Novel examples of NMD escape through alternative intronic polyadenylation

Maria Vlasenok[1,†], Antonina Kuznetsova[1,†], Dmitry A. Skvortsov[1,2], and Dmitri D. Pervouchine[1,*]

[1]Center for Molecular and Cellular Biology, Moscow 143026, Russia

[2]Moscow State University, Moscow 119991, Russia

[*]e-mail: pervouchine@gmail.com; tel/fax +7 (916) 220 69 99

[†]These authors contributed equally to the work

March 5, 2026

Abstract

The nonsense-mediated mRNA decay (NMD) surveillance system detects and selectively eliminates transcripts with premature stop codons. A stop codon is considered premature if it is followed by an exon-exon junction more than 50 nucleotides downstream. Pruning of the 3'-untranslated region containing such junctions through alternative polyadenylation may provide a mechanism of NMD escape. Here, we systematically examine a subclass of poi-

son exons that carry a premature stop codon for the presence of polyadenylation sites in the downstream intron. Using data from the GTEx consortium, we observed that poison exons are more often followed by an active polyadenylation site compared with cassette exons. We also identified tissue-specific switches between NMD-targeted and NMD-escape isoforms in several human genes, including the vaccinia-related kinase *VRK3*, nuclear transcription factor *NFX1*, Notch pathway regulator *TM2D3*, and RNA helicase *DDX31*. Blocking the cleavage and polyadenylation sites in these genes using antisense oligonucleotides in human cells led to a switch from NMD-escape to NMD-target isoform, accompanied by a decrease in gene expression levels. This study reveals that NMD escape via alternative polyadenylation is a widespread, yet currently overlooked post-transcriptional mechanism of gene expression regulation.

Keywords: NMD escape, splicing, polyadenylation, VRK3, NFX1.

Introduction

Alternative splicing (AS) and alternative polyadenylation (APA) are two major steps of post-transcriptional pre-mRNA processing that substantially alter the aminoacid sequence of the gene product [1, 2, 3]. While AS operates by selecting different exon combinations, APA involves the use of different polyadenylation sites (PAS) resulting not only in alterations of the 3'-untranslated regions (3'UTRs) but also in the generation of proteins with different C-termini. The coordination between AS and APA is important in many biological processes including development [4, 5], tumorigenesis [6], and functioning of the nervous system [7].

Interestingly, both AS and APA carry out important roles beyond merely diversifying the

proteome. APA significantly impacts mRNA stability, localization and translation through modulating the 3'UTR sequences, which contain *cis*-regulatory elements of microRNAs and RNA-binding proteins [8, 9]. AS is involved in the so-called unproductive splicing, a process in which a premature termination codon (PTC) is purposely introduced into the mRNA to trigger its degradation by the nonsense-mediated decay (NMD) pathway [10]. This way of regulation is particularly widespread among RNA-binding proteins [11]. Curiously, population genomics studies conclude that AS exerts even greater impact on NMD-induced changes in gene expression than on expanding protein diversity [12].

One of the current NMD models postulates that exon-exon junction complexes (EJCs) that are deposited on the mRNA during splicing are displaced after the first round of translation, while EJCs that remain bound outside of the reading frame serve as a signal that mRNA contains a PTC [13]. Yet, PTC-containing transcripts can evade degradation by NMD in a process called NMD escape [14, 15]. Several stress response pathways inhibit NMD by affecting the translation efficacy [16]. Inefficient translation termination at the PTC (readthrough), or reinitiation of translation downstream of the PTC, evicts downstream EJCs also leading to NMD escape, for instance in *ATRX* and *ATP7A* genes [17, 18]. Some RNA-binding proteins such as PABPC1, PTBP1 and hnRNPL are capable of coordinating with NMD factors to protect specific transcripts from degradation [19].

Interestingly, AS and APA can act synergistically to trigger NMD escape. If an exon containing a PTC (poison exon) is skipped, or if the sole intron downstream of a PTC is inefficiently spliced, the mRNA will no longer be subjected to NMD [14]. A similar outcome occurs when a PTC is followed by a PAS located within the downstream intron, because after cleavage and

polyadenylation, no EJCs remain downstream of the PTC, causing it to be recognized as a normal stop codon. Only a few examples of NMD escape through APA are currently known, including translation of the truncated form of the *Tau* mRNA in Alzheimer's disease (AD) patients [20], premature polyadenylation within the last post-PTC intron of the *HFE* gene [21], and activation of an intronic PAS downstream of a cryptic poison exon in the *CSF3R* gene [22].

NMD serves not only as an mRNA quality control system but also as a regulatory mechanism to maintain the abundances of physiological transcripts [23]. This prompted us to investigate whether NMD escape via APA might be a more widespread phenomenon than is currently believed. To address this question, we juxtaposed the catalogs of poison exons with the published PAS datasets and GTEx transcriptomic data to detect tissue-specific switches to the NMD isoform and simultaneous activation of an intronic PAS — a pattern that is expected for the proposed mechanism. We present statistical evidence for the fact that intronic polyadenylation happens more frequently after poison exons than after non-poison cassette exons, an observation implying that more cases of NMD escape through APA exist beyond the handful of currently known examples. An extension of this search paradigm towards cryptic poison exons and unannotated APA events further confirmed this tendency. As a proof of principle, we performed an experiment in human A549 and HeLa cells, in which antisense oligonucleotides (ASOs) targeting the intronic polyadenylation signal and cleavage site resulted in suppression of the APA isoform and upregulation of the NMD isoform in four human genes, for which NMD escape had not previously been reported.

Methods

Genome assembly, transcript annotation, and PAS

The GRCh38 assembly of the human genome and comprehensive CHESS transcript annotation v3.1.3 were downloaded from Genome Reference Consortium [24] and CHESS websites [25], respectively. The lists of PAS were obtained from the PolyASite 2.0 database [26] and from a study of short reads containing non-templated adenines (polyA reads) [27]. In the latter, the genomic coordinates of PAS from GTEx were lifted over to the GRCh38 assembly.

RNA-seq data

The RNA-seq data from 9,423 samples of the GTEx project v8 were obtained from the dbGaP portal and aligned to the human genome assembly GRCh38 (hg38) using STAR aligner v2.7.8a in paired-end mode [28]. The results of NMD inactivation experiments in human 293T cells, iPS-bronchioalveolar organoid, HeLa and A549 cells followed by RNA-seq were obtained from the Gene Expression Omnibus under the accession numbers GSE126520, GSE217099, GSE125086 and GSE270310, respectively. The data were downloaded in FASTQ format and aligned to the GRCh38 (hg38) assembly using STAR aligner v2.7.8a in paired-end mode [28]. Sequencing results of transcripts bound to EJCs in HEK293 cells (RIPiT protocol [29]) together with the corresponding control RNA-seq datasets were obtained from GEO under the accession number GSE115788 and processed in the same way.

Poison and cassette exons

Cassette exons were defined as exons that are included in a group of annotated protein-coding or NMD transcripts and skipped in another such group. Exons that share an upstream or downstream splice junctions with another non-terminal exon were not considered. Each cassette exon was characterized by the uniquely defined upstream, downstream and skipping splice junctions (I_1, I_2, and E, respectively, in Figure 1A). Annotated poison exons were defined as cassette exons that contain a stop codon 50 or more nts upstream of exon's 3'-end (the 50-nt rule). Out of 12,528 cassette exons, 1,388 contained a stop codon, and of those 813 obeyed the 50-nt rule. To define alternative terminal exons, we selected the annotated terminal exons that reside in an intron and carry a stop codon, excluding cases when the terminal exon shared the I_1 splice junction with another non-terminal exon ($n = 3,249$).

Read coverage and gene expression levels

The read coverage values (in BPMs, bins per million mapped reads) were calculated for each GTEx sample using the bamCoverage utility [30] considering only uniquely mapped reads. The read coverage values were averaged across samples within each detailed tissue (SMTSD) using the wiggletools utility [31]. These values were used for visualization using the Gviz package [32]. The mean read coverage was computed for 150-nt windows at the end and at the beginning of the upstream and the downstream constitutive exons flanking each cassette exon, respectively (referred to as e_1 and e_2). If the upstream or downstream exon was shorter than the window of 150 nts, the mean was computed over the exon's genomic coordinates. The read counts corresponding to gene expression values were computed for each GTEX sample based on the primary

GENCODE v43 annotation downloaded from the EMBL-EBI portal with featureCounts utility [33]. The read counts were normalized to TPM, and the resulting expression levels were averaged between samples within each detailed tissue type (SMTSD). In what follows, the term 'tissue' refers to SMTSD.

Predicted poison exons

To predict poison exons residing inside alternative terminal exons, we extracted split reads from GTEx data using the IPSA pipeline [34]. Overall, 1,030,425 splice junctions supported by at least 10 split reads were detected, of which 332,730 were fully annotated and 402,248 had one annotated and one unannotated splice site. The latter were intersected with the list of alternative terminal exons ($n = 3,249$) in order to select cases when (1) the acceptor splice site of the splice junction coincides with the end of the intron that contains the alternative terminal exon, and (2) the donor splice site of the splice junction is located >50 nts downstream of the annotated stop codon either within the exon or <550 nts downstream of the stop codon. The threshold of 550 nts was selected as the 95th percentile of the 3'UTR length distribution for annotated poison exons. Finally, we selected splice junctions that were collectively supported by at least 100 split reads in GTEx, which corresponds to the 10th percentile of the split read support distribution of I_2 for annotated poison exons (Figure S1). This yielded 260 predicted poison exons that overlap alternative terminal exons.

Splicing and polyadenylation metrics

To quantify exon inclusion levels, the split read counts corresponding to splice junctions (I_1, I_2, and E in Figure 1A) were obtained using the IPSA pipeline [35] and processed as described previously [34]. For the GTEx data, the split read counts were pooled within each tissue. In NMD inactivation experiments, the split read counts were pooled across biological replicates. Each exon was characterized by the following three metrics: $\psi_1 = \frac{I_1}{I_1+E}$, $\psi_2 = \frac{I_2}{I_2+E}$, and $\varphi = \frac{I_1}{I_1+I_2}$. The values of these metrics with a denominator below 10 were discarded. The counts of continuous short reads overlapping the splice sites (r_1, r_2) were extracted by the IPSA pipeline with the default settings and analyzed using the same metrics, where I_1 was replaced by r_1, and I_2 was replaced by r_2 [35].

To quantify the level of polyadenylation at an intronic PAS, we introduced $\tau = \frac{e_2}{e_1}$, where e_1 and e_2 are the read coverage values in the upstream and downstream exons, respectively. This metric is sensitive to the value of the denominator, hence it was used only when $e_1 > 0.2$, which excludes ~40% of tissue-specific cases for all cassette exons and ~25% for poison exons followed by a PAS (Figure S1C). By construction, τ can take arbitrary non-negative values, while ψ_1, ψ_2, and φ are restricted to the interval from 0 to 1. Since the values of $\tau > 1$ originate from the 3'-end bias or reflect the existence of an unannotated transcript isoform, we chose to introduce an upper limit on τ values by redefining them as $\min\{\tau, 1\}$, that is, by setting $\tau = 1$ when $\tau > 1$. This cap not only allows for uniformity of the analysis but also helps to remove outliers that produce spurious tissue-specific APA events in a wrong τ range.

To assess splicing and polyadenylation changes, we used $\Delta\varphi = \varphi_{exp} - \varphi_{ctrl}$, $\Delta\tau = \tau_{exp} - \tau_{ctrl}$, and $\Delta\psi_i = \varphi_{i,exp} - \varphi_{i,ctrl}$, $i = 1, 2$, where 'exp' and 'ctrl' refer to EJC vs. RNA-seq experiments

for RIPiT, and CHX treatment vs. untreated cells for the NMD inactivation experiments.

Statistical analysis

The data were analyzed using R statistics software version 3.6.3. One-sided non-parametric tests were performed using normal approximation with continuity correction. For proportion comparisons, hypergeometric tests were used with Poisson or normal approximations when appropriate. The odds ratio (*OR*) values were calculated as the fraction of poison exons followed by a PAS divided by the fraction of other cassette exons followed by a PAS, across various PAS groups. Throughout the paper, *P* and *cor* denote the *P*-value and Spearman correlation coefficient, respectively. To identify significant negative correlations between τ and ψ_1 in tissues, the respective one-sided *P*-value was estimated using asymptotic *t*-approximation. Multiple hypothesis testing was performed using Benjamini-Hochberg (BH) correction. In all figures, significance levels of 5%, 1% and 0.1% are denoted by *, ** and ***, respectively; whiskers denote 95% confidence intervals.

Antisense oligonucleotides

ASO with 2'O methyl modified bases on a phosphorothioate backbone (2'OMe) were designed based on recommendations from [36]. Synthesis of ASO was carried out by Syntol JSC (Moscow, Russia). The sequences of used ASO are listed in Table S1.

Cell cultures, transfections and treatments

Human A549 lung adenocarcinoma cells were maintained in Dulbecco's modified Eagle's medium/Nutrient Mixture F-12 (PanEco). HeLa S3 cells were cultured in Nutrient Mixture F12 medium (PanEco). Before use, the medium was supplemented with with 10% v/v fetal bovine serum (FBS), 1% GlutaMAX, 0.05 mg/ml streptomycin and 50 units/ml penicillin (all products from Thermo Fisher Scientific). Cells were cultured at 37°C in 5% CO_2. Cells were plated at a density of 200,000 cells per well in a 12-well plate. ASO treatment was performed using Lipofectamine RNAiMAX (Invitrogen) in accordance with the manufacturer's protocol on 70–80% confluent cells. Cells were harvested after 24h of treatment. To inactivate NMD, cycloheximide (CHX) was added to the cells 3h before harvest giving a final concentration of 300 mg/ml in the growth medium. Experiments were made in at least five independent biological replicates.

RNA purification and cDNA synthesis

Total RNA was isolated by a guanidinium thiocyanate-phenol-chloroform method using ExtractRNA Reagent (Evrogene) in accordance with the manufacturer's protocol. One microgram of total RNA was first subjected to RNase-free DNase I digestion (Thermo Fisher Scientific) at 37°C for 30 min to remove contaminating genomic DNA. Next, 500 ng of total RNA were used for complementary DNA (cDNA) synthesis using Magnus First Strand cDNA Synthesis Kit (Evrogene) for reverse transcription-quantitative PCR (RT-qPCR) to a final volume of 20 µl. cDNA was diluted 1:5 with nuclease-free water for quantitative PCR (qPCR) and endpoint PCR.

Endpoint PCR and qPCR

qPCR reactions were performed in triplicates in a final volume of 12 μl in 96-well plates with 420 nM gene-specific primers and 2 μl of cDNA using 5XqPCRmix-HS SYBR reaction mix (Evrogen). Primers for qPCR are listed in Table S2. Primers for endpoint PCR are listed in Table S3. A sample without reverse transcriptase enzyme was included as a control to verify the absence of genomic DNA contamination. Amplification of the targets was carried out on CFX96 Real-Time System (Bio-Rad), with the following parameters: 95°C for 5 min, followed by 39 cycles at 95°C for 20 s, the annealing temperature (see below) for 20 s and 72°C for 20 s, ending at 72°C for 5 min. The annealing temperatures were 58°C for GAPDH and VRK3, 60°C for NFX1, DDX31 and TM2D3 (NMD-escape isoform), and 64°C for TM2D3 (NMD-target and skip isoforms). For each pair of primers in qPCR analysis, the primer efficiencies were estimated using a calibration curve. The expression of isoforms was calculated by the efficiency method with primers efficiency more than 90% in all cases.

Endpoint PCR was carried out under the following conditions: denaturing at 95°C for 5 min, 40 cycles of denaturing at 95°C for 20 s, annealing at 60°C for 20 s, and extension at 72°C for 1 min, followed by extension at 72°C for 5 min. The resulting products were analyzed on a 2% agarose gel stained with ethidium bromide and visualized using ChemiDoc XRS+ (Bio-Rad).

Results

PAS are more frequent downstream of poison exons

If cleavage and polyadenylation occur in the intron downstream of the PTC, the resulting transcript will lack EJCs after the stop codon. Consequently, such a PTC will be interpreted as a regular stop codon, allowing the transcript to escape NMD. We consider three scenarios for NMD-escape events driven by intronic polyadenylation. In Case I, both the poison exon and the downstream PAS are annotated (Figure 1A, left). There is a protein-coding isoform in which the poison exon is skipped (skip), and a NMD isoform that contains the poison exon (NMD target). The annotated PAS in the downstream intron introduces a third possibility, a truncated transcript, where the extended poison exon becomes an alternative terminal exon, and the PTC becomes a normal stop codon (NMD escape).

Considering that transcript catalogs are often incomplete, we also consider Case II, in which the poison exon is annotated but the intronic PAS is not (Figure 1A, middle), and Case III, in which the PAS is annotated but the poison exon is not (Figure 1A, right). To identify unannotated NMD escape isoforms in Case II, we combine evidence from PolyASite 2.0 [26] and polyA read support [27] to detect putative PAS in the intron downstream of the annotated poison exon. To identify unannotated poison exons in Case III, we rely on splice junction support derived from RNA-seq that are located more than 50 nts downstream of the stop codon and connect the annotated alternative terminal exon to the downstream constitutive exon. Such splice junctions suggest the existence of an unannotated poison exon with PTC that satisfies the 50-nt rule.

To assess the extent to which these scenarios occur, we assembled a catalog of annotated

poison exons followed by an annotated PAS (Case I, $n = 64$). Poison exons were defined as cassette exons containing a stop codon that satisfies the 50-nt rule with respect to the coding frame of the major transcript isoform (a broader definition also includes generation of a PTC by a frameshift, but it is not considered here). This catalog was extended by adding the annotated poison exons followed by an intronic PAS from the PolyASite2.0 database [26] or an intronic PAS inferred from short reads with non-templated adenines (polyA reads) in GTEx [27] (Case II, $n = 127$). Finally, to detect unannotated poison exons, we considered the annotated alternative terminal exons and used split reads from GTEx to identify novel splice junctions connecting them to the downstream constitutive exon (Case III, $n = 260$). In total, this yielded 451 candidate events of NMD escape via intronic polyadenylation (Supplementary Data File 1).

As a result, 64 out 814 (7.9%) poison exons in the initial list contained an annotated PAS in the downstream intron, while only 261 out of 11,827 (2.2%) non-poison cassette exons did so ($OR = 3.6$, $P \approx 10^{-14}$) (Figure 1B, Case I). In Case II, a similar test for intronic PAS with substantial level of support (≥50 polyA reads or ≥0.5 tpm for PAS from PolyASite2.0) showed that 191 out of 814 poison exons (23.5%) were associated with a PAS in the downstream intron (Figure 1B), while only 2,280 out of 11,827 non-poison cassette cassette exons (19.3%) did so ($OR = 1.2$, $P = 0.005$). To reduce the level of noise in polyA read data, we imposed a series of stricter thresholds on the PAS support level and found even greater prevalence of PAS after poison exons, especially for PAS preceded by the canonical polyadenylation signal (Figure 1C left, Figure S2A,B). The same trend was observed when strengthening the threshold for PAS from the PolyASite2.0 database (Figure 1C, middle). Overall, metrics computed from different data sources unambiguously suggest that transcription termination occurs more frequently after

poison exons than after non-poison cassette exons (Figure 1C, right).

If transcription terminates within the poison exon or in the downstream intron, the level of support by split reads for splice junctions at the 5'- and the 3'-end must be distributed differently (I_1 and I_2 in Figure 1A). To characterize this difference, we introduced the metric $\varphi = \frac{I_1}{I_1+I_2}$, which captures the imbalance between the 5'- and the 3'-end support levels. For cassette exons, one would expect $I_1 \approx I_2$, and consequently, $\varphi \approx 0.5$. In the extreme cases, the value $\varphi \approx 0$ indicates that $I_2 \gg I_1$ and $\varphi \approx 1$ indicates that $I_1 \gg I_2$, which is characteristic of terminal exons.

When applying this metric to human tissue transcriptomes, we found $\varphi \approx 0.5$ for the majority of non-poison cassette exons (Figure S2C) and poison exons without a PAS downstream (Figure 1D, left), with a larger peak at $\varphi = 0$ than at $\varphi = 1$. A notably different pattern emerged for poison exons followed by a PAS, where the distribution displayed a pronounced peak at $\varphi \approx 1$ indicating a lack of I_2 reads, especially for poison exons followed by high-confidence, annotated transcript ends (Figure 1D, right). Moreover, the prevalence of I_1 over I_2 increased with higher PAS support (Figure S2D). Not unexpectedly, for alternative terminal exons followed by an unannotated I_2 junction (Case III), we observed a stronger dominance of I_1 over I_2 (Figure 1E). A similar analysis using continuous reads overlapping splice sites instead of split reads revealed more frequent retention at the 3'-end of poison exons followed by a PAS, further supporting their transformation into alternative terminal exons (Figure S3).

In summary, these findings indicate that poison exons are followed by a PAS in the downstream intron more frequently than non-poison cassette exons and that they tend to be extended at the 3'-end, thereby converting to alternative terminal exons. The combined lists for cases I, II, and III contain 451 annotated or predicted poison exons followed by an annotated or predicted

PAS, representing potential NMD escape events.

Tissue-specific switch between NMD-target and NMD-escape isoforms

To identify genes that evade NMD through intronic polyadenylation in a tissue-specific manner, we analyzed human transcriptomes from the GTEx project, focusing on tissue-specific switches between the isoform with a skipped poison exon (skip), the isoform containing the poison exon (NMD target), and the truncated isoform in which the poison exon forms part of an alternative terminal exon (NMD escape).

A widely used metric to quantify exon inclusion is the percent-spliced-in (PSI, ψ) ratio. It combines the split read counts from both sides of the exon ($\psi = \frac{I_1+I_2}{I_1+I_2+2E}$) and, therefore, is not optimal for distinguishing between NMD-target and NMD-escape isoforms. Following the intron-centric approach [35], we introduced two separate metrics, $\psi_1 = \frac{I_1}{I_1+E}$ and $\psi_2 = \frac{I_2}{I_2+E}$, which estimate the inclusion rate at the 5'- and 3'-end of the exon separately (Figure 2A). Since the NMD-target and NMD-escape isoforms share an upstream splice junction, ψ_1 measures their combined inclusion ratio, while ψ_2 represents the inclusion of the poison exon alone. However, NMD-sensitive isoforms are rapidly degraded when the NMD system is active, as evidenced by the low ψ_2 values in tissues (Figure S4A). Consequently, the only observable signature of NMD escape is a switch from the skip isoform to the NMD-escape isoform, both of which are stable.

In addition to ψ_1, we introduced another metric, τ, which doesn't rely on split reads to distinguish between NMD-target and NMD-escape isoforms. It is defined as $\tau = \frac{e_2}{e_1}$, where e_1 and e_2 are the read coverage values in the upstream and the downstream constitutive exons, respectively (Figure 2A). This metric serves as a proxy for measuring the polyadenylation rate by the read

coverage drop, which reflects the fraction of transcripts that were not cleaved and polyadenylated within the intron. A value $\tau \approx 0$ indicates a decrease in read coverage, suggesting that intronic polyadenylation predominates. Conversely, $\tau \approx 1$ indicates that the read coverage remains unchanged, implying that polyadenylation does not occur and transcription continues.

In what follows, we use the list of 451 poison exons followed by a PAS to identify cases in which an exon is skipped on one tissue but included as a terminal exon of the NMD-escape isoform in other tissues. These switches are manifested by anticorrelated tissue-specific changes in ψ_1 and τ, where ψ_1 increases, indicating splicing at the exon's 5'-end, while τ decreases, indicating termination at the intronic PAS. This logic is illustrated in Figure 2B.

To select a threshold for defining tissue-specific exon inclusion, we examined the ψ_1 values for cassette exons not followed by a PAS and computed their pairwise differences ($\Delta\psi_1$) between tissues (Figure S4B). The 90th percentile of this distribution was $\Delta\psi_1 \simeq 0.2$, which, when used as a threshold, identified 193 poison exons with $\Delta\psi_1 > 0.2$ in at least one pair of tissues. Similarly, to define a threshold for tissue-specific APA, we examined the τ values for cassette exons followed by an annotated PAS and computed their pairwise differences ($\Delta\tau$) between tissues (Figure S4C). A natural threshold was the 95th percentile of this distribution ($\Delta\tau \simeq 0.25$), which identified 145 poison exons exhibiting an absolute change in τ greater than 0.25 in at least one tissue pair.

The intersection of these two exon sets ($\max \Delta\psi_1 > 0.2$ and $\max \Delta\tau > 0.25$) comprised 94 exons. Imposing the additional requirement of a negative correlation between ψ_1 and τ reduced this number to 50 exons, representing high-confidence predictions of NMD escape events (Figures 2C, S6, and S7). We retained this set for subsequent analyses because increasing $\Delta\tau$ beyond

0.25 primarily reduces the number of candidates with minimal change in the fraction that also satisfies the other criteria, while increasing $\Delta\psi_1$ does increase this fraction, but at the expense of substantially reducing the size of the selected set (Figure S5). The characteristics of the selected exons and their associated diseases, as reported in MalaCards [37], are summarized in Table 1.

Response of NMD escape events to NMD inactivation

To experimentally validate the proposed mechanism, we chose a strategy of inhibiting the production of the NMD-escape isoform using antisense oligonucleotides (ASO) that block the polyadenylation signal and cleavage site. In response, we expect a switch toward the production of the NMD-target isoform and, consequently, a reduction in gene expression level. However, the isoform containing the poison exon is detectable only when the NMD system is inactivated. Therefore, we reanalyzed existing RNA-seq experiments in which NMD was inhibited using the translation inhibitor cycloheximide (CHX).

We focused on candidates expressing both NMD-escape and NMD-target isoforms containing poison exons (Figure 2C). NMD-target isoforms were selected based on $\Delta\psi_2$ (the change of ψ_2 under CHX treatment), with larger $\Delta\psi_2$ indicating higher expression of the NMD-target isoform. Additionally, a decrease in the φ metric ($\Delta\varphi$) under the CHX treatment was required to capture an increase in the relative abundance of NMD-target isoforms among the combined pool of NMD-target and NMD-escape isoforms. High φ values or low τ values in control experiment without CHX were used as indicators that NMD-escape isoforms are sufficiently abundant when the NMD system is active (polyA$_{\text{ctrl}}$ in Figure 2C).

To select thresholds, we examined the set of cassette exons containing a stop codon followed

by a PAS, which included both poison exons and non-poison exons with stop codons that do not satisfy the 50-nt rule. The φ distribution showed a pronounced peak at $\varphi = 1$, accompanied by a notable increase at $\varphi = 0.5$ upon CHX treatment for poison exons (Figures S8A). The $\Delta\psi_2$ distribution was skewed towards positive values, indicating that the inclusion rate of NMD-target isoform also increases (Figure S8B). These observations confirm that the relative abundance of the NMD-target isoform, with respect to both NMD-escape and skip isoforms, increases under NMD inhibition.

However, CHX treatment inhibits NMD by blocking the translation elongation phase, thereby affecting other pathways. An alternative approach to detect NMD-sensitive transcripts is to pull down transcripts bound to EJCs using the RIPiT protocol, which isolates post-splicing, pre-translational mRNA-EJC particles [29]. To increase the power of detecting NMD-target transcripts, we applied the same approach to analyze the RIPiT dataset obtained from HEK293 cells [29]. Although the RIPiT dataset yielded fewer genes expressing both NMD-escape and NMD-target isoforms, the results showed substantial overlap between experimental approaches (Figure S8C). For example, 56 of 85 poison exons with higher inclusion in the EJC-bound fraction also displayed increased inclusion upon CHX treatment ($\Delta_{CHX}\psi_2$ and $\Delta_{RIPiT}\psi_2$ sets, $P < 10^{-14}$). These findings support the robustness of our observations across different experimental methods (Figure S8D).

By combining the lists obtained for the CHX and RIPiT protocols with the results of the tissue-specific analysis, we obtained 11 genes expressing both the NMD-escape and NMD-target isoforms in cell lines and exhibiting a tissue-specific switch between them (*NFX1*, *TM2D3*, *VRK3*, *TPM2*, *BDP1*, *PITPNM2*, *DDX31*, *NMT2*, *UGT1A6*, *THAP6*, and *CDK12*). We sub-

sequently focused on the *VRK3*, *NFX1*, *TM2D3*, and *DDX31* genes (extended correlation plots in Figure S9) as they were expressed at sufficiently high levels in the available cell lines (A549 and HeLa).

Polyadenylation inhibition promotes NMD isoform expression

VRK3 encodes vaccinia-related kinase 3, whose kinase activity plays a role in cell cycle regulation [38]. Additionally, *VRK3* plays a role in preventing neuronal apoptosis caused by prolonged oxidative stress, with CDK5 activating neuroprotective signaling via VRK3 phosphorylation [39].

Our analysis indicates that approximately half of the *VRK3* transcripts undergo NMD escape in cervix and esophagus, as evidenced by lower read coverage in the downstream constitutive exon, a read coverage drop at the PAS, and low I_2 values (Figure 3A). In contrast, in adrenal glands and kidneys, the exon-skipping isoform predominates ($\psi_1 = 0.22$ and $\psi_1 = 0.25$, respectively), with no drop in read coverage in the downstream exon. This model potentially explains the increased VRK3 production in kidneys [40]. We found that *VRK3* also harbors an upstream PAS, which is preferentially activated in the brain and could represent a distinct mode of NMD escape in neural tissues (Figure S10). This observation is consistent with VRK3 supporting neuronal survival under oxidative stress via CDK5-mediated phosphorylation [39].

To test the proposed mechanism in *VRK3*, we designed ASO sequences that block the polyadenylation signal and the cleavage site through complementary base pairing (Figure 3B). These ASOs, along with a mock ASO targeting an off-target site, were transfected into human adenocarcinoma A549 cells, with and without NMD inhibition by CHX. Endpoint RT-PCR showed that transfec-

tion with 25 nM of the ASO targeting the cleavage and polyadenylation site, either alone or in combination with a second ASO blocking the polyadenylation signal, efficiently promoted inclusion of the poison exon and suppressed production of the NMD-escape isoform. In contrast, transfection with the control ASO targeting an off-target site didn't induce the NMD-target isoform (Figure 3C). The observed effects were more significant in cells treated with CHX.

Quantitative PCR (RT-qPCR) using isoform-specific primers confirmed an upregulation of the NMD-target isoform and downregulation of the NMD-escape isoform following ASO treatment (Figure 3D). The treatment with ASO masking the PAS was more effective than with ASO targeting the polyadenylation signal. Measurement of *VRK3* mRNA levels revealed a reduction of the overall gene expression upon treatment with ASO mix in cells not exposed to CHX (Figure S11). This outcome is expected, as mRNA degradation occurs more efficiently when the NMD system is active, confirming that the fraction of *VRK3* transcripts degraded via the NMD pathway can be increased by inhibiting an intronic PAS.

The next target, *NFX1*, encodes a transcriptional repressor that binds the X-box motif in the promoter regions of major histocompatibility complex class II genes and the telomerase reverse transcriptase (hTERT) gene. It is expressed in multiple transcript isoforms with distinct functions in human epithelial cells. The truncated isoform results in NFX1-91 protein that represses hTERT transcription in keratinocytes, playing a role in epidermis function [41]. The long isoform NFX1-123, in contrast, resides in the cytoplasm and is able to increase hTERT levels post-transcriptionally via binding to 5'UTR of its mRNA [42]. NFX1-123 contributes to cellular immortalization and, unlike NFX1-91, is upregulated in cervical cancers [43]. Evidence suggests that *NFX1* may also be involved in cerebellar development regulation [44].

Our analysis identified a 279-nt cryptic poison exon located within the alternative terminal exon of *NFX1* (Figure 4A). In accordance with [44], expression of this exon was observed in the cerebellum, cerebellar hemisphere, and skeletal muscle, possibly due to lower NMD efficiency in these tissues [45]. This exon harbors two tandemly arranged acceptor splice sites separated by three nucleotides, but we treated them as a single site for simplicity. Our analysis suggests a tissue-specific switch from tissues such as the pancreas, where the NMD-target isoform is predominantly expressed, to tissues such as the skin, where both APA and overall gene expression increase, indicating NMD escape (Figure 4A).

Since an ASO that blocked only the cleavage site was sufficient to achieve the switch from the NMD-escape to the NMD-target isoform in the previous example, we designed a similar ASO for *NFX1* (Figure 4B, right). Transfection of HeLa cells with 25 nM of the ASO targeting the cleavage site promoted the inclusion of the poison exon and suppression of the NMD-escape isoform compared to transfection with the mock ASO or the untreated control, both with and without CHX treatment (Figure 4B, left). The upregulation of the NMD-target isoform and the downregulation of the NMD-escape isoform following ASO treatment were further confirmed by RT-qPCR (Figure 4D).

Next, we examined two additional Case III targets: the Notch signaling regulator *TM2D3* and the DEAD-box RNA helicase *DDX31*. TM2D3 associates with Notch1 at a region distinct from its primary ligand-binding domain, thereby enhancing its presentation on the cell surface [46]. *DDX31* encodes a nucleolar RNA helicase involved in a wide range of fundamental cellular processes and diseases, including cancer [47, 48, 49]. According to our analysis, both genes harbor cryptic poison exons located within alternative terminal exons (Figures S12A and S13A).

As observed previously, treatment of A549 cells with 25 nM or 50 nM ASO targeting the cleavage sites in these genes promoted poison exon inclusion and suppressed the NMD-escape isoform, in contrast to treatment with a mock ASO or no treatment (Figures S12B and S13B). Notably, the effect was dose dependent: higher ASO concentrations led to increased levels of the NMD-target isoform, most notably in the presence of CHX.

Interestingly, our predictions also encompass *TPM2*. This well-characterized gene encodes β-tropomyosin, which stabilizes actin filaments and regulates muscle contraction. It has been implicated in cardiac diseases, myopathies, and atherosclerosis [50, 51]. According to our findings, *TPM2* transcript isoforms that skip the poison exon predominate in the cervix, prostate, and most other tissues, whereas the NMD-escape isoform containing an alternative terminal exon is predominant in muscle, consistent with the high expression level of *TPM2* in this tissue (Figure S14A). Notably, in the heart we observe significant upregulation of the NMD-target isoform, which may help explain the reduced expression of *TPM2* in cardiac tissue [40].

Blockade of the polyadenylation signal and cleavage site in *TPM2* using ASOs did not yield the intended outcome: suppression of the NMD-escape isoform led to skipping of the poison exon, whereas the NMD-target isoform was undetectable in any of the tested cell lines (A549, Huh7, HeLa, and PC3; data not shown). However, we identified anecdotal evolutionary evidence supporting the proposed mechanism. The murine ortholog of *TPM2* contains an unannotated polyadenylation site and an unannotated poison exon at precisely aligning positions, suggesting that these elements have been evolutionarily conserved for regulatory purposes (Figure S14). However, in the absence of a specialized cardiac cell line, we chose not to investigate this gene further, despite its significant clinical relevance.

Discussion

Early estimates suggested that up to one-third of alternatively spliced human transcripts contain a PTC and are thus susceptible to degradation by the NMD pathway [52]. However, accumulating evidence also indicates that not all PTC-containing transcripts are degraded, and a substantial fraction can undergo NMD escape [53]. For instance, a study of human genotyping datasets found that 52.3% of highly common nonsense variants can escape NMD [54]. Another study demonstrated that variability in NMD efficiency can only partially be explained by the 50-nt rule, with up to 30% of the variance remaining unexplained [55], suggesting that NMD escape is considerably more widespread than currently appreciated [14].

For a PTC-containing transcript, the two major routes of NMD escape are translation reinitiation [56, 57, 58, 59] and stop codon readthrough [60, 61, 62]. Both mechanisms enable the ribosome to traverse the mRNA, displace all EJCs, and simultaneously produce a protein from the nonsense allele. In this work, we consider yet another mechanism that also allows protein production, in which APA removes the portion of the 3'UTR signaling the presence of a PTC, thereby converting it into a normal stop codon. To date, only three genes were suggested to undergo NMD escape via this mechanism [21, 22, 20].

Our study provides strong evidence that intronic polyadenylation is a frequent mechanism of NMD escape. Part of the statistical support arises from the simple observation that annotated transcript ends are more abundant in introns following poison exons than in introns following non-poison cassette exons. Additional support comes from two novel RNA-seq-derived metrics: φ, which captures the imbalance between split read coverage at the 5'- and 3'-splice junctions, and τ, which reflects the drop in read coverage caused by intronic polyadenylation. Together,

these metrics demonstrate a tendency for poison exons to be followed by an active intronic PAS (Cases I and II), suggesting that NMD escape via intronic polyadenylation may contribute to the unexplained variability in NMD efficiency. Moreover, alternative terminal exons may harbor cryptic poison exons (Case III), which remain unannotated due to NMD-mediated degradation, further expanding the repertoire of NMD escape events.

Applying stringent filtering criteria, we analyzed tissue-specific splicing and polyadenylation and identified 50 genes exhibiting tissue-specific NMD escape via intronic polyadenylation. However, this number likely represents only the tip of the iceberg, as none of the few genes previously known to undergo this mechanism were detected. The poison exon in the *HFE* gene [21] was missed because it was not annotated, and furthermore we did not observe splicing upregulation under CHX treatment when specifically examining its 3'UTR. The remaining two cases of NMD escape, in the *CSF3R* and *Tau* genes, did not pass our tissue-specific filters because they are restricted to CD34+ immune cells and Alzheimer's disease patients, respectively. Moreover, the active NMD system in tissues limits the observable response of poison exons to APA-mediated escape, representing a major challenge in quantifying unproductive splicing rates in unperturbed physiological conditions [63, 64].

Intronic polyadenylation generates transcripts with alternative terminal exons, which can be classified into two categories: skipped terminal exons, which may function as internal terminal exons or be excluded, and composite terminal exons, which arise from polyadenylation within a retained intron [65]. The pipeline presented in this study (Figure 1) focuses on poison exons, representing only one subtype of skipped terminal exons, whereas the diversity of NMD isoforms is considerably broader. Intronic polyadenylation occurring in the intron upstream of a poison

exon may generate a composite terminal exon, providing an additional route for NMD escape. Using the same methodology, it can also be shown that introns flanking poison exons at the 5' end tend to contain more active PAS than introns flanking non-poison cassette exons. However, exploring this or other APA scenarios falls beyond the scope of the present study.

Several studies have compared the efficacy of blocking the cleavage site versus the polyadenylation signal. While both approaches are generally effective, blocking the polyadenylation signal is often slightly more efficient [66, 67]. In contrast, for *VRK3*, masking the cleavage site resulted in stronger suppression. The effectiveness of cis-element blocking can be context-dependent, influenced by partial compensation from surrounding non-canonical signals, local pre-mRNA conformation, and ASO accessibility [68]. Our results demonstrate that, for the four genes tested, blocking the cleavage site can as well be effective.

In conclusion, this study expands current knowledge of NMD escape via intronic polyadenylation from three experimentally validated cases (*HFE*, *CSF3R*, and *Tau*) to seven, with the addition of *VRK3*, *NFX1*, *TM2D3*, and *DDX31*. Further application and refinement of the analytical framework presented here should enable the identification of many additional events, thereby contributing to a more comprehensive understanding of polyadenylation-mediated NMD escape and its role in gene expression regulation.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by Russian Science Foundation grant 22-14-00330.

Authors' contributions

DP and MV designed the study; MV performed data analysis; AK and DS performed the experiments; MV and DP wrote the first draft of the manuscript. All authors edited the final version of the manuscript.

Acknowledgments

The authors cordially thank Prof. Vera Rybko for introducing us to this research area.

References

[1] Ule, J. and Blencowe, B. J. (2019) Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell,* **76**, 329–345.

[2] Tian, B. and Manley, J. L. (2017) Alternative polyadenylation of mRNA precursors. *Nat Rev Mol Cell Biol,* **18**, 18–30.

[3] Zhang, X., Liu, F., and Zhou, Y. (2024) Coupling of alternative splicing and alternative polyadenylation. *Acta Biochim Biophys Sin (Shanghai),* **57**, 22–32.

[4] Agarwal, V., Lopez-Darwin, S., Kelley, D. R., and Shendure, J. (2021) The landscape of alternative polyadenylation in single cells of the developing mouse embryo. *Nat Commun,* **12**, 5101.

[5] Sun, Y., Pang, Y., Wu, X., Zhu, R., Wang, L., Tian, M., He, X., Liu, D., and Yang, X. (2024) Landscape of alternative splicing and polyadenylation during growth and development of muscles in pigs. *Commun Biol,* **7**, 1607.

[6] Xiang, Y., Ye, Y., Lou, Y., Yang, Y., Cai, C., Zhang, Z., Mills, T., Chen, N.-Y., Kim, Y., Muge Ozguc, F., Diao, L., Karmouty-Quintana, H., Xia, Y., Kellems, R. E., Chen, Z., Blackburn, M. R., Yoo, S.-H., Shyu, A.-B., Mills, G. B., and Han, L. (2018) Comprehensive Characterization of Alternative Polyadenylation in Human Cancer. *J Natl Cancer Inst,* **110**, 379–389.

[7] Lee, S., Aubee, J. I., and Lai, E. C. (2023) Regulation of alternative splicing and polyadenylation in neurons. *Life Sci Alliance,* **6**, e202302000.

[8] Turner, R. E., Pattison, A. D., and Beilharz, T. H. (2018) Alternative polyadenylation in the regulation and dysregulation of gene expression. *Semin Cell Dev Biol,* **75**, 61–69.

[9] Mittleman, B. E., Pott, S., Warland, S., Zeng, T., Mu, Z., Kaur, M., Gilad, Y., and Li, Y. (2020) Alternative polyadenylation mediates genetic regulation of gene expression. *Elife,* **9**, e57492.

[10] Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C., and Brenner, S. E. (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature,* **446**, 926–9.

[11] Müller-McNicoll, M., Rossbach, O., Hui, J., and Medenbach, J. (2019) Auto-regulatory feedback by RNA-binding proteins. *J Mol Cell Biol,* **11**, 930–939.

[12] Fair, B., Buen Abad Najar, C. F., Zhao, J., Lozano, S., Reilly, A., Mossian, G., Staley, J. P., Wang, J., and Li, Y. I. (2024) Global impact of unproductive splicing on human gene expression. *Nat Genet,* **56**, 1851–1861.

[13] Hug, N., Longman, D., and Cáceres, J. F. (2016) Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res,* **44**, 1483–95.

[14] Dyle, M. C., Kolakada, D., Cortazar, M. A., and Jagannathan, S. (2020) How to get away with nonsense: Mechanisms and consequences of escape from nonsense-mediated RNA decay. *Wiley Interdiscip Rev RNA,* **11**, e1560.

[15] Sherlock, M. E., Baquero Galvis, L., Vicens, Q., Kieft, J. S., and Jagannathan, S. (2023) Principles, mechanisms, and biological implications of translation termination-reinitiation. *RNA,* **29**, 865–884.

[16] Goetz, A. E. and Wilkinson, M. (2017) Stress and the nonsense-mediated RNA decay pathway. *Cell Mol Life Sci,* **74**, 3509–3531.

[17] Howard, M. T., Malik, N., Anderson, C. B., Voskuil, J. L. A., Atkins, J. F., and Gibbons, R. J. (2004) Attenuation of an amino-terminal premature stop codon mutation in the ATRX gene by an alternative mode of translational initiation. *J Med Genet,* **41**, 951–6.

[18] Paulsen, M., Lund, C., Akram, Z., Winther, J. R., Horn, N., and Møller, L. B. (2006) Evidence that translation reinitiation leads to a partially functional Menkes protein containing two copper-binding sites. *Am J Hum Genet,* **79**, 214–29.

[19] Kishor, A., Ge, Z., and Hogg, J. R. (2019) hnRNP L-dependent protection of normal mRNAs from NMD subverts quality control in B cell lymphoma. *EMBO J,* **38**, e99128.

[20] Ngian, Z.-K., Tan, Y.-Y., Choo, C.-T., Lin, W.-Q., Leow, C.-Y., Mah, S.-J., Lai, M. K.-P., Chen, C. L.-H., and Ong, C.-T. (2022) Truncated Tau caused by intron retention is enriched in Alzheimer's disease cortex and exhibits altered biochemical properties. *Proc Natl Acad Sci U S A,* **119**, e2204179119.

[21] Martins, R., Proença, D., Silva, B., Barbosa, C., Silva, A. L., Faustino, P., and Romão, L. (2012) Alternative polyadenylation and nonsense-mediated decay coordinately regulate the human HFE mRNA levels. *PLoS One,* **7**, e35461.

[22] Druhan, L. J., Lance, A., Hamilton, A., Steuerwald, N. M., Tjaden, E., and Avalos, B. R. (2020) Altered splicing and intronic polyadenylation of CSF3R via a cryptic exon in acute myeloid leukemia. *Leuk Res,* **92**, 106349.

[23] Zavileyskiy, L. G. and Pervouchine, D. D. (2024) Post-transcriptional Regulation of Gene Expression via Unproductive Splicing. *Acta Naturae,* **16**, 4–13.

[24] Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., Albracht, D., Fulton, R. S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K. M., Auger, K., Chow, W.,

Collins, J., Harden, G., Hubbard, T., Pelan, S., Simpson, J. T., Threadgold, G., Torrance, J., Wood, J. M., Clarke, L., Koren, S., Boitano, M., Peluso, P., Li, H., Chin, C.-S., Phillippy, A. M., Durbin, R., Wilson, R. K., Flicek, P., Eichler, E. E., and Church, D. M. (2017) Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res,* **27**, 849–864.

[25] Pertea, M., Shumate, A., Pertea, G., Varabyou, A., Breitwieser, F. P., Chang, Y.-C., Madugundu, A. K., Pandey, A., and Salzberg, S. L. (2018) CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol,* **19**, 208.

[26] Herrmann, C. J., Schmidt, R., Kanitz, A., Artimo, P., Gruber, A. J., and Zavolan, M. (2020) PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3' end sequencing. *Nucleic Acids Res,* **48**, D174–D179.

[27] Vlasenok, M., Margasyuk, S., and Pervouchine, D. D. (2023) Transcriptome sequencing suggests that pre-mRNA splicing counteracts widespread intronic cleavage and polyadenylation. *NAR Genom Bioinform,* **5**, lqad051.

[28] Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics,* **29**, 15–21.

[29] Kovalak, C., Donovan, S., Bicknell, A. A., Metkar, M., and Moore, M. J. (2021) Deep sequencing of pre-translational mRNPs reveals hidden flux through evolutionarily conserved alternative splicing nonsense-mediated decay pathways. *Genome Biol,* **22**, 132.

[30] Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., and Manke, T. (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res,* **42**, W187–91.

[31] Zerbino, D. R., Johnson, N., Juettemann, T., Wilder, S. P., and Flicek, P. (2014) WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics,* **30**, 1008–9.

[32] Hahne, F. and Ivanek, R. (2016) Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol,* **1418**, 335–51.

[33] Liao, Y., Smyth, G. K., and Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* **30**, 923–30.

[34] Margasyuk, S., Kuznetsova, A., Zavileyskiy, L., Vlasenok, M., Skvortsov, D., and Pervouchine, D. D. (2024) Human introns contain conserved tissue-specific cryptic poison exons. *NAR Genom Bioinform,* **6**, lqae163.

[35] Pervouchine, D. D., Knowles, D. G., and Guigó, R. (2013) Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics,* **29**, 273–4.

[36] Naveed, A., Cooper, J. A., Li, R., Hubbard, A., Chen, J., Liu, T., Wilton, S. D., Fletcher, S., and Fox, A. H. (2021) NEAT1 polyA-modulating antisense oligonucleotides reveal opposing functions for both long non-coding RNA isoforms in neuroblastoma. *Cell Mol Life Sci,* **78**, 2213–2230.

[37] Rappaport, N., Twik, M., Plaschkes, I., Nudel, R., Iny Stein, T., Levitt, J., Gershoni, M., Morrey, C. P., Safran, M., and Lancet, D. (2017) MalaCards: an amalgamated human

disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res,* **45**, D877–D887.

[38] Park, C.-H., Ryu, H. G., Kim, S.-H., Lee, D., Song, H., and Kim, K.-T. (2015) Presumed pseudokinase VRK3 functions as a BAF kinase. *Biochim Biophys Acta,* **1853**, 1738–48.

[39] Song, H., Kim, W., Choi, J.-H., Kim, S.-H., Lee, D., Park, C.-H., Kim, S., Kim, D.-Y., and Kim, K.-T. (2016) Stress-induced nuclear translocation of CDK5 suppresses neuronal death by downregulating ERK activation via VRK3 phosphorylation. *Sci Rep,* **6**, 28634.

[40] Consortium, G. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat Genet,* **45**, 580–5.

[41] Chintala, S. and Katzenellenbogen, R. A. (2021) NFX1, Its Isoforms and Roles in Biology, Disease and Cancer. *Biology (Basel),* **10**, 279.

[42] Katzenellenbogen, R. A., Vliet-Gregg, P., Xu, M., and Galloway, D. A. (2009) NFX1-123 increases hTERT expression and telomerase activity posttranscriptionally in human papillomavirus type 16 E6 keratinocytes. *J Virol,* **83**, 6446–56.

[43] Vliet-Gregg, P. A., Robinson, K. L., Levan, J., Matsumoto, L. R., and Katzenellenbogen, R. A. (2019) NFX1-123 is highly expressed in cervical cancer and increases growth and telomerase activity in HPV 16E6 expressing cells. *Cancer Lett,* **449**, 106–113.

[44] Ha, T. J., Zhang, P. G. Y., Robert, R., Yeung, J., Swanson, D. J., Mathelier, A., Wasserman, W. W., Im, S., Itoh, M., Kawaji, H., Lassmann, T., Daub, C. O., Arner, E., Consortium, F., Carninci, P., Hayashizaki, Y., Forrest, A. R. R., and Goldowitz, D. (2019) Identification

of novel cerebellar developmental transcriptional regulators with motif activity analysis. *BMC Genomics,* **20**, 718.

[45] Palou-Márquez, G. and Supek, F. (2025) Variable efficiency of nonsense-mediated mRNA decay across human tissues, tumors and individuals. *Genome Biol,* **26**, 316.

[46] Masuda, W., Yamakawa, T., Ajima, R., Miyake, K., Umemiya, T., Azuma, K., Tamaru, J.-I., Kiso, M., Das, P., Saga, Y., Matsuno, K., and Kitagawa, M. (2023) TM2D3, a mammalian homologue of Drosophila neurogenic gene product Almondex, regulates surface presentation of Notch receptors. *Sci Rep,* **13**, 20913.

[47] Ho, T. L. F., Lee, M. Y., Goh, H. C., Ng, G. Y. N., Lee, J. J. H., Kannan, S., Lim, Y. T., Zhao, T., Lim, E. K. H., Phua, C. Z. J., Lee, Y. F., Lim, R. Y. X., Ng, P. J. H., Yuan, J., Chan, D. K. H., Lieske, B., Chong, C. S., Lee, K. C., Lum, J., Cheong, W. K., Yeoh, K. G., Tan, K. K., Sobota, R. M., Verma, C. S., Lane, D. P., Tam, W. L., and Venkitaraman, A. R. (2023) Domain-specific p53 mutants activate EGFR by distinct mechanisms exposing tissue-independent therapeutic vulnerabilities. *Nat Commun,* **14**, 1726.

[48] Daizumoto, K., Yoshimaru, T., Matsushita, Y., Fukawa, T., Uehara, H., Ono, M., Komatsu, M., Kanayama, H.-O., and Katagiri, T. (2018) A DDX31/Mutant-p53/EGFR Axis Promotes Multistep Progression of Muscle-Invasive Bladder Cancer. *Cancer Res,* **78**, 2233–2247.

[49] Fukawa, T., Ono, M., Matsuo, T., Uehara, H., Miki, T., Nakamura, Y., Kanayama, H.-o., and Katagiri, T. (2012) DDX31 regulates the p53-HDM2 pathway and rRNA gene

transcription through its interaction with NPM1 in renal cell carcinomas. *Cancer Res,* **72**, 5867–77.

[50] Tajsharghi, H., Ohlsson, M., Palm, L., and Oldfors, A. (2012) Myopathies associated with β-tropomyosin mutations. *Neuromuscul Disord,* **22**, 923–33.

[51] Zhang, J., Zhang, C., Miao, L., Meng, Z., Gu, N., and Song, G. (2021) Abnormal TPM2 expression is involved in regulation of atherosclerosis progression via mediating RhoA signaling ¡i¿in vitro¡/i¿. *Arch Med Sci,* **20**, 1197–1208.

[52] Lewis, B. P., Green, R. E., and Brenner, S. E. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A,* **100**, 189–92.

[53] Klonowski, J., Liang, Q., Coban-Akdemir, Z., Lo, C., and Kostka, D. (2023) aenmd: Annotating escape from nonsense-mediated decay for transcripts with protein-truncating variants. *bioRxiv,* p. 2023.03.17.533185.

[54] Jagannathan, S. and Bradley, R. K. (2016) Translational plasticity facilitates the accumulation of nonsense genetic variants in the human population. *Genome Res,* **26**, 1639–1650.

[55] Lindeboom, R. G. H., Vermeulen, M., Lehner, B., and Supek, F. (2019) The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat Genet,* **51**, 1645–1651.

[56] Buisson, M., Anczuków, O., Zetoune, A. B., Ware, M. D., and Mazoyer, S. (2006) The 185delAG mutation (c.68_69delAG) in the BRCA1 gene triggers translation reinitiation at a downstream AUG codon. *Hum Mutat,* **27**, 1024–9.

[57] Hamid, R., Hedges, L. K., Austin, E., Phillips, J. A., Loyd, J. E., and Cogan, J. D. (2010) Transcripts from a novel BMPR2 termination mutation escape nonsense mediated decay by downstream translation re-initiation: implications for treating pulmonary hypertension. *Clin Genet,* **77**, 280–6.

[58] Hulsebos, T. J. M., Kenter, S., Verhagen, W. I. M., Baas, F., Flucke, U., and Wesseling, P. (2014) Premature termination of SMARCB1 translation may be followed by reinitiation in schwannomatosis-associated schwannomas, but results in absence of SMARCB1 expression in rhabdoid tumors. *Acta Neuropathol,* **128**, 439–48.

[59] Stump, M. R., Gong, Q., Packer, J. D., and Zhou, Z. (2012) Early LQT2 nonsense mutation generates N-terminally truncated hERG channels with altered gating properties by the reinitiation of translation. *J Mol Cell Cardiol,* **53**, 725–33.

[60] Loughran, G., Jungreis, I., Tzani, I., Power, M., Dmitriev, R. I., Ivanov, I. P., Kellis, M., and Atkins, J. F. (2018) Stop codon readthrough generates a C-terminally extended variant of the human vitamin D receptor with reduced calcitriol response. *J Biol Chem,* **293**, 4434–4444.

[61] Yamaguchi, Y. and Baba, H. (2018) Phylogenetically Conserved Sequences Around Myelin P0 Stop Codon are Essential for Translational Readthrough to Produce L-MPZ. *Neurochem Res,* **43**, 227–237.

[62] Yamaguchi, Y., Hayashi, A., Campagnoni, C. W., Kimura, A., Inuzuka, T., and Baba, H. (2012) L-MPZ, a novel isoform of myelin P0, is produced by stop codon readthrough. *J Biol Chem,* **287**, 17765–17776.

[63] Steward, C. A., Roovers, J., Suner, M.-M., Gonzalez, J. M., Uszczynska-Ratajczak, B., Pervouchine, D., Fitzgerald, S., Viola, M., Stamberger, H., Hamdan, F. F., Ceulemans, B., Leroy, P., Nava, C., Lepine, A., Tapanari, E., Keiller, D., Abbs, S., Sanchis-Juan, A., Grozeva, D., Rogers, A. S., Diekhans, M., Guigó, R., Petryszak, R., Minassian, B. A., Cavalleri, G., Vitsios, D., Petrovski, S., Harrow, J., Flicek, P., Lucy Raymond, F., Lench, N. J., Jonghe, P. D., Mudge, J. M., Weckhuysen, S., Sisodiya, S. M., and Frankish, A. (2019) Re-annotation of 191 developmental and epileptic encephalopathy-associated genes unmasks de novo variants in ¡i¿SCN1A¡/i¿. *NPJ Genom Med,* **4**, 31.

[64] Mironov, A., Petrova, M., Margasyuk, S., Vlasenok, M., Mironov, A. A., Skvortsov, D., and Pervouchine, D. D. (2023) Tissue-specific regulation of gene expression via unproductive splicing. *Nucleic Acids Res,* **51**, 3055–3066.

[65] Tian, B., Pan, Z., and Lee, J. Y. (2007) Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. *Genome Res,* **17**, 156–65.

[66] Vickers, T. A., Wyatt, J. R., Burckin, T., Bennett, C. F., and Freier, S. M. (2001) Fully modified 2' MOE oligonucleotides redirect polyadenylation. *Nucleic Acids Res,* **29**, 1293–9.

[67] Marsollier, A.-C., Ciszewski, L., Mariot, V., Popplewell, L., Voit, T., Dickson, G., and Dumonceaux, J. (2016) Antisense targeting of 3' end elements involved in DUX4 mRNA processing is an efficient therapeutic strategy for facioscapulohumeral dystrophy: a new gene-silencing approach. *Hum Mol Genet,* **25**, 1468–78.

[68] Ma, F. and Tian, B. (2026) Modulating the cleavage and polyadenylation site: from research tools to therapeutic opportunities. *RNA,* p. rna.080927.125.

[69] Girotto, G., Abdulhadi, K., Buniello, A., Vozzi, D., Licastro, D., d'Eustacchio, A., Vuckovic, D., Alkowari, M. K., Steel, K. P., Badii, R., and Gasparini, P. (2013) Linkage study and exome sequencing identify a BDP1 mutation associated with hereditary hearing loss. *PLoS One,* **8**, e80323.

[70] Eerola, I., Boon, L. M., Mulliken, J. B., Burrows, P. E., Dompmartin, A., Watanabe, S., Vanwijck, R., and Vikkula, M. (2003) Capillary malformation-arteriovenous malformation, a new clinical and genetic disorder caused by RASA1 mutations. *Am J Hum Genet,* **73**, 1240–9.

[71] Popp, M. W.-L. and Maquat, L. E. (2013) Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet,* **47**, 139–65.

[72] Quereda, V., Bayle, S., Vena, F., Frydman, S. M., Monastyrskyi, A., Roush, W. R., and Duckett, D. R. (2019) Therapeutic Targeting of CDK12/CDK13 in Triple-Negative Breast Cancer. *Cancer Cell,* **36**, 545–558.e7.

[73] Ekumi, K. M., Paculova, H., Lenasi, T., Pospichalova, V., Bösken, C. A., Rybarikova, J., Bryja, V., Geyer, M., Blazek, D., and Barboric, M. (2015) Ovarian carcinoma CDK12 mu-

tations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. *Nucleic Acids Res,* **43**, 2575–89.

[74] Zhang, H., Zhou, Y., Feng, Y., Hou, W., Chen, Y., Xing, Z., Zhang, Y., Wei, Q., Yin, Y., Guo, J., and Hu, H. (2024) Cyclin-dependent kinase 12 deficiency reprogrammes cellular metabolism to alleviate ferroptosis potential and promote the progression of castration-resistant prostate cancer. *Clin Transl Med,* **14**, e1678.

[75] Mun, J.-Y., Shu, C., Gao, Q., Zhu, Z., Akman, H. O., Westhoff, M.-A., Karpel-Massler, G., and Siegelin, M. D. (2025) CDK12 regulates cellular metabolism to promote glioblastoma growth. *JCI Insight,* **10**, e190780.

[76] Xie, Y., Liu, Y., Ding, J., Li, G., Ni, B., Pang, H., Hu, X., and Wu, L. (2022) Identification of ¡i¿DDX31¡/i¿ as a Potential Oncogene of Invasive Metastasis and Proliferation in PDAC. *Front Cell Dev Biol,* **10**, 762372.

[77] Shen, M. M., Rummey, C., and Lynch, D. R. (2024) Phenotypic variation of FXN compound heterozygotes in a Friedreich ataxia cohort. *Ann Clin Transl Neurol,* **11**, 1110–1121.

[78] Dürr, A. (2002) Friedreich's ataxia: treatment within reach. *Lancet Neurol,* **1**, 370–4.

[79] Borie, C., Gasparini, F., Verpillat, P., Bonnet, A.-M., Agid, Y., Hetet, G., Brice, A., Dürr, A., Grandchamp, B., and disease genetic study group, F. P. (2002) Association study between iron-related genes polymorphisms and Parkinson's disease. *J Neurol,* **249**, 801–4.

[80] Schmiesing, J., Lohmöller, B., Schweizer, M., Tidow, H., Gersting, S. W., Muntau, A. C., Braulke, T., and Mühlhausen, C. (2017) Disease-causing mutations affecting surface residues of mitochondrial glutaryl-CoA dehydrogenase impair stability, heteromeric complex formation and mitochondria architecture. *Hum Mol Genet,* **26**, 538–551.

[81] Heliö, K., Mäyränpää, M. I., Saarinen, I., Ahonen, S., Junnila, H., Tommiska, J., Weckström, S., Holmström, M., Toivonen, M., Nikus, K., Hathaway, J., Siivonen, P., Muona, M., Sistonen, J., Salmenperä, P., Gentile, M., Paananen, J., Myllykangas, S., Alastalo, T.-P., Heliö, T., and Koskenvuo, J. (2021) GRINL1A Complex Transcription Unit Containing GCOM1, MYZAP, and POLR2M Genes Associates with Fully Penetrant Recessive Dilated Cardiomyopathy. *Front Genet,* **12**, 786705.

[82] Silveira, L. F. G., MacColl, G. S., and Bouloux, P. M. G. (2002) Hypogonadotropic hypogonadism. *Semin Reprod Med,* **20**, 327–38.

[83] Francke, S., Clement, K., Dina, C., Inoue, H., Behn, P., Vatin, V., Basdevant, A., Guy-Grand, B., Permutt, M. A., Froguel, P., and Hager, J. (1997) Genetic studies of the leptin receptor gene in morbidly obese French Caucasian families. *Hum Genet,* **100**, 491–6.

[84] Mammès, O., Aubert, R., Betoulle, D., Péan, F., Herbeth, B., Visvikis, S., Siest, G., and Fumeron, F. (2001) LEPR gene polymorphisms: associations with overweight, fat mass and response to diet in women. *Eur J Clin Invest,* **31**, 398–404.

[85] de Luca, C., Kowalski, T. J., Zhang, Y., Elmquist, J. K., Lee, C., Kilimann, M. W., Ludwig, T., Liu, S.-M., and Chua, S. C. (2005) Complete rescue of obesity, diabetes, and infertility in db/db mice by neuron-specific LEPR-B transgenes. *J Clin Invest,* **115**, 3484–93.

[86] Zanoteli, E., Laporte, J., Rocha, J. C. C., Kretz, C., Oliveira, A. S. B., Mandel, J.-L., Perez, A. B. A., Gabbai, A. A., and Buj-Bello, A. (2005) Deletion of both MTM1 and MTMR1 genes in a boy with myotubular myopathy. *Am J Med Genet A,* **134**, 338–40.

[87] Buj-Bello, A., Furling, D., Tronchère, H., Laporte, J., Lerouge, T., Butler-Browne, G. S., and Mandel, J.-L. (2002) Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in DM1 muscle cells. *Hum Mol Genet,* **11**, 2297–307.

[88] Chintala, S., Dankoski, M. A., Maina, T. K., Oduor, C. I., Quist, K. M., Bailey, J. A., and Katzenellenbogen, R. A. (2025) Reduction of NFX1-123 and HPV 16 E6 and E7 Decreased Telomerase and CENP-F in Cervical Cancer Cell Lines. *Cancers (Basel),* **17**, 2044.

[89] Ding, X., Zhao, S., Zhang, Q., Yan, Z., Wang, Y., Wu, Y., Li, X., Liu, J., Niu, Y., Zhang, Y., Zhang, M., Wang, H., Zhang, Y., Chen, W., Yang, X.-Z., Liu, P., Posey, J. E., Lupski, J. R., Wu, Z., Yang, X., Wu, N., and Wang, K. (2020) Exome sequencing reveals a novel variant in ¡i¿NFX1¡/i¿ causing intracranial aneurysm in a Chinese family. *J Neurointerv Surg,* **12**, 221–226.

[90] Cantuti-Castelvetri, L., Ojha, R., Pedro, L. D., Djannatian, M., Franz, J., Kuivanen, S., van der Meer, F., Kallio, K., Kaya, T., Anastasina, M., Smura, T., Levanov, L., Szirovicza, L., Tobi, A., Kallio-Kokko, H., Österlund, P., Joensuu, M., Meunier, F. A., Butcher, S. J., Winkler, M. S., Mollenhauer, B., Helenius, A., Gokce, O., Teesalu, T., Hepojoki, J., Vapalahti, O., Stadelmann, C., Balistreri, G., and Simons, M. (2020) Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science,* **370**, 856–860.

[91] Booth, K. T., Azaiez, H., Kahrizi, K., Simpson, A. C., Tollefson, W. T. A., Sloan, C. M., Meyer, N. C., Babanejad, M., Ardalani, F., Arzhangi, S., Schnieders, M. J., Najmabadi, H., and Smith, R. J. H. (2015) PDZD7 and hearing loss: More than just a modifier. *Am J Med Genet A,* **167A**, 2957–65.

[92] Ebermann, I., Phillips, J. B., Liebau, M. C., Koenekoop, R. K., Schermer, B., Lopez, I., Schäfer, E., Roux, A.-F., Dafinger, C., Bernd, A., Zrenner, E., Claustres, M., Blanco, B., Nürnberg, G., Nürnberg, P., Ruland, R., Westerfield, M., Benzing, T., and Bolz, H. J. (2010) PDZD7 is a modifier of retinal disease and a contributor to digenic Usher syndrome. *J Clin Invest,* **120**, 1812–23.

[93] Vona, B., Lechno, S., Hofrichter, M. A. H., Hopf, S., Läig, A. K., Haaf, T., Keilmann, A., Zechner, U., and Bartsch, O. (2016) Confirmation of PDZD7 as a Nonsyndromic Hearing Loss Gene. *Ear Hear,* **37**, e238–46.

[94] Wang, S. S., Esplin, E. D., Li, J. L., Huang, L., Gazdar, A., Minna, J., and Evans, G. A. (1998) Alterations of the PPP2R1B gene in human lung and colon cancer. *Science,* **282**, 284–7.

[95] Deng, Y., Wu, J., and Li, X. (2022) lncRNA RUNDC3A-AS1 Regulates Proliferation and Apoptosis of Thyroid Cancer Cells via the miR-151b/SNRPB Axis. *Int J Endocrinol,* **2022**, 9433434.

[96] Chen, P., Wang, W., Wong, S. W., Li, J., Wu, Q., Zhang, S.-D., Lin, Y., and Kwok, H. F. (2022) RUNDC3A regulates SNAP25-mediated chemotherapy resistance by binding AKT in gastric neuroendocrine carcinoma (GNEC). *Cell Death Discov,* **8**, 296.

[97] Jiang, X., Bergquist, A., Löscher, B.-S., Venkatesh, G., Mold, J. E., Holm, K., Laerdahl, J. K., Skånland, S. S., Maleki, K. T., Cornillet, M., Taskén, K., Franke, A., Karlsen, T. H., Björkström, N. K., and Melum, E. (2021) A heterozygous germline CD100 mutation in a family with primary sclerosing cholangitis. *Sci Transl Med,* **13**, eabb0036.

[98] Gabillard-Lefort, C., Martinez, C. S., Gueguen, N., Desquiret-Dumas, V., Wery, M., Legoff, L., Guimier, A., Rondeau, S., Barcia, G., Barnerias, C., Cogne, B., Besnard, T., Lorino, E., Douglas, J., Bodamer, O., Vetro, A., Guerrini, R., Balestrini, S., Conti, V., Siri, L., Chevrollier, A., Bris, C., Colin, E., Procaccio, V., Prunier-Mirebeau, D., Lenaers, G., Khiati, S., Nizon, M., and Baris, O. R. (2025) Bi-allelic variants in TM2D3 cause a severe syndromic neurodevelopmental disorder associated with endoplasmic reticulum and mitochondrial abnormalities. *Am J Hum Genet,* **112**, 1711–1721.

[99] Sun, W., Qiu, G., Zou, Y., Cai, Z., Wang, P., Lin, X., Huang, J., Jiang, L., Ding, X., and Hu, G. (2015) Knockdown of TMEM45A inhibits the proliferation, migration and invasion of glioma cells. *Int J Clin Exp Pathol,* **8**, 12657–67.

[100] Jiang, H., Chen, H., Wan, P., Liang, M., and Chen, N. (2021) Upregulation of TMEM45A Promoted the Progression of Clear Cell Renal Cell Carcinoma in vitro. *J Inflamm Res,* **14**, 6421–6430.

[101] Zhu, M., Jiang, B., Yan, D., Wang, X., Ge, H., and Sun, Y. (2020) Knockdown of TMEM45A overcomes multidrug resistance and epithelial-mesenchymal transition in human colorectal cancer cells through inhibition of TGF-β signalling pathway. *Clin Exp Pharmacol Physiol,* **47**, 503–516.

[102] Beck, A. E., McMillin, M. J., Gildersleeve, H. I. S., Kezele, P. R., Shively, K. M., Carey, J. C., Regnier, M., and Bamshad, M. J. (2013) Spectrum of mutations that cause distal arthrogryposis types 1 and 2B. *Am J Med Genet A,* **161A**, 550–5.

[103] Lehtokari, V.-L., Ceuterick-de Groote, C., de Jonghe, P., Marttila, M., Laing, N. G., Pelin, K., and Wallgren-Pettersson, C. (2007) Cap disease caused by heterozygous deletion of the beta-tropomyosin gene TPM2. *Neuromuscul Disord,* **17**, 433–42.

[104] Donner, K., Ollikainen, M., Ridanpää, M., Christen, H.-J., Goebel, H. H., de Visser, M., Pelin, K., and Wallgren-Pettersson, C. (2002) Mutations in the beta-tropomyosin (TPM2) gene–a rare cause of nemaline myopathy. *Neuromuscul Disord,* **12**, 151–8.

[105] Mokbel, N., Ilkovski, B., Kreissl, M., Memo, M., Jeffries, C. M., Marttila, M., Lehtokari, V.-L., Lemola, E., Grönholm, M., Yang, N., Menard, D., Marcorelles, P., Echaniz-Laguna, A., Reimann, J., Vainzof, M., Monnier, N., Ravenscroft, G., McNamara, E., Nowak, K. J., Laing, N. G., Wallgren-Pettersson, C., Trewhella, J., Marston, S., Ottenheijm, C., North, K. N., and Clarke, N. F. (2013) K7del is a common TPM2 gene mutation associated with nemaline myopathy and raised myofibre calcium sensitivity. *Brain,* **136**, 494–507.

Figures and Tables

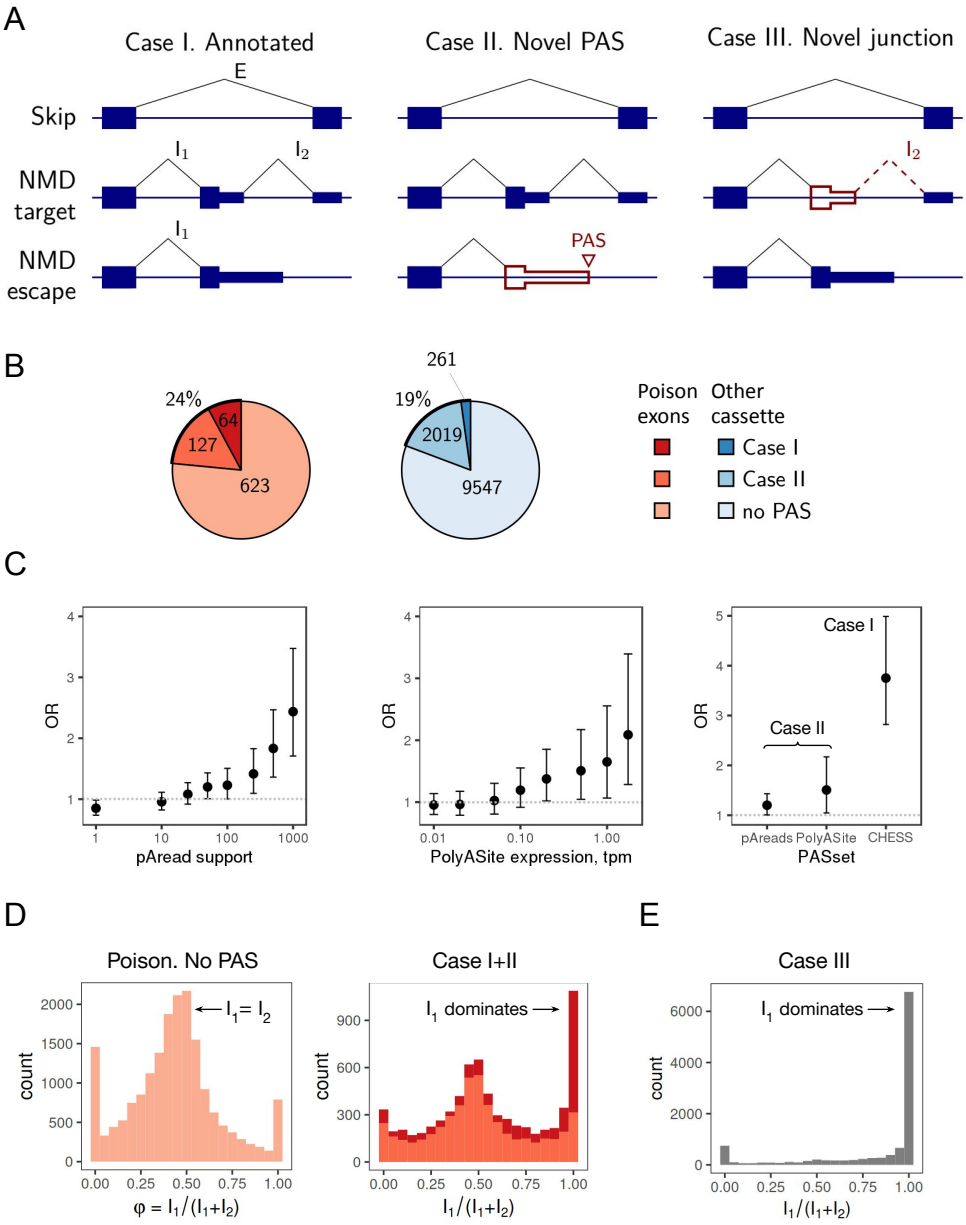

Figure 1: Poison exons are frequently followed by an intronic PAS. **(A)** Schematic representation of transcript isoforms in a gene containing a poison exon and a PAS within the downstream intron. In this study, we used CHESS annotations, incorporated novel PAS identified in two published databases, and included novel exon-exon junctions derived from GTEx transcriptomic data. Accordingly, a gene may contain both an annotated poison exon and PAS (Case I), or alternatively, either the PAS (Case II) or the poison exon (Case III) may be novel. Unannotated exons, PAS, and junctions are indicated with a red outline. **(B)** Number of poison and other cassette exons followed by a PAS. Color intensity indicates the annotation status of the PAS: annotated in the CHESS database (Case I), identified in the PolyASite database or predicted from poly(A) reads (Case II), or absent. The PAS derived from transcriptomic data were required to be supported by ≥50 poly(A) reads. The PAS from PolyASite were required to have an expression level of ≥0.5 TPM. **(C)** Odds ratio (OR) for the proportion of exons followed by a PAS in cassette versus poison exons. Left: *OR* as a function of polyA read support. Middle: *OR* as a function of PAS expression in TPM according to the PolyASite database. Right: *OR* for PAS identified from the PolyASite database, polyA reads, and annotated transcript ends (CHESS). Whiskers indicate 95% confidence intervals. **(D,E)** Distribution of $\varphi = \frac{I_1}{I_1+I_2}$ for poison exons with and without a PAS, calculated from GTEx transcriptomic data. Color coding follows the scheme in panel B.

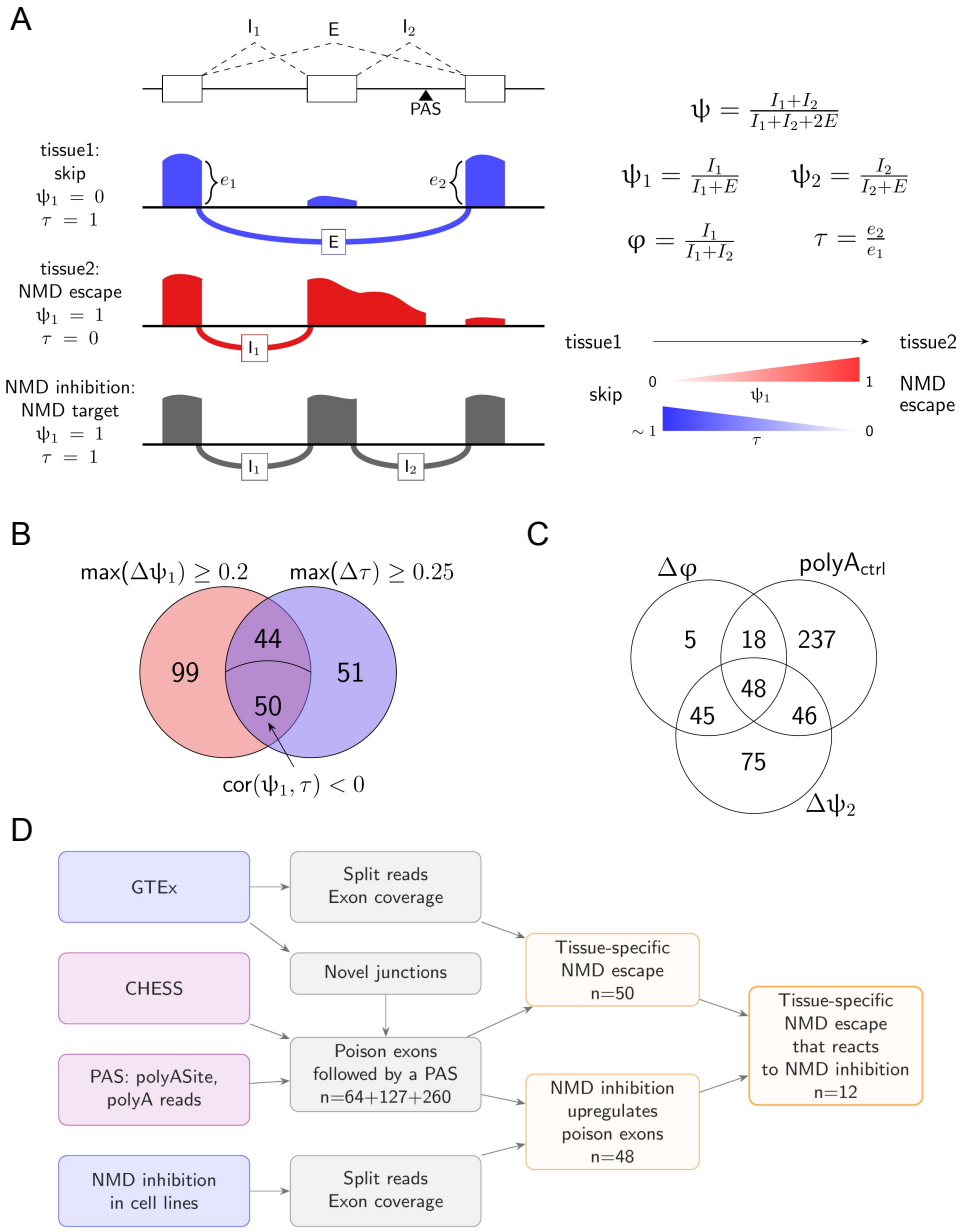

Figure 2: Identification of tissue-specific NMD escape events. **(A)** Left: Transcript isoforms and the metrics used for their assessment. e_1 and e_2 represent read coverage per nucleotide. I_1, I_2 and E denote the numbers of split reads supporting the corresponding exon-exon junctions. Right: Definition of the metrics and a schematic representation of changes in ψ_1 and τ that are characteristic of a switch from the skip to the NMD-escape isoform: an increase in ψ_1 accompanied by a decrease in τ. **(B)** Venn diagram of poison exons selected based on ψ_1 and τ metrics. The portion of the intersection exhibiting a negative correlation between ψ_1 and τ is indicated by an arrow. **(C)** Venn diagram illustrating the response of poison exons followed by a PAS to NMD inactivation. Δ indicates the changes in the respective metrics between NMD-inactivated and control conditions (see Methods). $polyA_{ctrl}$ refers to the category of exons with high φ_{ctrl} or low τ_{ctrl} values in untreated cells. **(D)** Flowchart of the analysis. Violet nodes represent database information and annotations, blue nodes correspond to transcriptomic data, gray nodes indicate processing steps, and orange nodes denote the output sets.

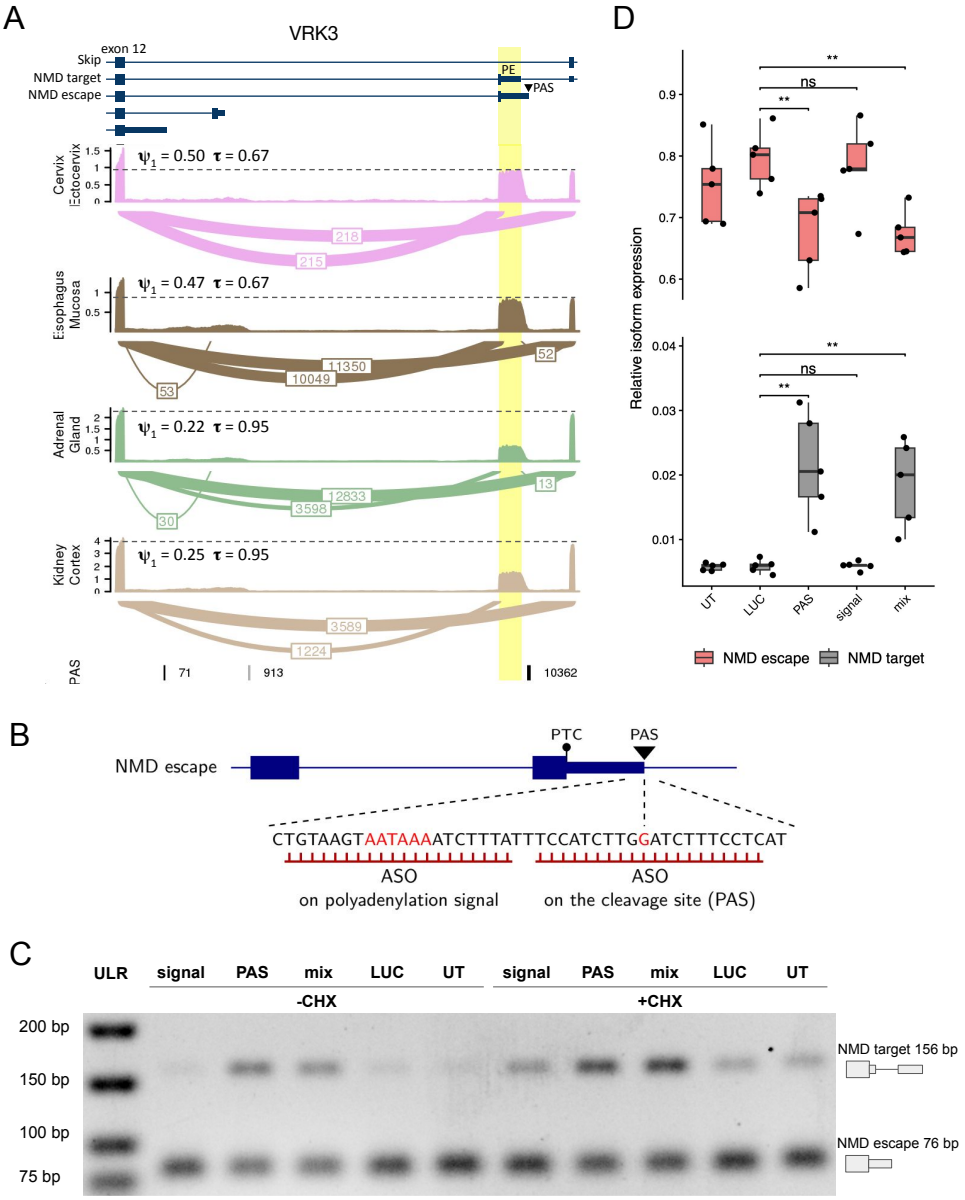

Figure 3: NMD escape in the *VRK3* gene. **(A)** Read coverage profiles and sashimi plots illustrating a tissue-specific switch from the skip to the NMD-escape isoform in the *VRK3* gene. The poison exon followed by a PAS is highlighted. The scale of the vertical axes represents read coverage density. Numbers on the sashimi plots indicate split read counts combined across samples from each tissue. Transcript models from CHESS are shown at the top. The PAS track indicates the number of poly(A) reads supporting each PAS. Dotted lines highlight the drop in read coverage between the upstream and downstream exons. **(B)** Genomic sequence surrounding the polyadenylation signal and the cleavage site (polyadenylation site, PAS) targeted by ASOs. **(C)** Endpoint RT-PCR analysis of untreated cells (UT), cells treated with ASO targeting luciferase (LUC), ASO targeting the polyadenylation signal (signal), ASO targeting the cleavage site (PAS), and their combination (mix), with (+CHX) or without (-CHX) cycloheximide treatment. **(D)** RT-qPCR analysis showing the fraction of NMD-escape (red) and NMD-target (grey) isoforms among all three isoforms under ASO treatment in the presence of CHX. ULR — ultra low range DNA ladder.

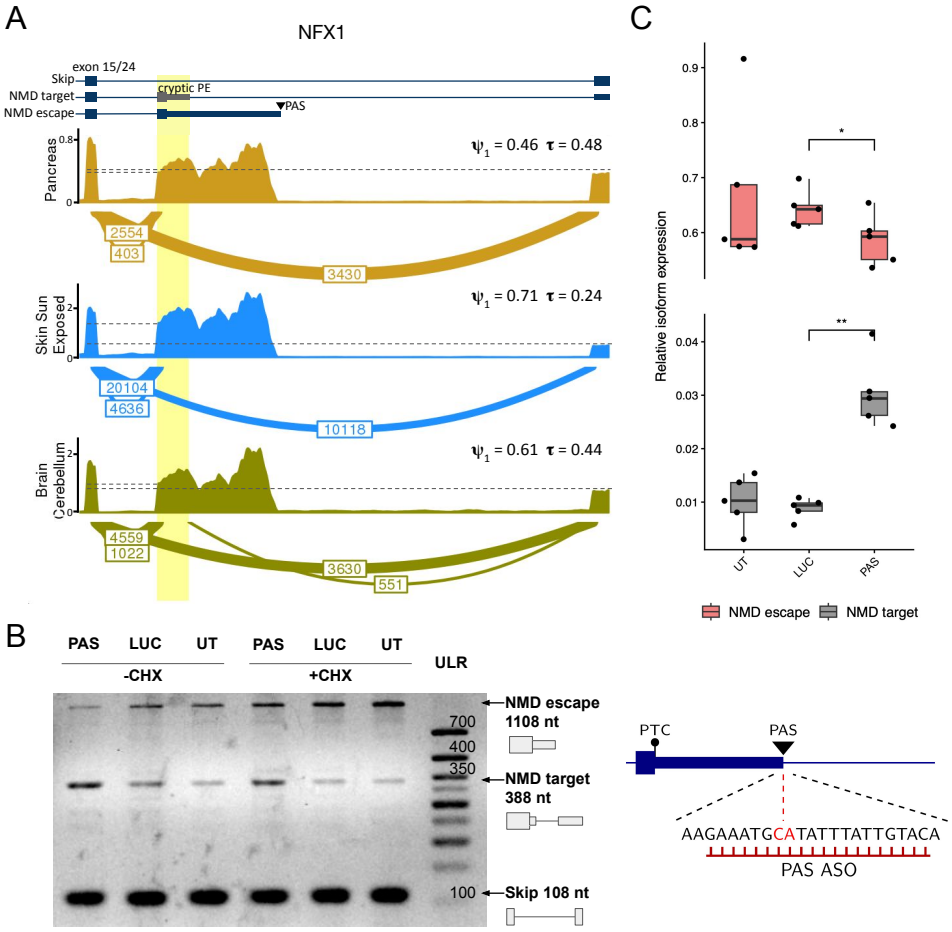

Figure 4: NMD escape in the *NFX1* gene. **(A)** Read coverage profiles and sashimi plots illustrating a tissue-specific switch from the skip to the NMD-escape isoform in the *NFX1* gene. The poison exon followed by a PAS is highlighted. The vertical axes and numbers on the sashimi plots are as described in Figure 3. Transcript models from CHESS are shown at the top in blue; the unannotated poison exon is shown in gray. Dotted lines indicate the drop in read coverage between the upstream and downstream exons. **(B)** Left: Endpoint RT-PCR analysis of untreated cells (UT), cells treated with ASO targeting luciferase (LUC), and ASO targeting the cleavage site (PAS), with (+CHX) or without (-CHX) cycloheximide treatment. Right: Genomic sequence surrounding the PAS targeted by the ASO. **(C)** RT-qPCR analysis showing the fraction of NMD-escape (red) and NMD-target (grey) isoforms among all three isoforms under ASO treatment in the presence of CHX. ULR — ultra low range DNA ladder.

Gene name	range(τ)	range(ψ_1)	cor(ψ_1,τ)	case	Associated disease
AKR1E2	0.32	0.26	-0.58	II	Cataract
ANKS1B	0.67	0.20	-0.81	III	AAMR
ANKS1B	0.67	0.96	-0.44	III	AAMR
ARMC7	0.31	0.38	-0.58	III	Breast Cancer (CNV)
BDP1	0.47	0.73	-0.44	II	Deafness [69]
BTN2A2	0.74	1.00	-0.77	III	Chronic Fatigue Syndrome
CCNH	0.32	0.93	-0.70	III	Capillary and Arteriovenous Malformation [70, 71]
CDK12	0.33	0.39	-0.64	III	Breast, Ovarian and Prostate Cancers [72, 73, 74], Glioma [75]
DDX31	0.41	0.46	-0.81	III	Bladder Cancer [48], Pancreatic Ductal Adenocarcinoma [76]
FXN	0.31	0.37	-0.75	III	Friedreich Ataxia, Atrial Standstill [77, 78], Parkinson's Disease [79]
GCDH	0.33	0.91	-0.42	III	Glutaric Acidemia I [80]
GCOM1	0.41	0.49	-0.69	I	Dilated Cardiomyopathy [81]
LEPR	0.31	0.29	-0.63	III	Hypogonadotropic Hypogonadism, Kallmann Syndrome [82], Leptin Receptor Deficiency and Obesity [83, 84, 85]
MTMR1	0.31	0.35	-0.55	I	Myopathy [86], Myotonic Dystrophy [87]
NECTIN3	0.70	0.69	-0.92	III	Combined Oxidative Phosphorylation Deficiency
NFX1	0.25	0.30	-0.79	III	Cervical Cancer [43, 88], Vascular Disease [89]
NRP1	0.37	0.48	-0.49	III	Covid-19 [90]
PDZD7	0.62	0.52	-0.86	III	Deafness [91], Fundus Dystrophy, Hereditary Retinal Dystrophy, Usher Syndrome [92], Nonsyndromic Hearing Loss [93]
PPP2R1B	0.74	0.79	-0.93	III	Lung Cancer [94]
RUNDC3A	0.34	0.65	-0.72	III	Thyroid Cancer [95], Neuroendocrine Carcinoma [96]
SEMA4D	0.92	0.96	-0.82	III	Primary Sclerosing Cholangitis [97]
TM2D3	0.29	0.25	-0.98	III	Neurocardiorenal Malformation Syndrome [98]
TMEM45A	0.30	0.24	-0.42	II	Glioma [99], Renal Cell Carcinoma [100], Colorectal cancer [101]
TPM2	0.58	0.99	-0.89	I	Distal Arthrogryposis [102], Cap Myopathy [103], Nemaline Myopathy [104, 105]
TUBGCP5	0.27	1.00	-0.48	II	AAMR
UGT1A6	0.39	0.93	-0.65	II	Bilirubin Metabolic Disorders including Crigler-Najjar and Gilbert Syndromes

Table 1: Characteristics of tissue-specific NMD escape events. See also Figure S6. The table includes events in genes that are either associated with diseases reported in the literature or listed as EliteGenes for a disease in MalaCards [37]. Abbreviations: AAMR — Alacrima, achalasia, and impaired intellectual development syndrome; CNV — copy number variation.

Novel examples of NMD escape through alternative intronic polyadenylation

Maria Vlasenok[1,†], Antonina Kuznetsova[1,†], Dmitry A. Skvortsov[1,2], and Dmitri D. Pervouchine[1,*]

[1]Center for Molecular and Cellular Biology, Moscow 143026, Russia

[2]Moscow State University, Moscow 119991, Russia

[*]e-mail: pervouchine@gmail.com; tel/fax +7 (916) 220 69 99

[†]These authors contributed equally to the work

March 5, 2026

List of Figures

S1 Background distributions for cryptic poison exons 3

S2 The odds ratio for poison versus cassette exons 4

S3 Continuous short reads overlapping splice sites 5

S4 Distributions of ψ_1, ψ_2, and τ values 6

S5 Sensitivity of tissue-specific NMD-escape event calls to thresholds 7

S6 τ and ψ_1 values for selected poison exons across tissues 10

S7 τ and e_1 values for selected poison exons across tissues 13

S8 Response of cassette exons followed by a PAS to NMD inactivation 14

S9 Correlations between τ and ψ_1 for *VRK3*, *NFX1*, *TM2D3*, and *DDX31* 15

S10 Additional read coverage profiles and sashimi plots for the *VRK3* gene 16

S11 *VRK3* expression following ASO treatment with and without NMD inhibition . 17
S12 NMD escape in the *TM2D3* gene 18
S13 NMD escape in the *DDX31* gene 19
S14 Predicted NMD escape in the *TPM2* gene 20

List of Tables

S1 ASO sequences 21
S2 Primers for qPCR. 21
S3 Primers for endpoint PCR (skip, NMD-target and NMD-escape isoforms). . . 22

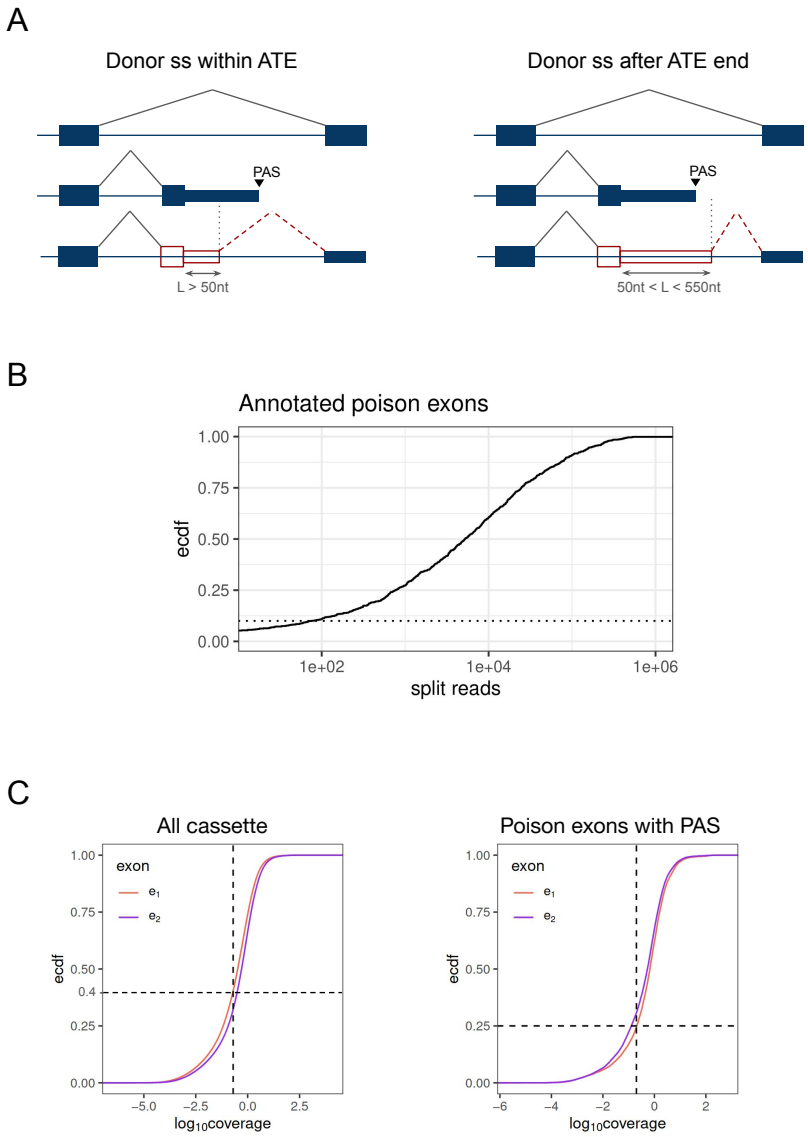

Figure S1: **(A)** Cryptic poison exons overlapping alternative terminal exons. Classification of cryptic poison exons according to the location of the donor splice site relative to annotated PAS. Only poison exons with fewer than 550 nucleotides between the PTC and the downstream donor splice site were considered (see Methods). Red dashed lines indicate the I_2 splice junction. Unannotated exons are outlined in red. **(B)** Empirical cumulative distribution function (eCDF) of split read support for I_2 splice junctions of annotated poison exons in GTEx. The line represents the 0.1 percentile for the annotated splice junctions, which corresponds approximately to 100 supporting reads. **(C)** eCDF of e_1 and e_2 coverage for all cassette exons (left) and poison exons with PAS (right) across GTEx tissues. The median e_1 value is lower than the median e_2 value in cassette exons and higher than the median e_2 value in poison exons with PAS. The $e_1 > 0.1$ threshold for τ computations excludes a quarter of exon-tissue pairs for poison exons with PAS (see Methods).

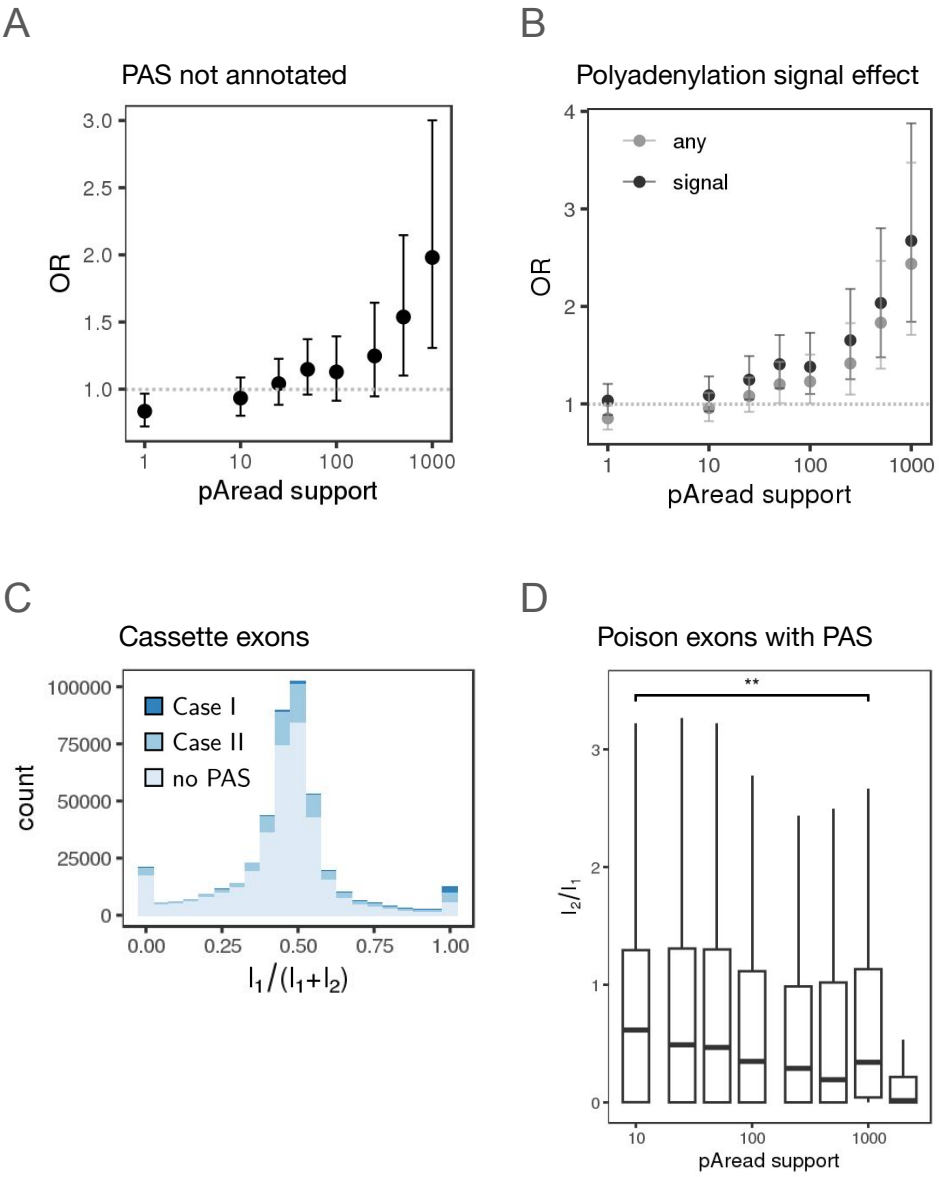

Figure S2: The odds ratio (OR) for the proportion of exons followed by a PAS in poison versus cassette exons. **(A)** OR for the unannotated PAS from the polyA read-based set as a function of polyA read support. **(B)** OR as a function of polyA read support for all PAS and PAS preceded by a polyadenylation signal. Whiskers represent 95% confidence intervals. **(C)** Distribution of $\varphi = I_1/(I_1 + I_2)$ values across GTEx tissues for cassette exons. Colors indicate whether the exon is followed by a PAS, as shown in Figure 1B. **(D)** Ratio of I_2 and I_1 as a function of polyA read support computed for poison exons. Asterisks denote statistical significance based on a two-sided Mann-Whitney U-test.

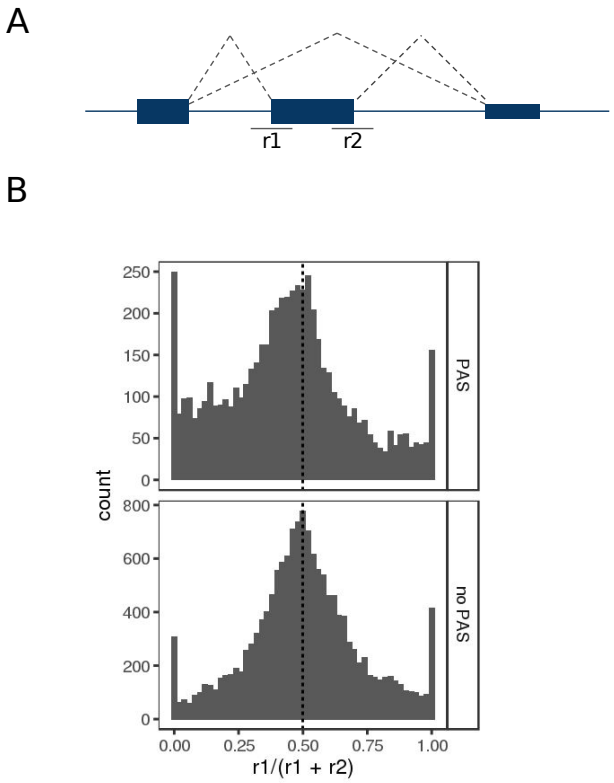

Figure S3: **(A)** A schematic illustrating continuous reads overlapping splice sites: r_1 denotes reads that overlap the 5' end of the cassette exon, and r_2 denotes reads that overlap the 3' end. **(B)** Distribution of the fraction of r_1 in the sum of r_1 and r_2 for poison exons followed by a PAS and poison exons not followed by a PAS. A low $r_1/(r_1 + r_2)$ ratio implies higher retention at the 3' end, which supports downstream PAS usage.

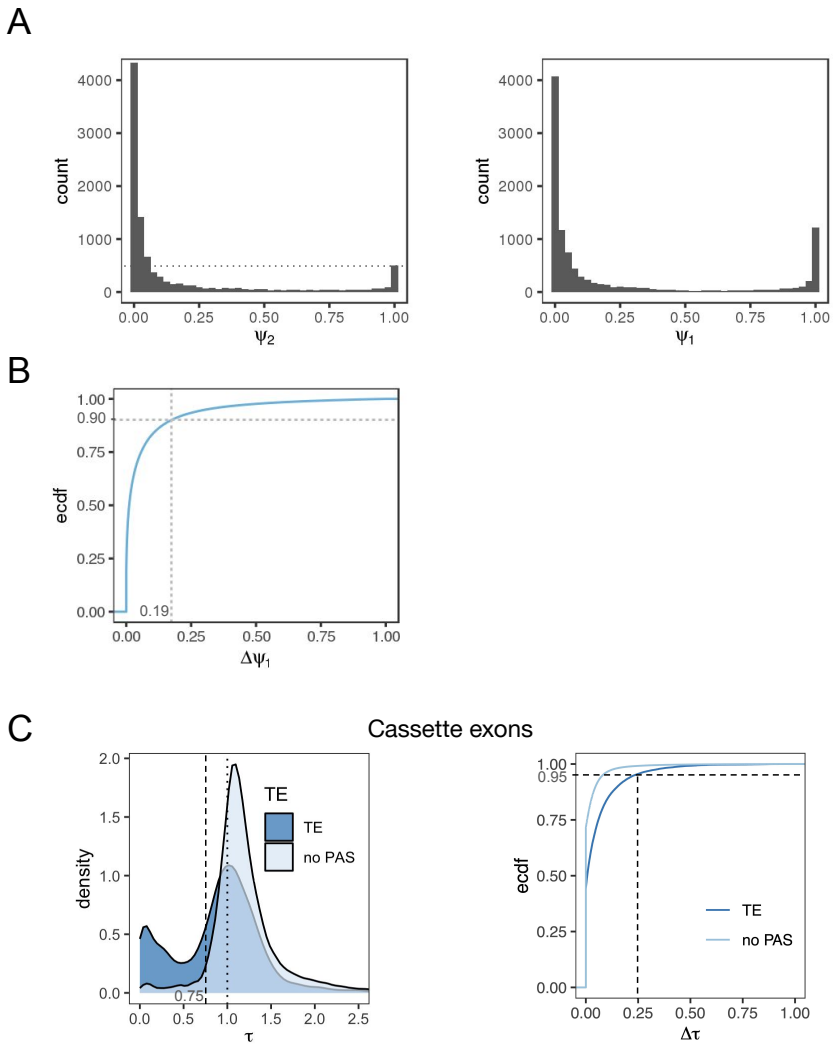

Figure S4: Distributions of ψ_1, ψ_2, and τ values. **(A)** Distribution of $\psi_1 = \frac{I_1}{I_1+E}$ and $\psi_2 = \frac{I_2}{I_2+E}$ for poison exons followed by a PAS in GTEx tissues. **(B)** Empirical cumulative distribution function (eCDF) of differences between ψ_1 values across different tissues for non-poison cassette exons not followed by a PAS. The line indicates the 90th percentile ($\Delta\psi_1 \approx 0.2$). This distribution was used to assess ψ_1 variability among tissues for standard tissue-specific cassette exons. The 90th percentile was set as a threshold to identify poison exons with tissue-specific inclusion of the junction upstream of the exon (Figure 2B). **(C)** left: Distribution of τ values across GTEx tissues for cassette exons followed by an annotated PAS (i.e., annotated alternative terminal exons, dark blue) and cassette exons not followed by any PAS (light blue). Exons with a PAS are clearly characterized by an additional peak at $\tau = 0$, supporting the validity of the metric. Right: eCDF of differences between τ values in different tissues for alternative terminal exons (dark blue) and cassette exons not followed by a PAS (light blue). The line shows the 95% percentile for the former set ($\Delta\tau \approx 0.25$).

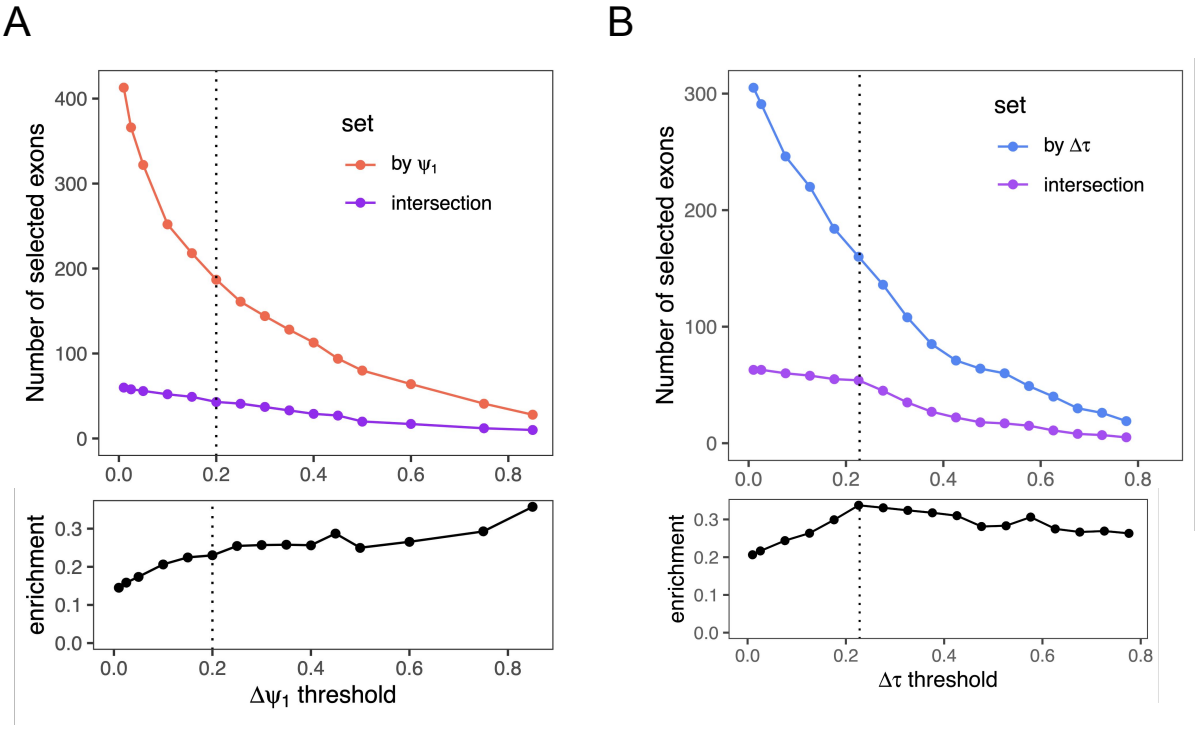

Figure S5: **(A)** Dependence on the splicing threshold $\Delta\psi_1$, with the polyadenylation threshold fixed at $\Delta\tau = 0.25$. The orange curve ("by ψ_1") represents the number of exons with $\Delta\psi_1$ above the indicated cutoff. The purple curve ("intersection") shows the subset that additionally satisfies $\Delta\tau \geq 0.25$ and exhibits a negative cross-tissue correlation between ψ_1 and τ. The bottom panel reports enrichment, defined as(intersection)/(by ψ_1). **(B)** Dependence on the polyadenylation threshold $\Delta\tau$, with the splicing threshold fixed at $\Delta\psi_1 = 0.2$. The blue curve ("by $\Delta\tau$") represents the number of exons with $\Delta\tau$ above the indicated cutoff. The purple curve ("intersection") shows the subset that additionally satisfies $\Delta\psi_1 \geq 0.2$ and exhibits a negative correlation between ψ_1 and τ. The bottom panel reports enrichment, defined as (intersection)/(by $\Delta\tau$). Dotted vertical lines indicate the thresholds used in the analysis: $\Delta\psi_1 = 0.2$ and $\Delta\tau = 0.25$.

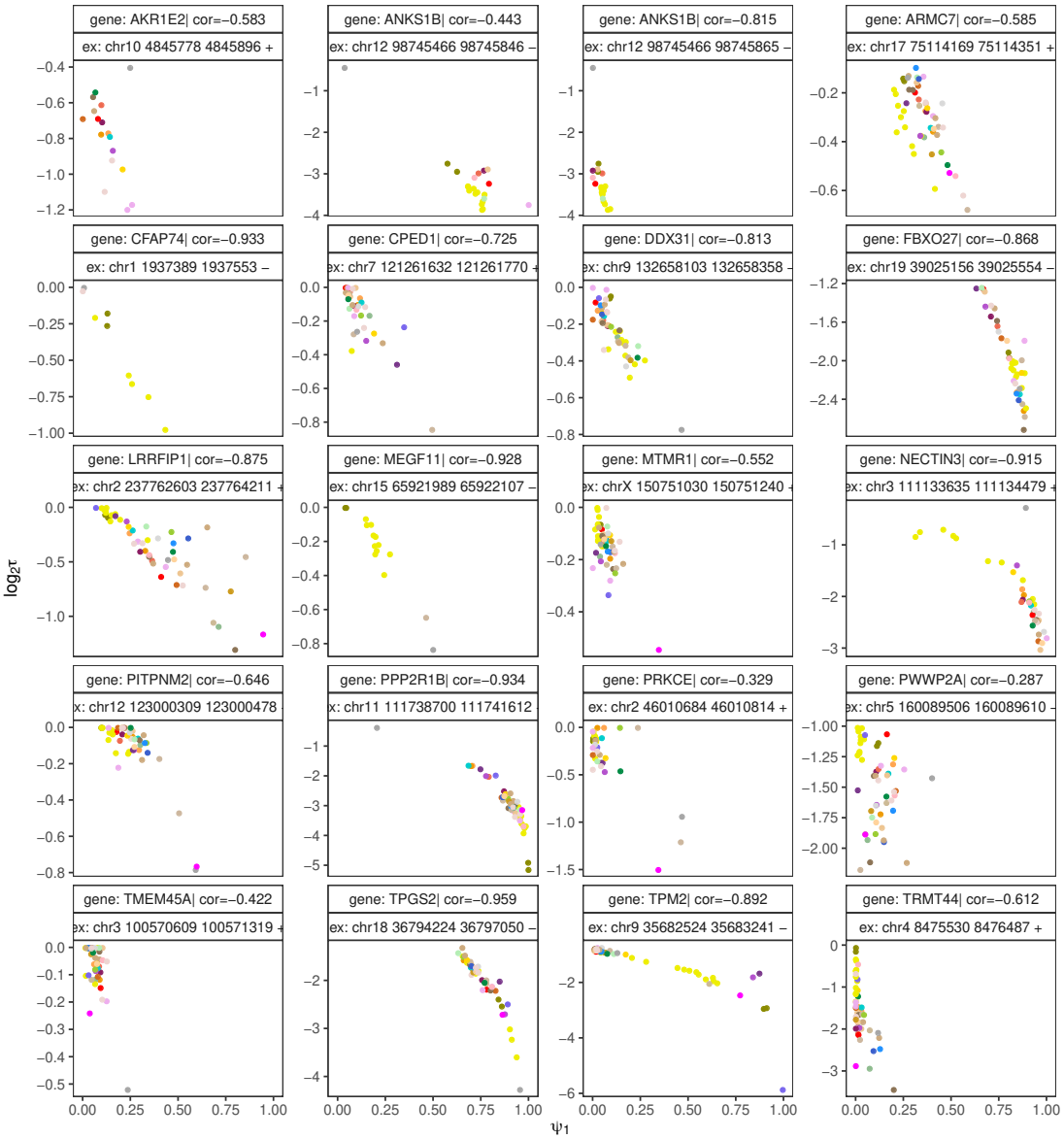

Figure S6: Part 1 — continued on the next page.

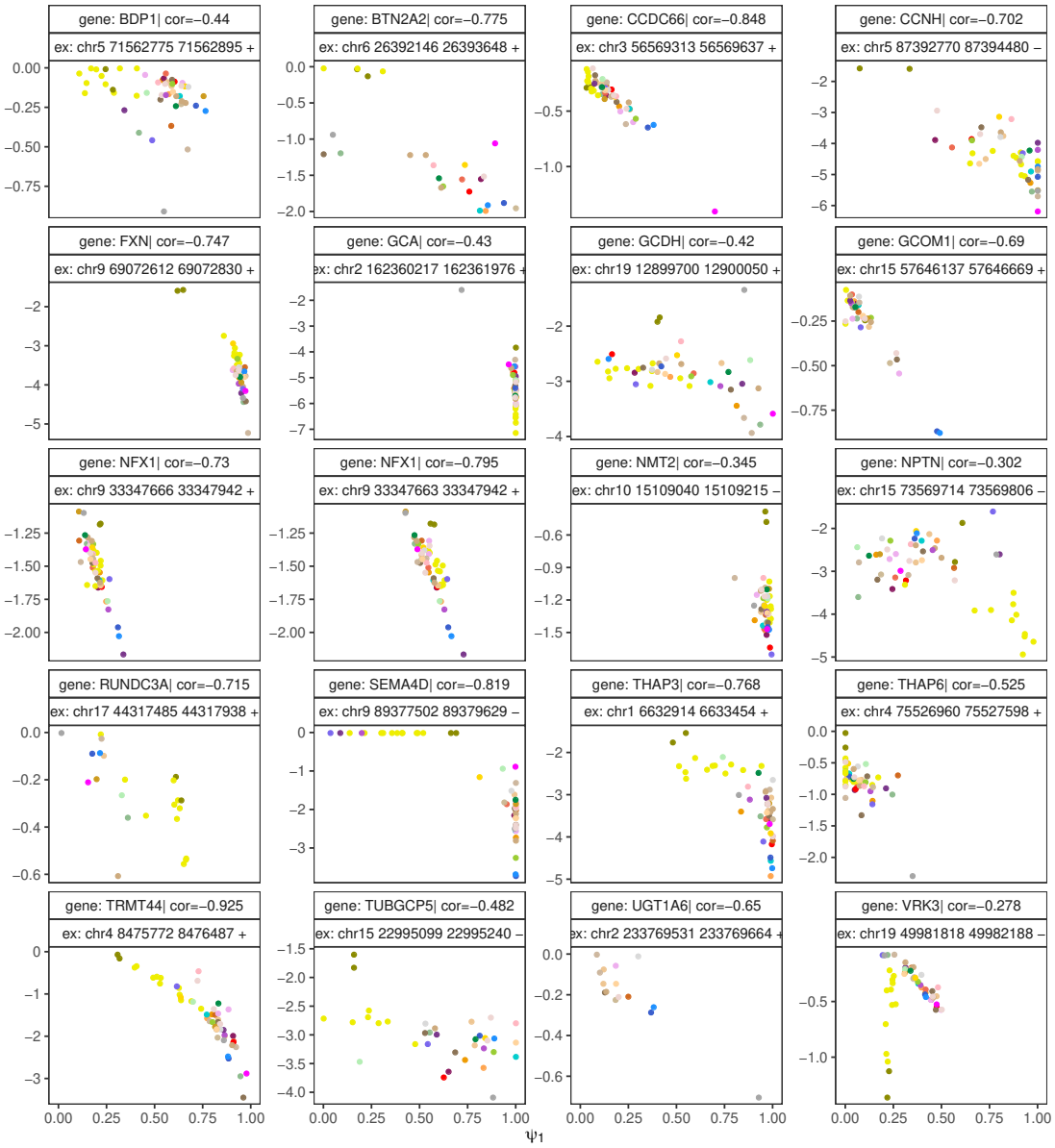

Figure S6: Part 2 — continued on the next page.

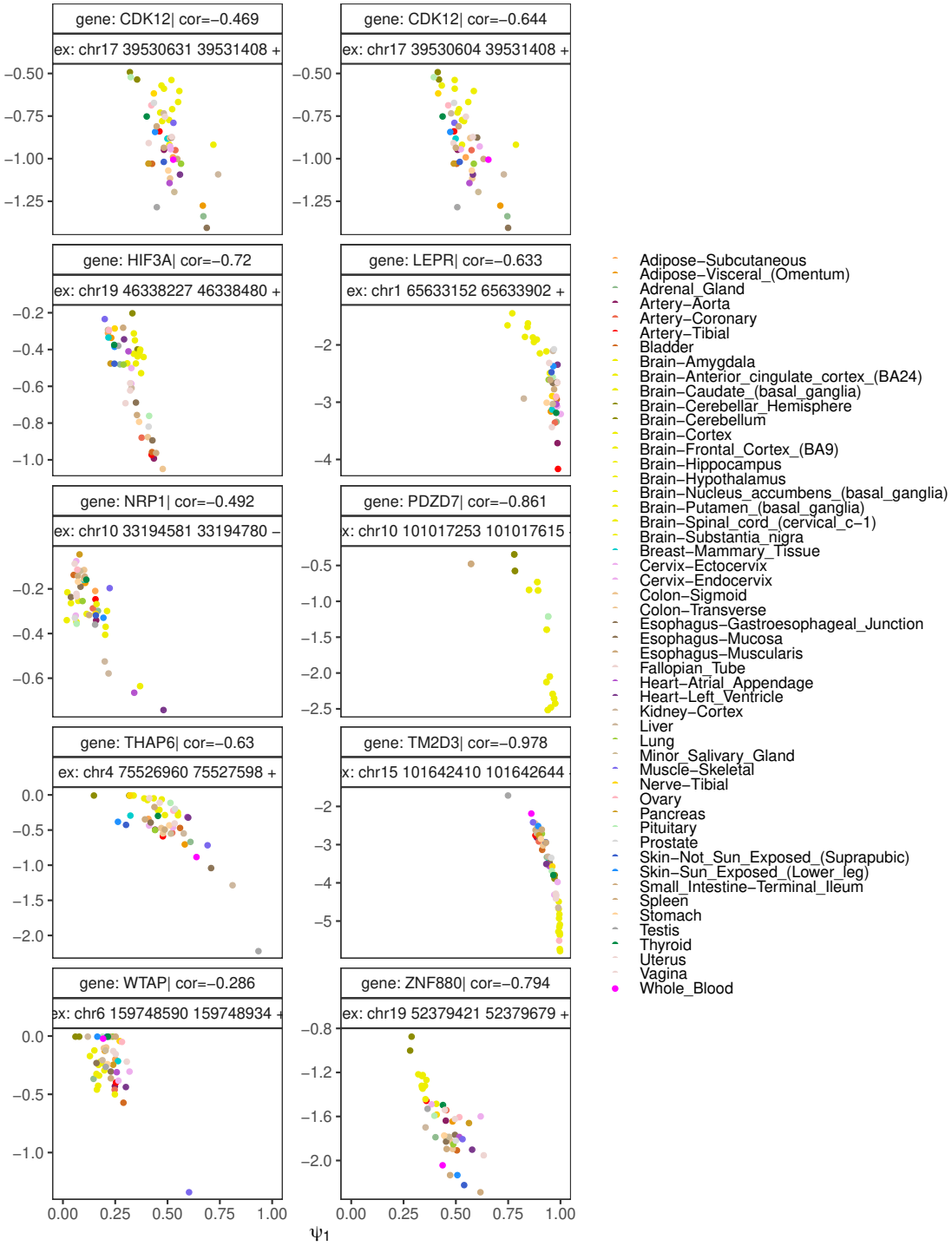

Figure S6: Part 3 — continued from the previous page. Mean τ and ψ_1 values across GTEx tissues for the selected poison exons exhibiting a tissue-specific switch between skip and NMD-escape isoforms. Poison exon coordinates, gene name, and the Spearman correlation coefficient between ψ_1 and τ are displayed above each dot plot.

Figure S7: Part 1 — continued on the next page.

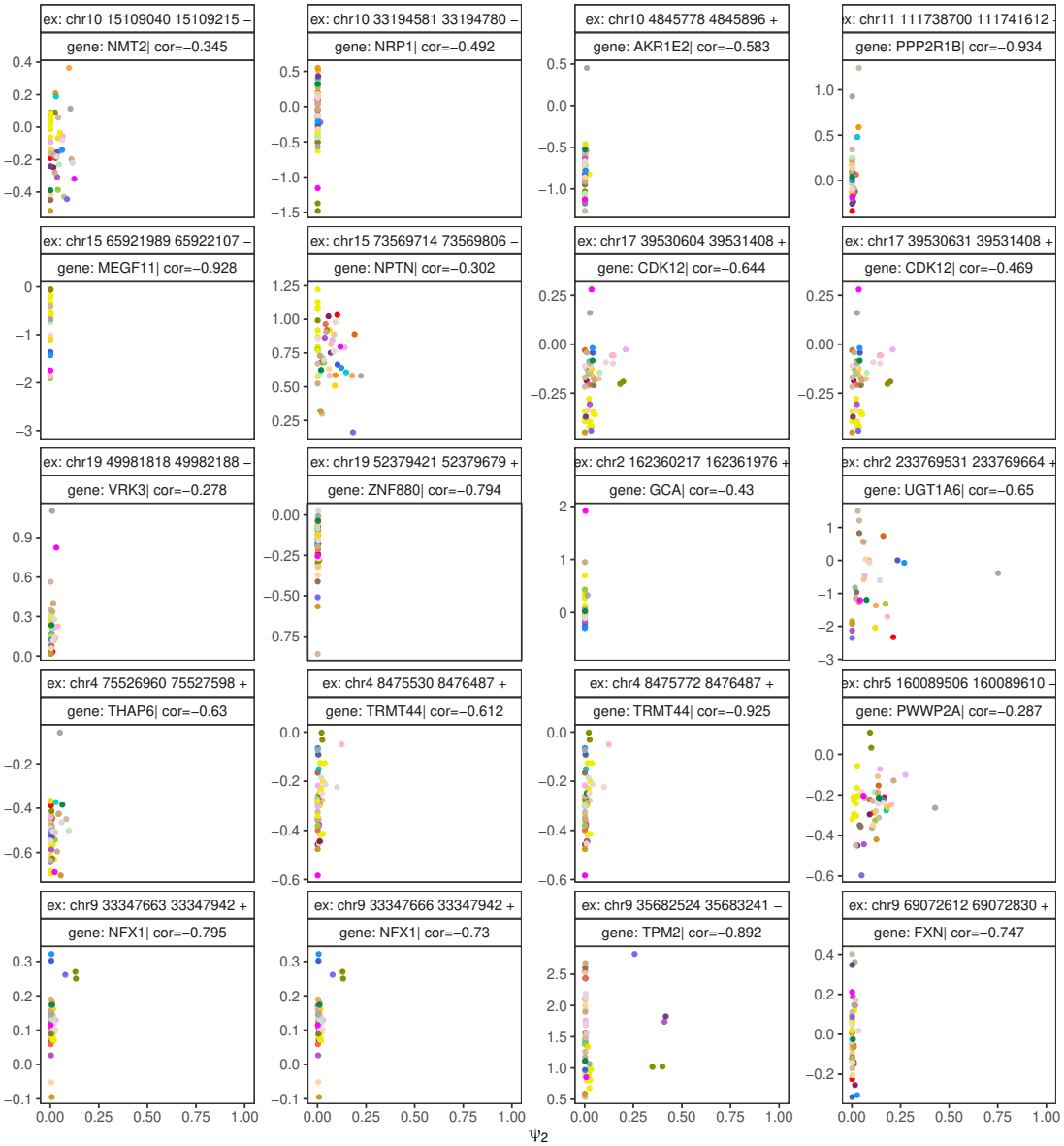

Figure S7: Part 2 — continued on the next page.

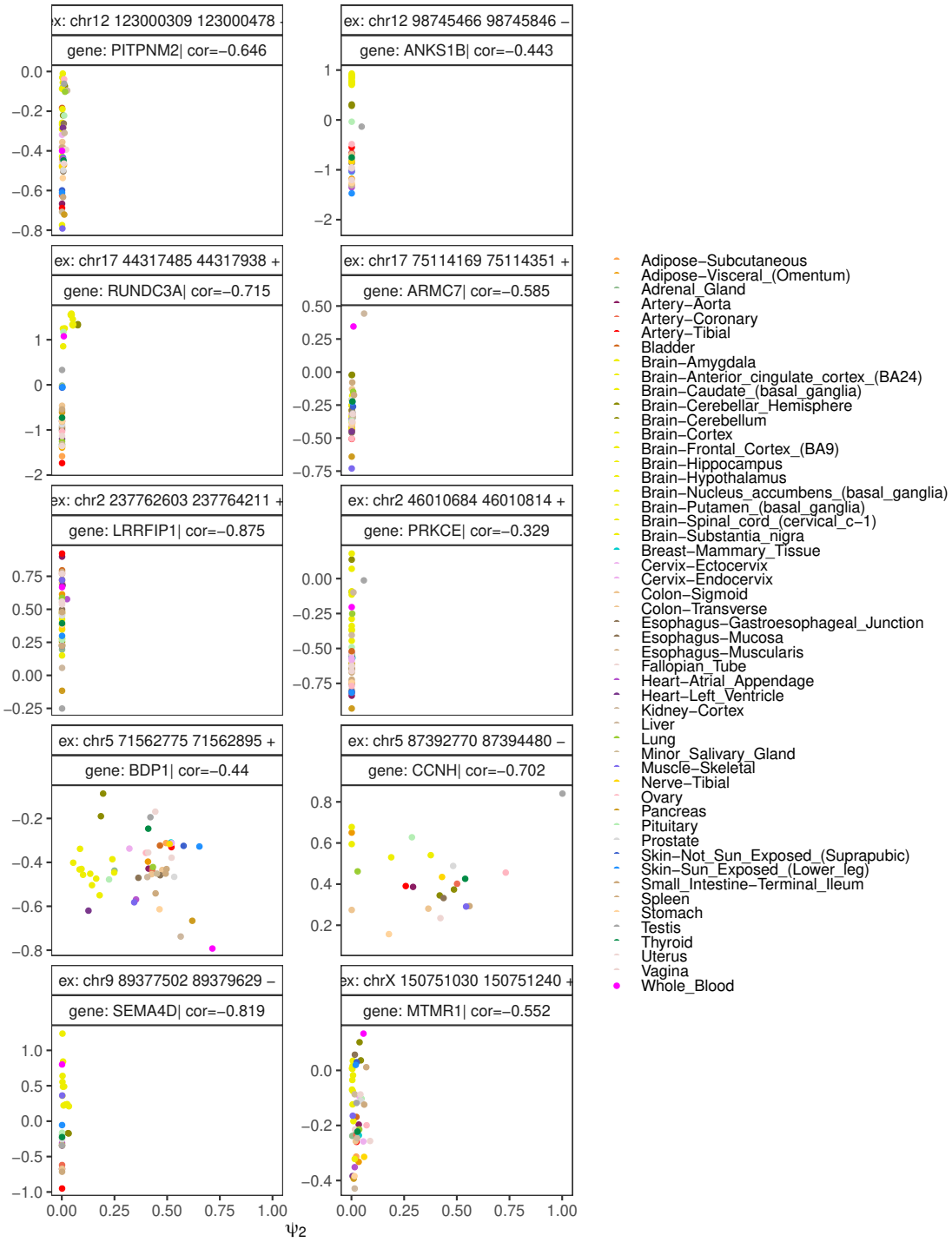

Figure S7: Part 3 — continued from the previous page. Mean τ and upstream exon coverage (e_1) across GTEx tissues for the selected poison exons exhibiting a tissue-specific switch between skip and NMD-escape isoforms. Poison exon coordinates, gene name, and the Spearman correlation between gene expression in TPM and τ are displayed above each dot plot.

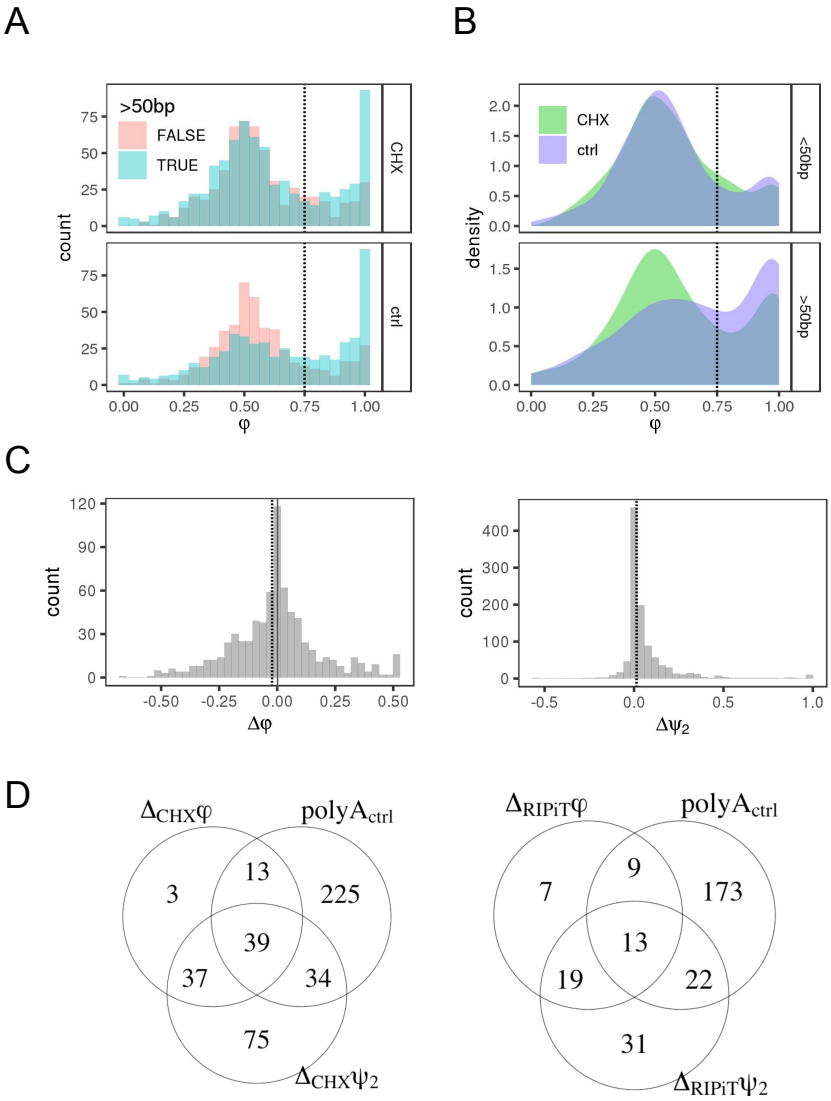

Figure S8: Response of cassette exons followed by a PAS to NMD inactivation. **(A,B)** Distribution of φ values for exons with a stop codon followed by a PAS before and after NMD inhibition with CHX. In (A), colors indicate whether the exon obeys the 50-nt rule (191 of 361 exons do). In (B), colors indicate whether the cells were treated with CHX. **(C)** Distribution of changes in φ and ψ_2 upon NMD inhibition: $\Delta\psi_2 = \psi_2^{CHX} - \psi_2^{ctrl}$, $\Delta\varphi = \varphi^{CHX} - \varphi^{ctrl}$, for poison exons followed by a PAS. Dashed lines indicate thresholds used to select candidates expressing both NMD-escape and NMD-target isoforms $\Delta\varphi \leq -0.025$, $\Delta\psi_2 \geq 0.015$ and to indicate polyadenylation in untreated cells $\varphi_{ctrl} \geq 0.75$ or $\delta_{ctrl} \leq 0.7$. **(D)** Venn diagrams illustrating the response of poison exons followed by a PAS to NMD inactivation. The category $polyA_{ctrl}$ denotes exons with high φ_{ctrl} or low τ_{ctrl} values in untreated cells. Δ represents the changes in the respective metrics between NMD inactivation and control conditions (see Methods). The left diagram corresponds to the dataset compiled from four cell lines treated with CHX, whereas the right diagram is based on the RIPiT dataset (see Methods).

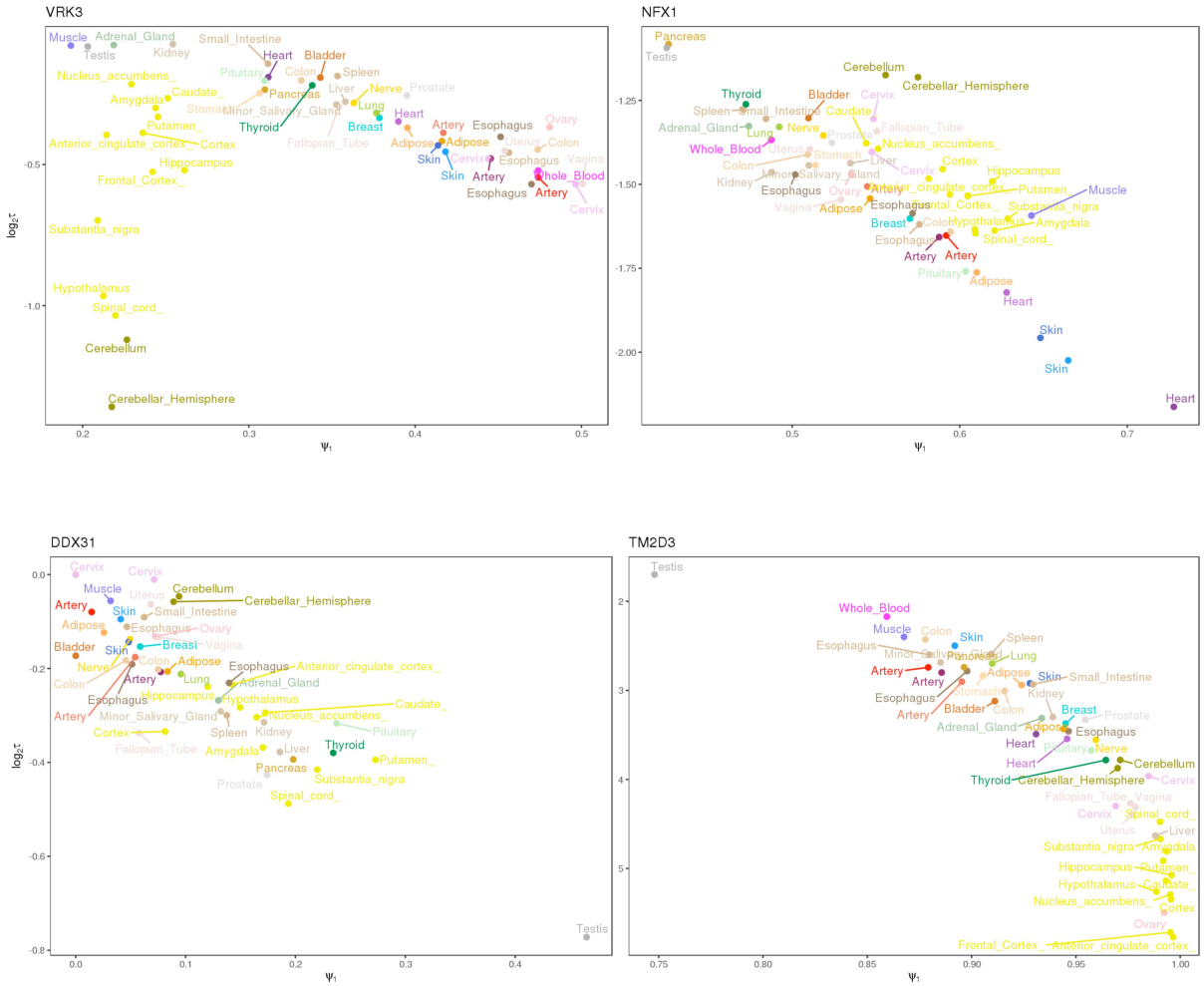

Figure S9: Correlations between τ and ψ_1 for *VRK3*, *NFX1*, *TM2D3*, and *DDX31* genes. Yellow and dark green colors correspond to different brain regions.

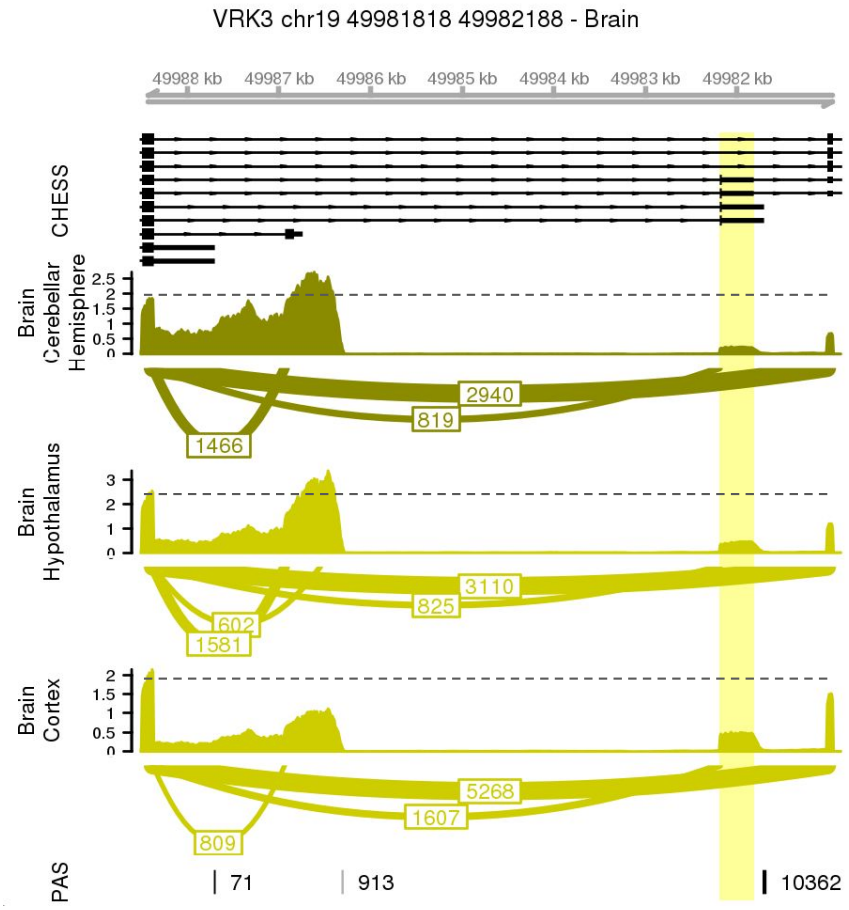

Figure S10: Read coverage profiles and sashimi plots for three brain tissues showing the poison exon in the *VRK3* gene. As in Figure 3, the poison exon followed by a PAS is highlighted. The vertical axes represent read coverage density, and the numbers on the sashimi plots indicate split read counts. Transcript models from CHESS are shown at the top. The PAS track displays the number of polyA reads supporting each PAS. Dotted lines illustrate the drop in read coverage between the upstream and downstream exons.

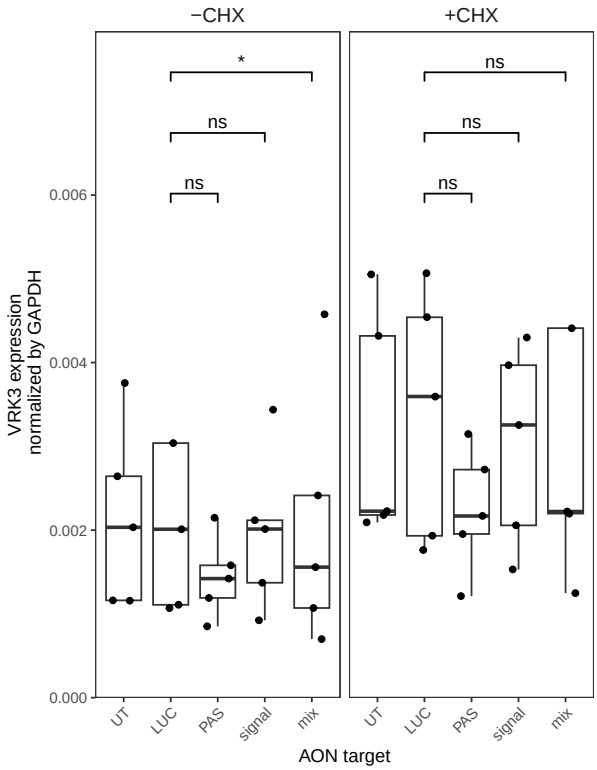

Figure S11: Total *VRK3* expression under ASO treatment, with and without NMD inhibition. Isoform-specific qRT-PCR results were normalized to GAPDH expression. An asterisk (*) indicates a statistically significant difference at the 5% level; "ns" denotes no statistically significant difference (one-sided Wilcoxon signed-rank test, $n = 5$ biological replicates).

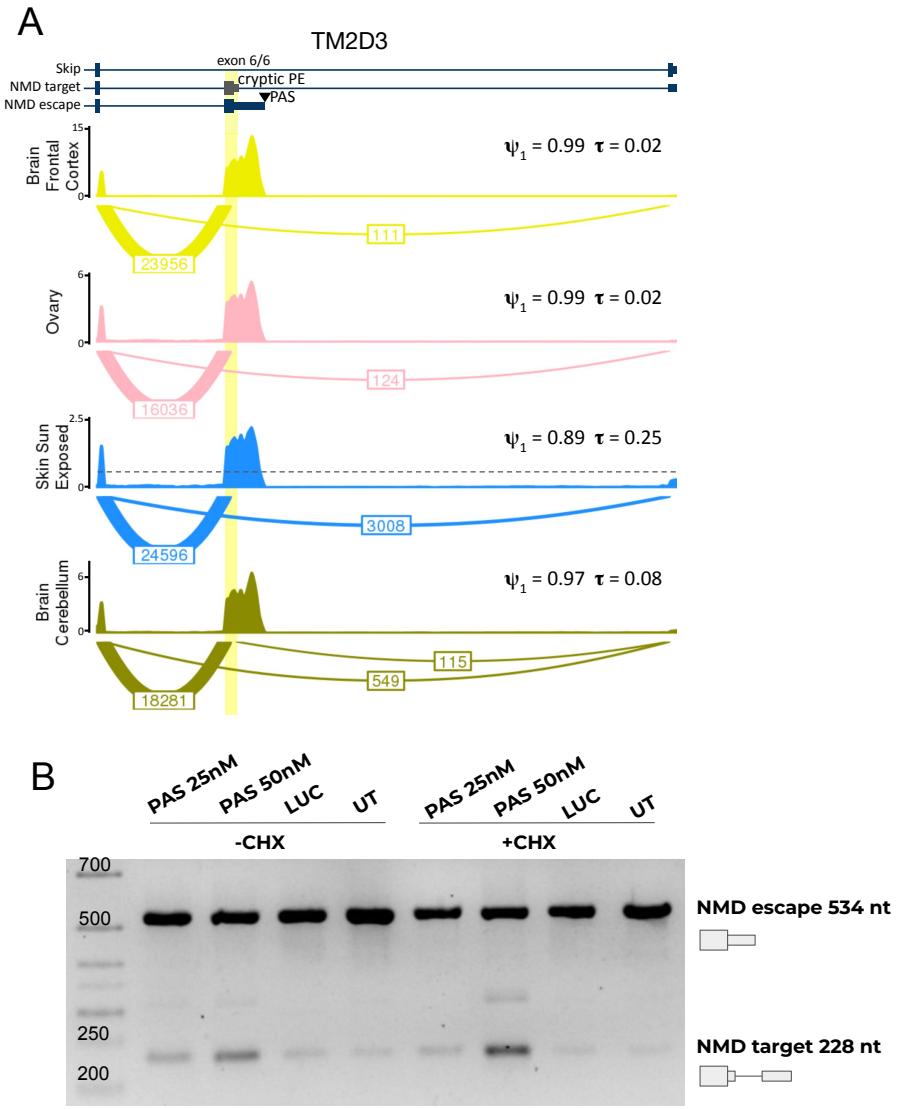

Figure S12: NMD escape in the *TM2D3* gene. **(A)** Read coverage profiles and sashimi plots illustrating a tissue-specific switch from the NMD-escape to the skip isoform in the *TM2D3* gene. The poison exon followed by a PAS is highlighted. The vertical axes represent read coverage density, and the numbers on the sashimi plots indicate split read counts summed across samples from each tissue. Transcript models from CHESS are shown at the top in blue, with the unannotated poison exon shown in gray. Dotted lines indicate the drop in read coverage between the upstream and downstream exons. **(B)** Endpoint RT-PCR analysis of untreated cells (UT), cells treated with ASO targeting luciferase (LUC), and ASO targeting the cleavage site (PAS) at two concentrations (25 and 50 nM), with (+CHX) or without (-CHX) cycloheximide treatment.

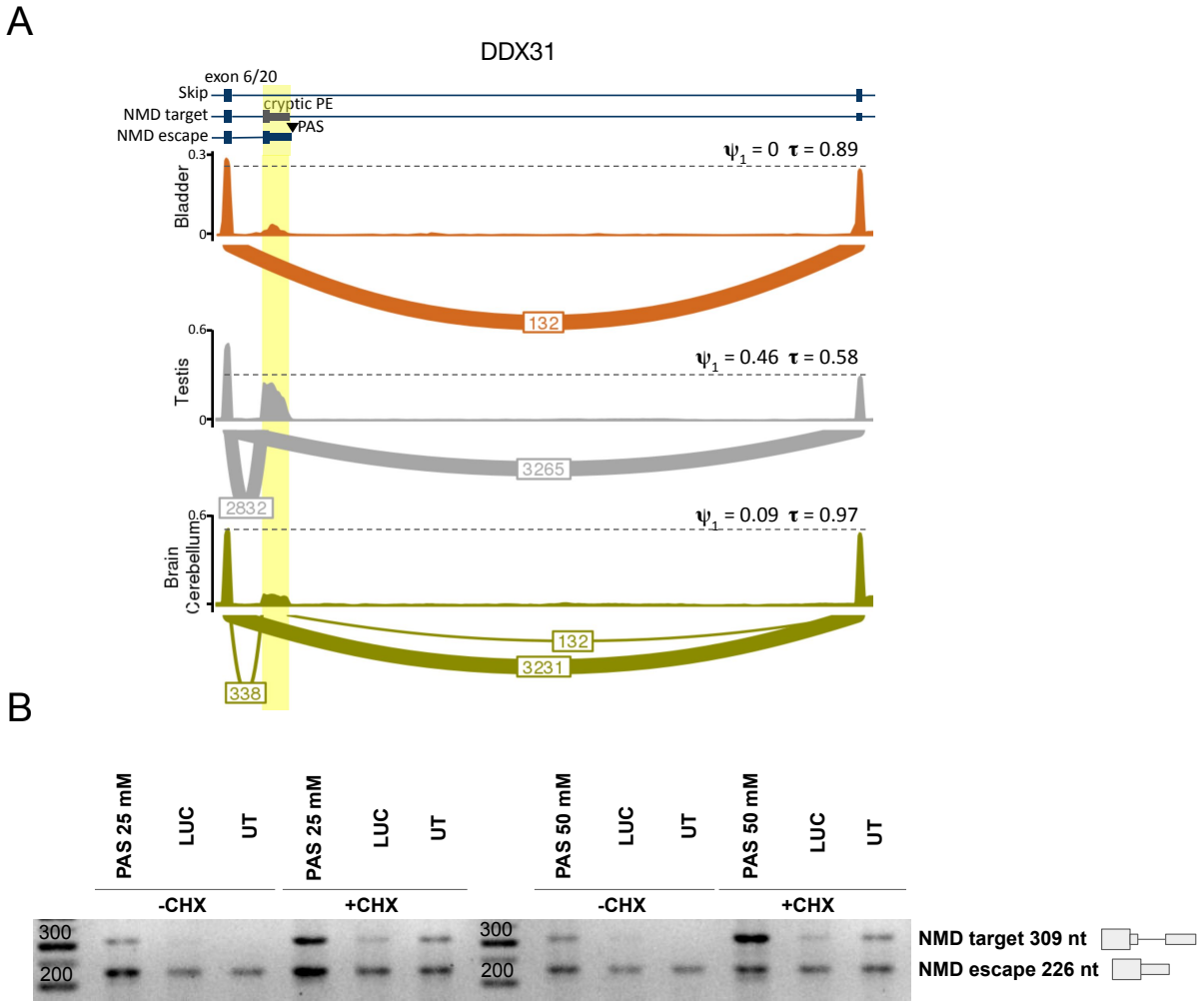

Figure S13: NMD escape in the *DDX31* gene. **(A)** Read coverage profiles and sashimi plots illustrating a tissue-specific switch from NMD-escape to skip isoform in the *DDX31* gene. The poison exon followed by the PAS is highlighted. Vertical axes, numbers on the sashimi plots, and transcript models are as described in Figure S12. Dotted lines indicate the drop in read coverage between the upstream and downstream exons. **(B)** Endpoint RT-PCR for untreated cells (UT) and cells treated with ASOs against luciferase (LUC) or the cleavage site (PAS) at two concentrations (25 and 50 nM), with (+CHX) and without (-CHX) cycloheximide treatment.

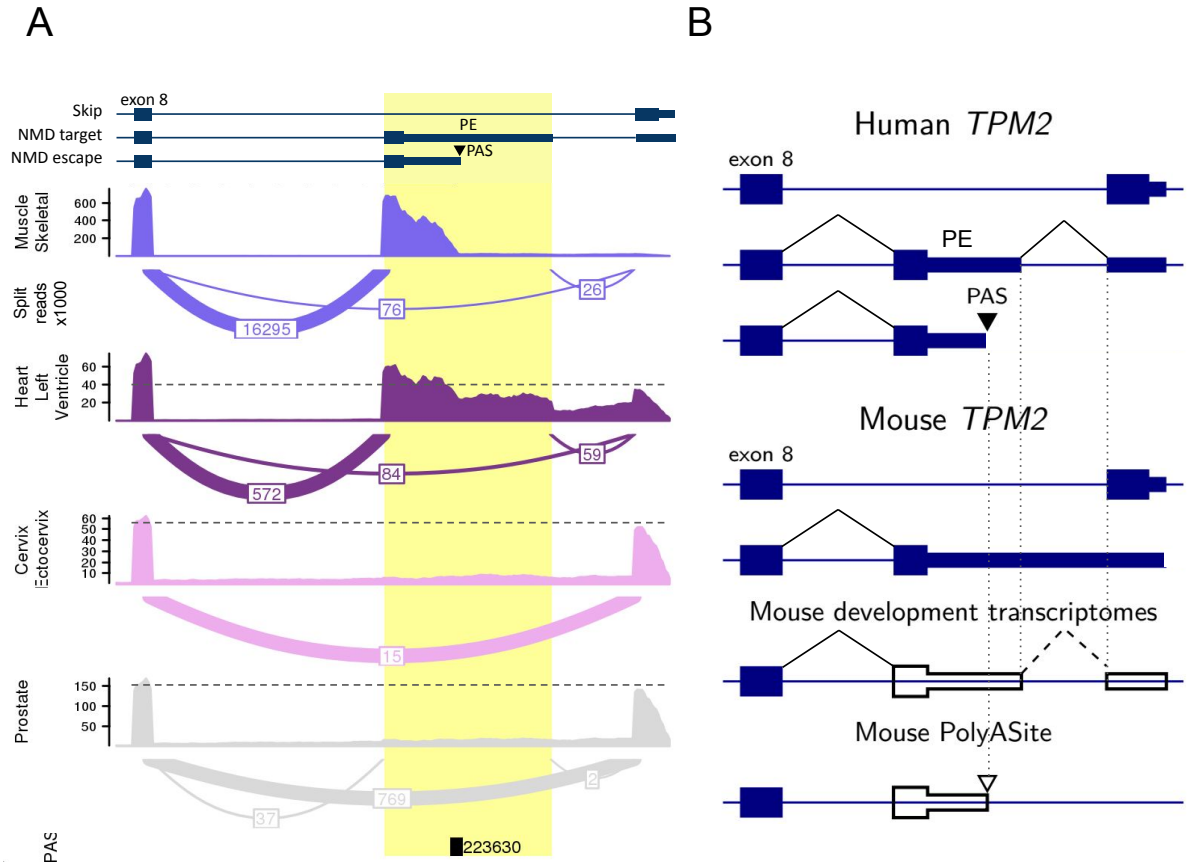

Figure S14: **(A)** Read coverage profiles and sashimi plots illustrating the predicted tissue-specific switch from NMD-escape to skip isoform in the *TPM2* gene. The poison exon followed by a PAS is highlighted. The vertical axes, numbers on the sashimi plots, and transcript models are as described in Figure S12. The PAS track indicates the number of poly(A) reads supporting each PAS. **(B)** Top: schematic representation of CHESS transcript models for the human *TPM2* gene. Middle: GENCODE (vM25) transcript models for the mouse *TPM2* gene. Bottom: transcript models derived from split reads in mouse embryonic development transcriptomes (ENCODE portal ID: ENCSR574CRQ) and from PAS in the mouse PolyASite database. The poison exon followed by the PAS is indicated (PE). Vertical dotted lines denote homologous positions. Unannotated exons are outlined, and dashed lines indicate unannotated splice junctions.

ASO	Sequence
VRK3 signal	mU*mA*mA*mA*mG*mA*mU*mU*mU*mU*mA*mU*mU*mA*mC*mU*mU*mA*mC*mA*
VRK3 PAS	mG*mA*mG*mG*mA*mA*mA*mG*mA*mU*mC*mC*mA*mA*mG*mA*mU*mG*mG*mA*
NFX1 PAS	mG*mU*mA*mC*mA*mA*mU*mA*mA*mA*mU*mA*mU*mG*mC*mA*mU*mU*mU*mC*
TM2D3 PAS	mG*mC*mC*mC*mA*mC*mA*mG*mU*mA*mU*mU*mU*mC*mU*mU*mA*mA*mC*mU*
DDX31 PAS	mA*mG*mC*mA*mU*mG*mA*mA*mU*mG*mG*mA*mU*mA*mG*mC*mU*mG*mC*mA*
LUC 2'OMe	mC*mU*mU*mA*mC*mG*mC*mU*mG*mA*mG*mU*mA*mC*mU*mU*mC*mG*mA*

Table S1: ASO sequences. RNA bases: G, A, U, and C. 2'-OMe bases: mG, mA, mU, and mC. Phosphorothioated RNA bases: G*, A*, U*, and C*.

Isoform	Primer	Sequence
Skip isoform	VRK3_F	AACAGAAGTTTGTTGATAAGCCGG
	VRK3_R	CTGAGGGCCTGATCCAGT
	NFX1_F	CAGAAGTGGTGCATGGGCAA
	NFX1_R	GTGCATCCCACATGGTAGCG
NMD-target isoform	VRK3_F_nmd	TCCCTGACCAGTGTTTGTTGAT
	VRK3_R_nmd	CTGAGGGCCTGATCCAGT
	NFX1_F_nmd	ATGGAAAACCAAACATCTTTCGGAG
	NFX1_R	GTGCATCCCACATGGTAGCG
NMD-escape isoform	VRK3_F_polyA	CACGAGCTCCCTGACCA
	VRK3_R_polyA	TGACCTCCCCACCCGT
	NFX1_F_polyA	TGTAAAAAATGTGATATCTATGAAATGC
	NFX1_R_polyA	TTTTTTTTTTTTTGCATTTCTTATTGGACTTC

Table S2: Primers for qPCR.

Primer	Sequence
VRK3_F_polyA	CACGAGCTCCCTGACCA
VRK3_R_nmd	TGTTCCTCAGCATGGCGTAG
VRK3_R_polyA	TGACCTCCCCACCCGT
NFX1_F	CAGAAGTGGTGCATGGGCAA
NFX1_R	GTGCATCCCACATGGTAGCG
NFX1_R_polyA	TTTTTTTTTTTTTGCATTTCTTATTGGACTTC
TM2D3_F_PE	ACGTCCTGCTCATTGGAGTT
TM2D3_R	GGACGCTCATCTGCAACAAC
TM2D3_R_polyA	CACATGACTGAAGCTGAGCC
DDX31_F_polyA	TCCTGGCTCCTGTGCTATGT
DDX31_F_PE	GCCATGCAGCAAGTAATGGAG
DDX31_R	ACACTCCAGGCACAATCCAG
DDX31_R_polyA	TTTTTTTTTTTTTAGCTGCAAAACCAAATCAC

Table S3: Primers for endpoint PCR (skip, NMD-target and NMD-escape isoforms).