# Poison exon splicing in the human brain: a new paradigm for understanding and targeting neurological disorders

Paolo Pigini[1,2,3,†], Huilin Xu[1,2,4,†], Yan Ji[1], Hannah Lindmeier[1], Henry R. Saltzman[1], Shuqi Yun[1], Christiano R. R. Alves[1,2], M. Catarina Silva[1,2], Dadi Gao[1,2,4*], Elisabetta Morini[1,2,3,*]

[1] Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

[2] Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

[3] Broad Institute of Harvard and MIT, Cambridge, MA, USA

[4] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

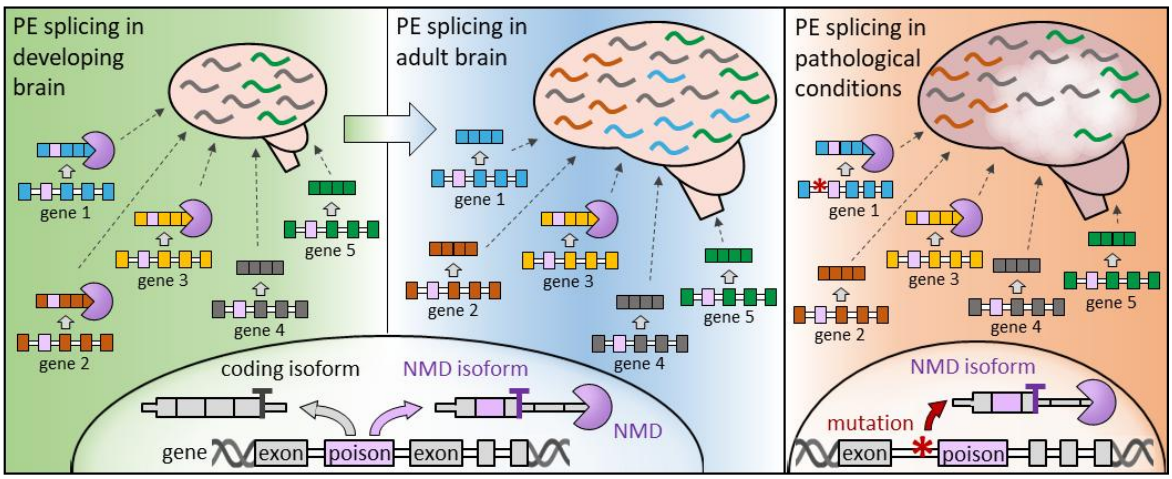
[†] These authors contributed equally to this work.

*Corresponding Authors:

Dadi Gao (dgao2@mgh.harvard.edu) and Elisabetta Morini (emorini@mgh.harvard.edu )

## Abstract

Poison exons (PE) are highly conserved exons whose inclusion creates premature termination codons (PTCs) and triggers nonsense-mediated decay (NMD) of the mature transcript. Despite their important role in post-transcriptional regulation, PEs remain poorly annotated due to the lack of systematic transcriptome-wide approaches. Here, we comprehensively investigate the function of PEs in the human brain under physiological and pathological conditions. A comparative analysis of 957 eukaryotic transcriptomes revealed that humans exhibit the highest enrichment of NMD-targeted isoforms. By annotating all splicing events that generate PTCs, we found 12,014 PEs in the human genome. Using GTEx and BrainSpan RNA-seq dataset, we compared PE splicing across tissues and developmental stages and identified 117 PEs uniquely found in the human brain, 1,214 PEs with brain-specific differential splicing, and 1,610 PEs differentially spliced during brain development. By integrating ClinVar variant annotations with SpliceAI predictions, we identified 1,877 pathogenic variants associated with neurological disorders that are predicted to affect the splicing of 891 PEs, and we functionally validated the impact of a subset of them using CRISPR prime-editing in human cells. Our findings highlight PEs as pivotal regulators of gene expression in the human brain and support the therapeutic targeting of PE splicing in neurological diseases.




**Graphical abstract.**

## Introduction

Alternative mRNA splicing is a fundamental post-transcriptional regulatory mechanism that significantly expands the coding potential of eukaryotic genomes. Recent studies suggest that disruptions in this process contribute to approximately 10-30% of human genetic disorders [1-4], with nearly half of these alterations occurring in the intronic regions [5]. Alternative splicing is particularly important in the brain, where it plays critical roles in neurodevelopment, neuronal differentiation, and synaptic function [6-11]. This regulatory mechanism helps define the differentiation trajectories of various cell types within the nervous system [9, 12, 13] and appears to have undergone substantial evolutionary diversification in vertebrates, especially primates [14]. Consequently, splicing alterations have been increasingly implicated in the pathogenesis of neurological disorders [15, 16], with non-coding variants emerging as particularly significant contributors [17].

Poison exons (PE) are naturally occurring, evolutionarily conserved exons that introduce premature termination codons (PTCs) when included in the mature transcript, thereby triggering nonsense-mediated decay (NMD) of the mature transcript and reducing protein

expression [18-25]. PEs act as finely tuned post-transcriptional regulators of gene expression [26] and have been implicated in both physiological and pathological processes [27]. In murine models, PE-induced NMD has been proposed to downregulate non-neuronal genes during brain development [28]. In humans, they contribute to key cellular functions, including lineage specification, neuronal differentiation, and neuroprotection [7, 29], and are also linked to disease mechanisms such as cancer [27]. In the human brain, PEs help establish cellular identity during differentiation and play critical roles in the temporal and cell type–specific regulation of synaptic proteins, neuron-specific RNA-binding proteins [24, 30, 31], and chromatin modifiers [32]. Notably, the splicing of PEs in the brain is regulated by a network of RNA-binding splicing regulators, including *RBFOX1*, *PTB1* and *NOVA1* [6, 7, 31, 33].

The dysregulation of PE splicing has been implicated in various neurological disorders [28]. For example, aberrant inclusion of the PE "20N" in the *SCN1A* gene is associated with Dravet syndrome (DS) [34-37]. An intronic variant disrupting normal skipping of a PE in the *FLNA* gene was found in a family segregating with periventricular nodular heterotopia (PVNH), a condition caused by abnormal neuronal migration during fetal brain development [7]. In amyotrophic lateral sclerosis (ALS), *TDP-43* gene dysfunction disrupts PE regulation in iPSC-derived neurons, potentially contributing to disease pathogenesis [38, 39]. Beyond their emerging role in disease etiology, PEs are also increasingly recognized as potential novel therapeutic targets for the treatment of neurological disorders [40-43].

Recent advances in computational approaches have improved identification of both PEs [44] and pathogenic variants affecting their splicing [2]. However, the field still lacks a comprehensive catalog of PE splicing events [45-47]. In this study, we conducted an evolutionary analysis of NMD-targeted transcripts and demonstrated that humans exhibit a striking enrichment of genes expressing transcripts containing PEs. We leveraged large RNA-seq datasets from GTEx and BrainSpan projects to identify human brain-specific and

neurodevelopmentally regulated PEs. By integrating ClinVar variant annotations with SpliceAI predictions, we identified pathogenic variants that disrupt PE splicing and are associated with neurodevelopmental and neurodegenerative disorders. To functionally validate the impact of a subset of these variants on PE splicing, we applied an optimized CRISPR prime editing workflow [48] in human cell lines. Together, this work provides new insights into the role of PEs in brain function and disease, and highlights their potential as therapeutic targets for neurological conditions.

## Results

### The human transcriptome exhibits a unique enrichment of NMD isoforms

Conventionally, NMD mRNA isoforms are defined by the presence of a PTC located more than 50 base pairs (bp) upstream of the last exon–exon junction [21-24] (Fig. 1A). Using these criteria, we systematically identified NMD isoforms across a dataset of 957 annotated eukaryotic genomes from Ensembl (Release 60 and 113), spanning the major clades of vertebrates, invertebrates, fungi, protists, and plants (Fig. 1B and Supplementary Table S1). This large-scale dataset enabled us to evaluate the evolutionary distribution and potential functional relevance of NMD across eukaryotes. On average, 2.67% of coding genes per genome were predicted to produce at least one NMD isoform (or NMD-regulated genes), though this proportion varied substantially across evolutionary lineages. Vertebrates (3.7%) and plants (3.7%) exhibited the highest fraction of coding genes producing at least one NMD isoform, followed by invertebrates (1.9%), while lower percentages were observed in fungi (0.6%) and protists (0.1%) (Fig. 1B and Supplementary Fig. S1A). A similar trend was observed when analysing the percentage of NMD isoforms over all isoforms in each species, with vertebrates (2.4%), invertebrates (2.0%), and plants (3.7%) showing higher values compared to protists (0.1%) and fungi (0.7%) (Supplementary Fig. S1B). We further evaluated

the average percentage of NMD isoforms per gene in each species. In this case, we found that vertebrates (60.5%) and invertebrates (80.2%) display the highest values, followed by plants (45.6%), fungi (30.3%) and finally protists (18.1%) (Supplementary Fig. S1C). These data suggest that alternative splicing leading to NMD is particularly prominent in multicellular lineages with complex developmental processes. Among all species analysed, humans emerged as the species with the highest number of NMD isoforms. We found that 39.6% of human protein-coding genes give rise to at least one NMD isoform. Furthermore, in humans, NMD isoforms represent 20.1% of all coding isoforms, while an average of 35.0% of isoforms for a given gene are predicted to undergo NMD. These findings strongly suggest that NMD plays a strikingly large role in regulating transcript stability in humans, potentially reflecting the complexity and regulatory precision of human gene expression.

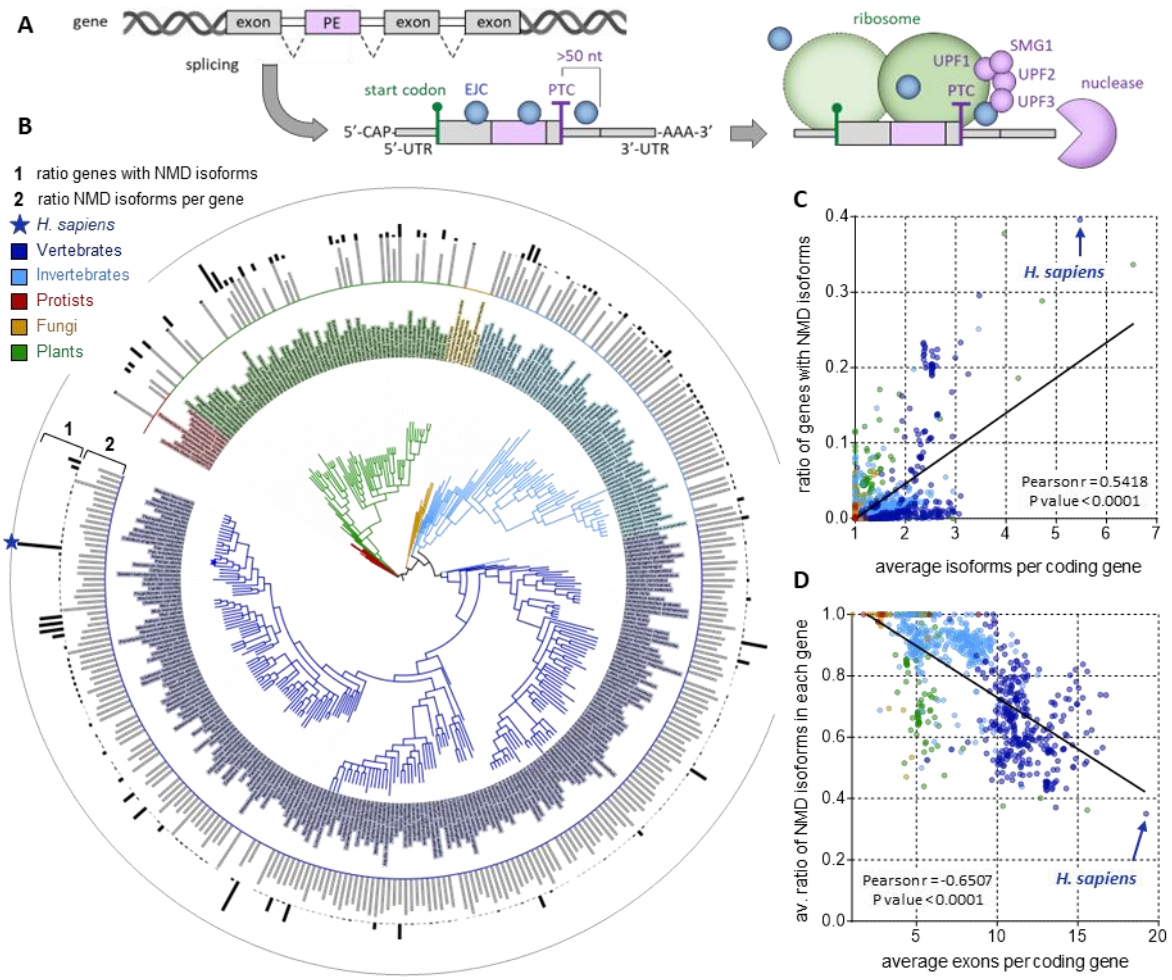



**Figure 1.** Evolutionary analysis of NMD isoforms across eukaryotes. (**A**) Schematic representation of a transcript carrying a PTC located more than 50 nt upstream of the last exon-exon junction, which triggers NMD. The ribosome stalls at the PTC and, unable to displace the exon-junction complex (EJC), triggers NMD via factors UPF1, UPF2, UPF3 and SMG1. (**B**) NMD isoforms annotated throughout the genomes of 957 eukaryotic species among vertebrates, invertebrates, protists, fungi and plants. For each species, bars indicate either the ratio of coding genes with at least one annotated NMD isoform (Track "1", highest value: 0.40), or the average ratio of annotated NMD isoforms over all isoforms of each coding gene (Track "2", highest value: 1.00). (**C**) The ratio of coding genes with at least one annotated NMD isoform in each species is compared with the average number of isoforms annotated for each coding gene. (**D**) The average ratio of NMD isoforms annotated for each coding gene is compared with the average number of exons per gene in each species. Colour codes in (C) and (D) are the same as in (B). Simple linear regression is shown for each scatter plot as a black line, with the Pearson correlation coefficient (r) and p-value (p).

To explore the evolutionary trend of NMD occurrence, we investigated whether it could be associated with the complexity of the genome and transcriptome. Surprisingly, we observed no striking correlation between the total number of coding genes and the percentage of NMD-

regulated genes across species (Supplementary Fig. S2A) or the frequency of NMD isoforms (Supplementary Fig. S2B and S2C). Notably, vertebrates often displayed higher percentages of NMD-regulated genes despite having fewer genes overall compared to plants, suggesting that genome size alone does not explain NMD frequency. In contrast, we found a positive association ($p < 0.0001$ in all comparisons) between transcriptome complexity and NMD abundance. Species with a greater number of isoforms per gene tended to have more NMD-regulated genes (Fig. 1C and Supplementary Fig. S2D) and more frequent NMD isoforms (Supplementary Fig. S2E and S2F). This pattern may reflect the fact that higher occurrence of alternative splicing may naturally increase the probability of producing NMD isoforms. We next examined whether exon density correlates with NMD occurrence. We observed a significant positive correlation ($p < 0.0001$ in all comparisons) between the average number of exons per gene and the number of NMD-regulated genes (Supplementary Fig. S2G) and the frequency of NMD isoforms (Fig. 1D and Supplementary Fig. S2H and S2I), consistent with the idea that splicing complexity increases opportunities for NMD-generating splicing events. Finally, we found a significant positive correlation ($p < 0.0001$ in all comparisons) between the average transcript length in each species and both the ratio of NMD-regulated genes (Supplementary Fig. S2J) and the frequency of NMD isoforms (Supplementary Fig. S2K and S2L), particularly when considering the average ratio of NMD isoforms per species. However, this could be explained by the previously observed correlation between NMD occurrence and the number of exons per gene.

Our evolutionary analysis highlights that NMD isoforms are broadly expressed across eukaryotes and are particularly enriched in vertebrates, especially in humans. Humans exhibit the highest density of NMD events among all surveyed species. Our findings also suggest that the occurrence of NMD isoforms has been favoured by increasingly complex transcriptomes over evolution.

**Comprehensive classification of alternative splicing events involving PEs in human transcriptome**

Building on the finding that humans exhibit a high frequency of NMD isoforms, we next sought to systematically annotate the full repertoire of PEs in the human transcriptome. PEs are highly conserved exon cassettes whose inclusion introduces PTCs in the mature transcript, thereby triggering NMD and reducing protein levels. To identify these exons at scale, we developed an *in silico* strategy using genomic sequences and exon annotations (Supplementary Fig. S3A, see Methods). We defined a PE as an exon that: (1) is ≤ 300 bp in length; and (2) its inclusion generates a PTC more than 150 bp downstream of the transcription start site and upstream of the last two exons. This stringent definition of PTC location was specifically chosen to minimize the likelihood of NMD escape [49], a mechanism by which some NMD transcripts avoid degradation. Using this approach, we identified 12,014 putative PEs, accounting for 21.8% of 55,150 annotated alternative splicing events in the human genome, each potentially producing NMD isoforms upon inclusion (Fig. 2A and Supplementary Table S2). A comparative analysis across model organisms, including *Mus musculus*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, and *Arabidopsis thaliana*, revealed species-specific differences in PE frequency, with humans having the highest number of putative PE splicing events (Fig. 2A).

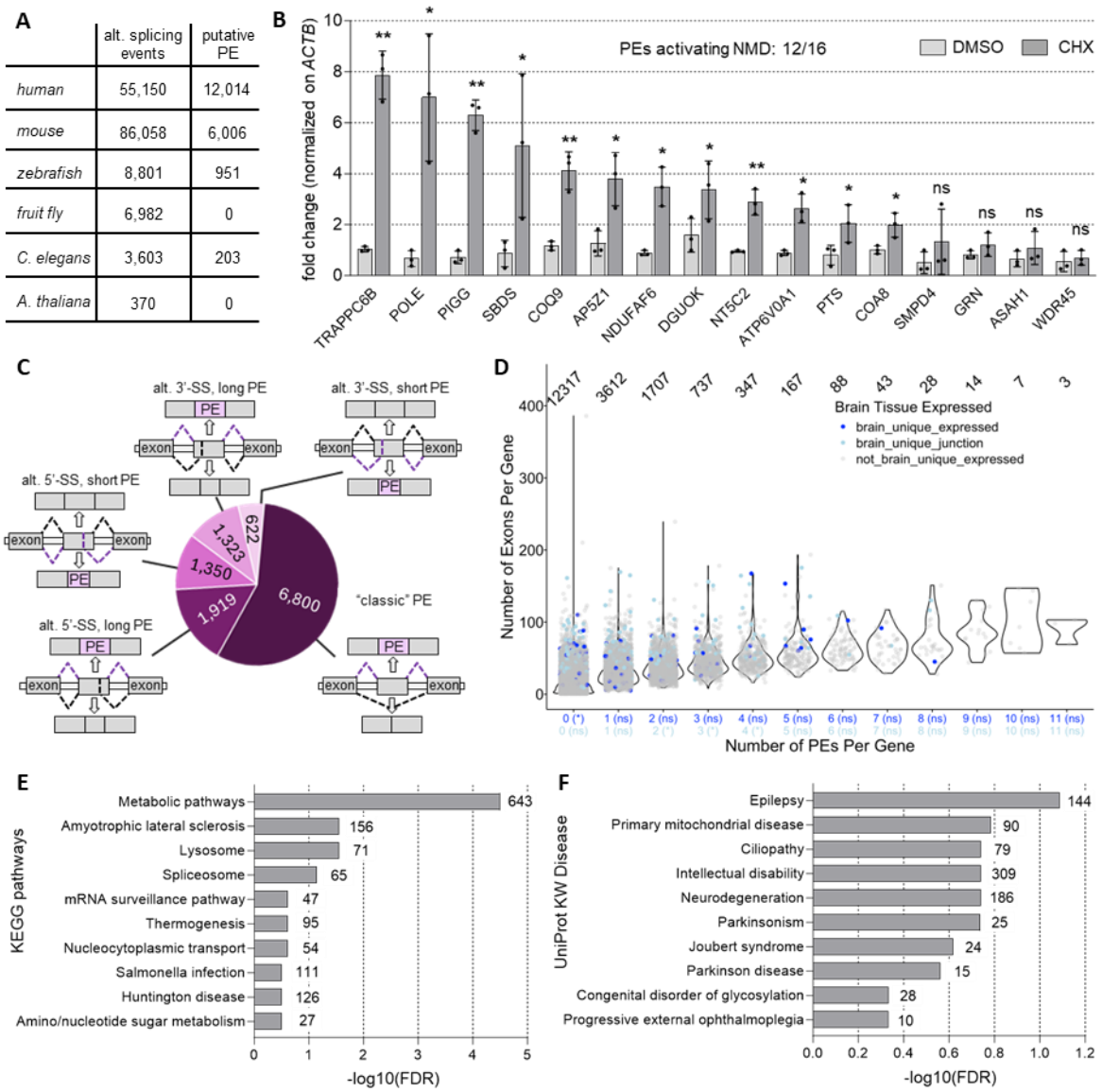



**Figure 2**. Analysis of putative PE splicing events. (**A**) Alternative splicing events involving putative PEs are counted for the genome annotation of different biological models. (**B**) Sensitivity to NMD assessed by CHX treatment in HEK293T cells for 16 putative PEs. Cells were treated with 50 µg/mL CHX (or DMSO), and the increase of transcript levels was evaluated by RT-qPCR. All values are internally normalized on *ACTB*, while transcript levels are represented as fold change compared to the DMSO. Statistical significance by Student's t-test is shown for every comparison. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$, ns: not significant. Data are shown as average ± standard deviation. Each data point represents one of three replicates. (**C**) Different types of alternative splicing events among putative PEs in humans. They include canonical PE exclusion or inclusion events and alternative 3'- and 5'-splice-sites. (**D**) Number of unique putative PEs counted for each annotated gene in the human genome (x-axis) against the total number of exons in each gene (y-axis). Genes are divided into three groups: 1) only expressed in the brain (dark blue); 2) expressed in multiple tissues but have a splice junction only used in the brain (light blue); and 3) others (grey). The brackets on the x-axis indicate whether PE inclusion? is significantly enriched in gene groups. (**E**) Pathway

and (**F**) disease enrichment analysis by DAVID tool for human genes that include PEs. Numbers indicate the total count of genes identified in each category.

To evaluate the robustness of our PE identification, we then assessed *in vitro* whether their inclusion indeed triggers NMD by measuring transcript levels after treating human cells with cycloheximide (CHX), which inhibits translation and stabilizes NMD-targeted mRNAs [28, 50, 51]. We performed this analysis in HEK293T cells for a subset of identified transcripts containing putative PEs. RT-qPCR following CHX or DMSO treatment revealed that 75% of transcripts containing PEs were significantly stabilized by CHX, confirming their degradation via NMD (Fig. 2B and Supplementary Fig. S3B). Transcripts without CHX-induced upregulation likely represent cases in which downstream splicing events restore the reading frame.

More than half of these PE splicing events (6,800 events, or 56.6%) arise from exons that are either included or skipped, while the remaining result from alternative 3′ or 5′ splice site usage (Fig. 2C and Supplementary Fig. S3C). PEs did not show a preferential position along the gene (Supplementary Fig. S3D). While genes expressed in the brain, particularly those involved in neuronal and synaptic functions, are well-known to be longer [52-54], we did not observe them to contain a higher number of PEs (Fig. 2D). In fact, brain-expressed genes had fewer PEs than non-brain genes (Fig. 2D), suggesting a requirement for strict PE regulation in the brain transcriptome.

To gain insight into the biological functions of genes producing transcripts with PEs, we conducted pathway and disease enrichment analyses. KEGG pathway analysis revealed a significant enrichment (FDR = 0.00003, 643 genes) in metabolic pathways (Fig. 2E and Supplementary Table S3). We also found an enrichment for functions related to the spliceosome (65 genes) and mRNA surveillance (47 genes). Several splicing regulators were among the enriched genes, such as SRS [55] factors and SF3 [56] factors. These findings are

consistent with the well-established data indicating that many splicing factors regulate their own expression through negative feedback loops mediated by PE inclusion [19, 27, 57]. Additionally, disease ontology enrichment using the UniProt KW Disease database (Fig. 2F and Supplementary Table S4) identified a preferential enrichment of genes containing PE in neurological disorders, most prominently epilepsy (144 genes), intellectual disability (309 genes) and neurodegeneration (186 genes), reinforcing the hypothesis that PE splicing plays a key regulatory role in brain functions [24, 30, 31].

**PEs exhibit brain-specific and spatio-temporal regulation during neurodevelopment**

To investigate the spatiotemporal regulation of PE splicing in the human brain, we leveraged large-scale transcriptomic datasets from both adult and developing tissues. We hypothesized that dynamic inclusion or exclusion of PEs may contribute to brain-specific gene regulation. We first analysed RNA-seq data from the Genotype-Tissue Expression (GTEx) Project, which encompasses 17,382 libraries derived from 54 distinct tissues, including 13 brain regions from 838 donors without major chronic diseases [58]. This analysis (see Methods) revealed that 1,214 PEs are differentially spliced in the brain compared to other tissues (Fig. 3A and Supplementary Table S5), and 117 PEs are uniquely detected in the brain. As anticipated, inclusion of these PEs correlated inversely with expression of their host genes (Supplementary Fig. S4), consistent with their role in post-transcriptional downregulation via NMD. To explore the functional implications of these brain-regulated PEs, we conducted gene ontology (GO) analyses by stratifying them based on PSI values, using 0.5 as a PSI cut-off between transcripts with low PE inclusion (PSI < 0.5) and high PE inclusion (PSI > 0.5), and found that genes showing low PE inclusion in the brain are significantly enriched in RNA metabolism and splicing-related pathways (Fig. 3B and Supplementary Table S6).

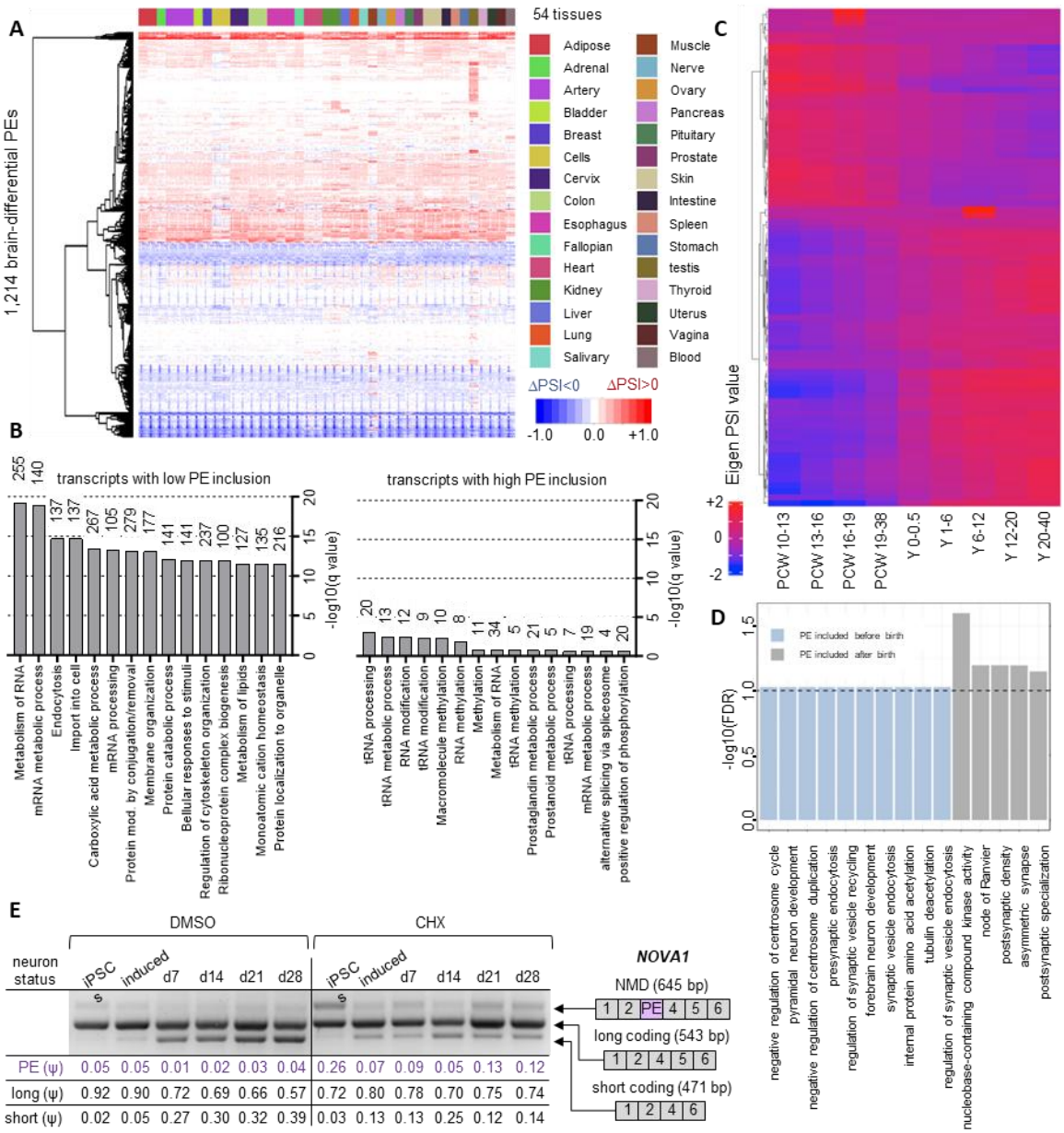



**Figure 3**. Analysis of PE splicing in the human brain. (**A**) Heatmap showing differential PSI (ΔPSI) of PEs across 54 GTEx tissues relative to brain. Each column represents a tissue, and each row represents one of the 1,214 PEs identified as differentially spliced in the brain. Negative values indicate higher PE inclusion in the brain, while positive values indicate lower inclusion. Hierarchical clustering is based on ΔPSI values. (**B**) GO and pathway enrichment analysis for host genes of with low PE inclusion (PSI < 0.5) or high PE inclusion (PSI > 0.5) in the brain. The top 15 enrichment groups (ranked by statistical q value) are shown for each analysis. (**C**) Heatmap of differential splicing of PEs across brain developmental stages (BrainSpan database). Each column represents a developmental stage, and each row represents a PE. Colours indicate splicing change (represented as eigen PSI value), with positive values indicating increased PE inclusion and negative values indicating decreased inclusion (see Methods). (**D**) GO and pathway enrichment analysis for host genes of PEs showing higher splicing inclusion before (light blue) or after (grey) birth. The top enrichment groups with FDR < 0.1 are shown for each analysis. (**E**) Splicing analysis of a brain-specific PE in the *NOVA1*

gene across neuronal differentiation, including undifferentiated iPSCs, induced neurons following 3 days of NGN2 overexpression, and neurons at 7, 14, 21 or 28 days of maturation. Splicing isoforms corresponding to each band are indicated on the right, and band identity was confirmed by Sanger sequencing. Cells were treated either with DMSO (vehicle) or 50 μg/ml CHX.

We then analysed the developmental regulation of PEs using the BrainSpan Atlas, a comprehensive resource comprising 607 libraries from 41 individuals across 16 distinct brain regions. These libraries span key developmental stages, ranging from 10 post-conception weeks (pcw) to 40 years of age. This analysis (see Methods) identified 1,610 differentially spliced PEs in at least one of the 16 brain regions (Supplementary Table S7). Clustering of these PEs revealed distinct developmental splicing patterns, characterized by higher PE inclusion before or after birth (Fig. 3C). GO enrichment of the clustered PEs with higher inclusion before birth revealed significant associations with neuronal development functions, while clustered PEs with higher inclusion after birth showed involvement in synapses development functions (Fig. 3D and Supplementary Table S8). It is worth noting that the PE inclusion showed both linear and non-linear trajectories and were region specific (Supplementary Fig. S5A-Q and Supplementary Table S9).

Finally, our analyses showed that splicing factors, including *RBFOX1*, *NOVA1*, *TARDBP*, and *UPF3B*, were significantly enriched for brain-specific PEs. We then investigated whether PE inclusion in the transcripts of these splicing factors indeed triggers NMD and changes during neuronal development, using iPSC-derived i3N neurons, a model of human neurons that differentiate via *NGN2* induction [59]. We specifically selected this neuronal cell model because the PSI values of the PE splicing better mirror those observed in the human brain (Supplementary Fig. S6A). After *NGN2*-mediated induction and a differentiation period of up to 28 days (Supplementary Fig. S6B and S6C), we assessed PE splicing by RT-PCR. We found that PE inclusion in *NOVA1* was higher in iPSCs but declined during neuronal maturation, coinciding with increased expression of coding isoforms [60], which was more evident after

CHX treatment (Fig. 3E). For splicing factor *TARDBP* we were able to validate that the inclusion of the PE triggered NMD, although no significant differentiation-specific splicing changes were observed, while for *RBFOX1* and *UPF3B* we could not detect the presence of the respective PEs at any differentiation stages (data not shown). Of note, consistently with the result in HEK293T cells, analysis of CHX-treated neurons confirmed that for the majority of putative PEs (75%) indeed trigger NMD (Supplementary Fig. S6D and S6E).

Our data reveal that PE splicing is highly dynamic across brain regions and developmental stages, suggesting a widespread mechanism of post-transcriptional regulation in the nervous system. The enrichment of PE-regulated genes in RNA metabolism and neurodevelopmental pathways, along with validation of key neuronal splicing regulators, supports a model in which PEs act as molecular switches to fine-tune gene dosage with temporal and spatial precision.

**Identification of pathogenic variants associated with neurological disorders and disrupting PE splicing**

To investigate the role of PE mis-splicing as a pathogenic mechanism for brain disorders, we sought to identify variants affecting PE splicing. We analysed all pathogenic and likely pathogenic variants in the ClinVar database [61] and used SpliceAI [2] to predict their impact on mRNA splicing. As expected, most pathogenic and likely pathogenic variants predicted to impact mRNA splicing (SpliceAI score > 0.2) were located in intronic regions: 45,691 intronic versus 22,991 exonic variants (Supplementary Fig. S7). For each variant, we assigned a positive or negative SpliceAI score to indicate the direction of the effect: positive scores indicate that the variant increased PE inclusion, and negative scores indicate that the variant decreased PE inclusion (significant absolute value > 0.2). We then compared? these predictions with PSI values from healthy brain tissues (GTEx) to prioritize variants significantly affecting PE splicing in the brain. We specifically considered two scenarios (see Methods): (1) a PE with

a PSI < 0.3 in the wild-type, for which the variant is predicted to increase inclusion; and (2) a PE with a PSI > 0.7 in the wild-type, for which the variant is predicted to decrease inclusion. Using this approach, we identified 1,877 pathogenic variants predicted to affect the splicing of 891 PEs (Fig. 4A and Supplementary Table S10). We then mapped these variants to associated diseases and clinical phenotypes using the ODiseA ontology. Notably, a large fraction of these variants was associated with seizures (665 variants) and epileptic syndromes (315 variants), including DS, consistent with previous findings of PE mis-splicing as a pathogenetic mechanism in these disorders[22,26] (Fig. 4B). In addition, we observed enrichment in genes associated with various neurodevelopmental (e.g. Joubert syndrome and microcephaly) and neurodegenerative conditions (e.g. Alzheimer's disease and Parkinson's disease). These findings suggest that altered PE splicing may represent a common mechanism contributing to a wide range of neurological disorders.

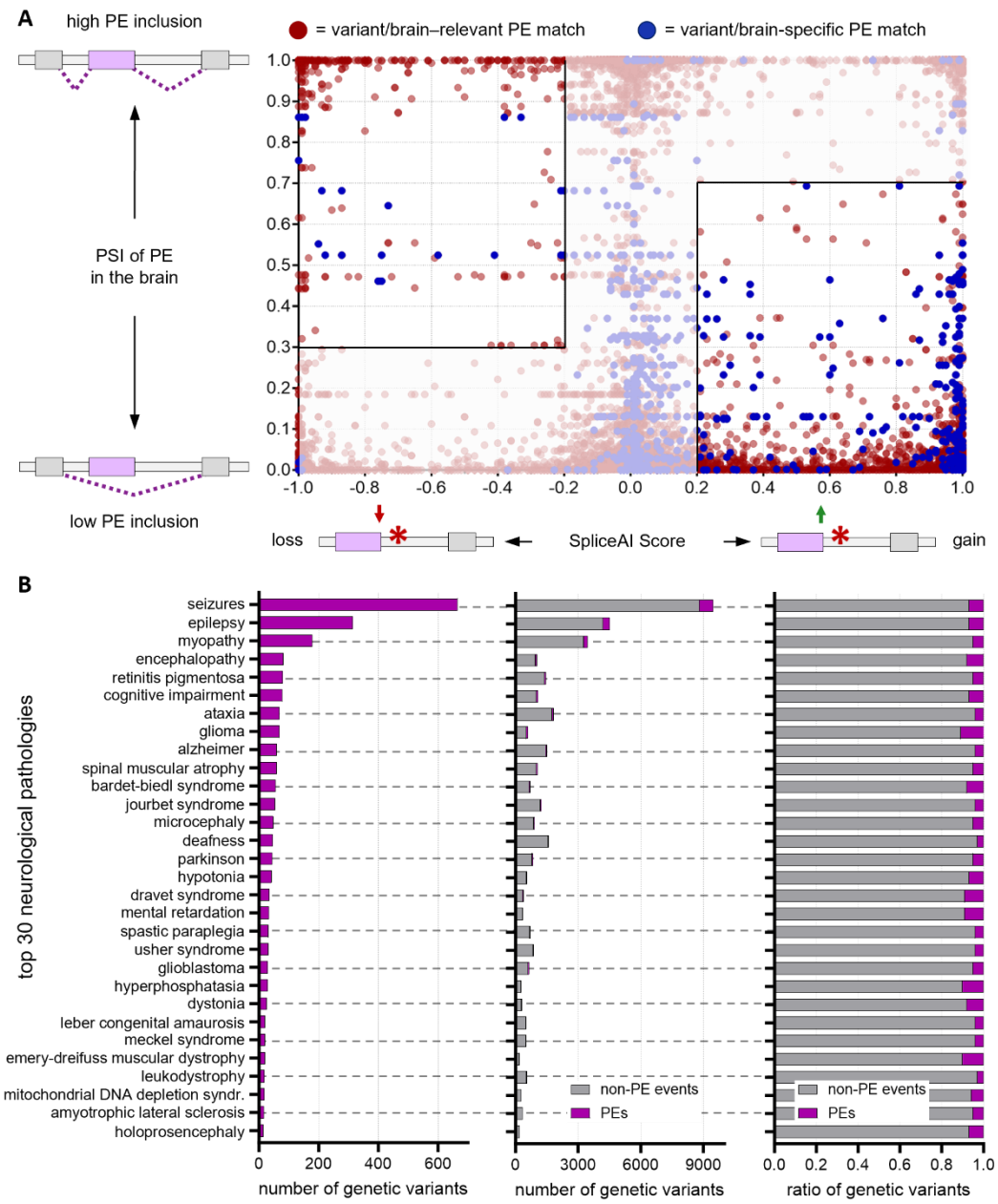



**Figure 4**. Pathogenic variants affecting PE splicing in neurological disorders. (**A**) Prioritizing pathogenic ClinVar variants based on its SpliceAI scores (x-axis) and the PSI values of PE splicing event in healthy human brain(y-axis). Each dot represents a variant/PE match. Blue dots represent PEs that are differentially spliced in the brain, while red dots represent PEs whose host gene is involved in neuronal functions as per GO. PSI were calculated in healthy brain tissues from the GTEx database. SpliceAI score have been assigned with a positive or a negative value depending on whether the mutation increases or decreases PE inclusion, respectively. Highlighted areas indicate variants that are highly likely to promote PE exclusion (GTEx PSI>0.7, SpliceAI score<-0.2) or inclusion (GTEx PSI<0.3, SpliceAI score>+0.2). (**B**) Neurological disorders associated with pathogenic variants affecting PE splicing are ranked by the number of such variants. Disease associations come from GO analysis and the ODiseA pathology database. The left panel shows the number of genetic variants affecting PEs. The other two panels demonstrate the number (middle) and the ratio (right) of variants affecting

PEs, compared to the number of those affecting non-PE splicing events. In all panels, pathologies are ranked based on the number of variants affecting PE splicing.

To functionally validate the effects of pathogenic variants on PE splicing, we developed and optimized a CRISPR prime editing workflow (Supplementary Fig. S8A, see Methods) [48, 62-64]. As proof-of-concept, we designed and tested 13 pegRNAs to introduce variants predicted to affect PE splicing in 12 different genes. Editing efficiency was assessed in HEK293T cells, with 10 of the 13 achieving >5% efficiency, and 9 exceeding 10% (average: 23.5%; Supplementary Fig. S8B). The effect of these variants on PE splicing was further assessed by RT-PCR in cells treated with CHX, allowing visualization of PE splicing events that would otherwise be reduced by NMD. Importantly, in all 10 selected gene, the introduction of the pathogenic variant produced the predicted change in PE inclusion (Supplementary Fig. S8C). Here, we report five representative examples, selected based on clinical relevance. Variants in these genes were predicted to have a distinct effect on PE splicing. Two mutations were predicted to increase PE inclusion. Of these, *TAFFAZIN* c.284+110G>A, located in intron 3 [65], was predicted to enhance PE inclusion(SpliceAI score: 0.57) and result in a +0.12 PSI change in a bulk-edited population with an average of 15% editing efficiency (Fig. 5A). *NDUFAF6* c.420+784C>T, associated with Leigh syndrome and located downstream of exon 3b [66], was predicted to increase PE inclusion (SpliceAI: 0.23) and was validated with a +0.13 PSI increase in a bulk-edited population with an average of 41% editing efficiency (Fig. 5B). We also validated two variants predicted to indirectly increase PE inclusion by disrupting canonical exon recognition, thereby favouring PE-containing isoforms. The *SLC12A6* c.2633-1G>A variant, located upstream of exon 23 had a SpliceAI score of 0.99 for loss of canonical splicing and produced a modest but reproducible increase in PE inclusion (+0.04 PSI, 21% editing efficiency; Fig. 5C). Similarly, *TSC2* c.4850-1G>A [67], was predicted to disrupt exon 33 splicing (SpliceAI: 0.99) and resulted in a +0.06 PSI change in PE inclusion (10% editing

efficiency; Fig. 5D). Finally, we also investigated one variant that results in decreased PE inclusion: *COQ9* c.521+1del in [68], disrupted the canonical donor site of exon 4. SpliceAI predicted a reduction in PE inclusion (score: 0.67), which was confirmed by a –0.28 PSI change in a bulk-edited population with an average of 29% editing efficiency (Fig. 5E). Together, these results highlight the diverse ways in which pathogenic variants can impact PE splicing, either by enhancing or suppressing PE inclusion, or by disrupting canonical exon recognition.

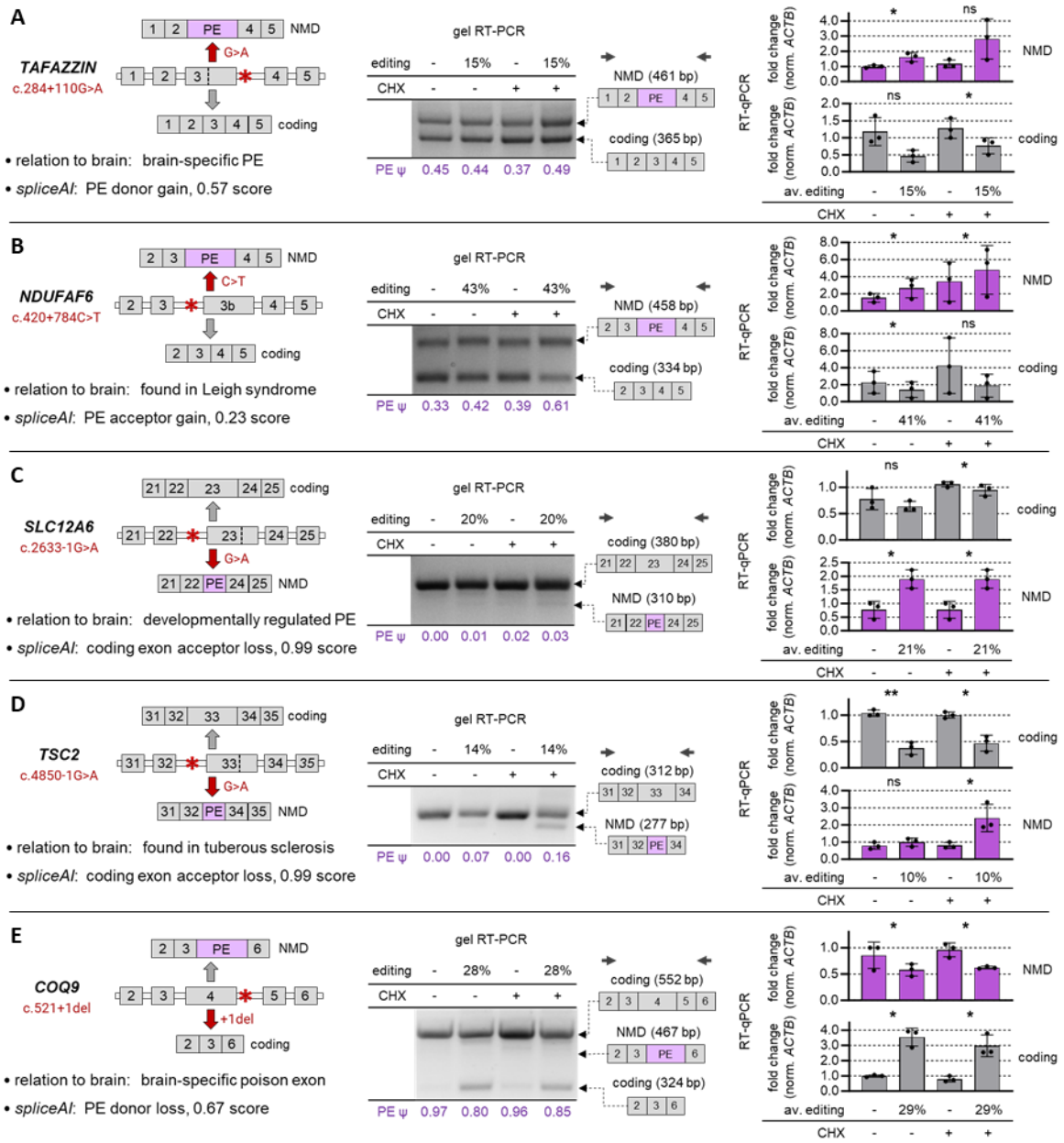


**Figure 5**. Validation of the effect of pathogenic variants on PE splicing. (**A**) c.284+110G>A, affecting E3 of *TAFFAZIN*; (**B**) c.420+784C>T, affecting E3b of *NDUFAF6*; (**C**) c.2633-1G>A, affecting E23 of *SCL12A6*; (**D**) c.4850-1G>A, affecting E33 of *TSC2*; (**E**) c.521+1del, affecting E4 of *COQ9*. The following information is reported for each variant: genomic position, disease association, spliceAI prediction, and expected RNA isoforms. Central panels show RT-PCR analysis of splicing in HEK293T cells carrying the respective variant introduced via prime editing. Cells were treated with CHX or DMSO (vehicle) 16 h prior to collection. The percentage of edited cells, measured by amplicon-NGS, is reported for each experiment. RNA isoforms with numbered exons corresponding to each PCR band are reported on the right, along with the PCR primers (black arrows). Band identity was confirmed by Sanger sequencing. PSI of the PE-containing isoform (calculated by densitometry analysis) is reported for each lane. Right panels indicate the quantitative analysis of splicing alternation by RT-qPCR. The average percentage of edited cells for each experiment. Primers specific for either the NMD isoform (purple bars) or the coding isoform (grey bars) were used. All values are internally normalized on the expression of *ACTB* gene, while transcript levels are represented as fold change compared to the DMSO condition with no editing. Statistical significance by Student's t-test is shown for each comparison. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$, ns: not significant. Data are shown as average ± standard deviation. Each data point represents one of three replicates.

We found of particular interest the intronic c.420+784C>T variant in the *NDUFAF6* gene, classified as "Pathogenic/Likely pathogenic" in ClinVar and associated with Leigh syndrome, a neurodevelopmental disease associated with mitochondrial complex I deficiency [66]. This deep intronic mutation causes retention of 124 intronic nucleotides and creates a premature stop codon in exon 4 [66] (Fig. 5B). *NDUFAF6* has recently been shown to play a role in mitochondrial complex I assembly in HEK293T cells [69]. To investigate the phenotypic consequences of *NDUFAF6* PE mis-splicing in neurons, we generated clonal i3N cell lines with different genotypes for c.420+784C>T (C/C, C/T and T/T; Fig. 6A) using the CRISPR prime editing workflow optimized in HEK293T cells. We assessed *NDUFAF6* PE splicing in undifferentiated iPSCs, induced neurons, and neurons at 1 and 21 days of maturation. In wild-type neurons the PE is weakly included with an average PSI of 0.13 and no significant changes across differentiation stages. Heterozygous (C/T) and homozygous (T/T) mutants showed increased PE inclusion, with average PSI of 0.37 and 0.94, respectively, which remained relatively stable throughout differentiation (Fig. 6B and 6C). These results confirmed the original prediction of "acceptor gain' for the PE (spliceAI score: 0.23). To assess the impact

on neuronal differentiation, we measured by RT-qPCR the expression of pluripotency markers *NANOG*, *SOX2* and *OCT4* [70], and neuronal differentiation markers *VGLUT1*, *MAP2*, *RBFOX2* and *TUBB3* [59] (S9A). Interestingly, we found that homozygous (T/T) mutant cells, but not heterozygous, display a significant higher level of *MAP2* and *RBFOX2* on day 21 of maturation (Fig 6D), possibly indicating an alteration in the differentiation status. Qualitative analysis by immunocytochemistry of differentiation markers MAP2, NeuN, TUJ1 and SMI-312 revealed no significant differences in their presence and intracellular localization (Fig. 6E and Supplementary Fig. S9B). However, we observed a difference in cell cluster formation, with heterozygous and homozygous mutant cells forming smaller clusters. This prompted us to perform a quantitative analysis of neuronal interconnectivity at early (day 1) and late (day 21) stages of maturation. Although we could still observe smaller and more fragmented cell-body clusters in homozygous mutant cells compared to wild-type and heterozygous cells, neurite formation and branch points appeared unaffected by PE mis-splicing (Supplementary Fig. S9C and S9D).

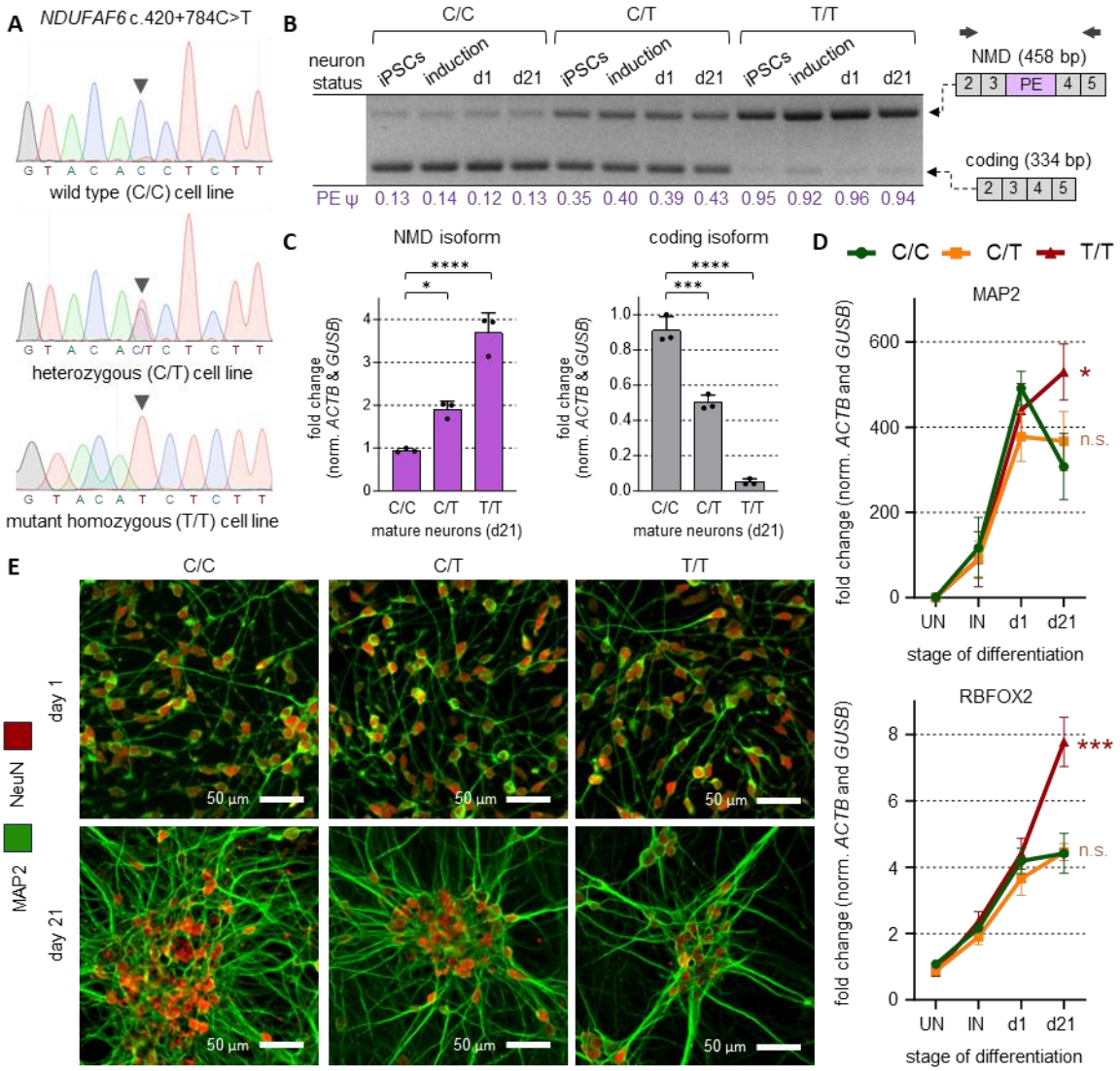



**Figure 6**. Functional characterization of *NDUFAF6* PE mis-splicing in neurons. (**A**) Genotyping by Sanger sequencing of clonal cell lines that have been prime edited to carry the mutation c.420+784C>T. Arrowheads indicate the edited base. Sequencing was performed on PCR amplicons obtained from genomic DNA. (**B**) Splicing analysis by RT-PCR of clonal cell lines with different genotypes for c.420+784C>T and at different stages of differentiation: undifferentiated iPSCs ("UN"), induced neurons following 3 days of NGN2 overexpression ("IN"), and neurons at 1 or 21 days of maturation. RNA isoforms with numbered exons corresponding to each PCR band are reported on the right, along with the PCR primers (black arrows). Band identity was confirmed by Sanger sequencing. PSI of the PE-containing isoform (calculated by band densitometry analysis) is reported for each lane. (**C**) Quantitative analysis of splicing alternation by RT-qPCR. Primers specific for either the NMD isoform (purple bars) or the coding isoform (grey bars) were used. All values are internally normalized on the expression of *ACTB* and *GUSB*, while transcript levels are represented as fold change compared to the DMSO condition with no editing. Statistical significance by Student's t-test is shown for each comparison. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$, ns: not significant. Data are shown as average ± standard deviation. Each data point represents one of three replicates. (**D**) Expression of pluripotency and neuronal differentiation markers is assessed by RT-qPCR for c.420+784C>T clonal cell lines at the stage of undifferentiated iPSCs

("UN"), induced neurons following 3 days of NGN2 overexpression ("IN"), and neurons at 1 or 21 days of maturation. All values are internally normalized on the expression of genes *ACTB* and *GUSB*, while expression values are represented as fold-change compared to the "UN" condition of the wild-type line. Statistical significance is indicated for the d21 stage of the mutant lines compared to wild-type. Data are shown as average of three replicates ± standard deviation. (**E**) Immunocytochemistry analysis of c.420+784C>T clonal cell lines, at 1 or 21 days of maturation, using antibodies for neuronal markers MAP2 and NeuN. Images are representative of multiple fields and replicates.

In summary, we show that pathogenic variants can significantly alter PE splicing in the brain, with implications for diverse neurological disorders. Our combined computational and experimental approach provides a scalable strategy for interpreting the splicing consequences of rare and clinically relevant pathogenic variants in the context of neurological disease. Overall, these findings establish PE dysregulation as a widespread, clinically relevant mechanism in the nervous system and highlight its potential as a therapeutic target.

## Discussion

Alternative splicing coupled with NMD is a central mechanism for regulating transcript stability and fine-tuning gene expression in eukaryotes. PE inclusion represents a specialized, highly conserved splicing event that triggers NMD and reduces protein expression. Although isolated examples of PEs have been described, their broader role in human brain development and disease remains poorly understood. Here, we provide a comprehensive characterization of the role of PE in human gene expression regulation by integrating evolutionary, transcriptomic, and disease-linked datasets with functional validations. Our study identified thousands of human PEs with tissue- and developmentally regulated splicing patterns in the brain, as well as hundreds of pathogenic variants affecting their splicing.

The comparative analysis of 957 eukaryotic transcriptomes revealed that NMD isoforms are particularly enriched in vertebrates, especially in humans. Nearly 40% of human coding genes give rise to at least one NMD transcript, and over a third of all protein-coding transcripts are

predicted to undergo NMD. Species with more complex splicing regulation have evolved to utilize NMD for fine-tuning gene expression at the post-transcriptional level. Within this context, PEs likely represent an evolutionarily refined strategy to control gene dosage in a spatial-temporal manner, especially in complex organs such as the brain.

We mapped the PE landscape, and we identified over 12,000 putative PEs across the human transcriptome, accounting for nearly 22% of all annotated alternative splicing events. These PEs were enriched in genes involved in RNA metabolism and in splicing factors, some of which are known to autoregulate their own expression through negative feedback loops mediated by PE inclusion [19, 27, 57]. While longer genes were more likely to contain PEs, those with brain-specific expression showed a significant depletion of PEs, suggesting selective pressure to preserve expression of critical neural genes. Experimental validation confirmed that PE inclusion typically leads to NMD, validating our predictive framework. The overlap of PE-containing genes with neurological disorder–associated pathways underscores the clinical significance of PE splicing as a key regulatory mechanism in the human brain.

We next investigated PE regulation in the context of the human brain. Using RNA-seq data from GTEx, we identified PEs that are differentially spliced in the brain compared to other human tissues. Analysis of BrainSpan RNA-seq data revealed dynamic changes in PE splicing across brain development, with distinct shifts occurring before and after birth. We observed significant changes in PE inclusion during neuronal differentiation, as for example in *NOVA1*, suggesting tight regulation of PE inclusion during neurogenesis. Together, our findings indicate that PEs act as temporally and spatially regulated molecular switches that fine-tune gene expression in both the developing and adult brain.

To assess the clinical impact of PE mis-splicing, we integrated ClinVar data with SpliceAI predictions and identified pathogenic variants predicted to affect PE splicing in neurological disorders. Many of these variants were located in intronic regions, reinforcing the critical role

of non-coding sequences in splicing regulation and their contribution to splicing disorders. Using CRISPR prime editing [48], we functionally validated representative variants selected based on their involvement in relevant neurological disorders and their effect on splicing, and confirmed that they induced alterations in endogenous mRNA splicing in human cells validating the accuracy of our predictions. These variants are associated with a broad spectrum of neurological conditions, including epilepsy, intellectual disability, and neurodegeneration, highlighting the role of PE mis-splicing in brain disorders. We further examined the consequences of the *NDUFAF6* c.420+784C>T variant**,** associated with Leigh syndrome, in the context of neuronal cells. After generating heterozygous and homozygous mutant i3N neurons, we observed an allele-dependent increase of the inclusion of a PE, confirming our original computational prediction. We additionally observed altered expression of the neuronal differentiation markers *MAP2* and *RBFOX2* in the homozygous mutant cells. Overall, our experimental pipeline provides a scalable and adaptable framework for interpreting and validating splicing variants affecting PE splicing within their native transcriptomic context.

This study establishes PEs as a widespread, evolutionarily conserved, and dynamically regulated mechanism of gene expression regulation in the human brain. By integrating computational prediction with functional validation, we show that PEs are critical regulators of gene expression during brain development and that their mis-splicing constitutes a recurrent pathogenetic mechanism in neurological disease. These findings highlight the therapeutic potential of targeting PE mis-splicing and provide a foundational resource for advancing the study of splicing regulation in the human brain and its contribution to disease.

## Materials and Methods

### Evolutionary analysis

Genome annotations of Eukaryotic species were downloaded in GTF format from Ensembl Genomes Release 60 (https://ftp.ebi.ac.uk/ensemblgenomes/pub/release-60/) and Ensemble Release 113 (for vertebrates, https://ftp.ensembl.org/pub/release-113/) using a Python-based script. Species are classified in this study based on their original group classification in the repository, which includes “vertebrates”, “metazoan” (here indicated as “invertebrates”), “protists”, “fungi” and “plants”.

A Python-based script was developed to systematically analyse each genome annotation for predicting NMD isoforms and collecting genomic information as follows: 1) unique isoforms are identified by their “transcript_id”, while coordinates of “exon” and “CDS” features are collected and annotated for each isoform; 2) the relative position of the last CDS coordinate (indicating the annotated stop codon) is analysed in relation to the position of the last exon-exon junction; 3) isoforms for which the stop codon is located more than 50 nt upstream of the las exon-exon junction are annotated as “NMD”; 4) additional information is collected from each genome annotation, including i) the total number of coding genes (uniquely identified by their “gene_id”, and for which at least one isoform with a “CDS” feature has been annotated), ii) the number of unique isoforms per coding gene or in total, iii) the number of unique “exon” feature per coding gene, iiii) the length of each coding isoform.

For phylogenetic tree reconstruction, taxonomic IDs of each species and lineage information were both retrieved from NCBI public repositories (respectively https://ftp.ncbi.nih.gov/genomes/ASSEMBLY_REPORTS/assembly_summary_refseq.txt and https://ftp.ncbi.nih.gov/pub/taxonomy/taxdump_archive/new_taxdump_2023-12-01.zip). The phylogenetic tree was generated with the ETE package [71], using the function ncbi.get_topology(taxids, intermediate_nodes=True) and lineage information from the file rankedlineage.dmp (retrieved in new_taxdump_2023-12-01.zip). The tree illustration was generated using iTOL online tool [72].

**Identification of putative PEs from genomic and transcriptomic reference**

This study focuses on defining the PE landscape by leveraging the annotated transcriptomes, based on the assumption that an annotated exon can be alternatively included as a PE in an otherwise protein-coding transcript and result in a downstream PTC. The derived PTC is not necessarily within the PE itself but has to be more than 150 bp from the transcription start site and not in the final two exons of the host gene, ensuring it doesn't escape NMD [49]. To search for PE from the transcriptome reference, non-redundant exons whose length ≤ 300 bp of a gene were first extracted from the GTF file. Each extracted exon was then computationally inserted into each transcript of the gene whose "transcript_biotype" is "protein-coding". If the insertion generated a PTC meeting the criteria above, the inserted exon would be marked as a putative PE. Note, this also means that whether an exon is "poisoning" or not depends on its transcript context. Depending on the relative position of the putative PE and the protein-coding transcript, the PE can be classified as one of "inclusion", "alternative 5'-ss" and "alternative 3'-ss" (Supplementary Fig. S3A).

**PSI value calculation for PEs**

PSI was used to measure the inclusion of an exon with potential for being a PE. To quantify the inclusion level of PEs, we computed the PSI value based on RNA-seq aligned read coverage. For each PE event, the read counts mapped to the alternative exon or intron retention region were extracted and denoted as $R_1$, $R_2$, and $R_3$, respectively. The PSI value was then calculated based on different PE modes. For PE inclusion, the PSI value was calculated as:

$$PSI = \frac{0.5 * (R_1 + R_2)}{0.5 * (R_1 + R_2) + R_3}$$

For alternative 5'/3'-ss mode, PSI was calculated as:

$$PSI = \frac{R_1}{(R_1 + R_3)}$$

Note, this should not be confused with the inclusion of an exon being explicitly used as a PE. From the short-read data used in this study, only the overall inclusion of an exon is measurable, regardless of whether it is included in a transcript as a poison or non-PE.

**Identification of tissue-specific PEs from GTEx**

The sample-wise PSI profiles were first built for all GTEx samples. For every pair of tissues, a generalized linear model (GLM) was fitted to examine the relationship between PSI and tissue type for every splicing event involving the exon with PE potential. The direction and significance of the coefficient for tissue type were used to judge whether the exon is differentially used.

**Identification of developmental-time-specific PEs from BrainSpan**

The sample-wise PSI profiles were first constructed for all BrainSpan samples. For each brain region, a generalized additive model (GAM) was applied to investigate the trajectories of PE inclusion over ages. In the GAM models, a smooth term for developmental time points was used to capture the temporal dynamics of PE inclusion. The significance of the smooth term indicates whether PE inclusion levels exhibit developmental trajectory. PE events with similar trajectories were clustered using Pearson correlation coefficients. Principal component analysis (PCA) was then applied to each cluster to extract an eigen PSI value, summarizing the pattern of splicing changes within the group.

**Analysis of pathogenic variants**

Pathogenic and likely-pathogenic mutations were extracted from ClinVar VCF file (version 20231112). SpliceAI [2] was applied to predict the effect on splicing for each mutation. A

Python-based workflow was developed to filter the variants analysed with SpliceAI, match them with PE inclusion data, and identify those affecting PE splicing. Only the most impacted putative ss with an absolute SpliceAI score change > 0.2 were retained. The genomic coordinates of these impacted ss were then matched with the coordinates of the PE's ss. When an exact match was found, the mutation would be considered to affect the splicing of the PE. Then, for each variant, we additionally assigned a positive or negative value to indicate directionality, where positive scores indicate increased PE inclusion, and negative scores indicate decreased inclusion (significant absolute value > 0.2). We paired these splicing-change predictions with the wild-type PSI values from GTEx brain tissues to prioritize variants that significantly affect PE splicing in the brain. The following two scenarios were of interest: 1) a PE's PSI in wild-type is < 0.3 and the mutation is predicted to increase its inclusion; and 2) a PE's PSI in wild-type is > 0.7 and the mutation is predicted to decrease its inclusion. Variants and genes associated with brain disorders were identified by looking for specific disease terminology in the ClinVar report (if available) or in the GO respectively. Gene ontologies were obtained using the Metascape online tool [73]. Specific terminology for brain disorders was retrieved from the ODiseA database [74].

**Cell culturing**

HEK293T cells (ATCC CRL-3216) were maintained in Dulbecco's Modified Eagle Medium (DMEM, Gibco™ 11965118) supplemented with 10% fetal bovine serum (FBS, HyClone™ SH3008003) and 1% penicillin-streptomycin (Gibco™ 15140122) at 37°C in a humidified incubator with 5% $CO_2$. For transfection, cells were seeded at 100,000 per well in 24-well plates one day prior. Transfections were performed using Lipofectamine 3000™ (Invitrogen™ L3000008) according to the manufacturer's protocol. Briefly, DNA plasmids and P3000 reagent™ were diluted in Opti-MEM™ (Gibco™ 31985070), combined with

Lipofectamine 3000™ diluted in Opti-MEM™, incubated for 15 minutes at room temperature, and added dropwise to the cells in DMEM with FBS but no antibiotics. Medium was replaced after 8 hours and doubled after 24 hours. Cells were collected 72 hours post-transfection for downstream analyses.

Human iPSCs carrying inducible NGN2 (Coriell Institute GM29370) were differentiated into i3N neurons following a multi-step protocol adapted from Wang et al. (2018) [59]. Plates were coated with Matrigel (Corning™ 354277) diluted in cold DMEM/F12 (Gibco™ 11320082) as per manufacturer's instructions, and incubated at room temperature for at least 3 hours. On day -3, iPSCs at ~80% confluency and cultured in mTeSR Plus™ medium (StemCell™ 100-0276) with 1% penicillin-streptomycin (Gibco™ 15140122) were dissociated with Accutase™ (StemCell™ 7920) and plated on Matrigel-coated dishes in induction medium, which was composed of KO-DMEM/F12 (Gibco™ 12660012) supplemented with N2 (Gibco™ 17502048), NEAA (Gibco™ 11140050), 10 ng/mL BDNF (PeproTech™ 450-02), 10 ng/mL NT3 (PeproTech™ 450-02), 0.2 μg/mL laminin (Corning™ 354239), 2 μg/mL doxycycline (Sigma-Aldrich™ D9891-1G), and 10 μM ROCK inhibitor (Sigma-Aldrich™ 688001). Medium was changed with complete induction medium on day -2, and with induction medium lacking ROCK inhibitor on day -1. On day 0, cells were dissociated, replated onto dishes coated with 50 μg/mL poly-D-lysine (Gibco™ A3890401) and 3 μg/mL laminin (Gibco™ 23017015), and cultured in maturation medium, which was composed of 50% DMEM/F12 (Gibco™ 11320082) and 50% Neurobasal-A™ (ThermoFisher™ 10888022), supplemented with 0.5× B27 (Gibco™ 17504044), N2 (Gibco™ 17502048), NEAA (Gibco™ 11140050), GlutaMax (Gibco™ 35050061), 1 μg/mL laminin (Corning™ 354239), 10 ng/mL BDNF (PeproTech™ 450-02), and 10 ng/mL NT3 (PeproTech™ 450-02). Half-medium changes were performed weekly starting from day 7. Neurons were cultured for up to 4 weeks before downstream assays. Cells were cultured at

37°C in a humidified incubator with 5% $CO_2$ at all stages. Live-cell imaging was performed using a Sartorius Incucyte SX5 system (Sartorius™) equipped with a 10× objective, and images were analysed using the integrated NeuroTrack software.

For CRISPR prime editing of iPSCs, cells were transfected with plasmids expressing the editing components using Lipofectamine™ Stem Transfection Reagent (Invitrogen™, STEM00001). GFP-positive cells were FACS sorted as single cells in 96-well plates using BigFoot™ Cell Sorter (Thermo Fisher™) to generate isogenic lines. Clonal lines were cultured for about 2 weeks before splitting them for genotyping by Sanger sequencing and downstream freezing. Cells were frozen using a solution of 50% DMEM/F12 (Gibco™ 11320082) and 50% heat-inactivated FBS (CellPro™ FB73), supplemented with 10% DMSO (Sigma-Aldrich™ D8418).

CHX treatment was performed both on HEK293T and i3N neurons by treating them for 16h with 50 μg/ml CHX (Thermo Scientific™ J66901.03) or an equivalent volume of DMSO (Sigma-Aldrich™ D8418) as control (0.5% of the culture medium volume). For HEK293T cells, medium was replaced with CHX-containing medium. For i3N neurons, culture medium was quickly removed from the cells, implemented with CHX, and placed back in culture.

**Plasmid design and preparation**

Plasmids expressing the CRISPR prime editor (Addgene 174820), the pegRNA (Addgene 223137), or the nicking gRNA (Addgene 65777) were prepared using competent DH5α E. coli (NEB™ C2987I) and QIAwave Plasmid Miniprep Kit (Qiagen™ 27204). pegRNAs and nicking gRNAs (Supplementary Table S11) were designed using the PRIDICT tool [64] and selecting the pegRNA and gRNA with top scores, and they were cloned using standard Type IIS restriction enzymes (NEB™). Sequences were confirmed by Sanger sequencing.

### *In vitro* splicing analysis

Total RNA was extracted from cultured cells using QIAzol (Qiagen™ 79306) reagent according to the manufacturer's instructions. After chloroform separation (Sigma-Aldrich™ C2432) and isopropanol precipitation (Sigma-Aldrich™ 109634), RNA pellets were washed in 75% ethanol (Sigma-Aldrich™ E7023), air-dried, and resuspended in nuclease-free water. RNA quantity and purity were assessed using a NanoDrop spectrophotometer (Thermo Scientific™ 2353-30-0010). For reverse transcription, 500 ng to 1 µg of total RNA was used with the SuperScript III Kit (Fisher Scientific™ 18080044), using random hexamer primers (Promega™ C1181) and following the recommended protocol.

PCR was performed using exon-specific primers (Supplementary Table S11) and GoTaq master mix (Promega™ M7123-Green) to amplify regions of interest from cDNA (25 µg per reaction). PCR products were resolved on 2% agarose gels (Fisher Scientific™ 16-550-100) stained with ethidium bromide (Sigma-Aldrich™ E1510) and visualized using a UV gel imaging system. Bands of expected size were excised and purified using the QIAquick Gel Extraction Kit (Qiagen™ 28706), followed by Sanger sequencing to confirm splicing isoforms and identify sequence variants. Images of gel bands were acquired using AlphaImager™ 2200, and densitometry analysis was performed in ImageJ [75].

qPCR was carried out using SYBR Green PCR Master Mix™ (Fisher Scientific™ A25742) in a LightCycler 480 II system™ (Roche™). Reactions were performed in technical duplicates using 10 µg cDNA per reaction and exon-specific or exon junction-specific primers (Supplementary Table S11). Relative gene expression was calculated using the ΔΔCt method [76], normalized to housekeeping genes. Melting curve analysis was used to ensure specificity and consistency across runs.

### Amplicon NGS

Genomic DNA was extracted from cultured cells using magnetic purification beads (Illumina™ 20060057) and following the manufacturer's specifications. Briefly, cells were lysed using the provided solution, and the lysate was incubated with purification beads to bind DNA. Magnetic separation was used to wash away impurities, and DNA was eluted in nuclease-free water. To quantify editing outcomes, targeted regions were amplified from genomic DNA using primers with overhangs compatible with Illumina adapters (Supplementary Table S11). Amplicons were cleaned with magnetic beads (Illumina™ 20060057) and pooled for multiplexed sequencing. Libraries were sequenced with MiSeq Reagent Micro Kit v2™ (Illumina™ MS-103-1002) on a MiSeq Syste™ (Illumina™) using 2×150 bp paired-end reads. Data were demultiplexed and analyzed using the CRISPResso2 tool [77].

**Immunochemistry**

Cells cultured in 96-well plates (Corning™ 3603) were fixed by adding an equal volume of 8% paraformaldehyde (Thermo Fisher™ 047347.9M) in PBS (Gibco™ 10010023) directly to the culture medium (final concentration 4% paraformaldehyde) and incubating for 30 min at room temperature. After fixation, cells were washed three times with DPBS without calcium and magnesium (Gibco™, 14190114), taking care to avoid drying, and permeabilized with 0.1% Triton X-100 (Thermo Scientific™ PI85111) in DPBS for 10 min at room temperature. Following three additional DPBS washes, cells were blocked in DPBS containing 5% normal goat serum (Sigma-Aldrich™ S26-100ML) and 0.01% Triton X-100 for 1 h at room temperature on a shaker. Primary antibodies diluted in blocking solution were applied overnight at 4 °C (see antibody details below). The following day, cells were washed three times with DPBS and incubated for 1 h at room temperature with the corresponding secondary antibodies diluted in blocking solution, protected from light. After washing, nuclei were

counterstained with 2 µg/mL Hoechst-33342 (BioRad™ 1351304EDU) for 10 min, followed by three DPBS washes. Wells were finally filled with DPBS for imaging. The following primary antibodies were used: NeuN (Millipore™ MAB377, clone 60), MAP2 (Millipore™ AB5543), βIII-tubulin (Synaptic Systems™ 302304), PSD95 (NeuroMab™ 75-028), and SMI-312 (Abcam™ ab254348). Secondary antibodies were: goat anti-mouse Alexa Fluor™ 594 (Thermo Fisher™ A11032), goat anti-chicken Alexa Fluor™ 488 (Thermo Fisher, A11039), goat anti-guinea pig Alexa Fluor™ 488 (Thermo Fisher™ A11073), and goat anti-rabbit Alexa Fluor™ 594 (Thermo Fisher™ A11012). Images were acquired using an IN Cell Analyzer 6000 (Cytiva™) with a 20× objective, capturing 30 fields per well, and analysed in ImageJ [75] for channel assignment, brightness/contrast adjustment.

**Statistical analysis**

The statistical analyses for the wet-lab work were performed using GraphPad Prism. The brain-unique PEs were defined as those with a minimal PSI > 0 in any brain tissue and a max PSI = 0 across all non-brain tissues. The brain-differential PEs were defined as those meeting all the following criteria when comparing a brain tissue to a non-brain tissue: 1) tissue coefficient ≥ 0.2 or ≤ -0.2, and 2) tissue coefficient's false discovery rate (FDR) < 0.1 from the Wald test of GLM. The developmental-stage-specific PEs were defined as those whose smooth term's FDR < 0.05 in the GAM. For the GO enrichment analyses, a threshold of FDR < 0.1 from Fisher's exact test was applied to select the significant terms. In all analyses performed with Student's t-test or ANOVA test, statistical significance is indicated as follows: "****" indicates P < 0.0001; "***" indicates P < 0.001; "**" indicates P < 0.01; "*" indicates P < 0.05; "ns" indicates non significance.

## Acknowledgements

We would like to express our sincere gratitude to Drs. Jacob Kitzman and Jennifer Yee from the University of Michigan Medical School for their invaluable initial contributions in identifying the optimal approach and models to study genetic mutations affecting RNA splicing. Their foundational work was instrumental in shaping the direction of this research. We also extend our sincere appreciation to Dr. Xander Nuttle, also from the University of Michigan Medical School, for his expert guidance in the editing of iPSCs and the subsequent selection for neuronal differentiation. His insights were crucial for the successful execution of key experimental procedures.

## Author contributions

Paolo Pigini: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. Huilin Xu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. Yan Ji: Investigation, Methodology. Hannah Lindmeier: Investigation, Methodology. Henry R. Saltzman: Investigation, Methodology. Shuqi Yun: Investigation, Methodology. Christiano R. R. Alves: Methodology, Resources. M. Catarina Silva: Conceptualization, Funding acquisition, Methodology, Resources, Supervision. Dadi Gao: Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. Elisabetta Morini: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.

## Conflict of interest

Paolo Pigini is an inventor on a patent application filed by the Agency of Science, Technology and Research of Singapore that describes a therapeutic technology based on modified small nuclear RNAs. Christiano R. R. Alves is an inventor on patent applications filed by Mass General Brigham that describe genome engineering technologies. Christiano R. R. Alves is also a consultant for Biogen and Ilios Therapeutics. Christiano R. R. Alves interests were reviewed and are managed by MGH and MGB in accordance with their conflict-of-interest policies. Elisabetta Morini serves on the scientific advisory board (SAB) of Revir Therapeutics. Elisabetta Morini and Dadi Gao are inventor on an International Patent Application Number PCT/US2021/012103, assigned to Massachusetts General Hospital and PTC Therapeutics entitled "RNA Splicing Modulation" related to the use of BPN-15477 in modulating splicing. Elisabetta Morini and Christiano R. R. Alves on a patent application number PCT/US25/26694, assigned to Massachusetts General Hospital, entitled “Gene Editing for Correcting Splicing Defect in Familial Dysautonomia”. This patent application covers optimized base editing and prime editing strategies to correct mutations in ELP1 and treat Familial Dysautonomia (FD).

## Funding

This work was funded by National Institutes of Health (NIH) grants (R01NS124561 to EM and MCS, R00NS118109 to DG, R21EY037018 to EM and CA), and by the Ellison Foundation (EM). CA was also supported by a MGH Physician/Scientist Development Award and NIH NINDS grant K01NS134784.

## References

1. Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. Proc Natl Acad Sci U S A. 2011;108(27):11093-8. Epub 20110617. doi: 10.1073/pnas.1101135108. PubMed PMID: 21685335; PubMed Central PMCID: PMCPMC3131313.
2. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019;176(3):535-48 e24. Epub 20190117. doi: 10.1016/j.cell.2018.12.015. PubMed PMID: 30661751.
3. Sakaguchi N, Suyama M. Pervasive occurrence of splice-site-creating mutations and their possible involvement in genetic disorders. NPJ Genom Med. 2022;7(1):22. Epub 20220318. doi: 10.1038/s41525-022-00294-0. PubMed PMID: 35304488; PubMed Central PMCID: PMCPMC8933504.
4. Camper SA, Smith C, Kitzman JO. Disruption of RNA Splicing Is an Important Contributor to Congenital Hypopituitarism and Other Human Genetic Diseases. Endocrinology. 2023;164(5). doi: 10.1210/endocr/bqad039. PubMed PMID: 36857601; PubMed Central PMCID: PMCPMC10282917.
5. Chiang HL, Chen YT, Su JY, Lin HN, Yu CA, Hung YJ, et al. Mechanism and modeling of human disease-associated near-exon intronic variants that perturb RNA splicing. Nat Struct Mol Biol. 2022;29(11):1043-55. Epub 20221027. doi: 10.1038/s41594-022-00844-1. PubMed PMID: 36303034.
6. Raj B, Blencowe BJ. Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron. 2015;87(1):14-27. doi: 10.1016/j.neuron.2015.05.004. PubMed PMID: 26139367.
7. Zhang X, Chen MH, Wu X, Kodani A, Fan J, Doan R, et al. Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. Cell. 2016;166(5):1147-62 e15. Epub 2016/08/28. doi: 10.1016/j.cell.2016.07.025. PubMed PMID: 27565344; PubMed Central PMCID: PMCPMC5248659.
8. Vuong CK, Black DL, Zheng S. The neurogenetics of alternative splicing. Nat Rev Neurosci. 2016;17(5):265-81. doi: 10.1038/nrn.2016.27. PubMed PMID: 27094079; PubMed Central PMCID: PMCPMC4861142.
9. Patowary A, Zhang P, Jops C, Vuong CK, Ge X, Hou K, et al. Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms. Science. 2024;384(6698):eadh7688. Epub 20240524. doi: 10.1126/science.adh7688. PubMed PMID: 38781356; PubMed Central PMCID: PMCPMC11960787.
10. Wen C, Margolis M, Dai R, Zhang P, Przytycki PF, Vo DD, et al. Cross-ancestry atlas of gene, isoform, and splicing regulation in the developing human brain. Science. 2024;384(6698):eadh0829. Epub 20240524. doi: 10.1126/science.adh0829. PubMed PMID: 38781368.
11. Aya F, Valcarcel J. Shaping human brain development and vulnerability through alternative splicing. Cell Genom. 2024;4(6):100584. doi: 10.1016/j.xgen.2024.100584. PubMed PMID: 38870907; PubMed Central PMCID: PMCPMC11228946.
12. Ha KCH, Sterne-Weiler T, Morris Q, Weatheritt RJ, Blencowe BJ. Differential contribution of transcriptomic regulatory layers in the definition of neuronal identity. Nat Commun. 2021;12(1):335. Epub 20210112. doi: 10.1038/s41467-020-20483-8. PubMed PMID: 33436550; PubMed Central PMCID: PMCPMC7804943.
13. Gonzalez-Iglesias A, Arcas A, Domingo-Muelas A, Mancini E, Galceran J, Valcarcel J, et al. Intron detention tightly regulates the stemness/differentiation switch in the adult neurogenic niche. Nat Commun. 2024;15(1):2837. Epub 20240402. doi: 10.1038/s41467-024-47092-z. PubMed PMID: 38565566; PubMed Central PMCID: PMCPMC10987655.

14. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, et al. The evolutionary landscape of alternative splicing in vertebrate species. Science. 2012;338(6114):1587-93. doi: 10.1126/science.1230612. PubMed PMID: 23258890.
15. Dredge BK, Polydorides AD, Darnell RB. The splice of life: alternative splicing and neurological disease. Nat Rev Neurosci. 2001;2(1):43-50. doi: 10.1038/35049061. PubMed PMID: 11253358.
16. Hervoso JL, Amoah K, Dodson J, Choudhury M, Bhattacharya A, Quinones-Valdez G, et al. Splicing-specific transcriptome-wide association uncovers genetic mechanisms for schizophrenia. Am J Hum Genet. 2024;111(8):1573-87. Epub 20240625. doi: 10.1016/j.ajhg.2024.06.001. PubMed PMID: 38925119; PubMed Central PMCID: PMCPMC11339621.
17. Frydas A, Wauters E, van der Zee J, Van Broeckhoven C. Uncovering the impact of noncoding variants in neurodegenerative brain diseases. Trends Genet. 2022;38(3):258-72. Epub 20210914. doi: 10.1016/j.tig.2021.08.010. PubMed PMID: 34535299.
18. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A. 2003;100(1):189-92. Epub 20021226. doi: 10.1073/pnas.0136770100. PubMed PMID: 12502788; PubMed Central PMCID: PMCPMC140922.
19. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature. 2007;446(7138):926-9. Epub 20070314. doi: 10.1038/nature05676. PubMed PMID: 17361132.
20. Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev. 2007;21(6):708-18. doi: 10.1101/gad.1525507. PubMed PMID: 17369403; PubMed Central PMCID: PMCPMC1820944.
21. Lykke-Andersen S, Jensen TH. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol. 2015;16(11):665-77. Epub 20150923. doi: 10.1038/nrm4063. PubMed PMID: 26397022.
22. He F, Jacobson A. Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. Annu Rev Genet. 2015;49:339-66. Epub 20151002. doi: 10.1146/annurev-genet-112414-054639. PubMed PMID: 26436458; PubMed Central PMCID: PMCPMC4837945.
23. Kurosaki T, Maquat LE. Nonsense-mediated mRNA decay in humans at a glance. J Cell Sci. 2016;129(3):461-7. Epub 20160119. doi: 10.1242/jcs.181008. PubMed PMID: 26787741; PubMed Central PMCID: PMCPMC4760306.
24. Lejeune F. Nonsense-Mediated mRNA Decay, a Finely Regulated Mechanism. Biomedicines. 2022;10(1). Epub 20220110. doi: 10.3390/biomedicines10010141. PubMed PMID: 35052820; PubMed Central PMCID: PMCPMC8773229.
25. Yi Z, Sanjeev M, Singh G. The Branched Nature of the Nonsense-Mediated mRNA Decay Pathway. Trends Genet. 2021;37(2):143-59. Epub 20200929. doi: 10.1016/j.tig.2020.08.010. PubMed PMID: 33008628; PubMed Central PMCID: PMCPMC7854845.
26. Pervouchine D, Popov Y, Berry A, Borsari B, Frankish A, Guigo R. Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. Nucleic Acids Res. 2019;47(10):5293-306. doi: 10.1093/nar/gkz193. PubMed PMID: 30916337; PubMed Central PMCID: PMCPMC6547761.
27. Leclair NK, Brugiolo M, Urbanski L, Lawson SC, Thakar K, Yurieva M, et al. Poison Exon Splicing Regulates a Coordinated Network of SR Protein Expression during Differentiation and Tumorigenesis. Mol Cell. 2020;80(4):648-65 e9. Epub 20201110. doi:

10.1016/j.molcel.2020.10.019. PubMed PMID: 33176162; PubMed Central PMCID: PMCPMC7680420.

28. Zhuravskaya A, Yap K, Hamid F, Makeyev EV. Alternative splicing coupled to nonsense-mediated decay coordinates downregulation of non-neuronal genes in developing mouse neurons. Genome Biol. 2024;25(1):162. Epub 20240620. doi: 10.1186/s13059-024-03305-8. PubMed PMID: 38902825; PubMed Central PMCID: PMCPMC11188260.

29. Lin JQ, Khuperkar D, Pavlou S, Makarchuk S, Patikas N, Lee FC, et al. HNRNPH1 regulates the neuroprotective cold-shock protein RBM3 expression through poison exon exclusion. EMBO J. 2023;42(14):e113168. Epub 20230530. doi: 10.15252/embj.2022113168. PubMed PMID: 37248947; PubMed Central PMCID: PMCPMC10350819.

30. Jaffrey SR, Wilkinson MF. Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease. Nat Rev Neurosci. 2018;19(12):715-28. doi: 10.1038/s41583-018-0079-z. PubMed PMID: 30410025; PubMed Central PMCID: PMCPMC6396682.

31. Carvill GL, Mefford HC. Poison exons in neurodevelopment and disease. Curr Opin Genet Dev. 2020;65:98-102. Epub 20200629. doi: 10.1016/j.gde.2020.05.030. PubMed PMID: 32615329; PubMed Central PMCID: PMCPMC8042789.

32. Yan Q, Weyn-Vanhentenryck SM, Wu J, Sloan SA, Zhang Y, Chen K, et al. Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. Proc Natl Acad Sci U S A. 2015;112(11):3445-50. Epub 20150303. doi: 10.1073/pnas.1502849112. PubMed PMID: 25737549; PubMed Central PMCID: PMCPMC4371929.

33. Eom T, Zhang C, Wang H, Lay K, Fak J, Noebels JL, et al. NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. Elife. 2013;2:e00178. Epub 20130122. doi: 10.7554/eLife.00178. PubMed PMID: 23359859; PubMed Central PMCID: PMCPMC3552424.

34. Carvill GL, Engel KL, Ramamurthy A, Cochran JN, Roovers J, Stamberger H, et al. Aberrant Inclusion of a Poison Exon Causes Dravet Syndrome and Related SCN1A-Associated Genetic Epilepsies. Am J Hum Genet. 2018;103(6):1022-9. doi: 10.1016/j.ajhg.2018.10.023. PubMed PMID: 30526861; PubMed Central PMCID: PMCPMC6288405.

35. Tang S, Stamberger H, Calhoun JD, Weckhuysen S, Carvill GL. Antisense oligonucleotides modulate aberrant inclusion of poison exons in SCN1A-related Dravet syndrome. JCI Insight. 2025;10(7). Epub 20250213. doi: 10.1172/jci.insight.188014. PubMed PMID: 39946203; PubMed Central PMCID: PMCPMC11981616.

36. Sparber P, Bychkov I, Pyankov D, Skoblov M. Functional investigation of SCN1A deep-intronic variants activating poison exons inclusion. Hum Genet. 2023;142(8):1043-53. Epub 20230425. doi: 10.1007/s00439-023-02564-y. PubMed PMID: 37186029.

37. Aziz MC, Schneider PN, Carvill GL. Targeting Poison Exons to Treat Developmental and Epileptic Encephalopathy. Dev Neurosci. 2021;43(3-4):241-6. Epub 20210510. doi: 10.1159/000516143. PubMed PMID: 33971653.

38. Brown AL, Wilkins OG, Keuss MJ, Hill SE, Zanovello M, Lee WC, et al. TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. Nature. 2022;603(7899):131-7. Epub 20220223. doi: 10.1038/s41586-022-04436-3. PubMed PMID: 35197628; PubMed Central PMCID: PMCPMC8891020.

39. Wilkins OG, Chien M, Wlaschin JJ, Barattucci S, Harley P, Mattedi F, et al. Creation of de novo cryptic splicing for ALS and FTD precision medicine. Science. 2024;386(6717):61-9. Epub 20241003. doi: 10.1126/science.adk2539. PubMed PMID: 39361759; PubMed Central PMCID: PMCPMC7616720.

40. Bakshi S, Isom LL. No more nonsense: evaluating poison exons as therapeutic targets in neurodevelopmental disorders. Curr Opin Genet Dev. 2025;92:102346. Epub 20250408. doi: 10.1016/j.gde.2025.102346. PubMed PMID: 40203733.

41. McLean ZL, Gao D, Correia K, Roy JCL, Shibata S, Farnum IN, et al. Splice modulators target PMS1 to reduce somatic expansion of the Huntington's disease-associated CAG repeat. Nat Commun. 2024;15(1):3182. Epub 20240412. doi: 10.1038/s41467-024-47485-0. PubMed PMID: 38609352; PubMed Central PMCID: PMCPMC11015039.

42. Bhattacharyya A, Trotta CR, Narasimhan J, Wiedinger KJ, Li W, Effenberger KA, et al. Small molecule splicing modifiers with systemic HTT-lowering activity. Nat Commun. 2021;12(1):7299. Epub 20211215. doi: 10.1038/s41467-021-27157-z. PubMed PMID: 34911927; PubMed Central PMCID: PMCPMC8674292.

43. Lim KH, Han Z, Jeon HY, Kach J, Jing E, Weyn-Vanhentenryck S, et al. Antisense oligonucleotide modulation of non-productive alternative splicing upregulates gene expression. Nat Commun. 2020;11(1):3501. Epub 20200709. doi: 10.1038/s41467-020-17093-9. PubMed PMID: 32647108; PubMed Central PMCID: PMCPMC7347940.

44. Felker SA, Lawlor JMJ, Hiatt SM, Thompson ML, Latner DR, Finnila CR, et al. Poison exon annotations improve the yield of clinically relevant variants in genomic diagnostic testing. Genet Med. 2023;25(8):100884. Epub 20230506. doi: 10.1016/j.gim.2023.100884. PubMed PMID: 37161864; PubMed Central PMCID: PMCPMC10524927.

45. Keegan NP, Fletcher S. A spotter's guide to SNPtic exons: The common splice variants underlying some SNP-phenotype correlations. Mol Genet Genomic Med. 2022;10(1):e1840. Epub 20211028. doi: 10.1002/mgg3.1840. PubMed PMID: 34708937; PubMed Central PMCID: PMCPMC8801146.

46. Margasyuk S, Kuznetsova A, Zavileyskiy L, Vlasenok M, Skvortsov D, Pervouchine DD. Human introns contain conserved tissue-specific cryptic poison exons. NAR Genom Bioinform. 2024;6(4):lqae163. Epub 20241211. doi: 10.1093/nargab/lqae163. PubMed PMID: 39664813; PubMed Central PMCID: PMCPMC11632617.

47. Toledano I, Supek F, Lehner B. Genome-scale quantification and prediction of pathogenic stop codon readthrough by small molecules. Nat Genet. 2024;56(9):1914-24. Epub 20240822. doi: 10.1038/s41588-024-01878-5. PubMed PMID: 39174735; PubMed Central PMCID: PMCPMC11387191.

48. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576(7785):149-57. Epub 20191021. doi: 10.1038/s41586-019-1711-4. PubMed PMID: 31634902; PubMed Central PMCID: PMCPMC6907074.

49. Klonowski J, Liang Q, Coban-Akdemir Z, Lo C, Kostka D. aenmd: Annotating escape from nonsense-mediated decay for transcripts with protein-truncating variants. bioRxiv. 2023. Epub 20230321. doi: 10.1101/2023.03.17.533185. PubMed PMID: 36993377; PubMed Central PMCID: PMCPMC10055276.

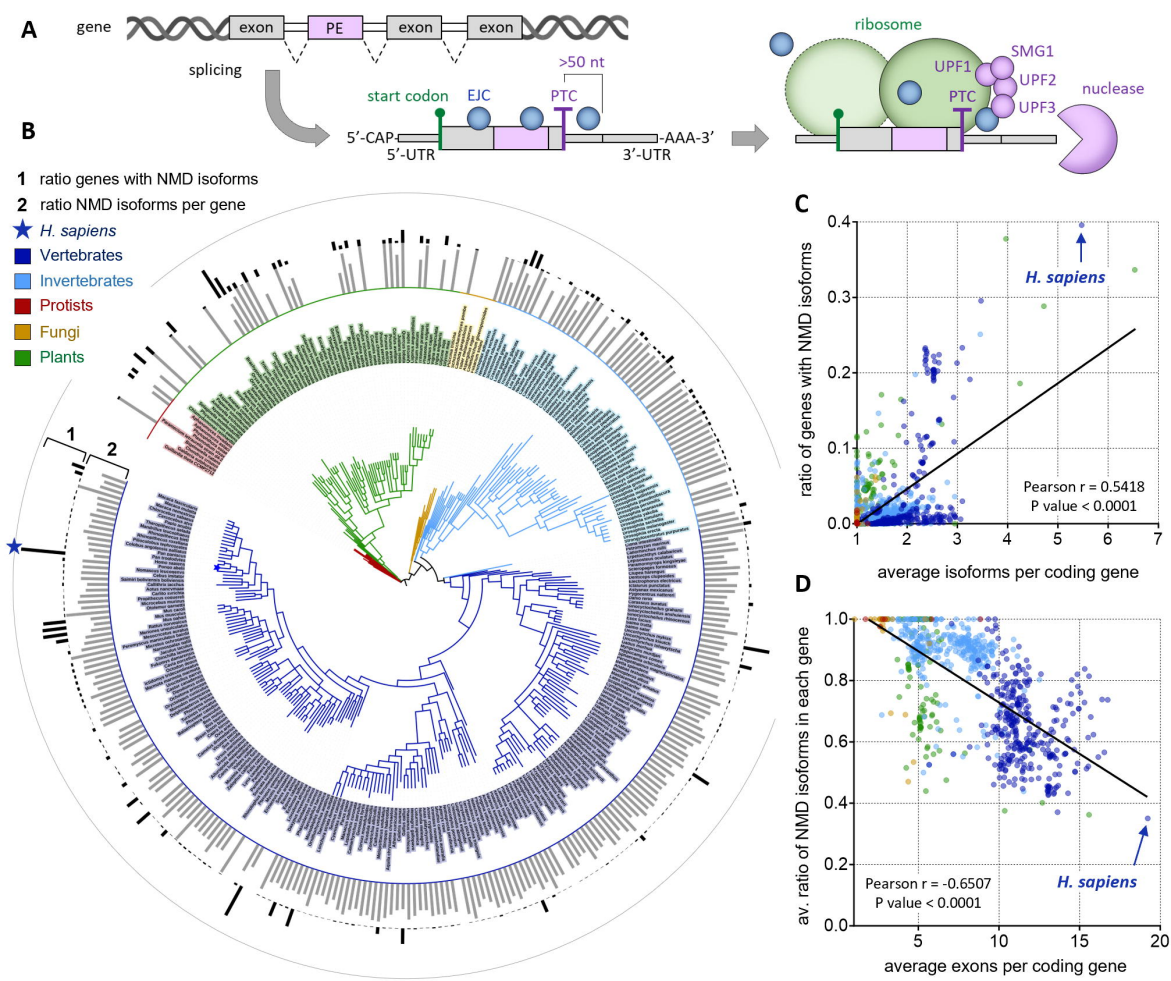
50. Thomas JD, Polaski JT, Feng Q, De Neef EJ, Hoppe ER, McSharry MV, et al. RNA isoform screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison exons. Nat Genet. 2020;52(1):84-94. Epub 20200107. doi: 10.1038/s41588-019-0555-z. PubMed PMID: 31911676; PubMed Central PMCID: PMCPMC6962552.

51. Belleville AE, Thomas JD, Tonnies J, Gabel AM, Borrero Rossi A, Singh P, et al. An autoregulatory poison exon in Smndc1 is conserved across kingdoms and influences organism growth. PLoS Genet. 2024;20(8):e1011363. Epub 20240816. doi: 10.1371/journal.pgen.1011363. PubMed PMID: 39150991; PubMed Central PMCID: PMCPMC11357089.

52. Zylka MJ, Simon JM, Philpot BD. Gene length matters in neurons. Neuron. 2015;86(2):353-5. doi: 10.1016/j.neuron.2015.03.059. PubMed PMID: 25905808; PubMed Central PMCID: PMCPMC4584405.
53. Barbash S, Sakmar TP. Length-dependent gene misexpression is associated with Alzheimer's disease progression. Sci Rep. 2017;7(1):190. Epub 20170315. doi: 10.1038/s41598-017-00250-4. PubMed PMID: 28298623; PubMed Central PMCID: PMCPMC5428066.
54. McCoy MJ, Fire AZ. Intron and gene size expansion during nervous system evolution. BMC Genomics. 2020;21(1):360. Epub 20200514. doi: 10.1186/s12864-020-6760-4. PubMed PMID: 32410625; PubMed Central PMCID: PMCPMC7222433.
55. Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. Biochem J. 2009;417(1):15-27. doi: 10.1042/BJ20081501. PubMed PMID: 19061484.
56. Lardelli RM, Thompson JX, Yates JR, 3rd, Stevens SW. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA. 2010;16(3):516-28. Epub 20100120. doi: 10.1261/rna.2030510. PubMed PMID: 20089683; PubMed Central PMCID: PMCPMC2822917.
57. Dalgliesh C, Aldalaqan S, Atallah C, Best A, Scott E, Ehrmann I, et al. An ultra-conserved poison exon in the Tra2b gene encoding a splicing activator is essential for male fertility and meiotic cell division. EMBO J. 2025;44(3):877-902. Epub 20250102. doi: 10.1038/s44318-024-00344-6. PubMed PMID: 39748121; PubMed Central PMCID: PMCPMC11791180.
58. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5. doi: 10.1038/ng.2653. PubMed PMID: 23715323; PubMed Central PMCID: PMCPMC4010069.
59. Wang C, Ward ME, Chen R, Liu K, Tracy TE, Chen X, et al. Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. Stem Cell Reports. 2017;9(4):1221-33. Epub 20170928. doi: 10.1016/j.stemcr.2017.08.019. PubMed PMID: 28966121; PubMed Central PMCID: PMCPMC5639430.
60. Lin JC, Chi YL, Peng HY, Lu YH. RBM4-Nova1-SRSF6 splicing cascade modulates the development of brown adipocytes. Biochim Biophys Acta. 2016;1859(11):1368-79. Epub 20160812. doi: 10.1016/j.bbagrm.2016.08.006. PubMed PMID: 27535496.
61. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42(Database issue):D980-5. Epub 20131114. doi: 10.1093/nar/gkt1113. PubMed PMID: 24234437; PubMed Central PMCID: PMCPMC3965032.
62. Chen PJ, Hussmann JA, Yan J, Knipping F, Ravisankar P, Chen PF, et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. Cell. 2021;184(22):5635-52 e29. Epub 20211014. doi: 10.1016/j.cell.2021.09.018. PubMed PMID: 34653350; PubMed Central PMCID: PMCPMC8584034.
63. Nelson JW, Randolph PB, Shen SP, Everette KA, Chen PJ, Anzalone AV, et al. Engineered pegRNAs improve prime editing efficiency. Nat Biotechnol. 2022;40(3):402-10. Epub 20211004. doi: 10.1038/s41587-021-01039-7. PubMed PMID: 34608327; PubMed Central PMCID: PMCPMC8930418.
64. Mathis N, Allam A, Kissling L, Marquart KF, Schmidheini L, Solari C, et al. Predicting prime editing efficiency and product purity by deep learning. Nat Biotechnol. 2023;41(8):1151-9. Epub 20230116. doi: 10.1038/s41587-022-01613-7. PubMed PMID: 36646933; PubMed Central PMCID: PMCPMC7614945.

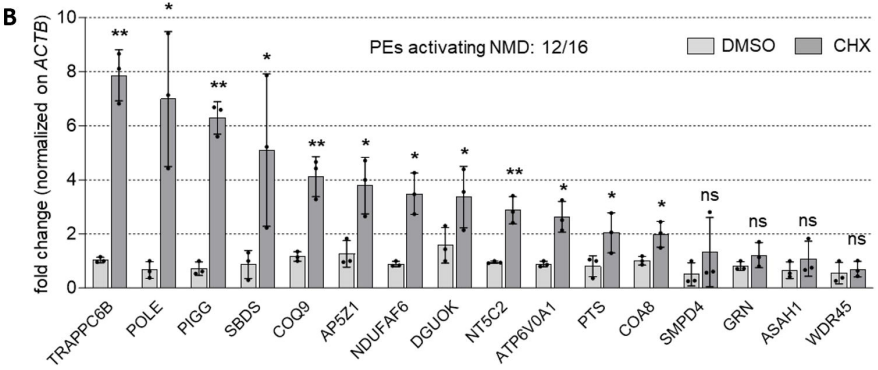
65. Sakamoto O, Ohura T, Katsushima Y, Fujiwara I, Ogawa E, Miyabayashi S, et al. A novel intronic mutation of the TAZ ( G4.5) gene in a patient with Barth syndrome: creation of a 5' splice donor site with variant GC consensus and elongation of the upstream exon. Hum Genet. 2001;109(5):559-63. Epub 20011013. doi: 10.1007/s00439-001-0612-3. PubMed PMID: 11735032.
66. Catania A, Ardissone A, Verrigni D, Legati A, Reyes A, Lamantea E, et al. Compound heterozygous missense and deep intronic variants in NDUFAF6 unraveled by exome sequencing and mRNA analysis. J Hum Genet. 2018;63(5):563-8. Epub 20180312. doi: 10.1038/s10038-018-0423-1. PubMed PMID: 29531337; PubMed Central PMCID: PMCPMC6071912.
67. Tyburczy ME, Dies KA, Glass J, Camposano S, Chekaluk Y, Thorner AR, et al. Mosaic and Intronic Mutations in TSC1/TSC2 Explain the Majority of TSC Patients with No Mutation Identified by Conventional Testing. PLoS Genet. 2015;11(11):e1005637. Epub 20151105. doi: 10.1371/journal.pgen.1005637. PubMed PMID: 26540169; PubMed Central PMCID: PMCPMC4634999.
68. Danhauser K, Herebian D, Haack TB, Rodenburg RJ, Strom TM, Meitinger T, et al. Fatal neonatal encephalopathy and lactic acidosis caused by a homozygous loss-of-function variant in COQ9. Eur J Hum Genet. 2016;24(3):450-4. Epub 20150617. doi: 10.1038/ejhg.2015.133. PubMed PMID: 26081641; PubMed Central PMCID: PMCPMC4755375.
69. Sung AY, Guerra RM, Steenberge LH, Alston CL, Murayama K, Okazaki Y, et al. Systematic analysis of NDUFAF6 in complex I assembly and mitochondrial disease. Nat Metab. 2024;6(6):1128-42. Epub 20240508. doi: 10.1038/s42255-024-01039-2. PubMed PMID: 38720117; PubMed Central PMCID: PMCPMC11395703.
70. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663-76. Epub 20060810. doi: 10.1016/j.cell.2006.07.024. PubMed PMID: 16904174.
71. Huerta-Cepas J, Dopazo J, Gabaldon T. ETE: a python Environment for Tree Exploration. BMC Bioinformatics. 2010;11:24. Epub 20100113. doi: 10.1186/1471-2105-11-24. PubMed PMID: 20070885; PubMed Central PMCID: PMCPMC2820433.
72. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23(1):127-8. Epub 20061018. doi: 10.1093/bioinformatics/btl529. PubMed PMID: 17050570.
73. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523. Epub 20190403. doi: 10.1038/s41467-019-09234-6. PubMed PMID: 30944313; PubMed Central PMCID: PMCPMC6447622.
74. Hekselman I, Kerber L, Ziv M, Gruber G, Yeger-Lotem E. The Organ-Disease Annotations (ODiseA) Database of Hereditary Diseases and Inflicted Tissues. J Mol Biol. 2022;434(11):167619. Epub 20220430. doi: 10.1016/j.jmb.2022.167619. PubMed PMID: 35504357.
75. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671-5. doi: 10.1038/nmeth.2089. PubMed PMID: 22930834; PubMed Central PMCID: PMCPMC5554542.
76. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8. doi: 10.1006/meth.2001.1262. PubMed PMID: 11846609.
77. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol.

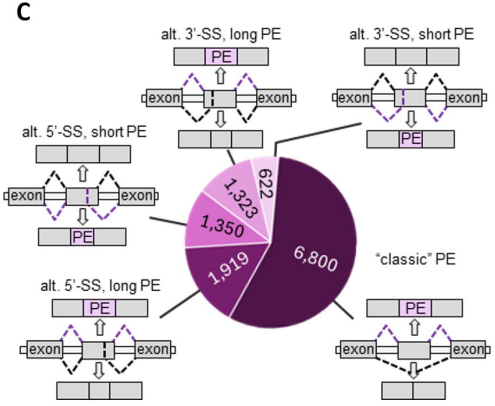
2019;37(3):224-6. doi: 10.1038/s41587-019-0032-3. PubMed PMID: 30809026; PubMed Central PMCID: PMCPMC6533916.

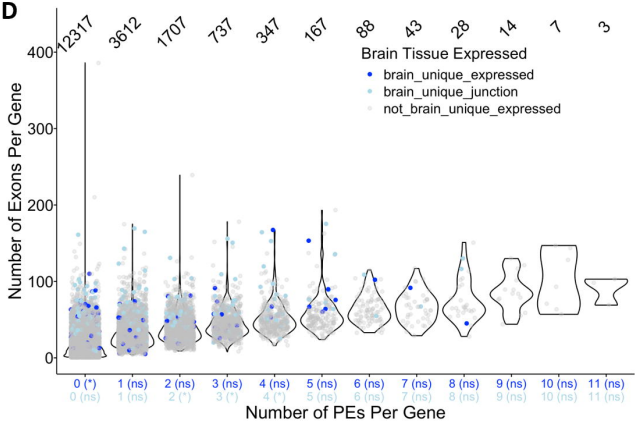

A
gene
exon
PE
exon
exon
splicing
>50 nt
start codon
EJC
PTC
5′-CAP
-AAA-3′
5′-UTR
3′-UTR
ribosome
SMG1
UPF1
UPF2
UPF3
PTC
nuclease
B
1 ratio genes with NMD isoforms
2 ratio NMD isoforms per gene
H. sapiens
Vertebrates
Invertebrates
Protists
Fungi
Plants
1
2
C
ratio of genes with NMD isoforms
average isoforms per coding gene
H. sapiens
Pearson r = 0.5418
P value < 0.0001
D
av. ratio of NMD isoforms in each gene
average exons per coding gene
Pearson r = -0.6507
P value < 0.0001
H. sapiens

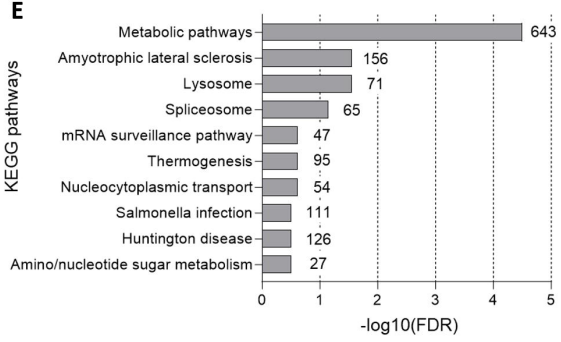
**A**

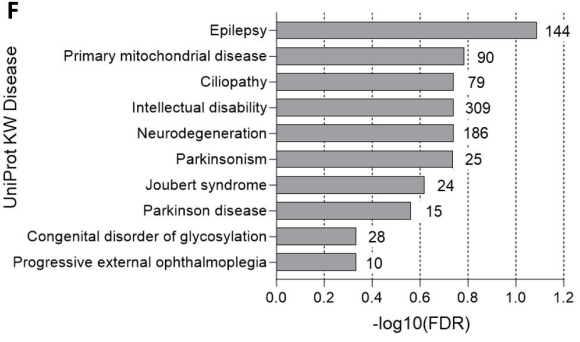
| | alt. splicing events | putative PE |
|---|---|---|
| *human* | 55,150 | 12,014 |
| *mouse* | 86,058 | 6,006 |
| *zebrafish* | 8,801 | 951 |
| *fruit fly* | 6,982 | 0 |
| *C. elegans* | 3,603 | 203 |
| *A. thaliana* | 370 | 0 |

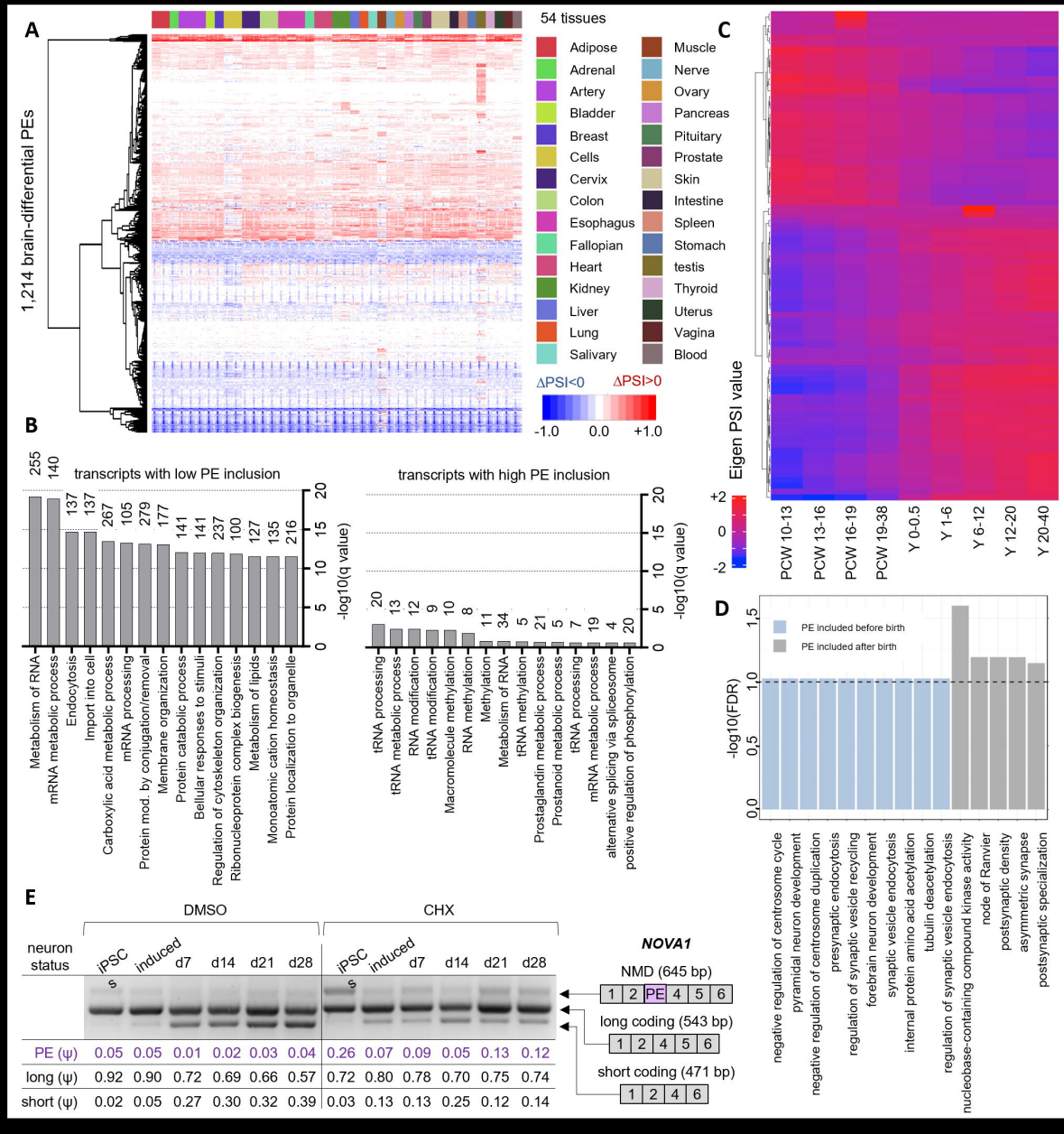


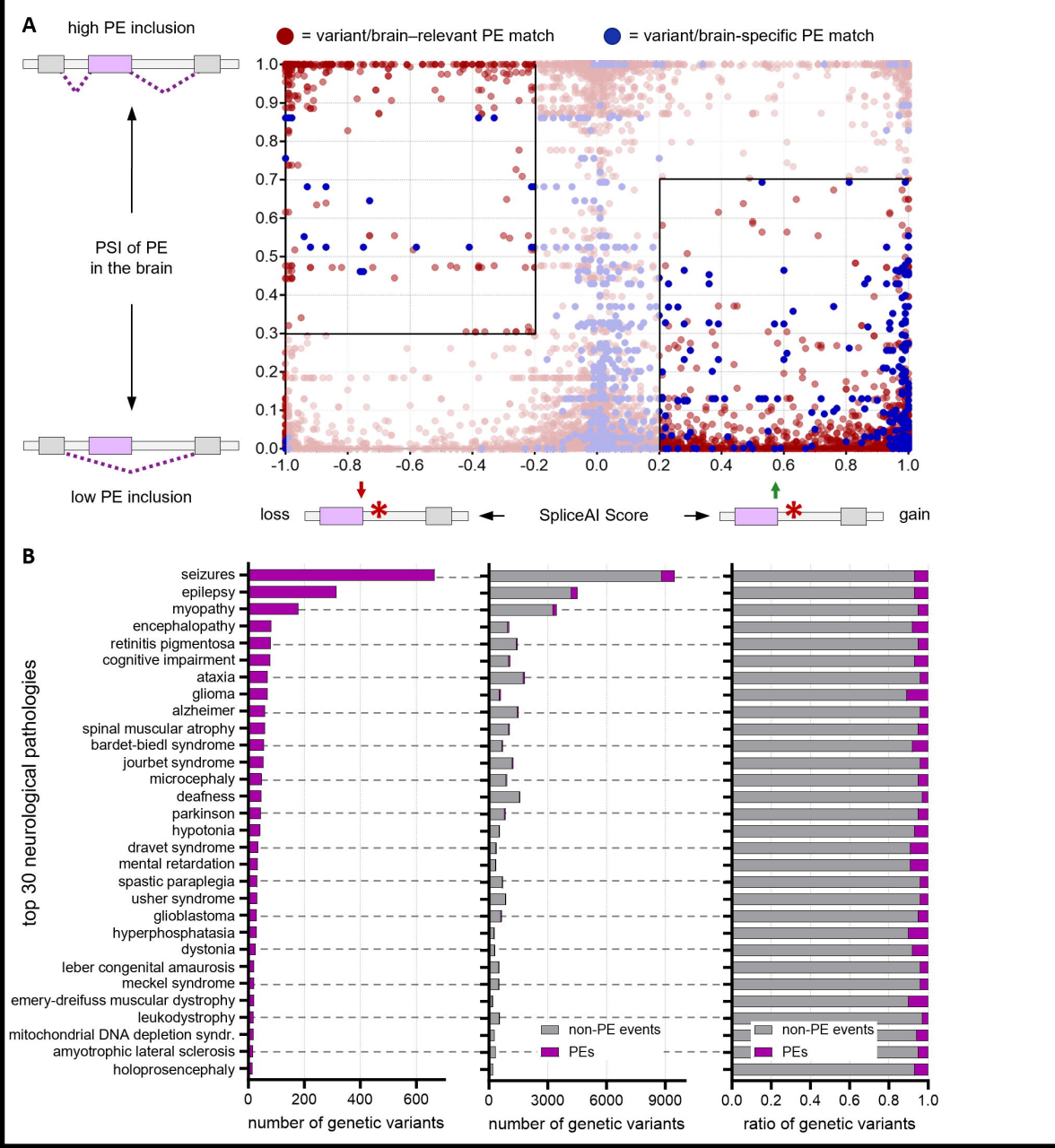



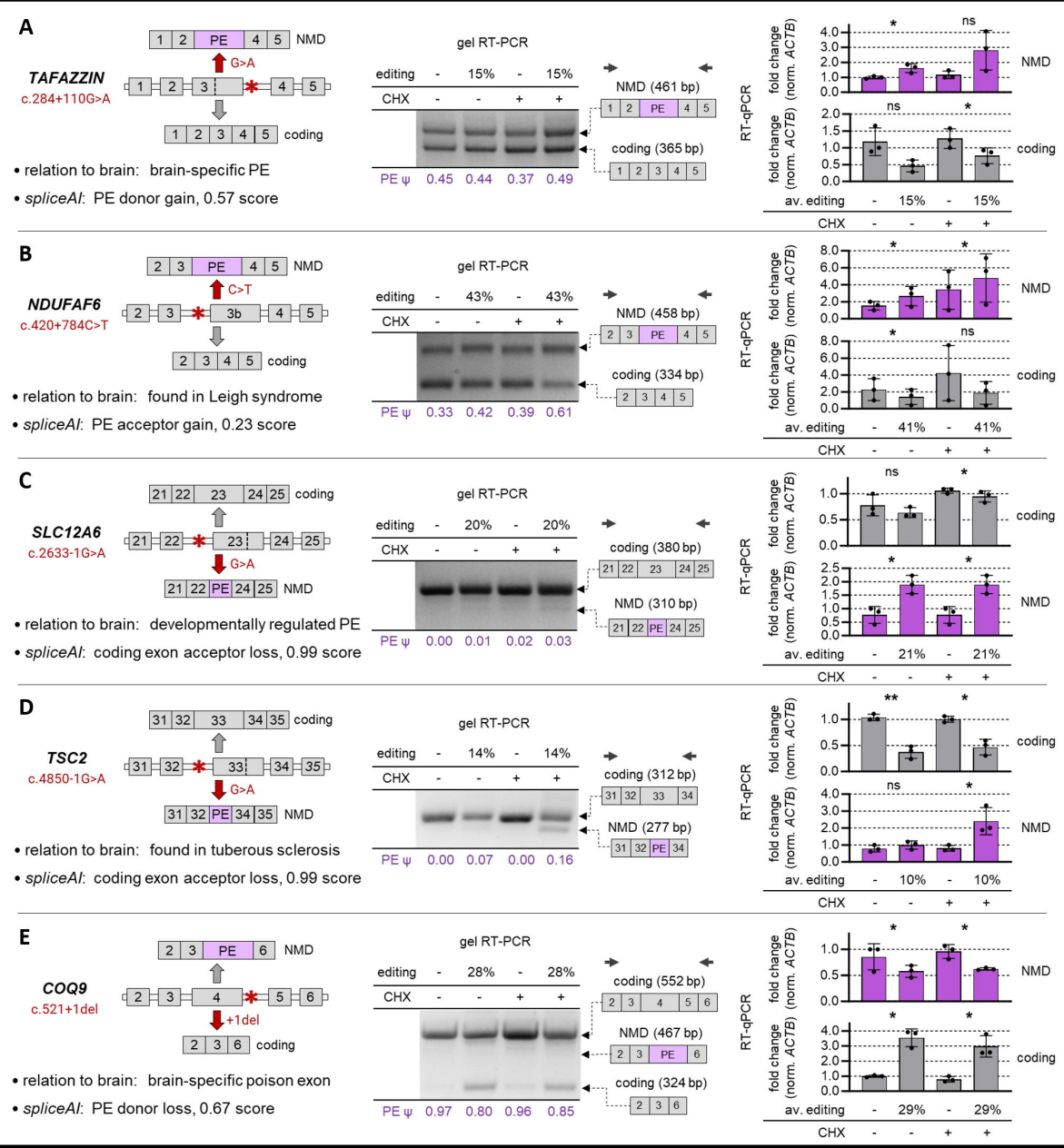



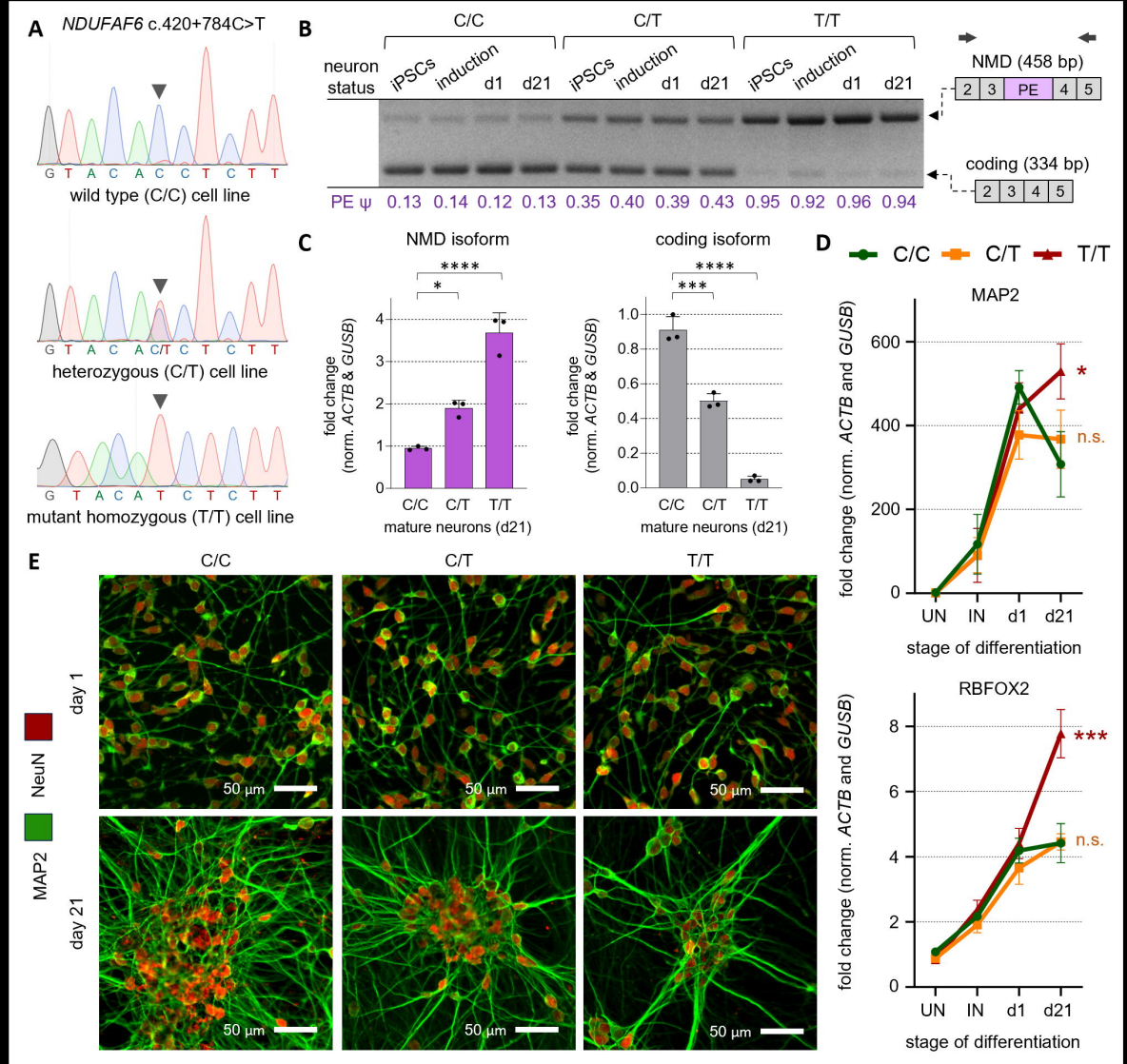




A
1,214 brain-differential PEs
54 tissues
Adipose
Adrenal
Artery
Bladder
Breast
Cells
Cervix
Colon
Esophagus
Fallopian
Heart
Kidney
Liver
Lung
Salivary
Muscle
Nerve
Ovary
Pancreas
Pituitary
Prostate
Skin
Intestine
Spleen
Stomach
testis
Thyroid
Uterus
Vagina
Blood
ΔPSI<0
ΔPSI>0
-1.0
0.0
+1.0
B
transcripts with low PE inclusion
255
140
137
137
267
105
279
177
141
141
237
100
127
135
216
Metabolism of RNA
mRNA metabolic process
Endocytosis
Import into cell
Carboxylic acid metabolic process
mRNA processing
Protein mod. by conjugation/removal
Membrane organization
Protein catabolic process
Bellular responses to stimuli
Regulation of cytoskeleton organization
Ribonucleoprotein complex biogenesis
Metabolism of lipids
Monoatomic cation homeostasis
Protein localization to organelle
transcripts with high PE inclusion
20
13
12
9
10
8
11
34
5
21
5
7
19
4
20
tRNA processing
tRNA metabolic process
RNA modification
tRNA modification
Macromolecule methylation
RNA methylation
Methylation
Metabolism of RNA
tRNA methylation
Prostaglandin metabolic process
Prostanoid metabolic process
tRNA processing
mRNA metabolic process
alternative splicing via spliceosome
positive regulation of phosphorylation
-log10(q value)
0
5
10
15
20
C
Eigen PSI value
+2
0
-2
PCW 10-13
PCW 13-16
PCW 16-19
PCW 19-38
Y 0-0.5
Y 1-6
Y 6-12
Y 12-20
Y 20-40
D
PE included before birth
PE included after birth
-log10(FDR)
0,0
0,5
1,0
1,5
negative regulation of centrosome cycle
pyramidal neuron development
negative regulation of centrosome duplication
presynaptic endocytosis
regulation of synaptic vesicle recycling
forebrain neuron development
synaptic vesicle endocytosis
internal protein amino acid acetylation
tubulin deacetylation
regulation of synaptic vesicle endocytosis
nucleobase-containing compound kinase activity
node of Ranvier
postsynaptic density
asymmetric synapse
postsynaptic specialization
E
DMSO
CHX
neuron status
iPSC
induced
d7
d14
d21
d28
s
NOVA1
NMD (645 bp)
1 2 PE 4 5 6
long coding (543 bp)
1 2 4 5 6
short coding (471 bp)
1 2 4 6
PE (ψ) 0.05 0.05 0.01 0.02 0.03 0.04 0.26 0.07 0.09 0.05 0.13 0.12
long (ψ) 0.92 0.90 0.72 0.69 0.66 0.57 0.72 0.80 0.78 0.70 0.75 0.74
short (ψ) 0.02 0.05 0.27 0.30 0.32 0.39 0.03 0.13 0.13 0.25 0.12 0.14


A
high PE inclusion
= variant/brain–relevant PE match
= variant/brain-specific PE match
PSI of PE
in the brain
low PE inclusion
loss
SpliceAI Score
gain
B
top 30 neurological pathologies
seizures
epilepsy
myopathy
encephalopathy
retinitis pigmentosa
cognitive impairment
ataxia
glioma
alzheimer
spinal muscular atrophy
bardet-biedl syndrome
jourbet syndrome
microcephaly
deafness
parkinson
hypotonia
dravet syndrome
mental retardation
spastic paraplegia
usher syndrome
glioblastoma
hyperphosphatasia
dystonia
leber congenital amaurosis
meckel syndrome
emery-dreifuss muscular dystrophy
leukodystrophy
mitochondrial DNA depletion syndr.
amyotrophic lateral sclerosis
holoprosencephaly
non-PE events
PEs
number of genetic variants
number of genetic variants
ratio of genetic variants


A
TAFAZZIN
c.284+110G>A
NMD
G>A
coding
• relation to brain: brain-specific PE
• spliceAI: PE donor gain, 0.57 score
gel RT-PCR
editing - 15% - 15%
CHX - - + +
NMD (461 bp)
coding (365 bp)
PE ψ 0.45 0.44 0.37 0.49
RT-qPCR
fold change (norm. ACTB)
av. editing - 15% - 15%
B
NDUFAF6
c.420+784C>T
NMD
C>T
coding
• relation to brain: found in Leigh syndrome
• spliceAI: PE acceptor gain, 0.23 score
gel RT-PCR
editing - 43% - 43%
NMD (458 bp)
coding (334 bp)
PE ψ 0.33 0.42 0.39 0.61
av. editing - 41% - 41%
C
SLC12A6
c.2633-1G>A
coding
G>A
NMD
• relation to brain: developmentally regulated PE
• spliceAI: coding exon acceptor loss, 0.99 score
gel RT-PCR
editing - 20% - 20%
coding (380 bp)
NMD (310 bp)
PE ψ 0.00 0.01 0.02 0.03
av. editing - 21% - 21%
D
TSC2
c.4850-1G>A
coding
G>A
NMD
• relation to brain: found in tuberous sclerosis
• spliceAI: coding exon acceptor loss, 0.99 score
gel RT-PCR
editing - 14% - 14%
coding (312 bp)
NMD (277 bp)
PE ψ 0.00 0.07 0.00 0.16
av. editing - 10% - 10%
E
COQ9
c.521+1del
NMD
+1del
coding
• relation to brain: brain-specific poison exon
• spliceAI: PE donor loss, 0.67 score
gel RT-PCR
editing - 28% - 28%
coding (552 bp)
NMD (467 bp)
coding (324 bp)
PE ψ 0.97 0.80 0.96 0.85
av. editing - 29% - 29%


A
NDUFAF6 c.420+784C>T
G T A C A C C T C T T
wild type (C/C) cell line
G T A C A C/T C T C T T
heterozygous (C/T) cell line
G T A C A T C T C T T
mutant homozygous (T/T) cell line
B
C/C
C/T
T/T
neuron status
iPSCs
induction
d1
d21
NMD (458 bp)
2 3 PE 4 5
coding (334 bp)
2 3 4 5
PE ψ 0.13 0.14 0.12 0.13 0.35 0.40 0.39 0.43 0.95 0.92 0.96 0.94
C
NMD isoform
coding isoform
fold change (norm. ACTB & GUSB)
mature neurons (d21)
C/C C/T T/T
D
C/C C/T T/T
MAP2
RBFOX2
fold change (norm. ACTB and GUSB)
UN IN d1 d21
stage of differentiation
n.s.
E
C/C
C/T
T/T
day 1
day 21
NeuN
MAP2
50 µm