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

Paolo Pigini[1,2,4], Huilin Xu[1,2,3,4], 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[1,2,3,*], Elisabetta Morini[1,2,*]

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

[4] 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 exon cassettes whose inclusion creates premature termination codons (PTCs) and triggers nonsense-mediated decay (NMD) of the mature transcript, therefore reducing protein expression. Despite their important role in post-transcriptional regulation, PEs remain poorly annotated due to the lack of systematic, transcriptome-wide approaches. In this study, we comprehensively investigated the function of PEs in the human brain and assessed the impact of pathogenic variants on their splicing. A comparative analysis of 957 eukaryotic transcriptomes revealed that humans exhibit the highest enrichment of NMD-targeted transcripts. To delineate the full landscape of PEs in the human transcriptome, we considered conserved (PhyloP score ≥ 20) alternatively spliced exons, less than 300 base pairs (bp) long, with the ability to introduce PTCs positioned more than 150 base pairs (bp) downstream of the transcription start site and outside the last two exons of the transcript. Our analysis identified 12,014 PEs in the human genome. For each PE, we calculated the percent-spliced-in (PSI) value across tissue types and developmental stages using GTEx and BrainSpan RNA-seq dataset, respectively. Overall, we identified 117 PEs uniquely found in the human brain, 1,214 PEs with brain-specific differential splicing compared to other tissues, and 1,610 PEs which splicing change during brain development. Integrating ClinVar and SpliceAI, we identified 1,877 annotated pathogenic variants predicted to affect the splicing of 891 PEs in the human brain. Notably, many of these variants were associated with neurodevelopmental and neurodegenerative disorders such as epilepsy, intellectual disability and frontotemporal dementia. We functionally validated the impact of selected variants using an optimized CRISPR prime-editing workflow in human cell lines. Our findings highlight PEs as pivotal regulators of gene expression in the human brain and establish a foundation for therapeutic strategies targeting PE mis-splicing in neurological diseases.

51 ## Introduction

52 Alternative mRNA splicing is a fundamental post-transcriptional regulatory mechanism that
53 significantly expands the coding potential of eukaryotic genomes. Recent studies suggest that
54 disruptions in this process contribute to approximately 10-30% of human genetic disorders[1,2],
55 with nearly half of these alterations occurring in the intronic regions[3]. Alternative splicing is
56 particularly important in the brain, where it plays critical roles in neurodevelopment, neuronal
57 differentiation, and synaptic function[4-9]. This regulatory mechanism helps define the
58 differentiation trajectories of various cell types within the nervous system[7,10,11] and appears to
59 have undergone substantial evolutionary refinement in vertebrates, especially primates[12].
60 Consequently, splicing alterations have been increasingly implicated in the pathogenesis of
61 neurological disorders[13,14], with non-coding genetic variants emerging as particularly
62 significant contributors[15].

63 Poison exons (PE) are naturally occurring, evolutionarily conserved exons that introduce
64 premature termination codons (PTCs) when included in the mature transcript, thereby reducing
65 protein expression[16-23]. PEs function as finely tuned regulatory elements that control gene
66 expression at the post-transcriptional level[24] and have been implicated in diverse physiological
67 processes including cellular differentiation and survival, as well as in pathological conditions
68 such as cancer[25]. Within the human brain, PEs are important for neurodevelopment and many
69 neuronal functions including regulation of synaptic proteins and neuron-specific RNA-binding
70 proteins[22,26,27]. Early studies revealed that PEs regulate chromatin modifiers in the mammalian
71 brain[28]. Subsequent analyses of alternative splicing in the cerebral cortex demonstrated that
72 PEs contribute to establishing cellular identity during differentiation[5]. PEs can also have
73 specialized functions, as exemplified by a PE in the *RBM3* gene that confers neuroprotection
74 in human induced pluripotent stem cell (iPSC)-derived neurons[29]. The splicing of PEs in the
75 brain is regulated by a network of RNA-binding proteins, including *RBFOX1*, *PTB1* and

76 *NOVA1*[5,27,30], which are also major regulators of splicing in the brain[4,5]. In murine models,
77 nonsense-mediated decay (NMD) has been proposed to downregulate non-neuronal genes
78 during brain development[31], further emphasizing its importance in establishing neural identity.
79 The dysregulation of PE splicing has been implicated in various neurological disorders[31].
80 For example, aberrant inclusion of the PE "20N" in the *SCN1A* gene is associated with Dravet
81 syndrome (DS)[32-35]. An intronic variant disrupting normal skipping of a PE in the *FLNA* gene
82 was found in a family segregating with periventricular nodular heterotopia (PVNH), a
83 condition caused by abnormal neuronal migration during fetal brain development[5]. In
84 amyotrophic lateral sclerosis (ALS), *TDP-43* dysfunction disrupts PE regulation in iPSC-
85 derived neurons, potentially contributing to disease pathogenesis[36]. Beyond their emerging role
86 in disease aetiology, PEs are also increasingly recognized as potential novel therapeutic targets
87 for the treatment of neurological disorders[37-40].
88 Recent advances in computational approaches have improved identification of both PEs[41]
89 and genetic variants affecting their splicing[2]. However, the field still lacks a comprehensive
90 catalog of PE splicing events[42-44]. In this study, we present an evolutionary analysis of NMD-
91 targated transcripts and demonstrated that humans exhibit a striking enrichment of genes
92 expressing transcripts containing PEs. We leveraged large RNASeq datasets from GTEx and
93 BrainSpan projects to identify brain-specific and neurodevelopmentally regulated PEs. By
94 integrating ClinVar variant annotations with SpliceAI predictions, we identified pathogenic
95 variants that disrupt PE splicing and are associated with neurodevelopmental and
96 neurodegenerative disorders. To functionally validate these findings, we applied an optimized
97 CRISPR prime editing workflow in human cell cultures to assess the molecular consequences
98 of PE mis-splicing. Together, this work provides new insights into the role of PEs in brain
99 function and disease, and highlights their potential as therapeutic targets for currently incurable
100 neurological conditions.

101

102 ## Results

103 **The human transcriptome exhibits a unique enrichment of NMD-associated transcripts**

104 Conventionally, NMD-associated transcripts are defined by the presence of a PTC located
105 more than 50 base pairs (bp) upstream of the last exon–exon junction[19-22] (Fig. 1A). Using this
106 criteria, we systematically identified NMD transcripts across a dataset of 957 annotated
107 eukaryotic genomes from Ensembl Genomes 61, spanning the major clades of vertebrates,
108 invertebrates, fungi, protists, and plants (Fig. 1B). This large-scale dataset enabled us to
109 evaluate the evolutionary distribution and potential functional relevance of NMD across
110 eukaryotes. On average, 2.67% of coding genes per genome were predicted to produce at least
111 one NMD isoform, though this proportion varied substantially across evolutionary lineages.
112 Vertebrates (3.72%) and plants (3.70%) exhibited the highest fraction of coding genes
113 producing at least one NMD transcript, followed by invertebrates (1.85%), while lower
114 percentages were observed in fungi (0.63%) and protists (0.11%) (Fig. 1B and Fig. S1A). A
115 similar trend was observed when analysing the average number of NMD transcripts per gene,
116 with vertebrates (1.25), invertebrates (1.58), and plants (1.21) showing higher values compared
117 to protists (0.18) and fungi (0.37) (Fig. S1B). We further evaluated the overall frequency of
118 NMD isoforms by calculating the ratio of NMD transcripts to all coding transcripts. Vertebrates
119 (0.61), invertebrates (0.80), and plants (0.46) displayed higher NMD frequencies than fungi
120 (0.30) and protists (0.18) (Fig. S1C). These data suggest that alternative splicing leading to
121 NMD is particularly prominent in multicellular lineages with complex developmental
122 processes. Among all species analysed, humans emerged as the species with the highest number
123 of NMD transcripts. We found that 39.6% of human protein-coding genes give rise to at least
124 one NMD transcript. Furthermore, humans exhibit an average of 2.77 NMD isoforms per
125 coding gene, and 35% of all annotated protein-coding transcripts are predicted to undergo

126 NMD. These findings strongly suggest that NMD plays a disproportionately large role in
127 regulating transcript stability in humans, potentially reflecting the complexity and regulatory
128 precision of human gene expression.

129 To explore the evolutionary and molecular drivers of NMD, we investigated whether several
130 genomic and transcriptomic features correlate with NMD frequency. Surprisingly, we observed
131 no significant correlation between the total number of coding genes and the percentage of genes
132 with NMD transcripts across species (Fig. S2A), nor between gene count and the average
133 number of NMD isoforms per gene (Fig. S2B) or their overall frequency (Fig. S2C). Notably,
134 vertebrates often displayed higher percentages of NMD-regulated genes despite having fewer
135 genes overall compared to plants, suggesting that genome size alone does not explain NMD
136 frequency. In contrast, we found a positive association between transcriptome complexity and
137 NMD abundance. Species with a greater number of transcripts per gene tended to produce more
138 NMD isoforms (Fig. 1C, S2D and S2E). This pattern may reflect both biological and technical
139 factors. Biologically, extensive alternative splicing may naturally increase the chance of
140 producing PTC-containing isoforms. Technically, more comprehensive annotations in well-
141 studied species (e.g., human, mouse, Arabidopsis) may lead to increased detection of NMD
142 candidates. Species with fewer transcripts per gene exhibited higher frequencies of NMD
143 among their annotated isoforms (Fig. S2F), possibly indicating annotation biases where only
144 dominant or easily detectable NMD isoforms are captured. We next examined whether exon
145 density correlates with NMD transcript. We observed a mild positive correlation between the
146 number of exons per gene and the number of NMD transcripts (Fig. 1D, S2G and S2H),
147 consistent with the idea that splicing complexity increases opportunities for NMD-generating
148 splicing events. However, similar to the trend for transcript numbers, species with lower exon
149 density also showed higher NMD frequencies (Fig. S2I), suggesting that detection sensitivity
150 and annotation completeness may influence these patterns.

151 Our evolutionary analysis highlights that NMD transcripts are not only widespread across
152 eukaryotes but are particularly enriched in vertebrates, and especially in humans. Humans
153 exhibit the highest density of NMD events among all surveyed species. This prompted us to
154 systematically characterize the landscape of human alternative splicing events involving PEs.
155

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

158 Building on the finding that humans exhibit a high frequency of NMD isoforms, we next
159 sought to systematically define the full repertoire of PEs in the human transcriptome. PEs are
160 highly conserved exon cassettes whose inclusion introduces PTCs in the mature transcript,
161 thereby triggering NMD and reducing protein levels. To identify these elements at scale, we
162 developed an *in silico* strategy using genomic sequences and exon annotations (Fig. S3A, see
163 Methods). We defined PEs as exons that: (1) are ≤ 300 bp in length; (2) have an average PhyloP
164 conservation score ≥ 20; and (3) generating a PTC more than 150 bp downstream of the
165 transcription start site and upstream of the last two exons. This more stringent definition of
166 PTC location was specifically chosen to minimize the likelihood of NMD escape[45]. Using this
167 approach, we identified 12,014 putative PEs, accounting for 21.8% of 55,150 annotated
168 alternative splicing events in the human genome, each predicted to produce NMD isoforms
169 upon inclusion (Fig. 2A). A comparative analysis across model organisms, including *Mus*
170 *musculus*, *Danio rerio*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Saccharomyces*
171 *cerevisiae*, and *Arabidopsis thaliana*, revealed species-specific differences in PE usage, with
172 humans having the highest number of putative PE splicing events (Fig. 2A).

173 To evaluate the robustness of our PE identification, we then assessed whether their inclusion
174 triggers NMD by measuring transcript levels after cycloheximide (CHX) treatment, which
175 inhibits translation and stabilizes NMD-targeted RNAs[31,46,47]. We performed this analysis in

176 HEK293T cells for a subset of transcripts. RT-qPCR following CHX or DMSO treatment
177 revealed that in the majority of cases (12 out of 16, 75%), transcripts containing the PE were
178 significantly stabilized by CHX, confirming their degradation via NMD (Fig. 2B and S3B).
179 Transcripts without CHX-induced upregulation likely reflect false positives or cases where
180 downstream splicing events restore the reading frame.

181 Analysis of common features of human PEs showed that more than half of these PE splicing
182 events (6,800 events, or 56.6%) arise from exons that are either included or skipped, while the
183 remainder result from alternative 3′ or 5′ splice site usage (Fig. 2C and S3C). While longer
184 genes tend to harbour more PEs, genes that are uniquely expressed in the brain, or that contain
185 brain-specific exons, contain significantly fewer PEs (Fig. 2D). Furthermore, our putative PEs
186 did not show a preferential position along the gene (Fig. S3D).

187 To gain insight into the biological functions of gene containing PEs, we conducted pathway
188 enrichment analyses. KEGG pathway analysis revealed significant enrichment in core cellular
189 processes such as metabolism and RNA splicing. Notably, several splicing regulators were
190 among the enriched genes. These findings are consistent with the well-established observation
191 that many splicing factors autoregulate their own expression through negative feedback loops
192 mediated by PE inclusion (Fig. 2E)[17,25,48]. Additionally, disease ontology enrichment using the
193 UniProt KW Disease database (Fig. 2F) identified a strong association between PE-containing
194 genes and neurological disorders, most prominently epilepsy, intellectual disability and
195 neurodegeneration, reinforcing the hypothesis that PE splicing plays a key regulatory role in
196 brain functions[22,26,27].

197 Overall, our analysis identifies over 12,000 PE splicing events as a widespread and
198 functionally diverse class of alternative splicing, enriched in regulatory and disease-associated
199 genes. Experimental validation confirms that PE inclusion commonly triggers NMD, with

200 strong enrichment in neurological pathways supporting their relevance in brain-specific
201 transcriptome regulation.
202
203 **PEs exhibit brain-specific inclusion patterns and dynamic regulation during**
204 **neurodevelopment**
205 To investigate the spatiotemporal regulation of PE splicing in the human brain, we leveraged
206 large-scale transcriptomic datasets from both adult and developing tissues. We hypothesized
207 that dynamic inclusion or exclusion of PEs may contribute to brain-specific transcriptome and
208 gene regulation. We first analysed RNAseq data from the Genotype-Tissue Expression (GTEx)
209 Project, which encompasses 17,382 libraries derived from 54 distinct tissues, including 13
210 brain regions from 838 donors without major chronic diseases. This analysis (see Methods)
211 revealed that 1,214 PEs are differentially spliced in the brain compared to other tissues (Fig.
212 3A), with an additional 117 PEs uniquely detected in the brain. As anticipated, inclusion of
213 these PEs correlated inversely with expression of their host genes (Fig. S4), consistent with
214 their role in post-transcriptional downregulation via NMD. To explore the functional
215 implications of these brain-regulated PEs, we conducted GO analyses by stratifying them based
216 on PSI values, using 0.5 as a PSI cut-off between upregulated transcripts (with PE exclusion,
217 or PSI < 0.5) and downregulated transcripts (with PE inclusion, or PSI > 0.5), and found that
218 genes showing PE exclusion a possibly upregulation in the brain are significantly enriched in
219 RNA metabolism and splicing-related pathways (Fig. 3B).
220 We then analysed the developmental regulation of PEs using the BrainSpan Atlas, a
221 comprehensive resource comprising 607 libraries from 41 individuals across 16 distinct brain
222 regions. These libraries cover key developmental stages of individuals, ranging from 10 post-
223 conception weeks (pcw) to 40 years-old. This analysis (see Methods) identified 1,610 PEs
224 differentially spliced PEs in at least one of the 16 brain regions. Clustering of these PEs

225 revealed distinct developmental splicing patterns, characterized by increased PE inclusion
226 either before or after birth (Fig. 3C). GO enrichment of the clustered PEs with higher inclusion
227 before birth revealed significant associations with neurodevelopmental and synaptic processes
228 (Fig. 3D), while clustered PEs with higher inclusion rate after birth showed involvement in
229 postsynaptic structures and functions (Fig. 3E). It is worth noting that the PE usage showed
230 both linear and non-linear trajectories and were region specific (Fig. S6A-Q).

231 Finally, given that the expression of splicing factors is known to be self-regulated via
232 alternative PE usage[25,49], we investigated whether PEs regulate splicing factors in the brain. To
233 do so, we analyzed well-known splicing regulators with key roles in the nervous system,
234 including *PTBP1*[50], *RBFOX1*[50], *RBFOX2*[50], *NOVA1*[50], *TARDBP*[50] and *ELAVL1*[51], as well as
235 core NMD regulators such as *UPF1*, *UPF2* and *UPF3*[23,52]. Among these, *RBFOX1*, *NOVA1*,
236 *TARDBP* and *UPF3B* contained brain-specific PEs. To validate the neuron-specific nature of
237 these PEs, we assessed their splicing in iPSC-derived i3N neurons, a model of human neurons
238 that mature through *NGN2* induction after 21 days of differentiation[53]. We specifically selected
239 this neuronal cell model because the PSI values of the predicted PEs closely mirror those
240 observed in the human brain (Fig. S5A). After *NGN*-mediated induction and a differentiation
241 period of 28 days (Fig. S5B and S5C), we assessed PE splicing by RT-PCR. We found that PE
242 inclusion in *NOVA1* was higher in iPSCs but declined during neuronal maturation, coinciding
243 with increased expression of coding isoforms[54], which was more evident with CHX treatment
244 (Fig. 3F). No significant differentiation-specific splicing changes were found for PEs in
245 *RBFOX1*, *TARDBP* and *UPF3B*, either with or without CHX treatment (data not shown). Of
246 note, CHX analysis on day-21 neurons, consistent with the result in HEK293T cells, confirmed
247 that for the majority of putative PEs (75%) indeed trigger NMD (Fig. S5D and S5E).

248 Our data reveal that PE splicing is highly dynamic across brain regions and developmental
249 stages, suggesting a widespread mechanism of post-transcriptional regulation in the nervous

250 system. The enrichment of PE-regulated genes in RNA metabolism and neurodevelopmental
251 pathways, along with validation in key splicing regulators, supports a model in which PEs act
252 as molecular switches to fine-tune gene dosage with temporal and spatial precision.
253

254 **Identification of pathogenic variants associated with neurological disorders and**
255 **disrupting poison exon splicing**

256 To identify variants affecting PE splicing, we analysed all pathogenic and likely pathogenic
257 variants in the ClinVar database and used SpliceAI to predict their impact on mRNA splicing.
258 As expected, most of the pathogenic variants predicted to impact mRNA splicing are located
259 in intronic regions (Fig. S7). For each variant, we calculated a splicing change score based on
260 SpliceAI predictions, where positive scores indicate increased PE inclusion, and negative
261 scores indicate decreased inclusion (significant absolute value > 0.2). We combined these
262 predictions with PSI values from GTEx brain tissues to prioritize variants significantly
263 affecting PE splicing in the brain (PSI < 0.3 for increased inclusion, or > 0.7 for decreased
264 inclusion). Using this approach, we identified 1,877 pathogenic variants predicted to affect the
265 splicing of 891 PEs (Fig. 4A). These include variants associated with neurological disorders,
266 variants in genes with known brain functions, and variants affecting brain-specific PE, as
267 previously defined. We then mapped these pathogenic variants to the associated diseases and
268 clinical phenotypes using the ODiseA ontology. A striking proportion of identified genes were
269 linked to seizures and epileptic syndromes, including Dravet syndrome, consistent with
270 previous findings that support a role for PE mis-splicing as a pathogenetic mechanism[22,26] (Fig.
271 4B). In addition to epilepsy-related conditions, we observed enrichments in genes associated
272 with various neurodevelopmental and neurodegenerative conditions, such as Bardet-Biedl
273 syndrome, Usher syndrome, and Parkinson's disease. These findings suggest that altered PE

274 splicing may represent a common mechanism contributing to a wide range of neurological
275 disorders.

276 To functionally validate the effects of pathogenic variants on PE splicing, we developed and
277 optimized a CRISPR prime editing workflow (Fig. S8A, see Methods). As a proof of concept,
278 we designed and tested 23 pegRNAs to introduce pathogenic variants predicted to affect PE
279 splicing. Editing efficiency was assessed in HEK293T cells, with 20 of the 23 achieving >5%
280 efficiency, and 17 exceeding 10% (average: 21.6%; Fig. S8B). The effect of these variants on
281 PE splicing was further assessed by RT-PCR, using cells treated with CHX in order to visualize
282 the full extent of PE splicing without the interference of NMD. After examining unedited
283 HEK293T controls, we found that only 10 of the 20 transcripts exhibit a baseline splicing
284 pattern that mirrors the brain, either showing detectable inclusion of the relevant PE or the
285 same isoform ratios seen in brain. The remaining 10 transcripts were excluded because they
286 either lacked detectable PE inclusion in HEK293T cells or displayed a splicing pattern that
287 does not recapitulate the brain-specific profile (Fig. S8C). By focusing on the subset of
288 transcripts whose splicing matches the brain context, we ensure that any observed shifts upon
289 editing reliably reflect the variant's effect on PE regulation. Importantly, in all 10 selected
290 gene, introduction of the pathogenic variant produced the predicted change in PE inclusion
291 (Fig. S8C). From these 10 genes, we selected five for in-depth analysis based on their clinical
292 relevance and the presence of clearly defined PE regulatory elements. Variants in these genes
293 were predicted to have a distinct effect on PE splicing. Two mutations were predicted to
294 increase PE inclusion. c.284+110G>A located in intron 3 of the *TAFFAZIN* gene was predicted
295 by SpliceAI (score: 0.57) to enhance PE inclusion and resulted in a +0.12 PSI change in a bulk-
296 edited population with an average of 15% editing efficiency (Fig. 5A). c.420+784C>T in
297 *NDUFAF6*, associated with Leigh syndrome and located downstream of exon 4, was predicted
298 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). Two variants were predicted to indirectly increase PE inclusion by disrupting canonical exon recognition, thereby favouring PE-containing isoforms. The c.2633-1G>A variant in *SLC12A6*, which affects the splice acceptor of exon 22, 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, c.4850-1G>A in *TSC2*, 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, one variant resulted in decreased PE inclusion. c.521+1del in *COQ9*, 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. Our combined computational and experimental approach provides a scalable strategy for interpreting the splicing consequences of rare and clinically relevant pathogenic variants, particularly in the context of neurological disease.

Discussion

Alternative splicing coupled with NMD is a central mechanism for regulating transcript stability and fine-tuning gene expression in eukaryotes. Among its many forms, PE inclusion represents a specialized, highly conserved splicing event that triggers NMD and reduces protein output. 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

324 results identified thousands of human PEs with tissue- and developmentally regulated splicing
325 patterns in the brain, as well as hundreds of pathogenic variants affecting their splicing.
326 Our comparative analysis of 957 eukaryotic transcriptomes revealed that NMD isoforms are
327 particularly enriched in vertebrates, especially in humans. Nearly 40% of human coding genes
328 give rise to at least one NMD transcript, and over a third of all protein-coding transcripts are
329 predicted to undergo decay. These trends were not explained by gene count or genome size,
330 but instead correlated with transcriptomic complexity, especially alternative splicing. This
331 suggests that species with more intricate splicing regulation have evolved to utilize NMD as a
332 widespread for fine-tuning gene expression at the post-transcriptional level. Within this
333 context, PEs likely represent an evolutionarily refined strategy to control gene dosage in a
334 context-specific manner, especially in complex organs such as the brain.
335 To map PE landscape genome-wide, we identified over 12,000 putative PEs across the
336 human transcriptome, accounting for nearly 22% of all annotated alternative splicing events.
337 These PEs were enriched in genes involved in RNA metabolism and frequently found in
338 splicing factors, which often autoregulating their own expression through negative feedback
339 loops involving PE inclusion[17,25,48]. While longer genes were more likely to contain PEs, those
340 with brain-specific expression showed a significant depletion of PEs, suggesting selective
341 pressure to preserve expression of critical neural genes. Experimental validation confirmed that
342 PE inclusion typically leads to NMD, validating our predictive framework. The overlap of PE-
343 containing genes with neurological disorder–associated pathways underscores the clinical
344 significance of PE splicing as a key regulatory mechanism in the human brain.
345 We next investigated PE regulation in the human brain and uncovered widespread tissue-
346 and developmental-specific splicing patterns. Using RNA-seq data from GTEx, we identified
347 PEs that are differentially spliced in the brain compared to other human tissues. We analyzed
348 BrainSpan RNA-seq data and found dynamic PE splicing changes across brain development,

349 with distinct shifts occurring before and after birth. Notably, we observed dramatic changes in
350 PE inclusion for genes such as NOVA1 during neuronal differentiation, suggesting tight
351 regulation of PE usage during neurogenesis. Together, our findings position PEs as temporally
352 and spatially regulated molecular switches that fine-tune gene expression in the developing and
353 adult brain.

354 To assess the clinical impact of PE mis-splicing, we integrated ClinVar data with SpliceAI
355 predictions and identified pathogenic variants predicted to affect PE splicing in neurological
356 disorders. Many of these variants were located in intronic region, reinforcing the critical role
357 of noncoding sequences in splicing regulation and their contribution to splicing-related
358 disorders. Using CRISPR prime editing, we functionally validated representative variants and
359 confirmed that they induced alterations in endogenous mRNA splicing, validating the accuracy
360 of our predictions. These variants are associated with a broad spectrum of neurological
361 conditions, including epilepsy, intellectual disability, and neurodegeneration, highlighting the
362 role of PE mis-splicing in brain disorders. Our experimental pipeline provides a scalable and
363 adaptable framework for interpreting and validating splicing variants within their native
364 transcriptomic context.

365 This study establishes poison exons as a widespread, evolutionarily conserved, and
366 dynamically regulated mechanism of transcriptome control in the human brain. By integrating
367 computational prediction with functional validation, we demonstrated that PEs not only play a
368 critical role in gene expression regulation during brain development but also represent a
369 common pathogenetic mechanism in neurological disease. Our findings highlight the
370 therapeutic potential of targeting PE mis-splicing and provide a foundational resource for
371 advancing the study of splicing regulation in the human brain and its contribution to disease.

372

373 **Materials and Methods**

374

375 **Evolutionary analysis**

376 A Python-based computational workflow was developed to retrieve genome annotations of
377 Eukaryotic species and predict NMD-prone transcripts. Genome annotations were downloaded
378 from Ensembl Genomes 61. Each genome annotation was analysed to collect all annotated
379 transcripts and predict whether they undergo NMD based on the position of the annotated
380 termination codon, which should be >50 bp upstream of the last exon–exon junction. The
381 phylogenetic tree was generated using the ETE package[55] based on taxonomic information
382 from NCBI, and it was visualized using iTOL[56]. Bar plots and scatter plots were generated
383 using GraphPad Prism.

384

385 **Identification of putative PE from transcriptomic reference**

386 This study focuses on defining the PE landscape by leveraging the annotated transcriptomes,
387 based on the assumption that an annotated exon can be alternatively included as a PE in an
388 otherwise protein-coding transcript and result in a downstream PTC. The derived PTC is not
389 necessarily within the PE itself but has to be more than 150 bp from the transcription start site
390 and not in the final two exons of the host gene, ensuring it doesn't escape NMD[45]. To search
391 for PE from the transcriptome reference, non-redundant exons whose length ≤ 300 bp of a gene
392 were first extracted from the GTF file. Each extracted exon was then computationally inserted
393 into each transcript of the gene whose "transcript_biotype" is "protein-coding". If the insertion
394 generated a PTC meeting the criteria above, the inserted exon would be marked as a putative
395 PE. Note, this also means that whether an exon is "poisoning" or not depends on its transcript
396 context. Depending on the relative position of the putative PE and the protein-coding transcript,
397 the PE can be classified as one of "inclusion", "altExonLen", and "IR" mode (Fig. S3A).

398

399 **PSI value calculation for PE**

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

405 $$PSI = \frac{0.5 * (R_1 + R_2)}{0.5 * (R_1 + R_2) + R_3}$$

406 For altExonLen mode, PSI was calculate as:

407 $$PSI = \frac{R_1}{(R_1 + R_3)}$$

408 For intron retention, PSI was calculated as:

409 $$PSI = \frac{\text{Median}\ \{R_i\}}{\text{Median}\ \{R_i\} + R_3}$$

410 Where i denotes the i_{th} reads that aligned to the IR region. (Fig. S3A). Note, this should not
411 be confused with the usage of an exon being explicitly used as a PE. From the short-read data
412 used in this study, only the overall usage of an exon is measurable, regardless of whether it is
413 included in a transcript as a poison or non-PE.

414

415 **Identification of tissue-specific PE from GTEx**

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

421

422 **Identification of developmental-time-specific PE from BrainSpan**

423 The sample-wise PSI profiles were first constructed for all BrainSpan samples. For each
424 tissue type, a generalized additive model (GAM) was applied to investigate the trajectories of
425 PE inclusion over ages. In the GAM models, a smooth term for developmental time points was
426 used to capture the temporal dynamics of PE inclusion. The significance of the smooth term
427 indicates whether PE inclusion levels exhibit developmental trajectory. PE events with similar
428 trajectories were clustered using Pearson correlation coefficients. Principal component analysis
429 (PCA) was then applied to each cluster to extract an eigen PSI value, summarizing the pattern
430 of splicing changes within the group.
431
432 **Analysis of pathogenic variants**
433 Pathogenic and likely-pathogenic mutations were extracted from ClinVar VCF file (version
434 20231112). For each such mutation, SpliceAI[2] was applied to predict the most influenced
435 putative splice sites (ss) nearby. Only the most impacted putative ss with an absolute SpliceAI
436 score change ≥ 0.2 were retained. The genomic coordinates of these impacted ss were then
437 matched with the coordinates of PE's ss. When an exact match was found, the mutation would
438 be considered to disrupt the splicing of PE.
439 To identify variants affecting PE splicing, we analysed all pathogenic and likely pathogenic
440 variants in the ClinVar database and used SpliceAI to predict their impact on RNA splicing
441 and, in particular, PE splicing. First we found that most of the pathogenic variants predicted to
442 impact RNA splicing were located in intronic regions (Fig. S7). Then, for each variant, we
443 calculated a splicing change score based on SpliceAI predictions, where positive scores
444 indicate increased PE inclusion, and negative scores indicate decreased inclusion (significant
445 absolute value > 0.2). We combined these predictions with PSI values from GTEx brain tissues
446 to prioritize variants that significantly affect PE splicing in the brain (PSI < 0.3 for increased
447 inclusion, or > 0.7 for decreased inclusion). Using this approach, we identified 1,877

448 pathogenic variants predicted to significantly affect the splicing of 891 PEs (Fig. 4A) such that
449 each variant/PE match satisfied at least one of the following criteria: A) the variant is associated
450 with neurological disorders; B) the gene is known to have relevant brain functions; C) the PE
451 is brain-specific, as previously determined.
452
453 **Cell culturing**
454 HEK293T cells (ATCC CRL-3216) were maintained in Dulbecco's Modified Eagle
455 Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-
456 streptomycin at 37°C in a humidified incubator with 5% CO_2. For transfection, cells were
457 seeded at 100,000 per well in 24-well plates one day prior. Transfections were performed using
458 Lipofectamine 3000 according to the manufacturer's protocol. Briefly, DNA plasmids and
459 P3000 reagent were diluted in Opti-MEM, combined with Lipofectamine 3000 diluted in Opti-
460 MEM, incubated for 15 minutes at room temperature, and added dropwise to the cells in
461 DMEM with FBS but no antibiotics. Medium was replaced after 8 hours and doubled after 24
462 hours. Cells were collected 72 hours post-transfection for downstream analyses.
463 Human iPSCs carrying inducible NGN (Coriell GM29370) were differentiated into i3N
464 neurons following a multi-step protocol adapted from Wang et al. (2018)[53]. Plates were pre-
465 coated with Matrigel diluted 1:100 in cold DMEM/F12 and incubated at room temperature for
466 at least 3 hours. Between days -4 and -1, plates were further coated with 50 µg/mL Poly-D-
467 Lysine followed by 3 µg/mL laminin immediately prior to seeding. On day -3, iPSCs at ~80%
468 confluency were dissociated with Accutase and plated on Matrigel-coated dishes in induction
469 medium (KO-DMEM/F12 supplemented with N2, NEAA, BDNF, NT3, 0.2 µg/mL laminin,
470 2 µg/mL doxycycline, and 10 µM ROCK inhibitor). Medium was changed with complete
471 induction medium on day -2, and with induction medium lacking ROCK inhibitor on day -1.
472 On day 0, cells were dissociated and replated onto PDL/laminin-coated dishes in maturation

473 medium (50% DMEM/F12 and 50% Neurobasal-A, supplemented with 0.5× B27, N2, NEAA,
474 GlutaMax, 1 μg/mL laminin, 10 ng/mL BDNF, and 10 ng/mL NT3). Half-medium changes
475 were performed weekly starting from day 7. Neurons were cultured for up to 4 weeks before
476 downstream assays. Cells were cultured at 37°C in a humidified incubator with 5% CO_2 at all
477 stages.

478 CHX treatment was performed both on HEK293T and i3N neurons by treating them for 16h
479 with 50 μg/ml CHX or an equivalent volume of DMSO as control (0.5% of the culture medium
480 volume). For HEK293T cells, medium was replaced with CHX-containing medium. For i3N
481 neurons, culture medium was quickly removed from the cells, implemented with CHX, and
482 placed back in culture.

483

484 **Plasmid design and preparation**

485 Plasmids expressing the CRISPR prime editor (Addgene 174820), the pegRNA (Addgene
486 223137), or the nicking gRNA (Addgene 65777) were prepared using competent DH5α E.
487 coli and Mini extraction kit. pegRNAs and nicking gRNAs were designed using the PRIDICT
488 tool[57] and selecting the pegRNA and gRNA with top scores, and they were cloned using
489 standard Type IIS restriction enzymes. Sequences were confirmed by Sanger sequencing.

490

491 ***In vitro* splicing analysis**

492 Total RNA was extracted from cultured cells using QIAzol reagent according to the
493 manufacturer's instructions. After chloroform separation and isopropanol precipitation, RNA
494 pellets were washed in 75% ethanol, air-dried, and resuspended in nuclease-free water. RNA
495 quantity and purity were assessed using a NanoDrop spectrophotometer. For reverse
496 transcription, 500 ng to 1 μg of total RNA was used with the SuperScript III Kit, using random
497 hexamer primers and following the recommended protocol.

498 PCR was performed using exon-specific primers and GoTaq master mix to amplify regions
499 of interest from cDNA (25 μg per reaction). PCR products were resolved on 2% agarose gels
500 stained with ethidium bromide and visualized using a UV gel imaging system. Bands of
501 expected size were excised and purified using the QIAquick Gel Extraction Kit, followed by
502 Sanger sequencing to confirm splicing isoforms and identify sequence variants.
503 qPCR was carried out using SYBR Green PCR Master Mix Roche LightCycler 480 II
504 system. Reactions were performed in technical duplicates using 10 μg cDNA per reaction and
505 exon-specific or exon junction-specific primers. Relative gene expression was calculated using
506 the ΔΔCt method, normalized to housekeeping genes. Primer efficiency and melt curve
507 analysis were used to ensure specificity and consistency across runs.
508
509 **PCR, qPCR and NGS**
510 Genomic DNA was extracted from cultured cells using magnetic purification beads. Cells
511 were lysed, and the lysate was incubated with purification beads to bind DNA. Magnetic
512 separation was used to wash away impurities, and DNA was eluted in nuclease-free water.
513 To quantify editing outcomes, targeted regions were amplified from genomic DNA using
514 primers with overhangs compatible with Illumina adapters. Amplicons were cleaned with
515 magnetic beads and pooled for multiplexed sequencing. Libraries were sequenced on an
516 Illumina MiSeq platform using 2×150 bp paired-end reads. Data were demultiplexed and
517 analyzed using the CRISPResso2 tool[58].
518
519 **Statistical analysis**
520 The statistical analyses for the wet-lab work were performed using GraphPad Prism. The
521 brain-unique PEs were defined as those with a minimal PSI > 0 in any brain tissue and a max
522 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 < 0.1 in the GAM. For the ontology enrichment analyses, a threshold of FDR < 0.1 from Fisher's exact test was applied to select the significant terms.

Funding

National Institutes of Health (NIH) grants (R01NS124561 to EM and MCS, R00NS118109 to DG, R21EY037018 to EM and CA). Ellison Foundation (EM).

References

1. Lim, K.H., Ferraris, L., Filloux, M.E., Raphael, B.J., and Fairbrother, W.G. (2011). Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. Proc Natl Acad Sci U S A *108*, 11093-11098. 10.1073/pnas.1101135108.
2. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. Cell *176*, 535-548 e524. 10.1016/j.cell.2018.12.015.
3. Chiang, H.L., Chen, Y.T., Su, J.Y., Lin, H.N., Yu, C.A., Hung, Y.J., Wang, Y.L., Huang, Y.T., and Lin, C.L. (2022). Mechanism and modeling of human disease-associated near-exon intronic variants that perturb RNA splicing. Nat Struct Mol Biol *29*, 1043-1055. 10.1038/s41594-022-00844-1.
4. Raj, B., and Blencowe, B.J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron *87*, 14-27. 10.1016/j.neuron.2015.05.004.
5. Zhang, X., Chen, M.H., Wu, X., Kodani, A., Fan, J., Doan, R., Ozawa, M., Ma, J., Yoshida, N., Reiter, J.F., et al. (2016). Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. Cell *166*, 1147-1162 e1115. 10.1016/j.cell.2016.07.025.
6. Vuong, C.K., Black, D.L., and Zheng, S. (2016). The neurogenetics of alternative splicing. Nat Rev Neurosci *17*, 265-281. 10.1038/nrn.2016.27.
7. Patowary, A., Zhang, P., Jops, C., Vuong, C.K., Ge, X., Hou, K., Kim, M., Gong, N., Margolis, M., Vo, D., et al. (2024). Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms. Science *384*, eadh7688. 10.1126/science.adh7688.

8. Wen, C., Margolis, M., Dai, R., Zhang, P., Przytycki, P.F., Vo, D.D., Bhattacharya, A., Matoba, N., Tang, M., Jiao, C., et al. (2024). Cross-ancestry atlas of gene, isoform, and splicing regulation in the developing human brain. Science *384*, eadh0829. 10.1126/science.adh0829.
9. Aya, F., and Valcarcel, J. (2024). Shaping human brain development and vulnerability through alternative splicing. Cell Genom *4*, 100584. 10.1016/j.xgen.2024.100584.
10. Ha, K.C.H., Sterne-Weiler, T., Morris, Q., Weatheritt, R.J., and Blencowe, B.J. (2021). Differential contribution of transcriptomic regulatory layers in the definition of neuronal identity. Nat Commun *12*, 335. 10.1038/s41467-020-20483-8.
11. Gonzalez-Iglesias, A., Arcas, A., Domingo-Muelas, A., Mancini, E., Galceran, J., Valcarcel, J., Farinas, I., and Nieto, M.A. (2024). Intron detention tightly regulates the stemness/differentiation switch in the adult neurogenic niche. Nat Commun *15*, 2837. 10.1038/s41467-024-47092-z.
12. Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. Science *338*, 1587-1593. 10.1126/science.1230612.
13. Dredge, B.K., Polydorides, A.D., and Darnell, R.B. (2001). The splice of life: alternative splicing and neurological disease. Nat Rev Neurosci *2*, 43-50. 10.1038/35049061.
14. Hervoso, J.L., Amoah, K., Dodson, J., Choudhury, M., Bhattacharya, A., Quinones-Valdez, G., Pasaniuc, B., and Xiao, X. (2024). Splicing-specific transcriptome-wide association uncovers genetic mechanisms for schizophrenia. Am J Hum Genet *111*, 1573-1587. 10.1016/j.ajhg.2024.06.001.
15. Frydas, A., Wauters, E., van der Zee, J., and Van Broeckhoven, C. (2022). Uncovering the impact of noncoding variants in neurodegenerative brain diseases. Trends Genet *38*, 258-272. 10.1016/j.tig.2021.08.010.
16. Lewis, B.P., Green, R.E., and Brenner, S.E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A *100*, 189-192. 10.1073/pnas.0136770100.
17. Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. Nature *446*, 926-929. 10.1038/nature05676.
18. Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E., and Ares, M., Jr. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev *21*, 708-718. 10.1101/gad.1525507.
19. Lykke-Andersen, S., and Jensen, T.H. (2015). Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. Nat Rev Mol Cell Biol *16*, 665-677. 10.1038/nrm4063.
20. He, F., and Jacobson, A. (2015). Nonsense-Mediated mRNA Decay: Degradation of Defective Transcripts Is Only Part of the Story. Annu Rev Genet *49*, 339-366. 10.1146/annurev-genet-112414-054639.
21. Kurosaki, T., and Maquat, L.E. (2016). Nonsense-mediated mRNA decay in humans at a glance. J Cell Sci *129*, 461-467. 10.1242/jcs.181008.
22. Lejeune, F. (2022). Nonsense-Mediated mRNA Decay, a Finely Regulated Mechanism. Biomedicines *10*. 10.3390/biomedicines10010141.
23. Yi, Z., Sanjeev, M., and Singh, G. (2021). The Branched Nature of the Nonsense-Mediated mRNA Decay Pathway. Trends Genet *37*, 143-159. 10.1016/j.tig.2020.08.010.

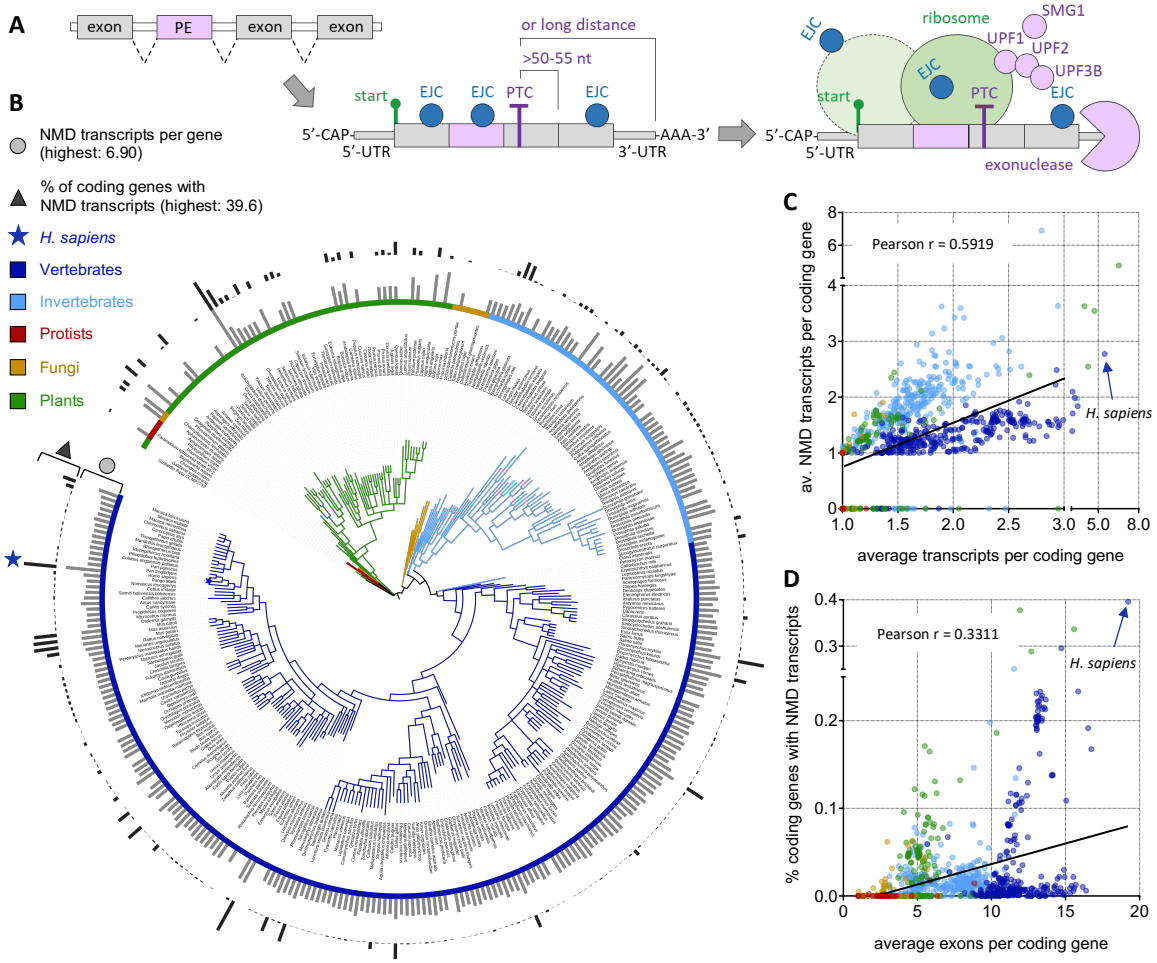
24. Pervouchine, D., Popov, Y., Berry, A., Borsari, B., Frankish, A., and Guigo, R. (2019). Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. Nucleic Acids Res *47*, 5293-5306. 10.1093/nar/gkz193.
25. Leclair, N.K., Brugiolo, M., Urbanski, L., Lawson, S.C., Thakar, K., Yurieva, M., George, J., Hinson, J.T., Cheng, A., Graveley, B.R., and Anczukow, O. (2020). Poison Exon Splicing Regulates a Coordinated Network of SR Protein Expression during Differentiation and Tumorigenesis. Mol Cell *80*, 648-665 e649. 10.1016/j.molcel.2020.10.019.
26. Jaffrey, S.R., and Wilkinson, M.F. (2018). Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease. Nat Rev Neurosci *19*, 715-728. 10.1038/s41583-018-0079-z.
27. Carvill, G.L., and Mefford, H.C. (2020). Poison exons in neurodevelopment and disease. Curr Opin Genet Dev *65*, 98-102. 10.1016/j.gde.2020.05.030.
28. Yan, Q., Weyn-Vanhentenryck, S.M., Wu, J., Sloan, S.A., Zhang, Y., Chen, K., Wu, J.Q., Barres, B.A., and Zhang, C. (2015). Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. Proc Natl Acad Sci U S A *112*, 3445-3450. 10.1073/pnas.1502849112.
29. Lin, J.Q., Khuperkar, D., Pavlou, S., Makarchuk, S., Patikas, N., Lee, F.C., Zbiegly, J.M., Kang, J., Field, S.F., Bailey, D.M., et al. (2023). HNRNPH1 regulates the neuroprotective cold-shock protein RBM3 expression through poison exon exclusion. EMBO J *42*, e113168. 10.15252/embj.2022113168.
30. Eom, T., Zhang, C., Wang, H., Lay, K., Fak, J., Noebels, J.L., and Darnell, R.B. (2013). NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. Elife *2*, e00178. 10.7554/eLife.00178.
31. Zhuravskaya, A., Yap, K., Hamid, F., and Makeyev, E.V. (2024). Alternative splicing coupled to nonsense-mediated decay coordinates downregulation of non-neuronal genes in developing mouse neurons. Genome Biol *25*, 162. 10.1186/s13059-024-03305-8.
32. Carvill, G.L., Engel, K.L., Ramamurthy, A., Cochran, J.N., Roovers, J., Stamberger, H., Lim, N., Schneider, A.L., Hollingsworth, G., Holder, D.H., et al. (2018). Aberrant Inclusion of a Poison Exon Causes Dravet Syndrome and Related SCN1A-Associated Genetic Epilepsies. Am J Hum Genet *103*, 1022-1029. 10.1016/j.ajhg.2018.10.023.
33. Tang, S., Stamberger, H., Calhoun, J.D., Weckhuysen, S., and Carvill, G.L. (2025). Antisense oligonucleotides modulate aberrant inclusion of poison exons in SCN1A-related Dravet syndrome. JCI Insight *10*. 10.1172/jci.insight.188014.
34. Sparber, P., Bychkov, I., Pyankov, D., and Skoblov, M. (2023). Functional investigation of SCN1A deep-intronic variants activating poison exons inclusion. Hum Genet *142*, 1043-1053. 10.1007/s00439-023-02564-y.
35. Aziz, M.C., Schneider, P.N., and Carvill, G.L. (2021). Targeting Poison Exons to Treat Developmental and Epileptic Encephalopathy. Dev Neurosci *43*, 241-246. 10.1159/000516143.
36. Brown, A.L., Wilkins, O.G., Keuss, M.J., Hill, S.E., Zanovello, M., Lee, W.C., Bampton, A., Lee, F.C.Y., Masino, L., Qi, Y.A., et al. (2022). TDP-43 loss and ALS-risk SNPs drive mis-splicing and depletion of UNC13A. Nature *603*, 131-137. 10.1038/s41586-022-04436-3.
37. Bakshi, S., and Isom, L.L. (2025). No more nonsense: evaluating poison exons as therapeutic targets in neurodevelopmental disorders. Curr Opin Genet Dev *92*, 102346. 10.1016/j.gde.2025.102346.

660 38. McLean, Z.L., Gao, D., Correia, K., Roy, J.C.L., Shibata, S., Farnum, I.N., Valdepenas-
661 Mellor, Z., Kovalenko, M., Rapuru, M., Morini, E., et al. (2024). Splice modulators
662 target PMS1 to reduce somatic expansion of the Huntington's disease-associated CAG
663 repeat. Nat Commun *15*, 3182. 10.1038/s41467-024-47485-0.
664 39. Bhattacharyya, A., Trotta, C.R., Narasimhan, J., Wiedinger, K.J., Li, W., Effenberger,
665 K.A., Woll, M.G., Jani, M.B., Risher, N., Yeh, S., et al. (2021). Small molecule splicing
666 modifiers with systemic HTT-lowering activity. Nat Commun *12*, 7299.
667 10.1038/s41467-021-27157-z.
668 40. Lim, K.H., Han, Z., Jeon, H.Y., Kach, J., Jing, E., Weyn-Vanhentenryck, S., Downs,
669 M., Corrionero, A., Oh, R., Scharner, J., et al. (2020). Antisense oligonucleotide
670 modulation of non-productive alternative splicing upregulates gene expression. Nat
671 Commun *11*, 3501. 10.1038/s41467-020-17093-9.
672 41. Felker, S.A., Lawlor, J.M.J., Hiatt, S.M., Thompson, M.L., Latner, D.R., Finnila, C.R.,
673 Bowling, K.M., Bonnstetter, Z.T., Bonini, K.E., Kelly, N.R., et al. (2023). Poison exon
674 annotations improve the yield of clinically relevant variants in genomic diagnostic
675 testing. Genet Med *25*, 100884. 10.1016/j.gim.2023.100884.
676 42. Keegan, N.P., and Fletcher, S. (2022). A spotter's guide to SNPtic exons: The common
677 splice variants underlying some SNP-phenotype correlations. Mol Genet Genomic Med
678 *10*, e1840. 10.1002/mgg3.1840.
679 43. Margasyuk, S., Kuznetsova, A., Zavileyskiy, L., Vlasenok, M., Skvortsov, D., and
680 Pervouchine, D.D. (2024). Human introns contain conserved tissue-specific cryptic
681 poison exons. NAR Genom Bioinform *6*, lqae163. 10.1093/nargab/lqae163.
682 44. Toledano, I., Supek, F., and Lehner, B. (2024). Genome-scale quantification and
683 prediction of pathogenic stop codon readthrough by small molecules. Nat Genet *56*,
684 1914-1924. 10.1038/s41588-024-01878-5.
685 45. Klonowski, J., Liang, Q., Coban-Akdemir, Z., Lo, C., and Kostka, D. (2023). aenmd:
686 Annotating escape from nonsense-mediated decay for transcripts with protein-
687 truncating variants. bioRxiv. 10.1101/2023.03.17.533185.
688 46. Thomas, J.D., Polaski, J.T., Feng, Q., De Neef, E.J., Hoppe, E.R., McSharry, M.V.,
689 Pangallo, J., Gabel, A.M., Belleville, A.E., Watson, J., et al. (2020). RNA isoform
690 screens uncover the essentiality and tumor-suppressor activity of ultraconserved poison
691 exons. Nat Genet *52*, 84-94. 10.1038/s41588-019-0555-z.
692 47. Belleville, A.E., Thomas, J.D., Tonnies, J., Gabel, A.M., Borrero Rossi, A., Singh, P.,
693 Queitsch, C., and Bradley, R.K. (2024). An autoregulatory poison exon in Smndc1 is
694 conserved across kingdoms and influences organism growth. PLoS Genet *20*,
695 e1011363. 10.1371/journal.pgen.1011363.
696 48. Dalgliesh, C., Aldalaqan, S., Atallah, C., Best, A., Scott, E., Ehrmann, I., Merces, G.,
697 Mannion, J., Badurova, B., Sandher, R., et al. (2025). An ultra-conserved poison exon
698 in the Tra2b gene encoding a splicing activator is essential for male fertility and meiotic
699 cell division. EMBO J *44*, 877-902. 10.1038/s44318-024-00344-6.
700 49. Tao, Y., Zhang, Q., Wang, H., Yang, X., and Mu, H. (2024). Alternative splicing and
701 related RNA binding proteins in human health and disease. Signal Transduct Target
702 Ther *9*, 26. 10.1038/s41392-024-01734-2.
703 50. Li, Y.I., Sanchez-Pulido, L., Haerty, W., and Ponting, C.P. (2015). RBFOX and PTBP1
704 proteins regulate the alternative splicing of micro-exons in human brain transcripts.
705 Genome Res *25*, 1-13. 10.1101/gr.181990.114.
706 51. Mulligan, M.R., and Bicknell, L.S. (2023). The molecular genetics of nELAVL in brain
707 development and disease. Eur J Hum Genet *31*, 1209-1217. 10.1038/s41431-023-
708 01456-z.

52. Deka, B., Chandra, P., and Singh, K.K. (2021). Functional roles of human Up-frameshift suppressor 3 (UPF3) proteins: From nonsense-mediated mRNA decay to neurodevelopmental disorders. Biochimie *180*, 10-22. 10.1016/j.biochi.2020.10.011.
53. Wang, C., Ward, M.E., Chen, R., Liu, K., Tracy, T.E., Chen, X., Xie, M., Sohn, P.D., Ludwig, C., Meyer-Franke, A., et al. (2017). Scalable Production of iPSC-Derived Human Neurons to Identify Tau-Lowering Compounds by High-Content Screening. Stem Cell Reports *9*, 1221-1233. 10.1016/j.stemcr.2017.08.019.
54. Lin, J.C., Chi, Y.L., Peng, H.Y., and Lu, Y.H. (2016). RBM4-Nova1-SRSF6 splicing cascade modulates the development of brown adipocytes. Biochim Biophys Acta *1859*, 1368-1379. 10.1016/j.bbagrm.2016.08.006.
55. Huerta-Cepas, J., Dopazo, J., and Gabaldon, T. (2010). ETE: a python Environment for Tree Exploration. BMC Bioinformatics *11*, 24. 10.1186/1471-2105-11-24.
56. Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics *23*, 127-128. 10.1093/bioinformatics/btl529.
57. Mathis, N., Allam, A., Kissling, L., Marquart, K.F., Schmidheini, L., Solari, C., Balazs, Z., Krauthammer, M., and Schwank, G. (2023). Predicting prime editing efficiency and product purity by deep learning. Nat Biotechnol *41*, 1151-1159. 10.1038/s41587-022-01613-7.
58. Clement, K., Rees, H., Canver, M.C., Gehrke, J.M., Farouni, R., Hsu, J.Y., Cole, M.A., Liu, D.R., Joung, J.K., Bauer, D.E., and Pinello, L. (2019). CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol *37*, 224-226. 10.1038/s41587-019-0032-3.

746 **Figures**

747



748

749 **Figure 1.** Evolutionary analysis of NMD transcripts across Eukaryotes. (**A**) Schematic
750 representation of how a PE triggers NMD. (**B**) NMD transcripts are predicted throughout the
751 annotated genome of 957 eukaryotic genomes (vertebrates = blue, invertebrates = light blue,
752 protists = red, fungi = yellow, plants = green. For each species the following is calculated: the
753 percentage of coding genes producing at least 1 NMD transcript (triangle); the average number
754 of NMD transcripts per gene (circle). (**C**) The average number of NMD transcripts per gene is
755 compared with the average number of total annotated transcripts per gene in each species. (**D**)
756 The percentage of coding genes producing at least 1 NMD transcript is compared with the

757 average number of exons per gene in each species. Linear regression, Pearson correlation
758 coefficient (r) and P values are shown for the scatter plots in (C) and (D).
759
760
761
762
763
764
765

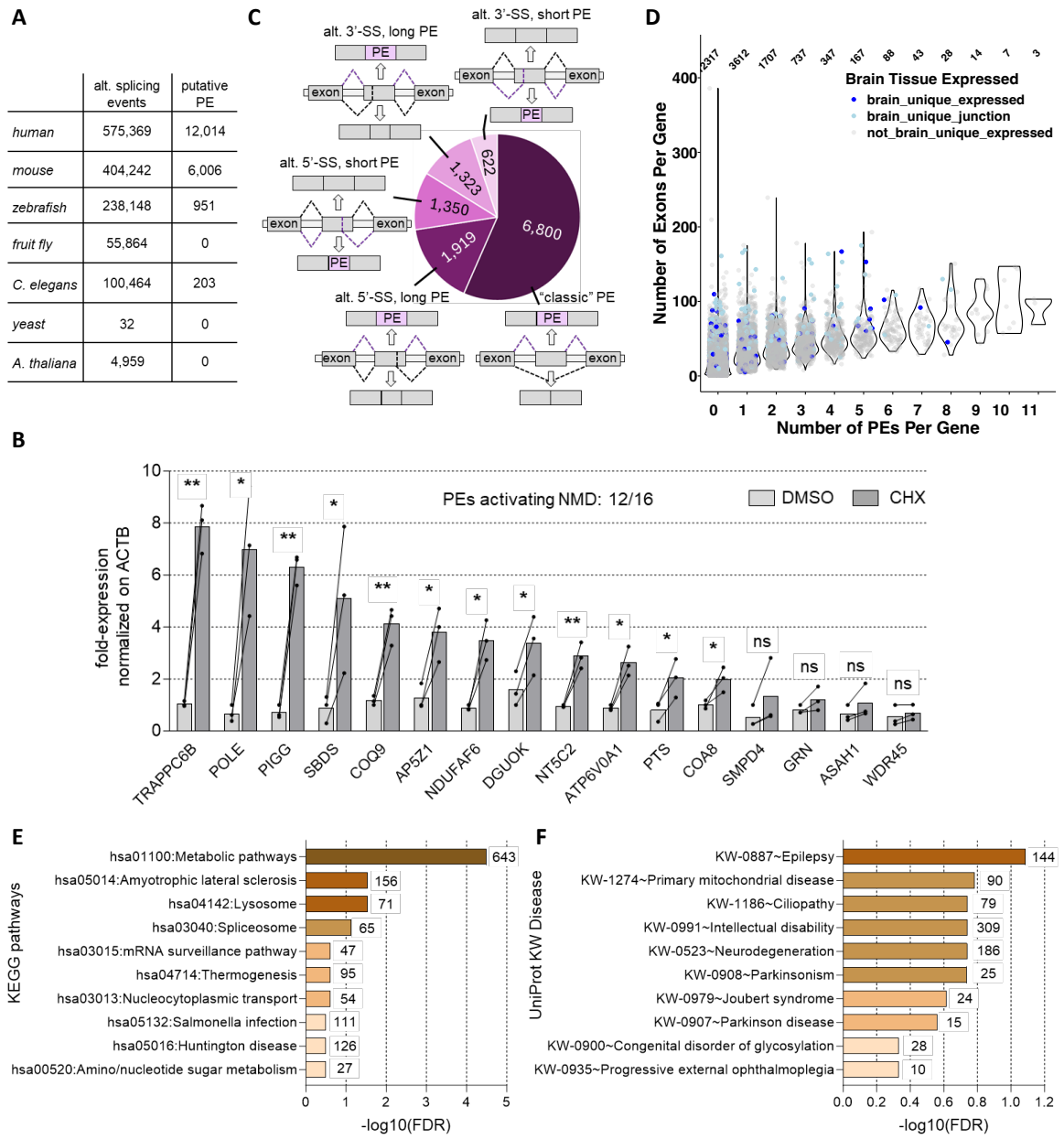


Figure 2. Prediction of alternative splicing events involving PEs. (**A**) PE splicing events are predicted for major biological models. (**B**) HEK293T cells are treated with CHX (or DMSO as control) to evaluate the activation of NMD when 16 predicted PEs are included. Transcript levels are measured by RT-qPCR, using primers specific for the exon-exon junctions that include the PE. (**C**) Counts of different types of PE splicing events in humans. They include canonical PE exclusion/inclusion as well as alternative 3'- and 5'-SSs. (**D**) Counts of PE splicing events per gene in humans. Pathway (**E**) and disease (**F**) enrichment for human genes producing PE splicing events. Numbers indicate total genes identified.

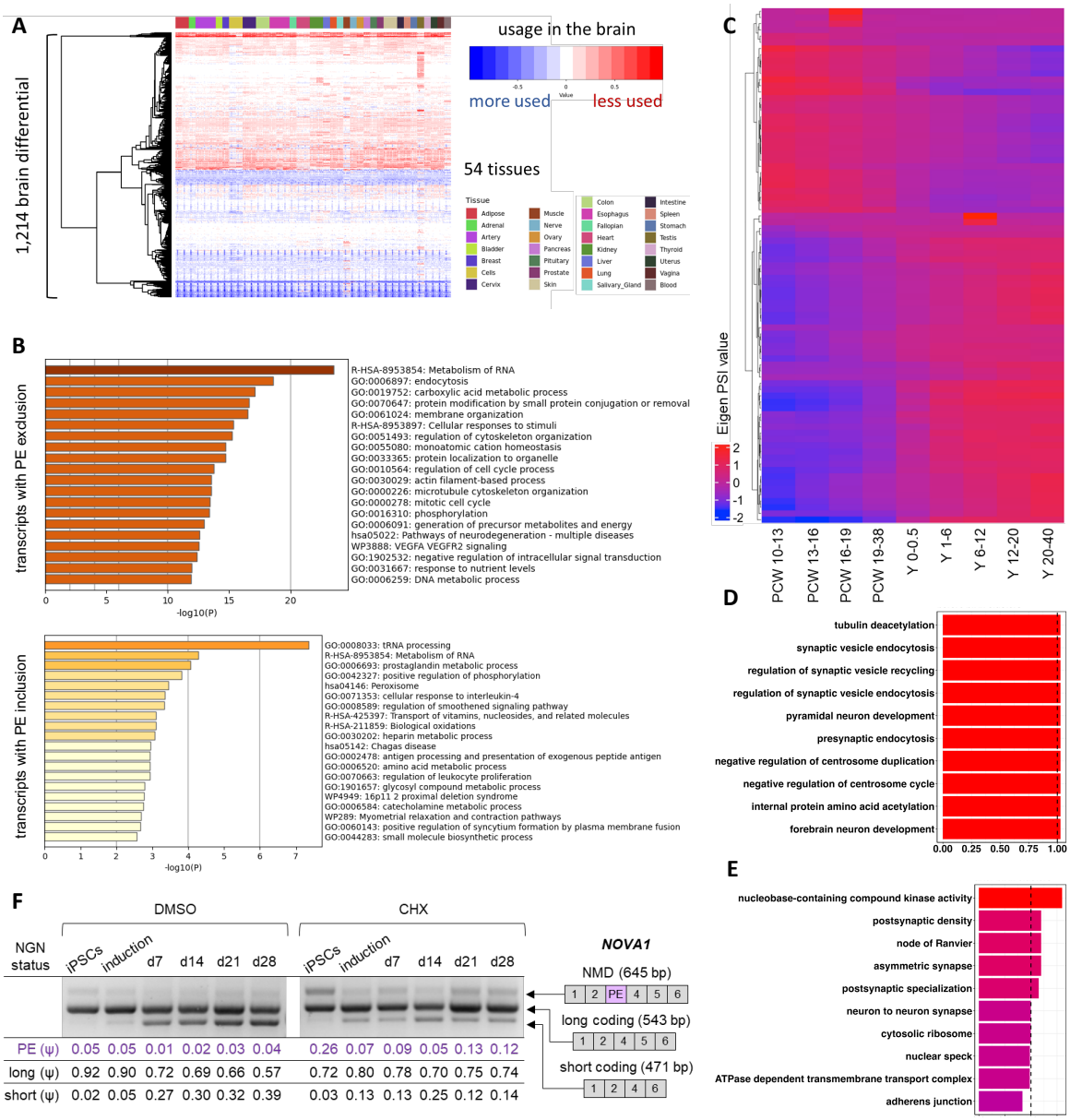


Figure 3. Analysis of PE splicing in human brain. (**A**) Heatmap showing the differential PSI of PEs found to be differentially spliced in the brain compared with other tissues in GTEX database. (**B**) Gene ontology and pathway enrichment analysis of PEs that are either spliced-out (PSI < 0.5) or spliced-in (PSI > 0.5) in the brain. (**C**) Heatmap showing the differential PSI of PEs found to be differentially spliced in the brain during development in BrainSpan database. (**D**) GO enrichment for host genes of PE events showing higher inclusion rate before birth. (**E**) GO enrichment for host genes of PE events showing higher inclusion rate after birth. (**F**) Splicing analysis in iP3 neurons for a brain-specific PE located in *NOVA1* gene. Changes

784 in the splicing patters are assessed by RT-PCR and agarose gel analysis, with 50 μg/ml CHX
785 or DMSO as control.
786
787
788
789
790
791
792
793
794

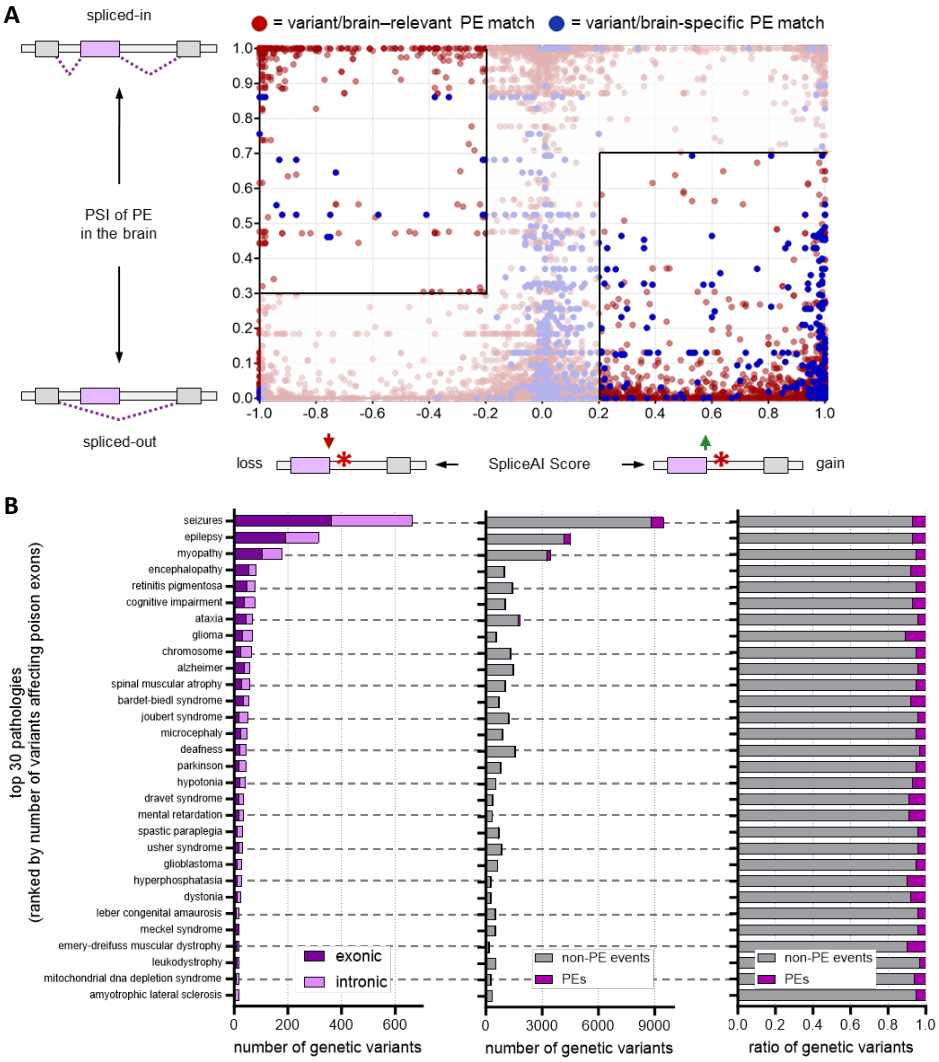


Figure 4. Prediction of pathogenic variants affecting poison exon splicing in neurological disorders. (**A**) Brain-specific PSI values from GTEx are integrated with SpliceAI-predicted splicing change scores for pathogenic ClinVar variants. A total of 101 brain-specific PEs are identified as significantly affected by pathogenic variants. (**B**) Affected genes are grouped by disease associations using the ODiseA ontology. The left panel shows the total number of genetic variants affecting PEs, divided by exonic and intronic variants. The middle and right panels compares as absolute numbers or relative ratios the variants affecting PEs with those affecting non-PE splicing events

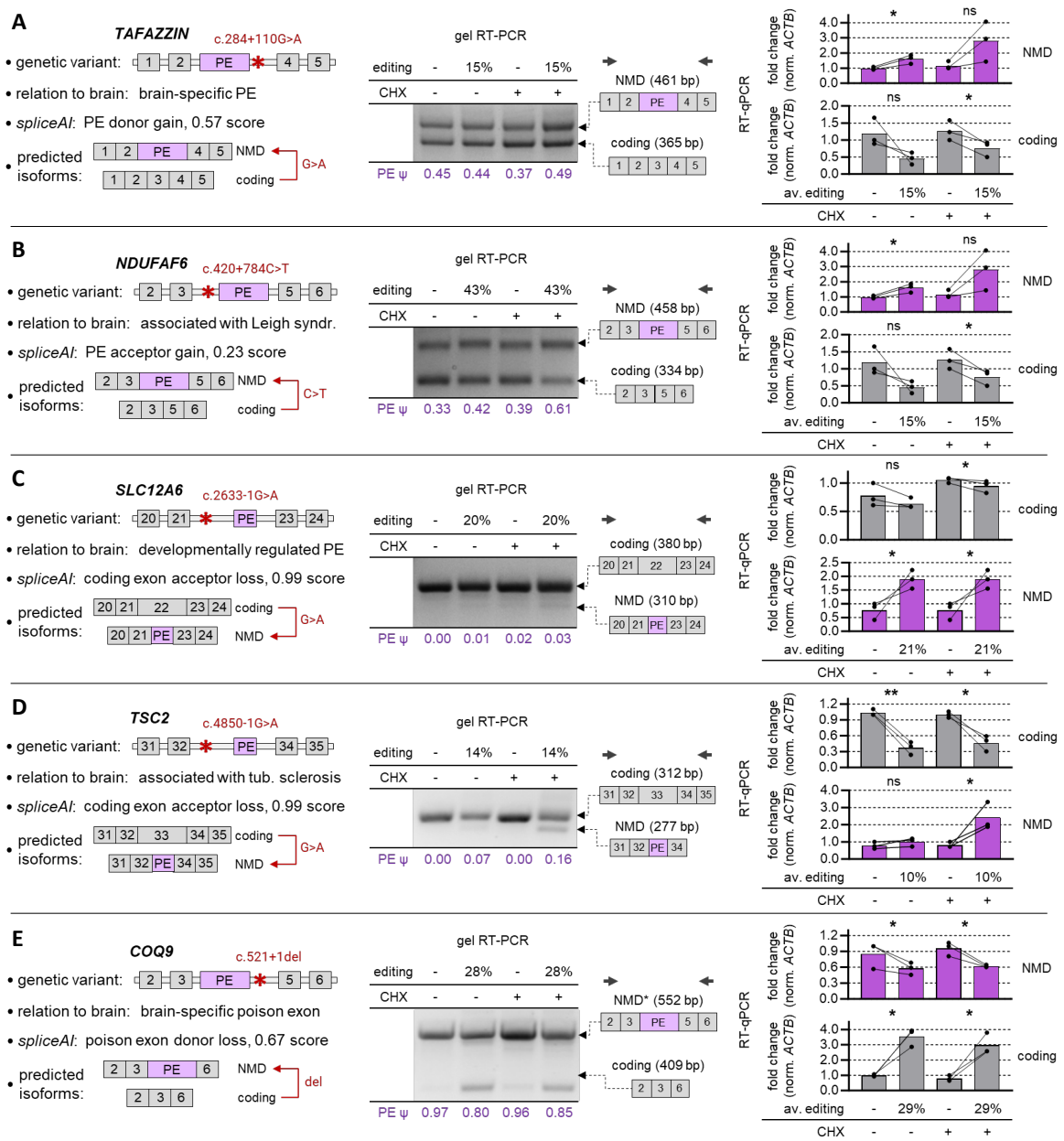


Figure 5. *In vitro* validation of the effect of select pathogenic genetic variants on PE splicing: (**A**) c.284+110G>A, affecting E3 of *TAFFAZIN*; (**B**) c.420+784C>T, affecting E4 of *NDUFAF6*; (**C**) c.2633-1G>A, affecting E22 of *SCL12A6*; (**D**) c.4850-1G>A, affecting E33 of *TSC2*; (**E**) c.521+1del, affecting E4 of *COQ9*. HEK293T cells were transfected with plasmid vectors expressing CRISPR prime editing components, and DNA and RNA were extracted 72h post-transfection for genome editing and RNA splicing analysis, respectively. Cells were treated either with DMSO or with CHX (50 µg/ml) 16 h prior to collection. The percentage of

813 edited cells was measured by amplicon-NGS, using PCR primers specific for the target genetic
814 region. The splicing pattern upon introduction of the target mutation was analyzed by RT-PCR
815 coupled with agarose gel analysis (central panels), using primers specific for flanking exons.
816 Quantitative analysis of splicing isoform alternation was performed by RT-qPCR (right panels)
817 using primers complementary to isoform-specific exon-exon junctions.