Hegde, M. (2018). Clinical utility of transcriptome sequencing: Toward a better diagnosis for mendelian disorders. *Clinical Chemistry*, *64*(6), 882–884. https://doi.org/10.1373/clinchem.2017.276980

Chapman-Davis, E., Zhou, Z. N., Fields, J. C., Frey, M. K., Jordan, B., Sapra, K. J., & Holcomb, K. M. (2021). Racial and Ethnic disparities in genetic testing at a hereditary breast and ovarian cancer center. *Journal of General Internal Medicine*, *36*(1), 35–42. https://doi.org/10.1007/s11606-020-06064-x

Chen, J., & Weiss, W. A. (2015). Alternative splicing in cancer: Implications for biology and therapy. *Oncogene*, *34*(1), 1–14. https://doi.org/10.1038/onc.2013.570

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Coley, A. B., Stahly, A. N., Kasukurthi, M. V., Barchie, A. A., Hutcheson, S. B., Houserova, D., & Borchert, G. M. (2021). MicroRNA-like snoRNA-derived RNAs (sdRNAs) promote castration resistant prostate cancer. *BioRxiv*, 473174. https://doi.org/10.1101/2021.12.17.473174

Conner, B. R., Hernandez, F., Souders, B., Landrith, T., Boland, C. R., & Karam, R. (2019). RNA analysis identifies pathogenic duplications in MSH2 in patients with Lynch syndrome. *Gastroenterology*, *156*(6), 1924–1925. https://doi.org/10.1053/j.gastro.2019.01.248

Craig, D. W., Keuren-Jensen, K. V., & Initiative, P. P. M. (2021). RNA sequencing of whole blood reveals early alterations in immune cells

and gene expression in Parkinson's disease. *Nature. Aging*, *1*, 734–747.

Culver, J. O., Brinkerhoff, C. D., Clague, J., Yang, K., Singh, K. E., Sand, S. R., & Weitzel, J. N. (2013). Variants of uncertain significance in BRCA testing: Evaluation of surgical decisions, risk perception, and cancer distress. *Clinical Genetics*, *84*(5), 464–472. https://doi.org/10.1111/cge.12097

Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., & MacArthur, D. G. (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Science Translational Medicine*, *9*(386), https://doi.org/10.1126/scitranslmed.aal5209

Danis, D., Jacobsen, J. O. B., Carmody, L. C., Gargano, M. A., McMurry, J. A., Hegde, A., & Robinson, P. N. (2021). Interpretable prioritization of splice variants in diagnostic next-generation sequencing. *American Journal of Human Genetics*, *108*(11), 2205. https://doi.org/10.1016/j.ajhg.2021.09.014

Distefano, R., Tomasello, L., Vinciguerra, G. L. R., Gasparini, P., Xiang, Y., Bagnoli, M., & Croce, C. M. (2022). A concurrent canonical and modified miRNAome pan-cancer study on TCGA and TARGET cohorts leads to an enhanced resolution in cancer. *bioRxiv*, 444694. https://doi.org/10.1101/2021.05.18.444694

Dobkin, C., Pergolizzi, R. G., Bahre, P., & Bank, A. (1983). Abnormal splice in a mutant human beta-globin gene not at the site of a mutation. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(5), 1184–1188. https://doi.org/10.1073/pnas.80.5.1184

Dutil, J., Teer, J. K., Golubeva, V., Yoder, S., Tong, W. L., Arroyo, N., & Monteiro, A. N. (2019). Germline variants in cancer genes in high-risk non-BRCA patients from Puerto Rico. *Scientific Reports*, *9*(1), 17769. https://doi.org/10.1038/s41598-019-54170-6

Ebrahimi, A. A., Ashoori, H., Vahidian, F., Mosleh, I. S., & Kamian, S. (2021). Long non-coding RNA panel as a molecular biomarker in glioma. *Journal of the Egyptian National Cancer Institute*, *33*(1), 31. https://doi.org/10.1186/s43046-021-00090-4

Ellingford, J. M., Ahn, J. W., Bagnall, R. D., Baralle, D., Barton, S., Campbell, C., & Whiffin, N. (2021). Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *medRxiv*, 21267792. https://doi.org/10.1101/2021.12.28.21267792

Ellingford, J. M., Barton, S., Bhaskar, S., Williams, S. G., Sergouniotis, P. I., O'Sullivan, J., & Black, G. C. (2016). Whole Genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. *Ophthalmology*, *123*(5), 1143–1150. https://doi.org/10.1016/j.ophtha.2016.01.009

Di Fiore, R., Suleiman, S., Felix, A., O'Toole, S. A., O'Leary, J. J., Ward, M. P., & Calleja-Agius, J. (2021). An overview of the role of long non-coding RNAs in human choriocarcinoma. *International Journal of Molecular Sciences*, *22*(12), https://doi.org/10.3390/ijms22126506

Fisher, S. A., & Kim, J. (2018). *Ontology of RNA Sequencing (ORNASEQ) and its application. bioRxiv*, 405720. https://doi.org/10.1101/405720

Frésard, L., Smail, C., Ferraro, N. M., Teran, N. A., Li, X., Smith, K. S., & Consortium, C. R. C. (2019). Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nature Medicine*, *25*(6), 911–919. https://doi.org/10.1038/s41591-019-0457-8

Gelfman, S., Wang, Q., McSweeney, K. M., Ren, Z., La Carpia, F., Halvorsen, M., & Goldstein, D. B. (2017). Annotating pathogenic non-coding variants in genic regions. *Nature Communications*, *8*(1), 236. https://doi.org/10.1038/s41467-017-00141-2

Ghatak, S., Mehrabi, S. F., Mehdawi, L. M., Satapathy, S. R., & Sjölander, A. (2022). Identification of a novel five-gene signature as a prognostic and diagnostic biomarker in colorectal cancers. *International Journal of Molecular Sciences*, *23*(2):793. https://doi.org/10.3390/ijms23020793

Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., & Dowling, J. J. (2019). Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *American Journal of Human Genetics*, *104*(3), 466–483. https://doi.org/10.1016/j.ajhg.2019.01.012

GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project, *Nature Genetics*, *45*(6), 580–585. https://doi.org/10.1038/ng.2653

Halperin, R. F., Carpten, J. D., Manojlovic, Z., Aldrich, J., Keats, J., Byron, S., & Craig, D. W. (2017). A method to reduce ancestry-related germline false positives in tumor only somatic variant calling. *BMC Medical Genomics*, *10*(1), 61. https://doi.org/10.1186/s12920-017-0296-8

Halperin, R. F., Hegde, A., Lang, J. D., Raupach, E. A., Legendre, C., Liang, W. S., & Group, C. R. R. (2021). Improved methods for RNAseq-based alternative splicing analysis. *Scientific Reports*, *11*(1), 10740. https://doi.org/10.1038/s41598-021-89938-2

Hasegawa, N., Kohsaka, S., Kurokawa, K., Shinno, Y., Takeda Nakamura, I., Ueno, T., & Mano, H. (2021). Highly sensitive fusion detection using plasma cell-free RNA in non-small-cell lung cancers. *Cancer Prevention Research*, *112*(10), 4393–4403. https://doi.org/10.1111/cas.15084

Hildebrandt, A., Kirchner, B., Nolte-'t Hoen, E. N. M., & Pfaffl, M. W. (2021). miREV: An online database and tool to uncover potential reference RNAs and biomarkers in small-RNA sequencing data sets from extracellular vesicles enriched samples. *Journal of Molecular Biology*, *433*(15), 167070. https://doi.org/10.1016/j.jmb.2021.167070

Hodes, R. J., & Buckholtz, N. (2016). Accelerating medicines partnership: Alzheimer's disease (AMP-AD) knowledge portal aids Alzheimer's drug discovery through open data sharing. *Expert Opinion on Therapeutic Targets*, *20*(4), 389–391. https://doi.org/10.1517/14728222.2016.1135132

Hong, J. H., Ko, Y. H., & Kang, K. (2018). RNA variant identification discrepancy among splice-aware alignment algorithms. *PLoS One*, *13*(8), e0201822. https://doi.org/10.1371/journal.pone.0201822

Huang, K., & Tang, Y. (2021). SChLAP1 promotes prostate cancer development through interacting with EZH2 to mediate promoter methylation modification of multiple miRNAs of chromosome 5 with a DNMT3a-feedback loop. *Cell Death & Disease*, *12*(2), 188. https://doi.org/10.1038/s41419-021-03455-8

Hurst, S. E., Liktor-Busa, E., Moutal, A., Parker, S., Rice, S., Szelinger, S., & Nelson, M. A. (2018). A novel variant. *Neuronal Signal*, *2*(3), NS20180141. https://doi.org/10.1042/NS20180141

Iaccarino, I., & Klapper, W. (2021). LncRNA as cancer biomarkers. *Methods in Molecular Biology*, *2348*, 27–41. https://doi.org/10.1007/978-1-0716-1581-2_2

Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J. F., Darbandi, S. F., Knowles, D., Li, Y. I., & Farh, K. K. (2019). Predicting splicing from primary sequence with deep learning. *Cell*, *176*(3), 535–548. e524. https://doi.org/10.1016/j.cell.2018.12.015

Jenkinson, G., Li, Y. I., Basu, S., Cousin, M. A., Oliver, G. R., & Klee, E. W. (2020). LeafCutterMD: An algorithm for outlier splicing detection in rare diseases. *Bioinformatics*, *36*(17), 4609–4615. https://doi.org/10.1093/bioinformatics/btaa259

Jian, X., Boerwinkle, E., & Liu, X. (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Research*, *42*(22), 13534–13544. https://doi.org/10.1093/nar/gku1206

Jung, H., Lee, K. S., & Choi, J. K. (2021). Comprehensive characterisation of intronic mis-splicing mutations in human cancers. *Oncogene*, *40*(7), 1347–1361. https://doi.org/10.1038/s41388-020-01614-3

Junqueira-Neto, S., Batista, I. A., Costa, J. L., & Melo, S. A. (2019). Liquid biopsy beyond circulating tumor cells and cell-free DNA. *Acta Cytologica*, *63*(6), 479–488. https://doi.org/10.1159/000493969

Karam, R., Conner, B., LaDuca, H., McGoldrick, K., Krempely, K., Richardson, M. E., & Chao, E. (2019). Assessment of diagnostic outcomes of RNA genetic testing for hereditary cancer. *JAMA Network Open*, *2*(10), e1913900. https://doi.org/10.1001/jamanetworkopen.2019.13900

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., & Consortium, G. A. D. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7

King, K., Flinter, F. A., Nihalani, V., & Green, P. M. (2002). Unusual deep intronic mutations in the COL4A5 gene cause X linked Alport syndrome. *Human Genetics*, *111*(6), 548–554. https://doi.org/10.1007/s00439-002-0830-3

Kinney, A. Y., Howell, R., Ruckman, R., McDougall, J. A., Boyce, T. W., Vicuña, B., & Walters, S. T. (2018). Promoting guideline-based cancer genetic risk assessment for hereditary breast and ovarian cancer in ethnically and geographically diverse cancer survivors: Rationale and design of a 3-arm randomized controlled trial. *Contemporary Clinical Trials*, *73*, 123–135. https://doi.org/10.1016/j.cct.2018.09.005

Koeppel, F., Bobard, A., Lefebvre, C., Pedrero, M., Deloger, M., Boursin, Y., & Lacroix, L. (2018). Added value of whole-exome and transcriptome sequencing for clinical molecular screenings of advanced cancer patients with solid tumors. *Cancer Journal*, *24*(4), 153–162. https://doi.org/10.1097/PPO.0000000000000322

Kopajtich, R., Smirnov, D., Stenton, S. L., Loipfinger, S., Meng, C., Scheller, I. F., & Prokisch, H. (2021). Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders. *medRxiv*, 2021.2003.2009 21253187. https://doi.org/10.1101/2021.03.09.21253187

Koster, R., Brandão, R. D., Tserpelis, D., van Roozendaal, C. E. P., van Oosterhoud, C. N., Claes, K. B. M., & Blok, M. J. (2021). Pathogenic neurofibromatosis type 1 (NF1) RNA splicing resolved by targeted RNAseq. *NPJ Genomic Medicine*, *6*(1), 95. https://doi.org/10.1038/s41525-021-00258-w

Kowalski, M. H., Qian, H., Hou, Z., Rosen, J. D., Tapia, A. L., Shan, Y., & Group, T. H. H. W. (2019). Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genetics*, *15*(12), e1008500. https://doi.org/10.1371/journal.pgen.1008500

Kremer, L. S., Wortmann, S. B., & Prokisch, H. (2018). "Transcriptomics": Molecular diagnosis of inborn errors of metabolism via RNA-sequencing. *Journal of Inherited Metabolic Disease*, *41*(3), 525–532. https://doi.org/10.1007/s10545-017-0133-4

Krishnan, P., & Damaraju, S. (2018). The challenges and opportunities in the clinical application of noncoding RNAs: The road map for miRNAs and piRNAs in cancer diagnostics and prognostics. *International Journal of Genomics*, *2018*, 5848046. https://doi.org/10.1155/2018/5848046

Lachmann, A., Torre, D., Keenan, A. B., Jagodnik, K. M., Lee, H. J., Wang, L., & Ma'ayan, A. (2018). Massive mining of publicly available RNA-seq data from human and mouse. *Nature Communications*, *9*(1), 1366. https://doi.org/10.1038/s41467-018-03751-6

Lam, J. K., Chow, M. Y., Zhang, Y., & Leung, S. W. (2015). siRNA versus miRNA as therapeutics for gene silencing. *Molecular Therapy. Nucleic Acids*, *4*, e252. https://doi.org/10.1038/mtna.2015.23

Landrith, T., Li, B., Cass, A. A., Conner, B. R., LaDuca, H., McKenna, D. B., & Karam, R. (2020). Splicing profile by capture RNA-seq identifies pathogenic germline variants in tumor suppressor genes. *NPJ Precision Oncology*, *4*, 4. https://doi.org/10.1038/s41698-020-0109-y

Landry, L. G., Ali, N., Williams, D. R., Rehm, H. L., & Bonham, V. L. (2018). Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice. *Health Affairs*, *37*(5), 780–785. https://doi.org/10.1377/hlthaff.2017.1595

Lanzafame, M., Bianco, G., Terracciano, L. M., Ng, C. K. Y., & Piscuoglio, S. (2018). The role of long non-coding RNAs in hepatocarcinogenesis. *International Journal of Molecular Sciences*, *19*(3), https://doi.org/10.3390/ijms19030682

Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., & Nelson, S. F. (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. *Journal of the American Medical Association*, *312*(18), 1880–1887. https://doi.org/10.1001/jama.2014.14604

Lee, H., Huang, A. Y., Wang, L. K., Yoon, A. J., Renteria, G., Eskin, A., & Network, U. D. (2020). Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genetics in Medicine*, *22*(3), 490–499. https://doi.org/10.1038/s41436-019-0672-1

Lee, S. C., & Abdel-Wahab, O. (2016). Therapeutic targeting of splicing in cancer. *Nature Medicine*, *22*(9), 976–986. https://doi.org/10.1038/nm.4165

Li, J., Kong, N., Han, B., & Sul, J. H. (2021). Rare variants regulate expression of nearby individual genes in multiple tissues. *PLoS Genetics*, *17*(6), e1009596. https://doi.org/10.1371/journal.pgen.1009596

Li, M. M., Datto, M., Duncavage, E. J., Kulkarni, S., Lindeman, N. I., Roy, S., & Nikiforova, M. N. (2017). Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A joint consensus recommendation of the association for molecular pathology, American Society of Clinical Oncology, and College of American Pathologists. *Journal of Molecular Diagnostics*, *19*(1), 4–23. https://doi.org/10.1016/j.jmoldx.2016.10.002

Li, X., Kim, Y., Tsang, E. K., Davis, J. R., Damani, F. N., & Chiang, C., GTEx Consortium. (2017). The impact of rare variation on gene expression across tissues. *Nature*, *550*(7675), 239–243. https://doi.org/10.1038/nature24267

Liu, X., Wu, C., Li, C., & Boerwinkle, E. (2016). dbNSFP v3.0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Human Mutation*, *37*(3), 235–241. https://doi.org/10.1002/humu.22932

Liu, Y., Teng, Y., Li, Z., Cui, J., Liang, D., & Wu, L. (2021). Increase in diagnostic yield achieved for 174 whole-exome sequencing cases reanalyzed 1-2 years after initial analysis. *Clinica Chimica Acta*, *523*, 163–168. https://doi.org/10.1016/j.cca.2021.09.015

Lu, N., Liu, J., Ji, C., Wang, Y., Wu, Z., Yuan, S., & Diao, F. (2021). MiRNA based tumor mutation burden diagnostic and prognostic prediction models for endometrial cancer. *Bioengineered*, *12*(1), 3603–3620. https://doi.org/10.1080/21655979.2021.1947940

Maddirevula, S., Kuwahara, H., Ewida, N., Shamseldin, H. E., Patel, N., Alzahrani, F., & Alkuraya, F. S. (2020). Analysis of transcript-deleterious variants in Mendelian disorders: Implications for RNA-based diagnostics. *Genome Biology*, *21*(1), 145. https://doi.org/10.1186/s13059-020-02053-9

Mantione, K. J., Kream, R. M., Kuzelova, H., Ptacek, R., Raboch, J., Samuel, J. M., & Stefano, G. B. (2014). Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Medical Science Monitor Basic Research*, *20*, 138–142. https://doi.org/10.12659/MSMBR.892101

Maoz, C., & Culver, J. (2016). Genetic, clinical and molecular characterization of a suspected CDH1 mutation. Collaborative Group of the Americas on Inherited Gastrointestinal Cancer (CGA-IGC) annual conference poster session.

Marco-Puche, G., Lois, S., Benítez, J., & Trivino, J. C. (2019). RNA-Seq perspectives to improve clinical diagnosis. *Frontiers in Genetics*, *10*, 1152. https://doi.org/10.3389/fgene.2019.01152

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., & Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA.

Science, *337*(6099), 1190–1195. https://doi.org/10.1126/science.1222794

McCullough, C. G., Szelinger, S., Belnap, N., Ramsey, K., Schrauwen, I., Claasen, A. M., & Craig, D. W. (2020). Utilizing RNA and outlier analysis to identify an intronic splice-altering variant in AP4S1 in a sibling pair with progressive spastic paraplegia. *Human Mutation*, *41*(2), 412–419. https://doi.org/10.1002/humu.23939

McQuerry, J. A., Mclaird, M., Hartin, S. N., Means, J. C., Johnston, J., Pastinen, T., & Younger, S. T. (2021). Massively parallel identification of functionally consequential noncoding genetic variants in undiagnosed rare disease patients. *medRxiv*. 21265771. https://doi.org/10.1101/2021.11.02.21265771

Mehmood, A., Laiho, A., Venäläinen, M. S., McGlinchey, A. J., Wang, N., & Elo, L. L. (2020). Systematic evaluation of differential splicing tools for RNA-seq studies. *Briefings in Bioinformatics*, *21*(6), 2052–2065. https://doi.org/10.1093/bib/bbz126

Mertes, C., Scheller, I. F., Yepez, V. A., Celik, M. H., Liang, Y., Kremer, L. S., & Gagneur, J. (2021). Detection of aberrant splicing events in RNA-seq data using FRASER. *Nature Communications*, *12*(1), 529. https://doi.org/10.1038/s41467-020-20573-7

Migeon, B. R. (2020). X-linked diseases: susceptible females. *Genetics in Medicine*, *22*(7), 1156–1174. https://doi.org/10.1038/s41436-020-0779-4

Montalban, G., Bonache, S., Moles-Fernández, A., Gisbert-Beamud, A., Tenés, A., Bach, V., & Gutiérrez-Enríquez, S. (2019). Screening of. *Journal of Medical Genetics*, *56*(2), 63–74. https://doi.org/10.1136/jmedgenet-2018-105606

Morello, G., Guarnaccia, M., Spampinato, A. G., Salomone, S., D'Agata, V., Conforti, F. L., & Cavallaro, S. (2019). Integrative multi-omic analysis identifies new drivers and pathways in molecularly distinct subtypes of ALS. *Scientific Reports*, *9*(1), 9968. https://doi.org/10.1038/s41598-019-46355-w

Mori, K., Oura, T., Noma, H., & Matsui, S. (2013). Cancer outlier analysis based on mixture modeling of gene expression data. *Computational and Mathematical Methods in Medicine*, *2013*, 693901. https://doi.org/10.1155/2013/693901

Mucaki, E. J., Shirley, B. C., & Rogan, P. K. (2020). Expression changes confirm genomic variants predicted to result in allele-specific, alternative mRNA splicing. *bioRxiv*, 549089. https://doi.org/10.1101/549089

Muller, I. B., Meijers, S., Kampstra, P., van Dijk, S., van Elswijk, M., Lin, M., & Cloos, J. (2021). Computational comparison of common event-based differential splicing tools: Practical considerations for laboratory researchers. *BMC Bioinformatics*, *22*(1), 347. https://doi.org/10.1186/s12859-021-04263-9

Mundt, E., Nix, P., Brown, K., Bowles, K. R., & Manley, S. (2017). Complexities of variant classification in clinical hereditary cancer genetic testing. *Journal of Clinical Oncology*, *35*(34), 3796–3799. https://doi.org/10.1200/JCO.2017.74.5182

Murdock, D. R., Dai, H., Burrage, L. C., Rosenfeld, J. A., Ketkar, S., Müller, M. F., & Network, U. D. (2021). Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *Journal of Clinical Investigation*, *131*(1), https://doi.org/10.1172/JCI141500

Mussack, V., Hermann, S., Buschmann, D., Kirchner, B., & Pfaffl, M. W. (2020). MIQE-compliant validation of MicroRNA biomarker signatures established by small RNA sequencing. *Methods in Molecular Biology*, *2065*, 23–38. https://doi.org/10.1007/978-1-4939-9833-3_3

National Research Council (US) Committee. (1996). The evaluation of forensic DNA evidence, *National Research Council (US) Committee on DNA Forensic Science: An update*. National Academies Press.

Ndugga-Kabuye, M. K., & Issaka, R. B. (2019). Inequities in multi-gene hereditary cancer testing: lower diagnostic yield and higher VUS rate in individuals who identify as Hispanic, African or Asian and Pacific Islander as compared to European. *Familial Cancer*, *18*(4), 465–469. https://doi.org/10.1007/s10689-019-00144-6

Nohr, E., Kunder, C. A., Jones, C., Sutton, S., Fung, E., Zhu, H., & Costa, H. A. (2019). Development and clinical validation of a targeted RNAseq panel (Fusion-STAMP) for diagnostic and predictive gene fusion detection in solid tumors. *bioRxiv*, 870634. https://doi.org/10.1101/870634

Osan, C., Chira, S., Nutu, A. M., Braicu, C., Baciut, M., Korban, S. S., & Berindan-Neagoe, I. (2021). The Connection between MicroRNAs and Oral Cancer Pathogenesis: Emerging biomarkers in oral cancer management, *Genes* (12). Basel. 12. https://doi.org/10.3390/genes12121989

Oto, J., Herranz, R., Plana, E., Sánchez-González, J. V., Pérez-Ardavín, J., Hervás, D., & Medina, P. (2021). Identification of miR-20a-5p as robust normalizer for urine microRNA studies in renal cell carcinoma and a profile of dysregulated microRNAs. *International Journal of Molecular Sciences*, *22*(15), https://doi.org/10.3390/ijms22157913

Park, E., Pan, Z., Zhang, Z., Lin, L., & Xing, Y. (2018). The expanding landscape of alternative splicing variation in human populations. *American Journal of Human Genetics*, *102*(1), 11–26. https://doi.org/10.1016/j.ajhg.2017.11.002

Parsons, M. T., Tudini, E., Li, H., Hahnen, E., Wappenschmidt, B., Feliubadaló, L., & Investigators, K. (2019). Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. *Human Mutation*, *40*(9), 1557–1578. https://doi.org/10.1002/humu.23818

Petrov, A. I., Kay, S. J. E., Gibson, R., Kulesha, E., Staines, D., & Bruford, E. A., RNAcentral Consortium. (2015). RNAcentral: An international database of ncRNA sequences. *Nucleic Acids Res*, *43(Database issue)*, D123–D129. https://doi.org/10.1093/nar/gku991

Petrovski, S., & Goldstein, D. B. (2016). Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biology*, *17*(1), 157. https://doi.org/10.1186/s13059-016-1016-y

Plon, S. E., Eccles, D. M., Easton, D., Foulkes, W. D., Genuardi, M., Greenblatt, M. S., & Group, I. U. G. V. W. (2008). Sequence variant classification and reporting: Recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human Mutation*, *29*(11), 1282–1291. https://doi.org/10.1002/humu.20880

Pohl, M., Bortfeldt, R. H., Grützmann, K., & Schuster, S. (2013). Alternative splicing of mutually exclusive exons: A review. *Biosystems*, *114*(1), 31–38. https://doi.org/10.1016/j.biosystems.2013.07.003

Psych, E. C., Akbarian, S., Liu, C., Knowles, J. A., Vaccarino, F. M., Farnham, P. J., & Sestan, N. (2015). The PsychENCODE project. *Nature Neuroscience*, *18*(12), 1707–1712. https://doi.org/10.1038/nn.4156

Qian, X., Wang, J., Wang, M., Igelman, A. D., Jones, K. D., Li, Y., & Chen, R. (2021). Identification of deep-intronic splice mutations in a large cohort of patients with inherited retinal diseases. *Frontiers in Genetics*, *12*, 647400. https://doi.org/10.3389/fgene.2021.647400

Rajagopalan, R., Gilbert, M. A., McEldrew, D. A., Nassur, J. A., Loomes, K. M., Piccoli, D. A., & Spinner, N. B. (2021). Genome sequencing increases diagnostic yield in clinically diagnosed Alagille syndrome patients with previously negative test results. *Genetics in Medicine*, *23*(2), 323–330. https://doi.org/10.1038/s41436-020-00989-8

Rehm, H. L., Harrison, S. M., & Martin, C. L. (2017). ClinVar is a critical resource to advance variant interpretation. *Oncologist*, *22*(12), 1562. https://doi.org/10.1634/theoncologist.2017-0246

Ren, J., Chen, S., Ye, F., Gong, X., Lu, Y., Cai, Q., & Chen, Y. (2021). Exploration of differentially-expressed exosomal mRNAs, lncRNAs and circRNAs from serum samples of gallbladder cancer and xantho-granulomatous cholecystitis patients. *Bioengineered*, *12*(1), 6134–6143. https://doi.org/10.1080/21655979.2021.1972780

Rentas, S., Rathi, K. S., Kaur, M., Raman, P., Krantz, I. D., Sarmady, M., & Tayoun, A. A. (2020). Diagnosing cornelia de Lange syndrome and related neurodevelopmental disorders using RNA sequencing. *Genetics in Medicine*, *22*(5), 927–936. https://doi.org/10.1038/s41436-019-0741-5

Richards, C. S., Aziz, N., Bale, S., Bick, D., Das, S., & Gastier-Foster, J., ACMG/AMP Interpretation of Sequence Variants Work Group 2015. (2018). Response to Biesecker and Harrison. *Genetics in Medicine*, *20*(12), 1689–1690. https://doi.org/10.1038/gim.2018.43

Roberts, M. E., Susswein, L. R., Janice Cheng, W., Carter, N. J., Carter, A. C., Klein, R. T., & Marshall, M. L. (2020). Ancestry-specific hereditary cancer panel yields: Moving toward more personalized risk assessment. *Journal of Genetic Counseling*, *29*(4), 598–606. https://doi.org/10.1002/jgc4.1257

Rofes, P., Menéndez, M., González, S., Tornero, E., Gómez, C., Vargas-Parra, G., & Lázaro, C. (2020). Improving genetic testing in hereditary cancer by RNA analysis: Tools to prioritize splicing studies and challenges in applying American College of Medical Genetics and Genomics Guidelines. *Journal of Molecular Diagnostics*, *22*(12), 1453–1468. https://doi.org/10.1016/j.jmoldx.2020.09.007

Rowlands, C. F., Taylor, A., Rice, G., Whiffin, N., Hall, H. N., Newman, W. G., Investigators, k, & (2022). MRSD: A quantitative approach for assessing suitability of RNA-seq in the investigation of mis-splicing in Mendelian disease. *American Journal of Human Genetics*, *109*(2), 210–222. https://doi.org/10.1016/j.ajhg.2021.12.014

Ryan, M. C., Cleland, J., Kim, R., Wong, W. C., & Weinstein, J. N. (2012). SpliceSeq: A resource for analysis and visualization of RNA-Seq data on alternative splicing and its functional impacts. *Bioinformatics*, *28*(18), 2385–2387. https://doi.org/10.1093/bioinformatics/bts452

Sanders, S. J., Schwartz, G. B., & Farh, K. K. (2020). Clinical impact of splicing in neurodevelopmental disorders. *Genome Medicine*, *12*(1), 36. https://doi.org/10.1186/s13073-020-00737-2

Sawaki, K., Kanda, M., & Kodera, Y. (2018). Review of recent efforts to discover biomarkers for early detection, monitoring, prognosis, and prediction of treatment responses of patients with gastric cancer. *Expert Review of Gastroenterology & Hepatology*, *12*(7), 657–670. https://doi.org/10.1080/17474124.2018.1489233

Scotti, M. M., & Swanson, M. S. (2016). RNA mis-splicing in disease. *Nature Reviews Genetics*, *17*(1), 19–32. https://doi.org/10.1038/nrg.2015.3

Shamseldin, H. E., AlAbdi, L., Maddirevula, S., Alsaif, H. S., Alzahrani, F., Ewida, N., & Alkuraya, F. S. (2021). Lethal variants in humans: Lessons learned from a large molecular autopsy cohort. *Genome Medicine*, *13*(1), 161. https://doi.org/10.1186/s13073-021-00973-0

Shao, L., Xing, F., Xu, C., Zhang, Q., Che, J., Wang, X., & Ouyang, Y. (2019). Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(12), 5653–5658. https://doi.org/10.1073/pnas.1820513116

Shen, S., Park, J. W., Lu, Z. X., Lin, L., Henry, M. D., Wu, Y. N., & Xing, Y. (2014). rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(51), E5593–E5601. https://doi.org/10.1073/pnas.1419161111

Shenker, S., Miura, P., Sanfilippo, P., & Lai, E. C. (2015). IsoSCM: improved and alternative 3' UTR annotation using multiple change-point inference. *RNA*, *21*(1), 14–27. https://doi.org/10.1261/rna.046037.114

Shi, Y., Chen, Z., Gao, J., Wu, S., Gao, H., & Feng, G. (2018). Transcriptome-wide analysis of alternative mRNA splicing signature in the diagnosis and prognosis of stomach adenocarcinoma. *Oncology Reports*, *40*(4), 2014–2022. https://doi.org/10.3892/or.2018.6623

Shvetsova, E., Sofronova, A., Monajemi, R., Gagalova, K., Draisma, H. H. M., White, S. J., & consortium, G. (2019). Skewed X-inactivation is common in the general female population. *European Journal of Human Genetics*, *27*(3), 455–465. https://doi.org/10.1038/s41431-018-0291-3

Slavin, T. P., Van Tongeren, L. R., Behrendt, C. E., Solomon, I., Rybak, C., Nehoray, B., & Weitzel, J. N. (2018). Prospective Study of Cancer Genetic Variants: Variation in Rate of Reclassification by Ancestry. *Journal of the National Cancer Institute*, *110*(10), 1059–1066. https://doi.org/10.1093/jnci/djy027

Solé, C., Mentxaka, G., & Lawrie, C. H. (2021). The ue of circRNAs as biomarkers of cancer. *Methods in Molecular Biology*, *2348*, 307–341. https://doi.org/10.1007/978-1-0716-1581-2_21

Stenton, S. L., & Prokisch, H. (2020). The clinical application of RNA sequencing in genetic diagnosis of Mendelian Disorders. *Clinics in Laboratory Medicine*, *40*(2), 121–133. https://doi.org/10.1016/j.cll.2020.02.004

Sun, Y., Chen, G., He, J., Huang, Z. G., Li, S. H., Yang, Y. P., & Wu, H. (2021). Clinical significance and potential molecular mechanism of miRNA-222-3p in metastatic prostate cancer. *Bioengineered*, *12*(1), 325–340. https://doi.org/10.1080/21655979.2020.1867405

Sun, Z. (2015). High-throughput long noncoding RNA profiling for diagnostic and prognostic markers in cancer: Opportunities and challenges. *Epigenomics*, *7*(7), 1075–1078. https://doi.org/10.2217/epi.15.69

Sun, Z., Nair, A., Chen, X., Prodduturi, N., Wang, J., & Kocher, J. P. (2017). UClncR: Ultrafast and comprehensive long non-coding RNA detection from RNA-seq. *Scientific Reports*, *7*(1), 14196. https://doi.org/10.1038/s41598-017-14595-3

Szelinger, S., Malenica, I., Corneveaux, J. J., Siniard, A. L., Kurdoglu, A. A., Ramsey, K. M., & Craig, D. W. (2014). Characterization of X chromosome inactivation using integrated analysis of whole-exome and mRNA sequencing. *PLoS One*, *9*(12), e113036. https://doi.org/10.1371/journal.pone.0113036

Tabury, K., Monavarian, M., Listik, E., Shelton, A. K., Choi, A. S., Quintens, R., & Mythreye, K. (2022). PVT1, a YAP1 dependent stress responsive lncRNA drives ovarian cancer metastasis and chemoresistance. *bioRxiv*, 475893. https://doi.org/10.1101/2022.01.11.475893

Tahiliani, J., Leisk, J., Aradhya, K., Ouyang, K., Aradhya, S., & Nykamp, K. (2020). Utility of RNA sequencing analysis in the context of genetic testing. *Curr Genet Med Rep*, *8*, 140–146.

Talhouarne, G. J. S., & Gall, J. G. (2018). Lariat intronic RNAs in the cytoplasm of vertebrate cells. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(34), E7970–E7977. https://doi.org/10.1073/pnas.1808816115

Tan, A. C., Fan, J. B., Karikari, C., Bibikova, M., Garcia, E. W., Zhou, L., & Chakravarti, A. (2008). Allele-specific expression in the germline of patients with familial pancreatic cancer: An unbiased approach to cancer gene discovery. *Cancer Biology & Therapy*, *7*(1), 135–144. https://doi.org/10.4161/cbt.7.1.5199

Tan, I. L., de Almeida, C., Modderman, R., Stachurska, R., Dekens, A., Barisani, J., & Withoff, D.S. (2021). Circulating miRNAs as potential biomarkers for celiac disease development. *Frontiers in Immunology*, (*12*):734763. https://doi.org/10.3389/fimmu.2021.734763

Tang, M., Abbas, H. A., Negrao, M. V., Ramineni, M., Hu, X., Hubert, S. M., & Kalhor, N. (2021). The histologic phenotype of lung cancers is associated with transcriptomic features rather than genomic characteristics. *Nature Communications*, *12*(1), 7081. https://doi.org/10.1038/s41467-021-27341-1

Tang, S. J., Shen, H., An, O., Hong, H., Li, J., Song, Y., & Chen, L. (2020). Cis- and trans-regulations of pre-mRNA splicing by RNA editing enzymes influence cancer development. *Nature Communications*, *11*(1), 799. https://doi.org/10.1038/s41467-020-14621-5

Tanner, M. K., Tang, Z., & Thornton, C. A. (2021). Targeted splice sequencing reveals RNA toxicity and therapeutic response in myotonic dystrophy. *Nucleic Acids Research*, *49*(4), 2240–2254. https://doi.org/10.1093/nar/gkab022

Tavtigian, S. V., Greenblatt, M. S., Goldgar, D. E., Boffetta, P., & Group, I. U. G. V. W. (2008). Assessing pathogenicity: Overview of results from the IARC Unclassified Genetic Variants Working Group. *Human Mutation*, *29*(11), 1261–1264. https://doi.org/10.1002/humu.20903

Thompson, K., Collier, J. J., Glasgow, R. I. C., Robertson, F. M., Pyle, A., Blakely, E. L., & Taylor, R. W. (2020). Recent advances in understanding the molecular genetic basis of mitochondrial disease. *Journal of Inherited Metabolic Disease*, *43*(1), 36–50. https://doi.org/10.1002/jimd.12104

Truty, R., Ouyang, K., Rojahn, S., Garcia, S., Colavin, A., Hamlington, B., & Aradhya, S. (2021). Spectrum of splicing variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation. *American Journal of Human Genetics*, *108*(4), 696–708. https://doi.org/10.1016/j.ajhg.2021.03.006

Tsang, E. S., Grisdale, C. J., Pleasance, E., Topham, J. T., Mungall, K., Reisle, C., & Loree, J. M. (2021). Uncovering clinically relevant gene fusions with integrated genomic and transcriptomic profiling of metastatic cancers. *Clinical Cancer Research*, *27*(2), 522–531. https://doi.org/10.1158/1078-0432.CCR-20-1900

Ungerleider, N., & Flemington, E. (2019). SpliceV: Analysis and publication quality printing of linear and circular RNA splicing, expression and regulation. *BMC Bioinformatics*, *20*(1), 231. https://doi.org/10.1186/s12859-019-2865-7

Urbanski, L. M., Leclair, N., & Anczuków, O. (2018). Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdisciplinary Reviews: RNA*, *9*(4), e1476. https://doi.org/10.1002/wrna.1476

Urbina-Jara, L. K., Rojas-Martinez, A., Martinez-Ledesma, E., Aguilar, D., Villarreal-Garza, C., & Ortiz-Lopez, R. (2019). Landscape of germline mutations in DNA repair genes for breast cancer in Latin America: Opportunities for PARP-Like inhibitors and immunotherapy, *Genes* (10). Basel. 10. https://doi.org/10.3390/genes10100786

Vaske, O. M., Bjork, I., Salama, S. R., Beale, H., Tayi Shah, A., Sanders, L., & Haussler, D. (2019). Comparative tumor RNA sequencing analysis for difficult-to-treat pediatric and young adult patients with cancer. *JAMA Network Open*, *2*(10), e1913968. https://doi.org/10.1001/jamanetworkopen.2019.13968

Vaz-Drago, R., Custodio, N., & Carmo-Fonseca, M. (2017). Deep intronic mutations and human disease. *Human Genetics*, *136*(9), 1093–1111. https://doi.org/10.1007/s00439-017-1809-4

Veneziano, D., Di Bella, S., Nigita, G., Laganà, A., Ferro, A., & Croce, C. M. (2016). Noncoding RNA: Current deep sequencing data analysis approaches and challenges. *Human Mutation*, *37*(12), 1283–1298. https://doi.org/10.1002/humu.23066

Virlogeux, V., Graff, R. E., Hoffmann, T. J., & Witte, J. S. (2015). Replication and heritability of prostate cancer risk variants: Impact of population-specific factors. *Cancer Epidemiology, Biomarkers and Prevention*, *24*(6), 938–943. https://doi.org/10.1158/1055-9965.EPI-14-1372

Volodarsky, M., Kerkhof, J., Stuart, A., Levy, M., Brady, L. I., Tarnopolsky, M., & Sadikovic, B. (2021). Comprehensive genetic sequence and copy number analysis for Charcot-Marie-Tooth disease in a Canadian cohort of 2517 patients. *Journal of Medical Genetics*, *58*(4), 284–288. https://doi.org/10.1136/jmedgenet-2019-106641

Wai, H. A., Lord, J., Lyon, M., Gunning, A., Kelly, H., & Cibin, P. (group 2020). Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genetics in Medicine*, *22*(6), 1005–1014. https://doi.org/10.1038/s41436-020-0766-9

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164. https://doi.org/10.1093/nar/gkq603

Wang, Y., Liu, J., Huang, B. O., Xu, Y. M., Li, J., Huang, L. F., & Wang, X. Z. (2015). Mechanism of alternative splicing and its regulation. *Biomedical Reports*, *3*(2), 152–158. https://doi.org/10.3892/br.2014.407

Wang, Y., Mashock, M., Tong, Z., Mu, X., Chen, H., Zhou, X., & Li, X. (2020). Changing technologies of RNA sequencing and their applications in clinical oncology. *Frontiers in Oncology*, (*10*):447. https://doi.org/10.3389/fonc.2020.00447

Wang, Z., Fan, X., Shen, Y., Pagadala, M. S., Signer, R., Cygan, K. J., & Huang, K. L. (2021). Non-cancer-related pathogenic germline variants and expression consequences in ten-thousand cancer genomes. *Genome Medicine*, *13*(1), 147. https://doi.org/10.1186/s13073-021-00964-1

Wen, J., Yang, T., Mallouk, N., Zhang, Y., Li, H., Lambert, C., & Li, G. (2021). Urinary exosomal CA9 mRNA as a novel liquid biopsy for molecular diagnosis of bladder cancer. *International Journal of Nanomedicine*, *16*, 4805–4811. https://doi.org/10.2147/IJN.S312322

Williams, H. L., Walsh, K., Diamond, A., Oniscu, A., & Deans, Z. C. (2018). Validation of the Oncomine. *Virchows Archiv*, *473*(4), 489–503. https://doi.org/10.1007/s00428-018-2411-4

Wong, L. S., & Wong, C. M. (2021). Decoding the roles of long noncoding RNAs in hepatocellular carcinoma. *International Journal of Molecular Sciences*, *22*(6), https://doi.org/10.3390/ijms22063137

Woolfe, A., Mullikin, J. C., & Elnitski, L. (2010). Genomic features defining exonic variants that modulate splicing. *Genome Biology*, *11*(2), R20. https://doi.org/10.1186/gb-2010-11-2-r20

Xi, X., Li, T., Huang, Y., Sun, J., Zhu, Y., Yang, Y., & Lu, Z. J. (2017). RNA biomarkers: Frontier of precision medicine for cancer. *Noncoding RNA*, *3*(1):3010009. https://doi.org/10.3390/ncrna3010009

Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nature Protocols*, *10*(10), 1556–1566. https://doi.org/10.1038/nprot.2015.105

Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., & Eng, C. M. (2014). Molecular findings among patients referred for clinical whole-exome sequencing. *Journal of the American Medical Association*, *312*(18), 1870–1879. https://doi.org/10.1001/jama.2014.14601

Yépez, V. A., Gusic, M., Kopajtich, R., Mertes, C., Smith, N. H., Alston, C. L., & Prokisch, H. (2021). Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *medRxiv*, 21254633. https://doi.org/10.1101/2021.04.01.21254633

Yuan, J., Kensler, K. H., Hu, Z., Zhang, Y., Zhang, T., Jiang, J., & Zhang, L. (2020). Integrative comparison of the genomic and transcriptomic landscape between prostate cancer patients of predominantly African or European genetic ancestry. *PLoS Genetics*, *16*(2), e1008641. https://doi.org/10.1371/journal.pgen.1008641

Zeuschner, P., Linxweiler, J., & Junker, K. (2020). Non-coding RNAs as biomarkers in liquid biopsies with a special emphasis on extracellular vesicles in urological malignancies. *Expert Review of Molecular Diagnostics*, *20*(2), 151–167. https://doi.org/10.1080/14737159.2019.1665998

Zhang, Y., Liu, X., MacLeod, J., & Liu, J. (2018). Discerning novel splice junctions derived from RNA-seq alignment: A deep learning approach. *BMC Genomics*, *19*(1), 971. https://doi.org/10.1186/s12864-018-5350-1

Zhang, Y., Qian, J., Gu, C., & Yang, Y. (2021). Alternative splicing and cancer: A systematic review. *Signal Transduction and Targeted Therapy*, *6*(1), 78. https://doi.org/10.1038/s41392-021-00486-7

Zhu, B., Gu, S., Wu, X., He, W., & Zhou, H. (2021). Bioinformatics analysis of tumor-educated platelet microRNAs in patients with hepatocellular carcinoma. *Bioscience Reports*, *41*(12), https://doi.org/10.1042/BSR20211420

How to cite this article: Postel, M. D., Culver, J. O., Ricker, C., & Craig, D. W. (2022). Transcriptome analysis provides critical answers to the "variants of uncertain significance" conundrum. *Human Mutation*, 43, 1590–1608. https://doi.org/10.1002/humu.24394