ISSN 0006-2979, Biochemistry (Moscow), 2025, Vol. 90, No. 6, pp. 683-699 © Pleiades Publishing, Ltd., 2025.
Russian Text © The Author(s), 2025, published in Biokhimiya, 2025, Vol. 90, No. 6, pp. 733-751.

REVIEW

Transcriptional Biomarkers in the Diagnosis of Genetic Disorders: Opportunities, Challenges, and Prospects for Application

Lidia N. Nefedova[1,a]* and Tatiana N. Krasnova[2]

[1]*Faculty of Biology, Lomonosov Moscow State University, 119991 Moscow, Russia*
[2]*Faculty of Fundamental Medicine, Lomonosov Moscow State University, 119991 Moscow, Russia*
[a]*e-mail: nefedova@mail.bio.msu.ru*

Received February 11, 2025
Revised April 17, 2025
Accepted May 21, 2025

Abstract—Quantitative analysis of gene transcription is widely used across various fields of biology and, in particular, in medicine, it serves as a tool for diagnostics and transcriptomic profiling of diseases. In recent years, transcriptome analysis methods based on large-scale next-generation sequencing have become widely adopted. Transcriptomic studies enable the identification of cellular processes that are active at specific time points, the investigation of transcriptome dynamics in different tissues or physiological states (such as during ontogenesis or adaptive responses) and the detection of differentially expressed genes in pathological conditions. A pronounced change in the transcription level of one or more genes under pathological conditions may be sufficient for diagnosis, serving as a transcriptional biomarker of disease. However, in some cases, altered transcription levels may indicate the presence of mutations, including those leading to disruption of splicing, activation of mobile elements, or pseudogenes. This review discusses cases in which transcriptional changes can provide insights into the genetic causes of disease, as well as the challenges that must be considered when using transcription as a diagnostic marker. In the future, specialized targeted panels based on transcriptome analysis are expected to be used not only as diagnostic and prognostic tools, but also as predictors of structural genomic abnormalities, thereby contributing to the development of novel strategies for effective disease treatment.

DOI: 10.1134/S0006297925600383

Keywords: genetic diseases, transcription, biomarkers

INTRODUCTION

Genetic disorders are caused by dysfunctions of the cellular genetic apparatus, often resulting from mutations in specific genes. To date, several thousand genes associated with more than 7500 monogenic inherited diseases [1] have been described, and the Human Gene Mutation Database (HGMD) contains information on over 200,000 pathogenic gene variants (alleles) [2]. Nevertheless, the causes of many diseases (mainly polygenic and multifactorial ones) remain unclear.

A major obstacle in identifying the causes of polygenic diseases is the substantial genetic diversity of the human population. The 1000 Genomes Project revealed that individual genomes differ from the reference genome by approximately 4.1 to 5 million single nucleotide polymorphisms (SNPs) [3]. The dbSNP database (https://www.ncbi.nlm.nih.gov/snp/) contains information on over 25 million distinct SNPs, of which only about 300,000 are located in exons, while the others are found in non-coding DNA regions and are most likely functionally neutral. According to the Exome Aggregation Consortium (ExAC) Project, which provides data on exome sequences from over 60,000 individuals, the SNPs' majority of coding regions are also predicted to be functionally neutral [4]. However, in practice, predicting which of the identified genetic variants are neutral and which may contribute to disease development is a complex task. Some neutral

* To whom correspondence should be addressed.

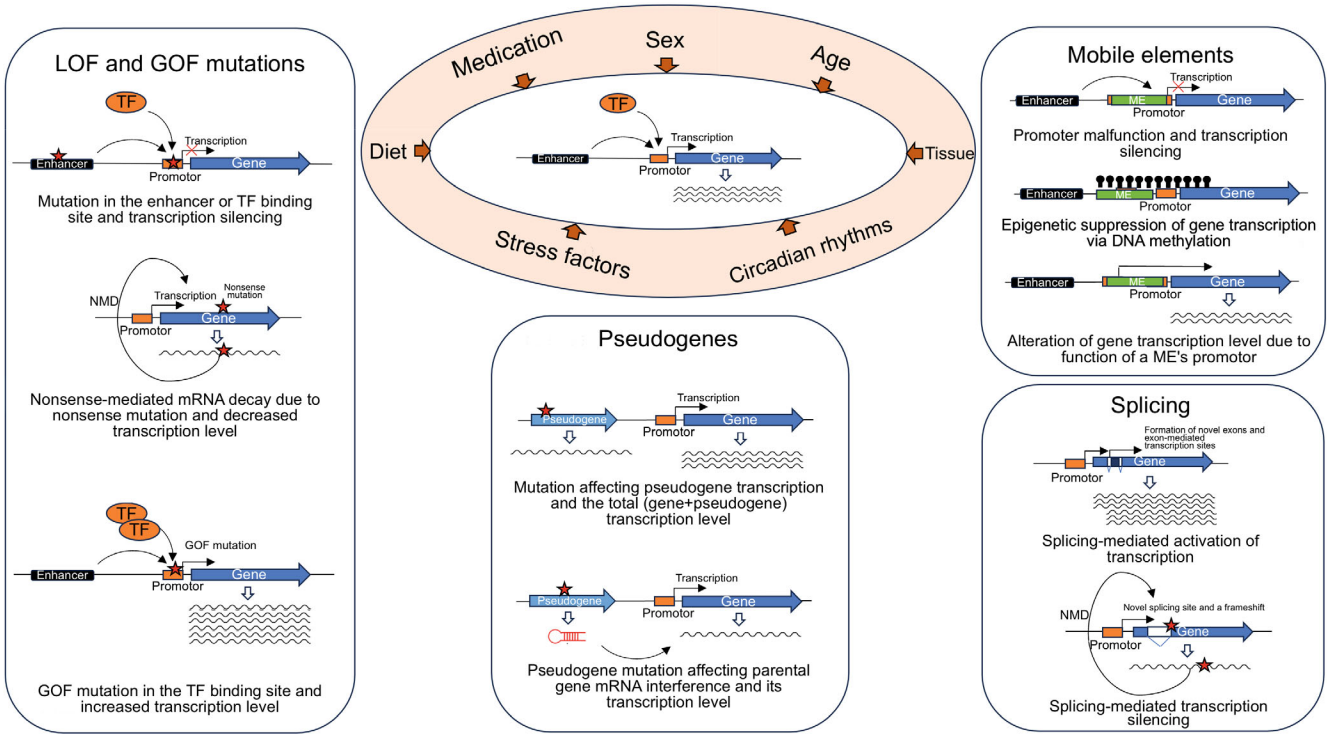

Fig. 1. Environmental and molecular-genetic mechanisms that may influence gene transcription levels. Designations: LOF, loss-of-function; GOF, gain-of-function; NMD, nonsense-mediated mRNA decay; TF, transcription factor; ME, mobile element.

variants may exhibit pathogenicity only under certain conditions, including environmental influences and/or interactions with other genetic variants.

To identify associations between SNPs and polygenic diseases, researchers use methods based on linking candidate genetic variants with mutant phenotypes, including genome-wide association studies (GWAS) [5]. GWAS typically involve searching for correlations between genotype variants (that is usually SNPs) and certain disease. This approach enables the estimation of disease risk (genetic predisposition) in individuals carrying specific gene variants (i.e., particular SNPs) compared to individuals who do not carry these variants.

On the other hand, gene dysfunction may result not from structural alterations in the gene itself, but from disruptions in its expression. Therefore, one approach to study mutant phenotypes involves analyzing changes in the transcription of specific candidate genes or performing a global analysis of alterations at the transcriptome level. Transcription is an indicator of gene function that is difficult to interpret. First, it represents an initial (but not the only) stage in the regulation of gene expression. Following transcription, additional regulatory mechanisms become activated. Disruptions at any stage of gene expression regulation, not necessarily at the transcriptional level, can affect gene function and lead to phenotypic changes. Second, transcription is a complex multistep process that involves both epigenetic regulation (chromatin remodeling, histone protein modifications, DNA methylation) and direct transcriptional regulation itself. The latter occurs through *cis*-regulatory gene sequences, such as enhancers, silencers, or insulators, as well as through *trans*-regulation by transcription factors (TFs). Third, transcription is significantly influenced by environment, which may include both external factors (abiotic or biotic) and internal ones (cell or tissue type, age, sex, and other physiological parameters) (Fig. 1).

Nevertheless, in some cases, transcription can serve as an informative indicator of specific structural alterations in the genome. Figure 1 shows the main molecular-genetic factors that may affect gene transcription levels: mutations, particularly those that alter splicing, as well as the transcriptional activity of mobile elements and pseudogenes.

To determine how gene transcription changes under the influence of various genetic and non-genetic factors, large-scale transcriptome analysis is now commonly used. Transcriptomic studies (RNA sequencing, RNA-seq) enable to identify which cellular processes are active at the time of RNA extraction [6]. One of the key advantages of RNA sequencing is its ability to reveal transcriptome dynamics across different tissues or states, such as during ontogenesis or physiological adaptation. It can also be used for comparative transcriptome analysis of biomedical samples obtained from diseased and healthy tissues.

The quality and predictive power of transcriptomic studies are influenced by numerous factors, since transcription is a labile process that enables cells to rapidly adapt to external and internal, environmental or physiological changes. Some genes (for example, housekeeping genes) are transcribed in a stable and condition-independent manner, showing minimal variation in transcription levels. Others are highly sensitive to environmental changes and exhibit substantial variability in transcription depending on the conditions.

A pronounced change in gene transcription levels under pathological conditions may be sufficient for diagnosis, serving as a transcriptional biomarker of disease. However, transcriptomic studies alone are not sufficient to fully understand the mechanisms underlying transcriptional alterations. In this context, combining genome-wide association studies (GWASs) with transcriptome-wide association studies (TWASs) has proven to be an effective approach [7]. The first approach (GWAS) focuses on identifying associations between SNPs and pathological genotypes, while the second (TWAS) aims to detect associations between gene transcription levels and disease, as well as to search for gene networks whose function is disrupted in disease. In the following sections, we will examine in detail the circumstances under which transcription levels may be useful for elucidating the genetic basis of disease, and the factors that must be taken into account when using transcription as a diagnostic tool. We will also discuss the challenges that complicate the interpretation of transcriptomic data.

GENE TRANSCRIPTION AS A BIOMARKER OF GENETIC DISORDERS: OPPORTUNITIES

Transcription levels can serve as indicators of both loss-of-function (LOF) and gain-of-function (GOF) mutations. LOF mutations include: nonsense-mediated mRNA decays (NMDs), which completely stop protein synthesis; missense mutations that negatively affect protein activity or stability; mutations in regulatory regions that impair gene expression [8]. Approximately one-third of mutant genes identified in monogenic and cancer-related disorders carry nonsense or frameshift mutations that result in the formation of premature termination codons [9]. The decay of mRNAs containing nonsense mutations is a cellular process that eliminates mRNA transcripts carrying premature stop codons, thereby preventing the synthesis of truncated and potentially dangerous proteins [10]. This mechanism ensures that only mRNAs capable of producing full-length proteins are translated. By reducing the levels of defective mRNAs, NMD decreases the expression of truncated proteins, some of which may exert dominant-negative effects, and results in the suppression of mutant allele transcription.

It is known that approximately 10% of patients with cystic fibrosis are homozygous for a mutation in the *CFTR* gene that introduces a premature termination codon (PTC). At the same time, the mRNA levels of *CFTR* contain a nonsense mutation are significantly reduced compared to wild-type *CFTR* mRNA not only in homozygous patients but also in healthy heterozygous carriers of the mutation [11]. It has been revealed that *CFTR* mRNA levels in cell lines harboring various nonsense mutations (Y122X, G542X, R1162X, and W1282X) are decreased by 50-80% relative to wild-type *CFTR* mRNA levels in the parental cell lines [12-14]. Other examples of transcript reduction due to NMD mechanism involve the following genes: *JAG1* (in Alagille syndrome), *DYRK1A* (in intellectual disability), and *ZIC2* (in holoprosencephaly); the involvement of the above-mentioned genes in haploinsufficiency disorders has been reported. To systematically investigate factors determining the efficiency of NMD in cancer, a database of somatic nonsense mutations in genes from 9769 cancer patients was constructed and integrated with mRNA expression data from The Cancer Genome Atlas (TCGA) [15]. Thus, decreased gene transcript levels may indicate the presence of nonsense mutations that affect transcription.

It is important to note that LOF mutations have different phenotypic consequences depending on whether a gene's function is dose-dependent or dose-independent. Recessive LOF mutations are typically found in genes encoding metabolic enzymes, suggesting that their function is not dependent on gene dosage. Genes with limited variability in transcript levels are known as dosage-sensitive genes [16]. Genes encoding regulatory proteins, transcription factors, receptors, or their ligands are dosage-sensitive. LOF mutations in these genes are usually dominant, since a single functional allele expression performance is insufficient to maintain normal operation. Functions, provided by the housekeeping genes, are also dosage-sensitive, and LOF mutations in these genes are generally dominant [17, 18].

Gain-of-function (GOF) mutations can lead to various alterations at the protein level, such as constitutive synthesis, substrate change, disruption of target-binding specificity, or protein aggregation [2]. All of the above make GOF mutations critically important in the development and progression of various cancer types. While most LOF mutations are localized within structured protein domains, GOF mutations are more frequently found in intrinsically disordered regions (IDRs) of proteins [19]. As a result, mutations in IDRs often disrupt molecular interactions and consequently affect signaling pathway function.

More than 90% of all variants associated with genetic diseases have been shown to be located in non-coding regions of the genome [20]. However, mutations in non-coding regions can affect gene function by disrupting the interaction between TFs and their binding sites in regulatory regions of genes [21]. As a result, LOF mutation may occur due to the loss of a binding site, or GOF mutation may arise due to the emergence of a new binding site (Fig. 1).

In the study by Fuxman Bass et al. [22], the authors investigated the impact of point mutations in non-coding regions on transcription factor (TF) binding to enhancers using a yeast-based system: most of the mutations resulted in a loss of interaction, but several dozen led to enhanced binding. Another example demonstrating how mutations in non-coding regions can affect gene function is provided by mutations in the promoter of the telomerase reverse transcriptase (*TERT*) gene, which are frequently observed in various types of cancer [23]. Two independent mutations in the *TERT* promoter were identified that generated *de novo* consensus binding motifs for the ETS transcription factor, leading to a 2-4-fold increase in the transcriptional activity of the gene [23].

Pathogenic GOF mutations in the *STAT1* gene are typically associated with elevated levels of phosphorylated STAT1 transcription factor, as well as increased transcription level of the gene itself. In the study by Zimmerman et al. [24], *STAT1* mRNA levels were analyzed in blood cells from healthy donors and patients carrying GOF mutations in the *STAT1* following induction with interferons γ and α. The results showed that the median transcription levels of *STAT1* were approximately three times higher in patients than in healthy donors.

Changes in transcriptional levels may indicate the presence of a mutation in an enhancer [25]. Most genes are regulated by more than one enhancer, and many enhancers control the transcription of multiple genes [26, 27]. Systematic analyses have provided evidence for the emergence of new gene functions through enhancer "reprogramming" during evolution, which occurs via the acquisition of novel transcription factor binding sites [28]. Super-enhancers are large clusters of enhancers, and disease-associated SNPs have been shown to be particularly enriched in super-enhancers of oncogenes in cancer cells [29]. For example, mutations in the super-enhancer of the *TAL1* gene, which is associated with T cell acute lymphoblastic leukemia, result in the emergence of MYB transcription factor binding site, leading to *TAL1* gene overexpression in the tumor [30].

Mutations in non-coding regions can also affect gene expression by disrupting interactions with microRNAs. For example, a mutation in the non-coding region of the E2F1:MIR136-5p locus disrupts microRNA-mediated regulation, resulting in increased activity of the E2F1 oncogene in colorectal cancer [31]. The SomamiR 2.0 database (http://compbio.uthsc.edu/SomamiR) contains data on somatic GOF mutations identified in cancer that potentially alter interactions between microRNAs and competing endogenous RNAs (ceRNAs), including mRNAs, circular RNAs (circRNAs), and long non-coding RNAs (lncRNAs) [32].

Thus, changes in gene transcription levels may indicate LOF or GOF mutations occurring not only within the coding sequences of these genes but also in regulatory regions, such as transcription factor binding sites or enhancers.

Transcription level as an indicator of mutations contributing to splicing defects. Splicing-disrupting mutations are a common cause of monogenic diseases. It has been estimated that up to 60% of all pathogenic SNPs may lead to splicing defects [33]. Point mutations within exons can disrupt the function of exonic splicing enhancers (ESEs) or produce new exonic splicing silencers (ESSs). It has been found that approximately 10% of ~5000 known pathogenic missense variants result in exon skipping [34]. If exon skipping does not result in a frameshift, the transcript may be translated into protein, and the pathogenicity of such an event is not always evident [35]. Conversely, a frameshift can provide to the formation of a stop codon, triggering NMD mechanism (Fig. 1).

Since splicing is highly tissue-specific [36], the tissue selection for analysis is of critical importance in mutation detection. A mRNA of the same gene can be spliced differently across tissues. Age-related changes in splicing include alternative splicing of aging-related genes, as well as alterations in the expression levels of core spliceosome genes and splicing regulatory factors [37].

Transcription and splicing are closely interconnected. Splicing provides feedback on transcription initiation, influencing the gene's transcriptional profile. A recently described phenomenon, exon-mediated activation of transcription starts (EMATS), demonstrates that splicing of internal exons can regulate transcription initiation and activate cryptic promoters [38]. Genes containing EMATS have been shown to be linked to numerous genetic diseases (neurodevelopmental disorders, immunodeficiency, cancer, deafness, and others) [39]. Thus, transcription can be modulated via the efficiency of internal exon splicing, and changes in transcription levels may indicate splicing mutations (Fig. 1).

Transcription of mobile elements as a marker of genetic disorders. Mobile elements (MEs) are an essential component of the human genome, accounting for approximately half of its content, with the majority presented by retrotransposons [40]. ME transposition is a potentially deleterious operation that can lead

to genomic rearrangements. In addition, rearrangements may also result from recombination processes between ME copies. Although many MEs have lost their transpositional activity over the course of evolution, individual copies of Alu, L1, SVA, and human endogenous retroviruses (HERVs) remain transcriptionally active, and some of these copies are still capable of transposition within the human genome [41].

Altered transcriptional profiles of mobile elements and their impact on nearby genes have been observed in various diseases, including cancer and neurodegenerative disorders [42]. Through their own promoters, MEs can either suppress or enhance the expression of neighboring genes (Fig. 1). For example, in Hodgkin's lymphoma, transcription of the proto-oncogene *CSF1R* is initiated from the long terminal repeat (LTR) of a *THE1B* element, a member of the *MaLR* LTR-retrotransposon family [43]. Activation of *THE1B* transcription may serve as a potential diagnostic and/or prognostic marker for Hodgkin's lymphoma.

The RNA sequencing data from The Cancer Genome Atlas (http://cancergenome.nih.gov/) have been used to quantitatively assess ME expression in colorectal cancer. As a result, ME expression was shown to function as a prognostic marker for patients with colorectal cancer [44].

Many human epithelial cancers, especially those associated with *TP53* mutations, are characterized by elevated expression of *L1* [45, 46]. Increased *L1* expression has been reported in ovarian, esophageal, colorectal, lung, breast, and pancreatic cancers, and it correlates with disease severity [47, 48]. The data evidence that *L1* expression analysis can be applied as a prognostic tool.

Approximately 80% of all long non-coding RNAs (lncRNAs) in the human transcriptome contain MEs' sequences [49]. Evidence suggests that ME-derived lncRNAs are involved in melanoma progression [50], contribute to tumor progression, metastasis, or chemoresistance in breast cancer [51], pancreatic cancer, and hepatocellular carcinoma [52].

The examples described above demonstrate that the analysis of transcriptional activity of MEs can use as a diagnostic and prognostic marker for various cancers. Moreover, MEs are also markers of age-related changes. In humans, HERV-K (HML-2) and HERV-W provide differential expression patterns in young and elderly individuals [53]. The expression of HERV-H and HERV-W has been shown to correlate significantly with age [54]. In particular, HERV-W expression markedly increases in individuals over the age of 40 years, that is a range that coincides with the development of neurodegenerative diseases, such as multiple sclerosis. Data about ME transcription levels may also facilitate a diagnosis of inflammatory brain disorders [55]. Several studies have shown that HERV-H, HERV-K, HERV-L, and HERV-W are activated in Alzheimer's disease [56]. Moreover, HERV-H expression is significantly elevated in patients with autism spectrum disorders (ASD), particularly in individuals with severe disease progression [57]. Thus, analysis of HERV-H transcription levels may provide a promising marker for ASD diagnosis; however, the authors emphasize that additional studies on larger patient sampling are required to confirm this hypothesis.

Thus, the overall level of ME transcriptional activity contributes to biodiagnostic management of many genetic disorders, especially cancers. Since ME transcription is strictly regulated by the host genome, disruptions in genome function are estimated to underlie the changes in ME expression levels. However very few studies have been carried out to date.

Pseudogene transcription as a marker of genetic disorders. Pseudogenes were historically considered merely nonfunctional copies of protein-coding genes that had lost their protein-coding capacity due to the accumulation of deleterious mutations. However, while the majority of human pseudogenes are indeed nonfunctional, approximately 20% exhibit transcriptional activity, and some are even capable of producing protein products [58]. With the advent of high-throughput sequencing technologies, thousands of pseudogenes have been identified and implicated in the etiology and pathogenesis of various diseases. Increasing evidence suggests that pseudogenes are integral components of the complex regulatory networks that control gene expression [59].

Some pseudogenes contribute to the regulation of gene expression and therefore should be considered as "functional" genes (Fig. 1). Recent studies have demonstrated that pseudogene RNAs can enhance the transcription of their parental genes by competing for binding with regulatory microRNAs, thereby alleviating microRNA-mediated repression of target genes [60].

Most pseudogenes are co-expressed with their parental genes, and their expression is critical for the function of the parental genes. For example, loss of *PTENP1* function, a processed pseudogene of the phosphatase and tensin homolog gene (*PTEN*), can lead to a significant decrease in *PTEN* transcription levels [61]. It has been shown that both *PTEN* and *PTENP1* can be deleted in melanoma [62], indicating that the functions of both the parental gene and its pseudogene are necessary under normal conditions.

Conversely, some pseudogenes exhibit expression patterns that differ entirely from those of their parental genes. A systematic analysis of pseudogene transcription revealed that pseudogenes are transcribed differentially depending on their presence in cancerous or normal tissues [63]. Some pseudogenes can be classified as cancer-specific [64].

Depending on certain pathological conditions, these pseudogenes produce unique expression profiles, which are considered to be potential biomarkers for clinical application. For instance, transcription of *SUMO1P*, a pseudogene of the ubiquitin-like modifier 1 gene *SUMO1*, is significantly elevated in gastric cancer tissues compared to adjacent non-tumorous tissues, and its expression level correlates with tumor size, differentiation, lymphatic metastasis, and invasion [65]. Expression of the pseudogene *INTS6P1*, derived from the integrator complex subunit 6 gene *INTS6*, is significantly reduced in the plasma of patients with hepatocellular carcinoma compared to healthy individuals [66]. The transcript levels of *FTH1P3*, a pseudogene of the ferritin 1 heavy chain gene *FTH1*, are increased in cell lines and tissues of uveal melanoma [67]. Moreover, it has been shown that expression of the *Foxo3* gene, which encodes a forkhead family TF, is regulated by its pseudogene *Foxo3P* [68]. Ectopic expression of *Foxo3P*, circular RNA *Foxo3*, and *Foxo3* mRNA has been demonstrated to suppress tumor growth, as well as cancer cell proliferation and survival.

In addition to cancer, changes in pseudogene expression levels have been observed in various other pathological conditions, e.g., neurodegenerative diseases [69], cardiovascular diseases [70], and diabetes [71]. Therefore, pseudogene transcription is suggested to be highly informative diagnostic biomarker for these disorders. Despite the recent identification of numerous pseudogenes, researchers typically focus only on the expression of their parental genes, excluding transcription of the corresponding pseudogenes. However, accounting for pseudogene expression is essential for accurately measuring parental gene transcription levels and for determining the contribution of pseudogenes to overall transcriptional activity. This consideration is particularly important when selecting transcriptional biomarkers.

Transcription analysis can be useful for identifying gene networks and discovering genetic modifiers. Genes always function in interaction with other genes that influence their activity to some extent, forming genetic networks. Therefore, even monogenic diseases can have diverse genetic causes, which explains their genetic heterogeneity and variable phenotypic manifestations.

Genetic modifiers form a group of genes that can alter the phenotypic effects of disease-causing genes. They may affect the expression of genes exhibiting haploinsufficiency or modify the phenotype in haploinsufficiency contexts. A significant allelic imbalance in transcription, observed for 88% of genes in human tissues, is presumably caused by genetic modifiers [72]. In dominantly inherited diseases caused by haploinsufficiency, allelic imbalance can either enhance the expression of the normal allele, compensating for haploinsufficiency, or reduce its expression, thereby exacerbating the condition [73, 74].

Genetic compensation of mutant allele expression can be achieved either through the presence of additional gene copies in the genome, where loss of function of one gene is compensated by the activity of other genes with similar functions, or through changes in the expression pattern of the single normal allele, as demonstrated in several model organisms [75]. This process, known as transcriptional adaptation, modulates the expression of compensatory genes, thereby preventing or reducing the severity of the mutant phenotype [76]. For example, knockout or knockdown of the histone deacetylase 1 gene (*HDAC-1*) leads to increased expression of its homolog, *HDAC-2*, and *vice versa*, as shown in several cell lines and in both human and mouse tissues [77, 78].

Numerous examples have been described demonstrating the influence of genetic modifiers on disease severity, with a particularly large body of research focused on identifying genetic modifiers of cystic fibrosis progression [79]. A study was conducted to investigate the association of phenotypic manifestations of cystic fibrosis in patients homozygous for the F508del mutation with transcription levels and allelic variants of the *STAT3*, *IL1B*, and *IFNGR1* genes [80]. The interaction of the products of these genes determines the balance between inflammation, antiviral defense, and tissue repair: *STAT3* encodes a transcription factor of the JAK–STAT signaling pathway; *IFNGR1* encodes the receptor for interferon-γ, which activates the JAK–STAT pathway; *IL1B* encodes the pro-inflammatory cytokine IL-1β, which activates the NF-κB pathway. Expression of all three genes was elevated in patients with cystic fibrosis, and the data demonstrated associations between allelic variants of *STAT3*, *IL1B*, and *IFNGR1* (determining their transcription levels) and disease severity [80].

A common intronic mutation in the *CFTR* gene is the c.3718-2477C>T variant, which is one of the most frequent mutations in the Polish cystic fibrosis patient population [81]. Patients carrying this mutation often exhibit a mild disease phenotype. It has been found that disease severity inversely correlates with a specific type of splicing transcript that facilitates the restoration of protein function. Studies have shown that increased expression of the splicing factors HTRA2-β1 and SC35 in the presence of the c.3718-2477C>T mutation promotes correct splicing of *CFTR* pre-mRNA, highlighting the role of splicing regulation as a significant modifier of cystic fibrosis clinical progression in the context of intronic mutations [82].

Thus, transcriptomic data can be used to identify genetic modifiers associated with specific genetic disorders.

GENE TRANSCRIPTION AS A BIOMARKER OF GENETIC DISEASES: ACTUAL CHALENGES

Not all LOF and GOF mutations lead to altered transcription level. As noted above, the NMD pathway eliminates mRNA transcripts carrying premature stop codons. However, transcripts containing nonsense mutations located within the last 50-55 nucleotides of the penultimate exon or within the final exon may be able to avoid NMD action [83]. For instance, in the *SOX10* gene, which encodes a transcription factor involved in neural crest development, some nonsense mutations arise outside the regions typically trigger NMD. As a result, these mutant transcripts escape degradation and produce truncated proteins with dominant-negative activity, leading to severe neurological disease [84]. Conversely, nonsense mutations in *SOX10* that occur within NMD-targeted regions result in recognition and degradation of the mutant transcripts, causing a milder phenotype due to haploinsufficiency [84, 85]. Thus, while reduced transcript levels may indicate gene inactivation in the case of LOF mutations, a presence of a LOF mutation by itself does not necessarily lead to decreased transcription.

Moreover, the efficiency of the NMD pathway may vary between different cell types. This has been demonstrated in Schmid metaphyseal chondrodysplasia, which is linked to a nonsense mutation in the *COL10A1* collagen gene [86]. In patients with this condition, the mutant mRNA is effectively degraded by the NMD mechanism in chondrocytes, but is poorly degraded (or not degraded at all) in lymphoblasts and osteoblasts. However, in this particular case, the issue about the cell type-specific differences in NMD efficiency remains to be elucidated [86].

More than 20% of LOF variants have been demonstrated to be located in exons that are frequently skipped during splicing and, therefore, do not provide a mutant phenotype [87]. In monogenic cardiomyopathies caused by LOF variants in the titin gene (*TTN*), transcript-level analysis revealed that nonsense mutation variants are predominantly found in exons that are absent in the most highly expressed alternative transcripts. Consequently, these variants do not produce the deleterious phenotypic effects typically associated with nonsense mutations [88].

Blood cell transcription is a convenient but not obligatory informative diagnostic marker. Numerous studies have demonstrated correlations between the expression of marker genes in blood cells of patients and both the presence of tumors and disease severity. For instance, blood cell transcriptome profiling has been used for the early diagnosis of colorectal cancer [89], resulting in the development of a targeted expression panel based on the transcription of 29 genes. Such an advance proved to be valuable for testing asymptomatic cases and predicting disease severity. Another research, aimed at the blood transcriptome of patients with metastatic renal cell carcinoma (when some patients are characterized by absence of immune response to checkpoint inhibitors), provide to identify a minimal gene set of 14 transcripts that changed in response to treatment. A gene expression panel was proposed that can accurately classify responders to therapy [90]. Similarly, transcriptional biomarkers based on the analysis of blood cell transcriptomes have also been suggested for lung cancer diagnosis [91].

In addition to cancer, various inherited diseases can be diagnosed through the analysis of transcriptomic alterations in blood cells. For example, transcriptional biomarkers for Parkinson's disease have been proposed based on gene expression data from blood, with 29 candidate genes identified for diagnostic purposes at the transcriptional level [92].

However, not all genetic diseases can be diagnosed solely based on gene transcription profiles in blood cells. For instance, a study performing RNA sequencing of whole blood and skin fibroblasts from 115 patients with various phenotypes but no established genetic diagnosis found that only 17% of patients demonstrated a unique transcriptional profile of a specific gene set associated with a particular disease [93]. Comparative analysis of transcriptomes from the two tissues, blood and fibroblasts, showed that fibroblasts produced higher and more consistent expression of disease-associated genes, while only genes related to immunodeficiency conditions exhibited higher expression in blood compared to fibroblasts [93].

Moreover, the blood transcriptome cannot be used to study tissue-specific diseases. For example, most genes whose regulation is typically disrupted in muscle disorders are weakly expressed in blood, suggesting that RNA-seq from blood cells may be insufficient to detect relevant transcriptional changes in muscle-specific genes [87].

It should also be noted that whole blood is not a representative material for studying sex-related differences, since blood cells contain only 12.9% of all sex-associated transcripts [94].

Therefore, while transcriptomic analysis of blood can aid in the diagnosis of certain conditions, it is not sufficient to serve as a universal diagnostic approach.

Gene transcription levels depend on the environment. As described above, gene transcription is strongly influenced by environmental factors; here, we highlight only a few of them.

Thousands of genes show age-related transcriptional changes [95]. However, the extent of these changes and the existence of transcriptional programs controlling aging remain unresolved issues [96].

Sex is another important factor determining the level and tissue specificity of gene transcription. Significant sex-associated differences in gene expression have been revealed, with genes showing sex-differential expression involved in various biological processes, such as drug and hormone response, embryonic development, tissue morphogenesis, fertilization, sexual reproduction, lipid metabolism, and immune response [94].

Humans, like many other organisms, exhibit temporal rhythms in gene expression (circadian rhythms) that regulate daily physiological cycles. The genetic regulation of circadian rhythms is generally conserved across all living organisms [97]. Circadian rhythms of mRNA transcription are followed by the combined action of an autonomous circadian oscillator, system signals, and other temporal signaling, such as feeding and fasting cycles [98]. Although it is commonly believed that about 10% of genes exhibit cyclicity at the protein production level, nearly 50% of genes expressed in the liver are characterized by cyclic mRNA levels [99], therefore, a considerable part of the rhythmic proteome is assumed to be regulated at the translational or post-translational level [100]. A recent study investigated sex- and age-dependent 24-hour rhythms of gene transcription across 46 tissues and identified two waves of expression – morning and evening [101]. These waves are regulated by factors related to the biological clock, immunity, carbohydrate metabolism, and cell proliferation.

It has been shown that changes in gene transcription can be triggered not only by mutations but also by external stimuli. For example, the increased transcription of the receptor tyrosine kinase gene *RET* is induced not only by the action of the glial cell line-derived neurotrophic factor (GDNF) and the GPI-anchored co-receptor GFRα1, but also depends on the concentration of interleukin IL-8 in the blood [102]. Increased transcription of the nerve growth factor receptor *NGFR* gene can be caused both by mutations within the gene itself and by systemic inflammatory diseases, such as osteoarthritis, psoriasis, inflammatory and degenerative disorders of the central nervous system [103]. Increased expression of the *MTOR* gene, which encodes a serine/threonine protein kinase involved in the regulation of cellular metabolism, growth, and cell survival, can be induced by various inflammatory cytokines (for example, TNF-α) [104]. It has also been demonstrated that interferon γ can stimulate the upregulation of the *MAPK1* gene, encoding mitogen-activated protein kinase 1. Such effect may be due to inflammation that is in progress or with the administration of interferon-based immunomodulatory therapy during genetic testing [105].

The microbiome is an essential source of genetic modification that has a great impact on the host transcriptome. For example, mutations in the gene encoding mannose-binding lectin (*MBL*) correlate with more severe progression of cystic fibrosis in chronic *Pseudomonas aeruginosa* infection [106]. Screening of commensal bacterial strains from respiratory tract microbiomes in cystic fibrosis patients identified strains capable of reducing the severity of inflammatory responses induced by *P. aeruginosa* [107]. Transcriptomic analysis of a model system of mono- and co-infection with *P. aeruginosa* and *Streptococcus* revealed downregulation of several signaling pathways involved in inflammatory responses during co-infection; protective genes of *Streptococcus* were identified [107].

Many of the aforementioned factors influencing transcription may also affect disease penetrance [108]. Incomplete penetrance can disrupt the interpretation of gene transcription analyses similarly to the presence of allelic variants, since control groups may include individuals exhibiting transcriptional profiles characteristic of the mutant phenotype but lacking phenotypic expression for various reasons.

Thus, the environment can influence transcription; however, although gene-environment interactions are evident, proving them remains extremely challenging because comprehensive and systematic collection of data on interactions between the human transcriptome and environmental factors is currently practically impossible.

Issues of sample size and reliability of differential gene expression assessment. Transcriptomic studies are limited by patient sample size. For rare diseases, obtaining a sufficient number of samples is understandably difficult. Even for common diseases, achieving appropriate sample sizes remains problematic. Due to high costs, many early next-generation sequencing studies typically included no more than three replicates per sample [109-112]. More recent studies have shown that at least 12 biological replicates are needed to reliably detect most differentially expressed genes (DEGs) [113]. Comparative transcriptomic analyses require large numbers of replicates due to genomic and transcriptomic plasticity. Moreover, DEG analysis can be complicated by poor reproducibility of RNA sequencing experiments, which, in turn, is provided by not only biological but also technical factors.

An important challenge in RNA sequencing studies is the detection of genes with low transcription levels, which requires substantial sequencing depth. To identify rare transcripts or analyze differential expression at the isoform level, both sequencing depth and the number of replicates must be increased. Using three replicates allows identification of 20-40% of significantly differentially expressed genes, whereas detecting 85% of all differentially expressed genes,

including those with less than two-fold changes, requires more than 20 replicates [113].

The false discovery rate (FDR) is another critical factor in RNA sequencing experiments. It has been shown that the FDR threshold is approximately 2^{-r}, where r is the number of replicates, varying from 0.25 for two replicates to 0.007 for seven replicates [114].

However, the optimal number of replicates for each experiment may vary due to dependence on factors, such as dispersion, library size, and the biological conditions being compared. Obtaining reliable estimates of dispersion for each gene, which is essential for DEG analysis, is also complicated by the small sample sizes that is typical for most RNA sequencing experiments [115].

As described above, genes with narrow variability in transcription levels are more suitable as diagnostic markers, since their use contributes to better discrimination between the study and the control groups. Such genes are often housekeeping genes and other haploinsufficient genes. However, many housekeeping genes are expressed at low levels, meaning that changes in their transcription may fall below the resolution limit of RNA sequencing methods. Furthermore, a twofold decrease in gene transcription, which can result from the action of the NMD mechanism, may also go undetected in transcriptomic analyses.

Moreover, not all environmental conditions can be assessed through RNA sequencing. Unaccounted factors contribute to the variability in transcription levels, which complicates their interpretation. Identifying the range of transcriptional variability under normal conditions is essential for reliably detecting expression changes associated with pathological states. However, accomplishing this requires generating large volumes of new data, which remains challenging at present.

GENE TRANSCRIPTION AS A BIOMARKER OF GENETIC DISEASES: PROSPECTS

Over the past decade, transcriptomics has become a powerful tool for studying human diseases at the molecular level. Transcriptomic profiling facilitates the identification of DEGs that may serve as disease biomarkers or therapeutic targets, thereby advancing the development of personalized treatment approaches.

However, transcription remains a challenging stage of gene expression to interpret due to its dynamic nature and sensitivity to external factors. Accurate assessment of gene expression dynamics under normal conditions is crucial for analyzing transcriptional changes associated with pathological states. Currently, such information can be obtained, among other sources, from publicly available databases containing RNA sequencing data from various human tissues across different ages and sexes. The most valuable publicly accessible resource is the Genotype-Tissue Expression (GTEx) project [116], which provides transcriptomic data from 54 tissues collected from nearly 1000 individuals. Within these tissues, expression quantitative trait loci (eQTLs) have been identified, showing significant correlations with gene expression variation.

Since 2021, the Developmental Genotype-Tissue Expression (dGTEx) project has been initiated to create an analytical resource for studying gene expression regulatory mechanisms during ontogenesis, the genetic basis of pediatric diseases, and their progression with age (https://www.genome.gov/Funded-Programs-Projects/Developmental-Genotype-Tissue-Expression/). Samples were collected from 120 relatively healthy pediatric donors across three age groups. Although the sample size is currently limited, the value of these data is expected to grow as larger datasets become available.

In the future, with the accumulation of large-scale genomic and transcriptomic datasets, it will become possible to use transcription not only as a disease-associated marker but also to predict the presence of specific genomic mutations. Transcriptomic analysis serves as the initial step in such studies, enabling the identification of DEGs.

The choice of tissue for gene expression analysis is critical. As discussed above, blood is a convenient material for such analyses. DEGs identified in blood can be used as biomarkers for various genetic diseases, including cancer. Blood is considered to be the most favorable tissue for assessing environmental influences. A recent study analyzing transcriptomes of blood cells from over 3000 adults, combined with phenotypic data, such as medical history, medication use, lifestyle factors, and body mass index, demonstrated the outstanding potential of transcriptomic diagnostics [117]. However, it is crucial to consider that gene expression patterns in blood cells may not accurately represent those in other tissues.

For this reason, diagnostic assays implemented in clinical practice are designed to test gene expression in tumor tissue. For example, in early-stage (I or II) breast cancer patients whose tumors are hormone receptor-positive and HER2-negative by histological and immunohistochemical assessment, the 21-gene expression assay Oncotype DX is used on tumor specimens [118]. This gene expression analysis enables prediction of disease course, assessment of recurrence risk, and evaluation of whether chemotherapy will reduce that risk. A similar assay, MammaPrint, evaluates the expression of 70 genes associated with breast cancer [119].

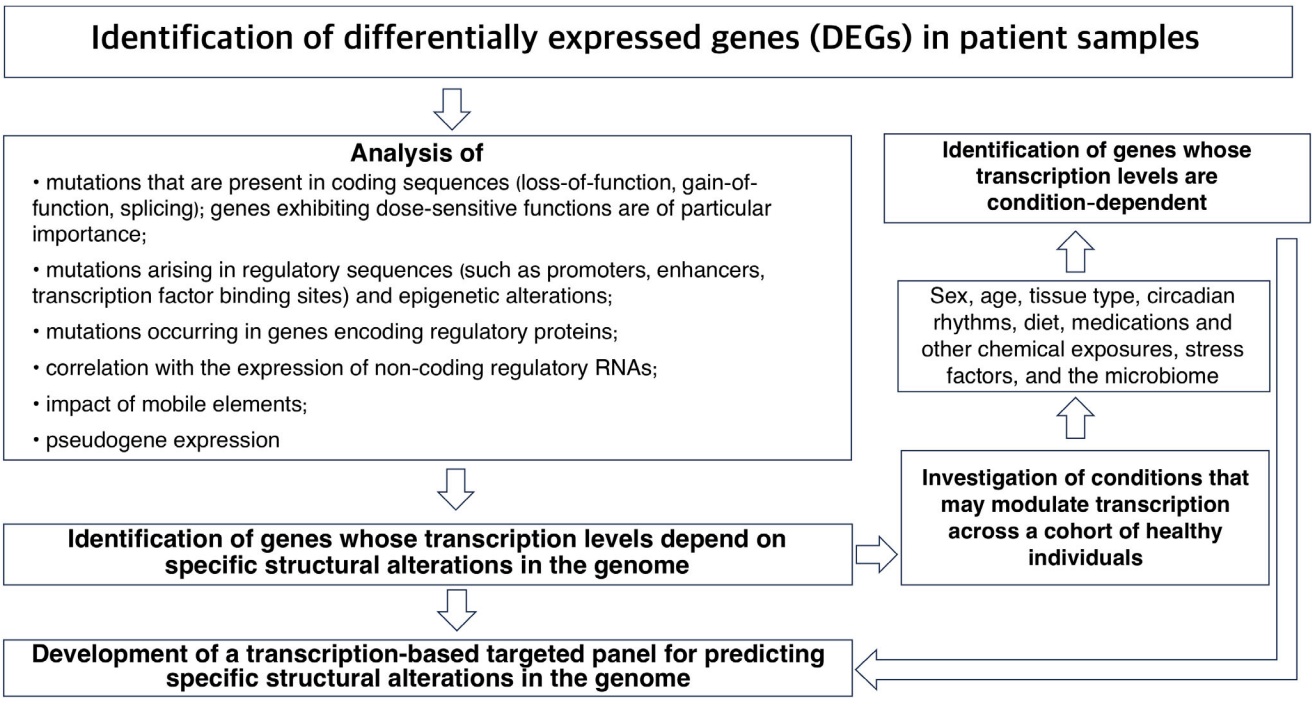

Fig. 2. Algorithm for the development of a targeted gene-expression panel for disease diagnosis and prediction of specific structural variants in the genome.

Developing a transcription-based diagnostic panel requires only the identification of genes whose expression is reproducibly altered under pathological conditions. However, to understand the underlying genetic mechanisms of the pathology, this is not sufficient. The next step must be a comprehensive investigation of the structural alterations that provide the observed expression changes. Figure 2 shows an algorithm for the development of a targeted gene-expression panel that can be used not only for diagnosis but also to predict specific structural variants in the genome.

Clearly, to elucidate the molecular-genetic mechanisms underlying pathology, transcriptional analysis must be complemented by structural genomic profiling. Ideally, the integration of GWAS and TWAS data will reveal correlations between genetic variants and expression changes, and will facilitate the identification of modifier genes that critically influence disease penetrance and may serve as novel therapeutic targets. Characterizing and analyzing the transcription of such modifier genes will advance our understanding of disease penetrance and broaden our insight into the architecture and dynamics of gene networks.

In summary, transcriptomic analysis is a powerful tool that can substantially optimize and enhance diagnostic workflows. It is clear that the accuracy of differential gene expression assessment will improve as sample sizes increase, RNA-sequencing studies expand, and computational algorithms for sequence data analysis advance. Further integration of genomic and transcriptomic datasets will enable the development not only of diagnostic but also of predictive and prognostic targeted panels, thereby facilitating novel, effective strategies for the treatment of genetic diseases.

Abbreviations. DEGs, differentially expressed genes; GOF, gain-of-function; HERVs, human endogenous retroviruses; LOF, loss-of-function; ME, mobile element; NMD, nonsense-mediated mRNA decay; SNP, single nucleotide polymorphism; TF, transcription factor.

Contributions. L.N.N. and T.N.K prepared the manuscript; L.N.N. edited the manuscript.

Funding. This work was supported by the MSU Development Program (project no. 23-SH-04-34).

Ethics approval and consent to participate. This work does not contain any studies involving human and animal subjects.

Conflict of interest. The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. Haendel, M., Vasilevsky, N., Unni, D., Bologa, C., Harris, N., Rehm, H., Hamosh, A., Baynam, G., Groza, T., McMurry, J., Dawkins, H., Rath, A., Thaxton, C.,

Bocci, G., Joachimiak, M. P., Köhler, S., Robinson, P. N., Mungall, C., and Oprea, T. I. (2020) How many rare diseases are there? *Nat. Rev. Drug Discov.*, **19**, 77-78, https://doi.org/10.1038/d41573-019-00180-y.

2. Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Hussain, M., Phillips, A. D., and Cooper, D. N. (2017) The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies, *Hum. Genet.*, **136**, 665-677, https://doi.org/10.1007/s00439-017-1779-6.
3. 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A., and Abecasis, G. R. (2015) A global reference for human genetic variation, *Nature*, **526**, 68-74, https://doi.org/10.1038/nature15393.
4. Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D. N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans, *Nature*, **536**, 285-291, https://doi.org/10.1038/nature19057.
5. Uffelmann, E., Huang, Q. Q., Munung, N. S., De Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., and Posthuma, D. (2021) Genome-wide association studies, *Nat. Rev. Methods Primers*, **1**, 59, https://doi.org/10.1038/s43586-021-00056-9.
6. Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics, *Nat. Rev. Genet.*, **10**, 57-63, https://doi.org/10.1038/nrg2484.
7. Pattee, J., Zhan, X., Xiao, G., and Pan, W. (2020) Integrating germline and somatic genetics to identify genes associated with lung cancer, *Genet. Epidemiol.*, **44**, 233-247, https://doi.org/10.1002/gepi.22275.
8. Høie, M. H., Cagiada, M., Beck Frederiksen, A. H., Stein, A., and Lindorff-Larsen, K. (2022) Predicting and interpreting large-scale mutagenesis data using analyses of protein stability and conservation, *Cell Rep.*, **38**, 110207, https://doi.org/10.1016/j.celrep.2021.110207.
9. Mort, M., Ivanov, D., Cooper, D. N., and Chuzhanova, N. A. (2008) A meta-analysis of nonsense mutations causing human genetic disease, *Hum. Mutat.*, **29**, 1037-1047, https://doi.org/10.1002/humu.20763.
10. Han, X., Wei, Y., Wang, H., Wang, F., Ju, Z., and Li, T. (2018) Nonsense-mediated mRNA decay: a 'nonsense' pathway makes sense in stem cell biology, *Nucleic Acids Res.*, **46**, 1038-1051, https://doi.org/10.1093/nar/gkx1272.
11. Sanderlin, E. J., Keenan, M. M., Mense, M., Revenko, A. S., Monia, B. P., Guo, S., and Huang, L. (2022) CFTR mRNAs with nonsense codons are degraded by the SMG6-mediated endonucleolytic decay pathway, *Nat. Commun.*, **13**, 2344, https://doi.org/10.1038/s41467-022-29935-9.
12. Spinner, N. B., Colliton, R. P., Crosnier, C., Krantz, I. D., Hadchouel, M., and Meunier-Rotival, M. (2001) Jagged mutations in Alagille syndrome, *Hum. Mutat.*, **17**, 18-33, https://doi.org/10.1002/1098-1004(2001)17:1<18::AID-HUMU3>3.0.CO,2-T.
13. Van Bon, B., Hoischen, A., Hehir-Kwa, J., De Brouwer, A., Ruivenkamp, C., Gijsbers, A. C., Marcelis, C. L., de Leeuw, N., Veltman, J. A., Brunner, H. G., and de Vries, B. B. (2011) Intragenic deletion in DYRK1A leads to mental retardation and primary microcephaly, *Clin. Genet.*, **79**, 296-299, https://doi.org/10.1111/j.1399-0004.2010.01544.x.
14. Brown, L. Y. (2001) Holoprosencephaly due to mutations in ZIC2: alanine tract expansion mutations may be caused by parental somatic recombination, *Hum. Mol. Genet.*, **10**, 791-796, https://doi.org/10.1093/hmg/10.8.791.
15. Lindeboom, R. G. H., Supek, F., and Lehner, B. (2016) The rules and impact of nonsense-mediated mRNA decay in human cancers, *Nat. Genet.*, **48**, 1112-1118, https://doi.org/10.1038/ng.3664.
16. Birchler, J. A., and Auger, D. L. (2013) Biological Consequences of Dosage Dependent Gene Regulation in Multicellular Eukaryotes, in *Madame Curie Bioscience Database*, Landes Bioscience.
17. Eisenberg, E., and Levanon, E. Y. (2013) Human housekeeping genes, revisited, *Trends Genet.*, **29**, 569-574, https://doi.org/10.1016/j.tig.2013.05.010.
18. Hundt, N., Preller, M., Swolski, O., Ang, A. M., Mannherz, H. G., Manstein, D. J., and Müller, M. (2014) Molecular mechanisms of disease-related human β-actin mutations p.R183W and p.E364K, *FEBS J.*, **281**, 5279-5291, https://doi.org/10.1111/febs.13068.
19. Li, Y., Zhang, Y., Li, X., Yi, S., and Xu, J. (2019) Gain-of-function mutations: an emerging advantage for cancer biology, *Trends Biochem. Sci.*, **44**, 659-674, https://doi.org/10.1016/j.tibs.2019.03.009.
20. Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., Reynolds, A. P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A. K., Canfield, T. K., Giste, E., Diegel, M., Bates, D., Hansen, R. S., Neph, S., Sabo, P. J., Heimfeld, S., Raubitschek, A., et al. (2012) Systematic localization of common disease-associated variation in regulatory DNA, *Science*, **337**, 1190-1195, https://doi.org/10.1126/science.1222794.
21. Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M. A., and Gerstein, M. (2016) Role of non-coding sequence variants in cancer, *Nat. Rev. Genet.*, **17**, 93-108, https://doi.org/10.1038/nrg.2015.17.

22. Fuxman Bass, J. I., Sahni, N., Shrestha, S., Garcia-Gonzalez, A., Mori, A., Bhat, N., Yi, S., Hill, D. E., Vidal, M., and Walhout, A. J. M. (2015) Human gene-centered transcription factor networks for enhancers and disease variants, *Cell*, **161**, 661-673, https://doi.org/10.1016/j.cell.2015.03.003.
23. Huang, F. W., Hodis, E., Xu, M. J., Kryukov, G. V., Chin, L., and Garraway, L. A. (2013) Highly recurrent TERT promoter mutations in human melanoma, *Science*, **339**, 957-959, https://doi.org/10.1126/science.1229259.
24. Zimmerman, O., Olbrich, P., Freeman, A. F., Rosen, L. B., Uzel, G., Zerbe, C. S., Rosenzweig, S. D., Kuehn, H. S., Holmes, K. L., Stephany, D., Ding, L., Sampaio, E. P., Hsu, A. P., and Holland, S. M. (2019) STAT1 gain-of-function mutations cause high total STAT1 levels with normal dephosphorylation, *Front. Immunol.*, **10**, 1433, https://doi.org/10.3389/fimmu.2019.01433.
25. Kvon, E. Z., Waymack, R., Gad, M., and Wunderlich, Z. (2021) Enhancer redundancy in development and disease, *Nat. Rev. Genet.*, **22**, 337, https://doi.org/10.1038/s41576-021-00361-9.
26. Sanyal, A., Lajoie, B. R., Jain, G., and Dekker, J. (2012) The long-range interaction landscape of gene promoters, *Nature*, **489**, 109-113, https://doi.org/10.1038/nature11279.
27. Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y., Yen, C. A., Schmitt, A. D., Espinoza, C. A., and Ren, B. (2013) A high-resolution map of the three-dimensional chromatin interactome in human cells, *Nature*, **503**, 290-294, https://doi.org/10.1038/nature12644.
28. Flores, M. A., and Ovcharenko, I. (2018) Enhancer reprogramming in mammalian genomes, *BMC Bioinformatics*, **19**, 316, https://doi.org/10.1186/s12859-018-2343-7.
29. Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H., and Young, R. A. (2013) Transcriptional super-enhancers connected to cell identity and disease, *Cell*, **155**, 934-947, https://doi.org/10.1016/j.cell.2013.09.053.
30. Mansour, M. R., Abraham, B. J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A. D., Etchin, J., Lawton, L., Sallan, S. E., Silverman, L. B., Loh, M. L., Hunger, S. P., Sanda, T., Young, R. A., and Look, A. T. (2014) Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element, *Science*, **346**, 1373-1377, https://doi.org/10.1126/science.1259037.
31. Lopes-Ramos, C. M., Barros, B. P., Koyama, F. C., Carpinetti, P. A., Pezuk, J., Doimo, N. T. S., Habr-Gama, A., Perez, R. O., and Parmigiani, R. B. (2017) E2F1 somatic mutation within miRNA target site impairs gene regulation in colorectal cancer, *PLoS One*, **12**, e0181153, https://doi.org/10.1371/journal.pone.0181153.
32. Bhattacharya, A., and Cui, Y. (2016) SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions, *Nucleic Acids Res.*, **44**, D1005-D1010, https://doi.org/10.1093/nar/gkv1220.
33. López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigó, R. (2005) Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett.*, **579**, 1900-1903, https://doi.org/10.1016/j.febslet.2005.02.047.
34. Soemedi, R., Cygan, K. J., Rhine, C. L., Wang, J., Bulacan, C., Yang, J., Bayrak-Toydemir, P., McDonald, J., and Fairbrother, W. G. (2017) Pathogenic variants that alter protein code often disrupt splicing, *Nat. Genet.*, **49**, 848-855, https://doi.org/10.1038/ng.3837.
35. De la Hoya, M., Soukarieh, O., López-Perolio, I., Vega, A., Walker, L. C., van Ierland, Y., Baralle, D., Santamariña, M., Lattimore, V., Wijnen, J., Whiley, P., Blanco, A., Raponi, M., Hauke, J., Wappenschmidt, B., Becker, A., Hansen, T. V., Behar, R., KConFaB Investigators, Niederacher, D., Arnold, N., Dworniczak, B., Steinemann, D., Faust, U., Rubinstein, W., Hulick, P. J., Houdayer, C., Caputo, S. M., Castera, L., et al. (2016) Combined genetic and splicing analysis of *BRCA1* c.[594-2A>C, 641A>G] highlights the relevance of naturally occurring in-frame transcripts for developing disease gene variant classification algorithms, *Hum. Mol. Genet.*, **25**, 2256-2268, https://doi.org/10.1093/hmg/ddw094.
36. Yeo, G., and Burge, C. B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals, *J. Comput. Biol.*, **11**, 377-394, https://doi.org/10.1089/1066527041410418.
37. Angarola, B. L., and Anczuków, O. (2021) Splicing alterations in healthy aging and disease, *Wiley Interdiscip. Rev. RNA*, **12**, e1643, https://doi.org/10.1002/wrna.1643.
38. Chathoth, K. T., Barrass, J. D., Webb, S., and Beggs, J. D. (2014) A splicing-dependent transcriptional checkpoint associated with prespliceosome formation, *Mol. Cell*, **53**, 779-790, https://doi.org/10.1016/j.molcel.2014.01.017.
39. Uriostegui-Arcos, M., Mick, S. T., Shi, Z., Rahman, R., and Fiszbein, A. (2023) Splicing activates transcription from weak promoters upstream of alternative exons, *Nat. Commun.*, **14**, 3435, https://doi.org/10.1038/s41467-023-39200-2.
40. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., et al. (2001) Initial sequencing and analysis of the human genome, *Nature*, **409**, 860-921, https://doi.org/10.1038/35057062.

41. Mills, R. E., Bennett, E. A., Iskow, R. C., and Devine, S. E. (2007) Which transposable elements are active in the human genome? *Trends Genet.*, **23**, 183-191, https://doi.org/10.1016/j.tig.2007.02.006.
42. Mao, J., Zhang, Q., and Cong, Y.-S. (2021) Human endogenous retroviruses in development and disease, *Comput. Struct. Biotechnol. J.*, **19**, 5978-5986, https://doi.org/10.1016/j.csbj.2021.10.037.
43. Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Köchert, K., Bouhlel, M. A., Richter, J., Soler, E., Stadhouders, R., Jöhrens, K., Wurster, K. D., Callen, D. F., Harte, M. F., Giefing, M., Barlow, R., Stein, H., Anagnostopoulos, I., Janz, M., Cockerill, P. N., Siebert, R., Dörken, B., Bonifer, C., and Mathas, S. (2010) Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma, *Nat. Med.*, **16**, 571-579, https://doi.org/10.1038/nm.2129.
44. Zhu, X., Fang, H., Gladysz, K., Barbour, J. A., and Wong, J. W. H. (2021) Overexpression of transposable elements is associated with immune evasion and poor outcome in colorectal cancer, *Eur. J. Cancer*, **157**, 94-107, https://doi.org/10.1016/j.ejca.2021.08.003.
45. McKerrow, W., Wang, X., Mendez-Dorantes, C., Mita, P., Cao, S., Grivainis, M., Ding, L., LaCava, J., Burns, K. H., Boeke, J. D., and Fenyö, D. (2022) LINE-1 expression in cancer correlates with P53 mutation, copy number alteration, and S phase checkpoint, *Proc. Natl. Acad. Sci. USA*, **119**, e2115999119, https://doi.org/10.1073/pnas.2115999119.
46. Rodić, N., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., Hruban, R. H., Iacobuzio-Donahue, C. A., Maitra, A., Torbenson, M. S., Goggins, M., Shih, Ie. M., Duffield, A. S., Montgomery, E. A., Gabrielson, E., Netto, G. J., Lotan, T. L., De Marzo, A. M., Westra, W., Binder, Z. A., Orr, B. A., Gallia, G. L., Eberhart, C. G., Boeke, J. D., Harris, C. R., and Burns, K. H. (2014) Long interspersed element-1 protein expression is a hallmark of many human cancers, *Am. J. Pathol.*, **184**, 1280-1286, https://doi.org/10.1016/j.ajpath.2014.01.007.
47. Pisanic, T. R., Asaka, S., Lin, S.-F., Yen, T.-T., Sun, H., Bahadirli-Talbott, A., Wang, T. H., Burns, K. H., Wang, T. L., and Shih, I. M. (2019) Long interspersed nuclear element 1 retrotransposons become deregulated during the development of ovarian cancer precursor lesions, *Am. J. Pathol.*, **189**, 513-520, https://doi.org/10.1016/j.ajpath.2018.11.005.
48. Sato, S., Gillette, M., de Santiago, P. R., Kuhn, E., Burgess, M., Doucette, K., Feng, Y., Mendez-Dorantes, C., Ippoliti, P. J., Hobday, S., Mitchell, M. A., Doberstein, K., Gysler, S. M., Hirsch, M. S., Schwartz, L., Birrer, M. J., Skates, S. J., Burns, K. H., Carr, S. A., and Drapkin, R. (2023) LINE-1 ORF1p as a candidate biomarker in high grade serous ovarian carcinoma, *Sci. Rep.*, **13**, 1537, https://doi.org/10.1038/s41598-023-28840-5.
49. Göke, J., and Ng, H. H. (2016) CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome, *EMBO Rep.*, **17**, 1131-1144, https://doi.org/10.15252/embr.201642743.
50. Leucci, E., Vendramin, R., Spinazzi, M., Laurette, P., Fiers, M., Wouters, J., Radaelli, E., Eyckerman, S., Leonelli, C., Vanderheyden, K., Rogiers, A., Hermans, E., Baatsen, P., Aerts, S., Amant, F., Van Aelst, S., van den Oord, J., de Strooper, B., Davidson, I., Lafontaine, D. L., Gevaert, K., Vandesompele, J., Mestdagh, P., and Marine, J. C. (2016) Melanoma addiction to the long non-coding RNA SAMMSON, *Nature*, **531**, 518-522, https://doi.org/10.1038/nature17161.
51. Zhan, H.-X., Wang, Y., Li, C., Xu, J.-W., Zhou, B., Zhu, J. K., Han, H. F., Wang, L., Wang, Y. S., and Hu, S. Y. (2016) LincRNA-ROR promotes invasion, metastasis and tumor growth in pancreatic cancer through activating ZEB1 pathway, *Cancer Lett.*, **374**, 261-271, https://doi.org/10.1016/j.canlet.2016.02.018.
52. Takahashi, K., Yan, I. K., Kogure, T., Haga, H., and Patel, T. (2014) Extracellular vesicle-mediated transfer of long non-coding RNA ROR modulates chemosensitivity in human hepatocellular cancer, *FEBS Open Bio*, **4**, 458-467, https://doi.org/10.1016/j.fob.2014.04.007.
53. Nevalainen, T., Autio, A., Mishra, B. H., Marttila, S., Jylhä, M., and Hurme, M. (2018) Aging-associated patterns in the expression of human endogenous retroviruses, *PLoS One*, **13**, e0207407, https://doi.org/10.1371/journal.pone.0207407.
54. Balestrieri, E., Pica, F., Matteucci, C., Zenobi, R., Sorrentino, R., Argaw-Denboba, A., Cipriani, C., Bucci, I., and Sinibaldi-Vallebona, P. (2015) Transcriptional activity of human endogenous retroviruses in human peripheral blood mononuclear cells, *Biomed Res. Int.*, **2015**, 164529, https://doi.org/10.1155/2015/164529.
55. Dumitrache, L. C., and McKinnon, P. J. (2022) Out of LINE: transposons, genome integrity, and neurodegeneration, *Neuron*, **110**, 3217-3219, https://doi.org/10.1016/j.neuron.2022.09.026.
56. Li, M., Radvanyi, L., Yin, B., Rycaj, K., Li, J., Chivukula, R., Lin, K., Lu, Y., Shen, J., Chang, D. Z., Li, D., Johanning, G. L., and Wang-Johanning, F. (2017) Downregulation of human endogenous retrovirus type K (HERV-K) viral Env RNA in pancreatic cancer cells decreases cell proliferation and tumor growth, *Clin. Cancer Res.*, **23**, 5892-5911, https://doi.org/10.1158/1078-0432.CCR-17-0001.
57. Balestrieri, E., Arpino, C., Matteucci, C., Sorrentino, R., Pica, F., Alessandrelli, R., Coniglio, A., Curatolo, P., Rezza, G., Macciardi, F., Garaci, E., Gaudi, S., and Sinibaldi-Vallebona, P. (2012) HERVs expression in autism spectrum disorders, *PLoS One*, **7**, e48831, https://doi.org/10.1371/journal.pone.0048831.

58. Cheetham, S. W., Faulkner, G. J., and Dinger, M. E. (2020) Overcoming challenges and dogmas to understand the functions of pseudogenes, *Nat. Rev. Genet.*, **21**, 191-201, https://doi.org/10.1038/s41576-019-0196-1.
59. Grandér, D., and Johnsson, P. (2016) Pseudogene-expressed RNAs: emerging roles in gene regulation and disease, *Curr. Top. Microbiol. Immunol.*, **394**, 111-126, https://doi.org/10.1007/82_2015_442.
60. Guo, X., Deng, L., Deng, K., Wang, H., Shan, T., Zhou, H., Liang, Z., Xia, J., and Li, C. (2016) Pseudogene PTENP1 suppresses gastric cancer progression by modulating PTEN, *Anticancer Agents Med. Chem.*, **16**, 456-464, https://doi.org/10.2174/1871520615666150507121407.
61. Travis, G., McGowan, E. M., Simpson, A. M., Marsh, D. J., and Nassif, N. T. (2023) PTEN, PTENP1, microRNAs, and ceRNA networks: precision targeting in cancer therapeutics, *Cancers*, **15**, 4954, https://doi.org/10.3390/cancers15204954.
62. Poliseno, L., Haimovic, A., Christos, P. J., Vega y Saenz de Miera, E. C., Shapiro, R., Pavlick, A., Berman, R. S., Darvishian, F., and Osman, I. (2011) Deletion of PTENP1 pseudogene in human melanoma, *J. Invest. Dermatol.*, **131**, 2497-2500, https://doi.org/10.1038/jid.2011.232.
63. An, Y., Furber, K. L., and Ji, S. (2017) Pseudogenes regulate parental gene expression via ceRNA network, *J. Cell Mol. Med.*, **21**, 185-192, https://doi.org/10.1111/jcmm.12952.
64. Kalyana-Sundaram, S., Kumar-Sinha, C., Shankar, S., Robinson, D. R., Wu, Y.-M., Cao, X., Asangani, I. A., Kothari, V., Prensner, J. R., Lonigro, R. J., Iyer, M. K., Barrette, T., Shanmugam, A., Dhanasekaran, S. M., Palanisamy, N., and Chinnaiyan, A. M. (2012) Expressed pseudogenes in the transcriptional landscape of human cancers, *Cell*, **149**, 1622-1634, https://doi.org/10.1016/j.cell.2012.04.041.
65. Mei, D., Song, H., Wang, K., Lou, Y., Sun, W., Liu, Z., Ding, X., and Guo, J. (2013) Up-regulation of SUMO1 pseudogene 3 (SUMO1P3) in gastric cancer and its clinical association, *Med. Oncol.*, **30**, 709, https://doi.org/10.1007/s12032-013-0709-2.
66. Lui, K. Y., Peng, H.-R., Lin, J.-R., Qiu, C.-H., Chen, H.-A., Fu, R.-D., Cai, C.-J., and Lu, M.-Q. (2016) Pseudogene integrator complex subunit 6 pseudogene 1 (INTS6P1) as a novel plasma-based biomarker for hepatocellular carcinoma screening, *Tumour Biol.*, **37**, 1253-1260, https://doi.org/10.1007/s13277-015-3899-8.
67. Zheng, X., Tang, H., Zhao, X., Sun, Y., Jiang, Y., and Liu, Y. (2017) Long non-coding RNA FTH1P3 facilitates uveal melanoma cell growth and invasion through miR-224-5p, *PLoS One*, **12**, e0184746, https://doi.org/10.1371/journal.pone.0184746.
68. Yang, W., Du, W. W., Li, X., Yee, A. J., and Yang, B. B. (2016) Foxo3 activity promoted by non-coding effects of circular RNA and Foxo3 pseudogene in the inhibition of tumor growth and angiogenesis, *Oncogene*, **35**, 3919-3931, https://doi.org/10.1038/onc.2015.460.
69. Costa, V., Esposito, R., Aprile, M., and Ciccodicola, A. (2012) Non-coding RNA and pseudogenes in neurodegenerative diseases: "The (Un)Usual Suspects", *Front. Gene*, **3**, 231, https://doi.org/10.3389/fgene.2012.00231.
70. Qi, Y., Wang, X., Li, W., Chen, D., Meng, H., and An, S. (2021) Pseudogenes in cardiovascular disease, *Front. Mol. Biosci.*, **7**, 622540, https://doi.org/10.3389/fmolb.2020.622540.
71. Chiefari, E., Iiritano, S., Paonessa, F., Le Pera, I., Arcidiacono, B., Filocamo, M., Foti, D., Liebhaber, S. A., and Brunetti, A. (2010) Pseudogene-mediated posttranscriptional silencing of HMGA1 can result in insulin resistance and type 2 diabetes, *Nat. Commun.*, **1**, 40, https://doi.org/10.1038/ncomms1040.
72. GTEx Consortium (2017) Genetic effects on gene expression across human tissues, *Nature*, **550**, 204-213, https://doi.org/10.1038/nature24277.
73. Glazier, A. A., Thompson, A., and Day, S. M. (2019) Allelic Imbalance and haploinsufficiency in MYBPC3-linked hypertrophic cardiomyopathy, *Pflug. Arch.*, **471**, 781-793, https://doi.org/10.1007/s00424-018-2226-9.
74. Hesson, L. B., Packham, D., Kwok, C.-T., Nunez, A. C., Ng, B., Schmidt, C., Fields, M., Wong, J. W., Sloane, M. A., and Ward, R. L. (2015) Lynch syndrome associated with two MLH1 promoter variants and allelic imbalance of MLH1 expression, *Hum. Mutat.*, **36**, 622-630, https://doi.org/10.1002/humu.22785.
75. Rouf, M. A., Wen, L., Mahendra, Y., Wang, J., Zhang, K., Liang, S., Wang, Y., Li, Z., Wang, Y., and Wang, G. (2023) The recent advances and future perspectives of genetic compensation studies in the zebrafish model, *Genes Dis.*, **10**, 468-479, https://doi.org/10.1016/j.gendis.2021.12.003.
76. Rossi, A., Kontarakis, Z., Gerri, C., Nolte, H., Hölper, S., Krüger, M., and Stainier, D. Y. R. (2015) Genetic compensation induced by deleterious mutations but not gene knockdowns, *Nature*, **524**, 230-233, https://doi.org/10.1038/nature14580.
77. Hagelkruys, A., Lagger, S., Krahmer, J., Leopoldi, A., Artaker, M., Pusch, O., Zezula, J., Weissmann, S., Xie, Y., Schöfer, C., Schlederer, M., Brosch, G., Matthias, P., Selfridge, J., Lassmann, H., Knoblich, J. A., and Seiser, C. (2014) A single allele of Hdac2 but not Hdac1 is sufficient for normal mouse brain development in the absence of its paralog, *Development*, **141**, 604-616, https://doi.org/10.1242/dev.100487.
78. Jurkin, J., Zupkovitz, G., Lagger, S., Grausenburger, R., Hagelkruys, A., Kenner, L., and Seiser, C. (2011) Distinct and redundant functions of histone deacetylases HDAC1 and HDAC2 in proliferation and tumorigenesis, *Cell Cycle*, **10**, 406-412, https://doi.org/10.4161/cc.10.3.14712.

79. Gallati, S. (2014) Disease-modifying genes and monogenic disorders: experience in cystic fibrosis, *Appl. Clin. Genet.*, **7**, 133-146, https://doi.org/10.2147/TACG.S18675.
80. Labenski, H., Hedtfeld, S., Becker, T., Tümmler, B., and Stanke, F. (2011) Initial interrogation, confirmation and fine mapping of modifying genes: *STAT3*, *IL1B* and *IFNGR1* determine cystic fibrosis disease manifestation, *Eur. J. Hum. Genet.*, **19**, 1281-1288, https://doi.org/10.1038/ejhg.2011.129.
81. Sobczyńska-Tomaszewska, A., Ołtarzewski, M., Czerska, K., Wertheim-Tysarowska, K., Sands, D., Walkowiak, J., Bal, J., and Mazurczak, T. (2013) NBS CF working group newborn screening for cystic fibrosis: polish 4 years' experience with CFTR sequencing strategy, *Eur. J. Hum. Genet.*, **21**, 391-396, https://doi.org/10.1038/ejhg.2012.180.
82. Nissim-Rafinia, M., Aviram, M., Randell, S. H., Shushi, L., Ozeri, E., Chiba-Falek, O., Eidelman, O., Pollard, H. B., Yankaskas, J. R., and Kerem, B. (2004) Restoration of the cystic fibrosis transmembrane conductance regulator function by splicing modulation, *EMBO Rep.*, **5**, 1071-1077, https://doi.org/10.1038/sj.embor.7400273.
83. Sun, B., and Chen, L. (2023) Mapping genetic variants for nonsense-mediated mRNA decay regulation across human tissues, *Genome Biol.*, **24**, 164, https://doi.org/10.1186/s13059-023-03004-w.
84. Inoue, K., Khajavi, M., Ohyama, T., Hirabayashi, S., Wilson, J., Reggin, J. D., Mancias, P., Butler, I. J., Wilkinson, M. F., Wegner, M., and Lupski, J. R. (2004) Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations, *Nat. Genet.*, **36**, 361-369, https://doi.org/10.1038/ng1322.
85. Miller, J. N., and Pearce, D. A. (2014) Nonsense-mediated decay in genetic disease: friend or foe? *Mutat. Res. Rev. Mutat. Res.*, **762**, 52-64, https://doi.org/10.1016/j.mrrev.2014.05.001.
86. Bateman, J. F. (2003) Tissue-specific RNA surveillance? Nonsense-mediated mRNA decay causes collagen X haploinsufficiency in Schmid metaphyseal chondrodysplasia cartilage, *Hum. Mol. Genet.*, **12**, 217-225, https://doi.org/10.1093/hmg/ddg054.
87. Cummings, B. B., Karczewski, K. J., Kosmicki, J. A., Seaby, E. G., Watts, N. A., Singer-Berk, M., Mudge, J. M., Karjalainen, J., Satterstrom, F. K., O'Donnell-Luria, A. H., Poterba, T., Seed, C., Solomonson, M., Alföldi, J., Genome Aggregation Database Production Team/Genome Aggregation Database Consortium, Daly, M. J., and MacArthur, D. G. (2020) Transcript expression-aware annotation improves rare variant interpretation, *Nature*, **581**, 452-458, https://doi.org/10.1038/s41586-020-2329-2.
88. Begay, R. L., Graw, S., Sinagra, G., Merlo, M., Slavov, D., Gowan, K., Jones, K. L., Barbati, G., Spezzacatene, A., Brun, F., Di Lenarda, A., Smith, J. E., Granzier, H. L., Mestroni, L., Taylor, M., and Familial Cardiomyopathy Registry (2015) Role of titin missense variants in dilated cardiomyopathy, *J. Am. Heart Assoc.*, **4**, e002645, https://doi.org/10.1161/JAHA.115.002645.
89. Ciarloni, L., Ehrensberger, S. H., Imaizumi, N., Monnier-Benoit, S., Nichita, C., Myung, S. J., Kim, J. S., Song, S. Y., Kim, T. I., van der Weg, B., Meier, R., Borovicka, J., Beglinger, C., Vallet, C., Maerten, P., Rüegg, C., and Dorta, G. (2016) Development and clinical validation of a blood test based on 29-gene expression for early detection of colorectal cancer, *Clin. Cancer Res.*, **22**, 4604-4611, https://doi.org/10.1158/1078-0432.CCR-15-2057.
90. Nagumo, Y., Kandori, S., Kojima, T., Hamada, K., Nitta, S., Chihara, I., Shiga, M., Negoro, H., Mathis, B. J., and Nishiyama, H. (2022) Whole-blood gene expression profiles correlate with response to immune checkpoint inhibitors in patients with metastatic renal cell carcinoma, *Cancers*, **14**, 6207, https://doi.org/10.3390/cancers14246207.
91. Urbarova, I., Skogholt, A. H., Sun, Y.-Q., Mai, X.-M., Grønberg, B. H., Sandanger, T. M., Sætrom, P., and Nøst, T. H. (2023) Increased expression of individual genes in whole blood is associated with late-stage lung cancer at and close to diagnosis, *Sci. Rep.*, **13**, 20760, https://doi.org/10.1038/s41598-023-48216-z.
92. Augustine, J., and Jereesh, A. S. (2022) Blood-based gene-expression biomarkers identification for the non-invasive diagnosis of Parkinson's disease using two-layer hybrid feature selection, *Gene*, **823**, 146366, https://doi.org/10.1016/j.gene.2022.146366.
93. Murdock, D. R., Dai, H., Burrage, L. C., Rosenfeld, J. A., Ketkar, S., Müller, M. F., Yépez, V. A., Gagneur, J., Liu, P., Chen, S., Jain, M., Zapata, G., Bacino, C. A., Chao, H. T., Moretti, P., Craigen, W. J., Hanchard, N. A., Undiagnosed Diseases Network, and Lee, B. (2021) Transcriptome-directed analysis for mendelian disease diagnosis overcomes limitations of conventional genomic testing, *J. Clin. Invest.*, **131**, e141500, https://doi.org/10.1172/JCI141500.
94. Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A. D. H., Cotter, D. J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., Castel, S. E., Mohammadi, P., Aguet, F., Zou, Y., Khramtsova, E. A., Skol, A. D., Garrido-Martín, D., Reverter, F., Brown, A., Evans, P., Gamazon, E. R., Payne, A., Bonazzola, R., Barbeira, A. N., Hamel, A. R., et al. (2020) The impact of sex on gene expression across human tissues, *Science*, **369**, eaba3066, https://doi.org/10.1126/science.aba3066.
95. Cellerino, A., and Ori, A. (2017) What have we learned on aging from omics studies? *Semin. Cell Dev. Biol.*, **70**, 177-189, https://doi.org/10.1016/j.semcdb.2017.06.012.

96. De Magalhães, J. P., Curado, J., and Church, G. M. (2009) Meta-analysis of age-related gene expression profiles identifies common signatures of aging, *Bioinformatics*, **25**, 875-881, https://doi.org/10.1093/bioinformatics/btp073.
97. Feng, D., and Lazar, M. A. (2012) Clocks, metabolism, and the epigenome, *Mol. Cell*, **47**, 158-167, https://doi.org/10.1016/j.molcel.2012.06.026.
98. Kornmann, B., Schaad, O., Bujard, H., Takahashi, J. S., and Schibler, U. (2007) System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock, *PLoS Biol.*, **5**, e34, https://doi.org/10.1371/journal.pbio.0050034.
99. Mauvoisin, D., Wang, J., Jouffe, C., Martin, E., Atger, F., Waridel, P., Quadroni, M., Gachon, F., and Naef, F. (2014) Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver, *Proc. Natl. Acad. Sci. USA*, **111**, 167-172, https://doi.org/10.1073/pnas.1314066111.
100. Deery, M. J., Maywood, E. S., Chesham, J. E., Sládek, M., Karp, N. A., Green, E. W., Charles, P. D., Reddy, A. B., Kyriacou, C. P., Lilley, K. S., and Hastings, M. H. (2009) Proteomic analysis reveals the role of synaptic vesicle cycling in sustaining the suprachiasmatic circadian clock, *Curr. Biol.*, **19**, 2031-2036, https://doi.org/10.1016/j.cub.2009.10.024.
101. Talamanca, L., Gobet, C., and Naef, F. (2023) Sex-dimorphic and age-dependent organization of 24-hour gene expression rhythms in humans, *Science*, **379**, 478-483, https://doi.org/10.1126/science.add0846.
102. Rusmini, M., Griseri, P., Matera, I., Pontarini, E., Ravazzolo, R., Mavilio, D., and Ceccherini, I. (2014) Expression variability and function of the RET gene in adult peripheral blood mononuclear cells, *J. Cell. Physiol.*, **229**, 2027-2037, https://doi.org/10.1002/jcp.24660.
103. Huang, J., Yu, Q., Zhou, Y., Chu, Y., Jiang, F., Zhu, X., Zhang, J., and Wang, Q. (2020) A positive feedback loop formed by NGFR and FOXP3 contributes to the resistance of non-small cell lung cancer to icotinib, *Transl. Cancer Res.*, **9**, 1044-1052, https://doi.org/10.21037/tcr.2019.12.60.
104. Marques-Ramos, A., and Cervantes, R. (2023) Expression of mTOR in normal and pathological conditions, *Mol. Cancer*, **22**, 112, https://doi.org/10.1186/s12943-023-01820-z.
105. Iannone, F. (2002) Increased expression of nerve growth factor (NGF) and high affinity NGF receptor (P140 TrkA) in human osteoarthritic chondrocytes, *Rheumatology*, **41**, 1413-1418, https://doi.org/10.1093/rheumatology/41.12.1413.
106. Winstanley, C., O'Brien, S., and Brockhurst, M. A. (2016) Pseudomonas aeruginosa evolutionary adaptation and diversification in cystic fibrosis chronic lung infections, *Trends Microbiol.*, **24**, 327-337, https://doi.org/10.1016/j.tim.2016.01.008.
107. Tony-Odigie, A., Wilke, L., Boutin, S., Dalpke, A. H., and Yi, B. (2022) Commensal bacteria in the cystic fibrosis airway microbiome reduce *P. aeruginosa* induced inflammation, *Front. Cell Infect. Microbiol.*, **12**, 824101, https://doi.org/10.3389/fcimb.2022.824101.
108. Kingdom, R., and Wright, C. F. (2022) Incomplete penetrance and variable expressivity: from clinical studies to population cohorts, *Front. Genet.*, **13**, 920390, https://doi.org/10.3389/fgene.2022.920390.
109. Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., Jaffrézic, F., and French StatOmique Consortium (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis, *Briefings Bioinform.*, **14**, 671-683, https://doi.org/10.1093/bib/bbs046.
110. Kvam, V. M., Liu, P., and Si, Y. A. (2012) Comparison of statistical methods for detecting differentially expressed genes from RNA-seq data, *Am. J. Bot.*, **99**, 248-256, https://doi.org/10.3732/ajb.1100340.
111. Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments, *BMC Bioinformatics*, **11**, 94, https://doi.org/10.1186/1471-2105-11-94.
112. Soneson, C., and Delorenzi, M. (2013) A comparison of methods for differential expression analysis of RNA-Seq data, *BMC Bioinform.*, **14**, 91, https://doi.org/10.1186/1471-2105-14-91.
113. Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M., and Barton, G. J. (2016) How many biological replicates are needed in an RNA-Seq experiment and which differential expression tool should you use? *RNA*, **22**, 839-851, https://doi.org/10.1261/rna.053959.115.
114. Lamarre, S., Frasse, P., Zouine, M., Labourdette, D., Sainderichin, E., Hu, G., Le Berre-Anton, V., Bouzayen, M., and Maza, E. (2018) Optimization of an RNA-Seq differential gene expression analysis depending on biological replicate number and library size, *Front. Plant Sci.*, **9**, 108, https://doi.org/10.3389/fpls.2018.00108.
115. Yoon, S., and Nam, D. (2017) Gene dispersion is the key determinant of the read count bias in differential expression analysis of RNA-Seq data, *BMC Genomics*, **18**, 408, https://doi.org/10.1186/s12864-017-3809-0.
116. GTEx Consortium (2013) The genotype-tissue expression (GTEx) project, *Nat. Genet.*, **45**, 580-585, https://doi.org/10.1038/ng.2653.
117. Schmidt, M., Hopp, L., Arakelyan, A., Kirsten, H., Engel, C., Wirkner, K., Krohn, K., Burkhardt, R., Thiery, J., Loeffler, M., Loeffler-Wirth, H., and

Binder, H. (2020) The human blood transcriptome in a large population cohort and its relation to aging and health, *Front. Big Data*, **3**, 548873, https://doi.org/10.3389/fdata.2020.548873.

118. Green, N., Al-Allak, A., and Fowler, C. (2019) Benefits of introduction of oncotype DX testing, *Ann. R. Coll Surg. Engl.*, **101**, 55-59, https://doi.org/10.1308/rcsann.2018.0173.
119. Knauer, M., Mook, S., Rutgers, E. J., Bender, R. A., Hauptmann, M., van de Vijver, M. J., Koornstra, R. H., Bueno-de-Mesquita, J. M., Linn, S. C., and van't Veer, L. J. (2010) The predictive value of the 70-gene signature for adjuvant chemotherapy in early breast cancer, *Breast Cancer Res. Treat.*, **120**, 655-661, https://doi.org/10.1007/s10549-010-0814-2.