# Abnormal expression of splicing regulators RBFOX and NOVA is associated with aberrant splicing patterns at the Neurexin-3 gene in a monogenic autism spectrum disorder

**Pedro K. K. Forti[1,2], Lilian L. Depieri[1], Breno B. Hernandes[1,2], João G. Bueno[1], Gustavo D. Verçosa[1,2], Israel C. Vasconcelos[3], Mario H. Bengtson[3,4], Marcelo F. Carazzolle[1], Antonio P. Camargo[5], Fabio Papes[1,*]**

[1] Department of Genetics, Evolution, Microbiology and Immunology, Institute of Biology, University of Campinas, Campinas, Sao Paulo, 13083-862, Brazil

[2] Graduate Program in Genetics and Molecular Biology, Institute of Biology, University of Campinas, Campinas, Sao Paulo, 13083-862, Brazil

[3] Center for Medicinal Chemistry, University of Campinas, Campinas, Sao Paulo, 13083-875, Brazil

[4] Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas, Campinas, Sao Paulo, 13083-862, Brazil

[5] Department of Biochemistry, Institute of Chemistry, University of Sao Paulo, Sao Paulo, 05508-000, Brazil

*** Corresponding author's contact info:**

*Fabio Papes, Ph.D*
Department of Genetics, Evolution, Microbiology and Immunology, Institute of Biology, University of Campinas; Rua Monteiro Lobato, 255, Campinas, SP, 13083-862, Brazil

E-mail address: papesf@unicamp.br

## Abstract

Autism spectrum disorders are diseases characterized by a combination of cognitive, behavioral and neurological symptoms. A complex interplay between environmental factors and a multitude of genetic determinants, most of them composed of low-risk variants, impose difficulties in understanding the molecular and cellular underpinnings of these conditions. In some cases, autistic patients have been shown to display alterations in splicing patterns of several genes, but the extent to which this phenomenon is common in the context of these multifactorial disorders is unknown, nor is it known if monogenic cases of autism also display dysregulation of splicing. Moreover, very few studies have investigated the causal links between splicing alterations in specific genes and the phenotypic characteristics of the neural tissue in autism patients. In this study, we have focused on a monogenic type of autism caused by haploinsufficiency of the Transcription Factor 4 gene, known as Pitt-Hopkins Syndrome. We show that neurons and organoids derived from patients with this disease have altered expression of splicing regulators in the NOVA and RBFOX families, accompanied by aberrant splicing patterns in a substantial number of genes involved in neural tissue development and synapse organization. We focused on a gene encoding a member of the Neurexin Family, the splicing of which normally leads to the production of transcript variants coding for both transmembrane and secreted protein isoforms. In Pitt-Hopkins Syndrome neurons, we detected an aberrant splicing pattern that results in lower expression of the transcript that encodes the secreted isoform, a phenomenon that may explain why the neural tissue in these patients have decreased electrical activity through impaired synapse organization. Our data shed light on the role splicing regulation plays in a monogenic type of autism.

## INTRODUCTION

RNA splicing plays a critical role in regulating transcription and producing the wide diversity of transcripts and protein isoforms that characterize the eukaryotic cell. Alternative splicing results in the joining of exons and exclusion of introns in different configurations, being classified as exon skipping (SE), intron retention (RI), alternative 5' or 3' splice site (A5SS or A3SS), and mutually exclusive exon (MXE) splicing events[1]. RNA splicing is tightly regulated by RNA-binding proteins, several of which are important to neurogenesis, axon guidance, synaptogenesis, and synaptic physiology during neurodevelopment[2].

Alternative splicing is very prevalent in the human brain, particularly during neuronal differentiation[3], affecting events such as synaptic function and stability[4]. Aberrant splicing patterns at genes involved in synaptic plasticity such as *NLGN3*, *NLGN4*, *NRXN1*, *SHANK3*, and *CADPS2* are present in individuals with autism-spectrum disorders[5], which has been taken as a suggestion that several of these genes may have a putative role in the molecular mechanisms driving autism.

Out of the many RNA-binding proteins regulating splicing, two families have been linked to autism in several studies – RNA Binding Fox-1 Homolog (RBFOX) and Neuro-Oncological Ventral Antigen (NOVA) proteins. RBFOX family members RBFOX1, RBFOX2, and RBFOX3 are expressed in neurons[6], regulating alternative splicing patterns particularly during early neurogenesis and for synaptic plasticity (RBFOX1 and RBFOX2) or in mature neurons (RBFOX3)[7]. RBFOX1 controls the skipping of neuron-specific exons in many genes, particularly in those related to synapse formation, like *NRCAM* and *GRIN1*, which are dysregulated in samples of autistic children[8]. Likewise, NOVA1 has been shown to bind introns flanking neuron-specific exons, leading to exon inclusion or skipping[9]. In the brains of autistic children or individuals with other types of neurodevelopmental disorders, some studies have identified downregulation of RBFOX and NOVA family members[1,10], and several direct RBFOX and NOVA target genes are involved in synaptic transmission or ion channel regulation[1,11]. Additionally, RBFOX variants have been associated with several conditions characterized by neuropsychiatric symptoms[12].

Despite recent advances in understanding the role played by the dysregulation of RNA splicing regulators in the pathogenesis of autism, little is known about how mutations in specific genes causing autism-spectrum disorders lead to RBFOX and NOVA alterations and aberrant splicing patterns in downstream genes. This gap of knowledge prompted us to investigate a

monogenic type of autism. Here, we have studied the occurrence of aberrant alternative splicing patterns in a monogenic autism disease called Pitt-Hopkins Syndrome (PTHS; MIM 610954)[13–16]. PTHS is caused by de novo heterozygous mutations in the Transcription Factor 4 gene (*TCF4*; OMIM 602272), which encodes a bHLH transcription factor expressed in the brain[17–20] and important for neurodevelopment, neural lineage commitment, and neuronal physiology[21–26]. PTHS individuals exhibit severe cognitive deficit, motor delay, hypotonia, respiratory abnormalities, autistic behaviors, chronic constipation, and distinctive facial features[27,28]. Interestingly, variants of the *TCF4* gene have also been associated with schizophrenia, bipolar disorder, post-traumatic stress disorder, and major depressive disorder[29–33], suggesting that part of the underlying molecular mechanisms may be common between these conditions and autism. Our choice to study aberrant splicing in PTHS is due to its monogenetic nature, which allows us to dissect its pathological molecular mechanisms and more precisely correlate them with the cellular abnormalities found in the PTHS neural tissue.

We have previously generated and analyzed neural progenitor cells, neurons, and brain organoids from patients with PTHS to analyze the molecular phenotypes of the diseased neural tissue[34,35]. We showed that PTHS neural progenitors cultured using traditional 2D techniques and neural progenitors in PTHS organoids display a deficit in proliferation and severely diminished differentiation rate into neurons. Strikingly, we detected that PTHS organoids representative of the cerebral cortex are aberrant in size and structure, containing a higher percentage of neural progenitor cells and fewer cortical neurons of many subtypes, including deep-layer and superficial-layer excitatory neurons; moreover, PTHS organoids representative of the subpallial region exhibit a deficient content of cortical-type GABAergic interneurons[34]. Importantly, PTHS neurons in 2D culture and the neural network of PTHS organoids are characterized by severely impaired electrical activity, as judged by patch-clamp and multi-electrode array assays, respectively. Additionally, assessment of neuronal activity via FOS expression revealed that FOS+ neurons are fewer in PTHS organoids, concomitant with a lower expression of excitatory neuron's synaptic markers vGLUT1 and vGLUT2[34], suggesting that synapse formation is altered in this type of autism.

In this paper, we studied the expression of RBFOX and NOVA splicing regulators in PTHS neurons, showing that the lower expression of those regulators is accompanied by distinct splicing profiles in a range of genes, in comparison with neurons of normal genotype. These aberrant splicing patterns affect several genes coding for proteins important for synapse physiology, including Neurexin genes. We found that the Neurexin-3 gene (*NRNX3*) displays an

aberrant pattern of splicing involving an exon skipping event, where PTHS neurons have relatively lower abundance of a transcript variant that includes an exon leading to a premature stop codon and producing a soluble secreted form of Neurexin. On the other hand, neurons from donors of normal genotype exhibit lower abundance of the other transcript variant (excluding the exon), which produces a transmembrane Neurexin thought to be involved in synapse formation and stability[36]. Our findings show the novel information that a severe form of autism is accompanied by altered splicing of a gene encoding a synaptic component, leading to an imbalance between different Neurexin neuronal isoforms, which may contribute to the PTHS neural tissue's impaired electrical activity.

120 ## RESULTS

121 ### Altered expression of splicing regulators *RBFOX1/2/3* and *NOVA1* in Pitt-Hopkins
122 Syndrome neurons in 2D culture

123 We have previously shown that the neural tissue in Pitt-Hopkins Syndrome (PTHS) is
124 characterized by poor proliferation of neural progenitor cells and impaired differentiation of
125 cortical neurons, in both traditional 2D neuronal cultures and in patient-derived cortical
126 organoids[34]. We also detected lower neural activity in PTHS neurons, which could be in part due
127 to impaired synaptogenesis. Importantly, we analyzed the gene expression profiles of both
128 PTHS neurons and organoids via bulk RNA sequencing followed by differential gene expression
129 analysis[34] to identify dysregulated genes and pathways that could potentially be causally
130 implicated with these phenotypes. These transcriptomic data were obtained from 4 PTHS
131 patients harboring heterozygous mutations in the *TCF4* gene, causally related to the disease[34].

132 Interestingly, we noticed that several entries in the list of differentially expressed genes between
133 PTHS samples and same-sex parental controls code for RNA splicing regulators, such as RNA
134 Binding Fox-1 Homolog 1 (RBFOX1), RBFOX2, and RBFOX3, as well as Neuro-Oncological
135 Ventral Antigen 1 (NOVA1) and NOVA2 (Supplementary Table S1)[34]. Even though *NOVA2* is
136 also downregulated in PTHS neurons in 2D culture, its overall expression levels across parental
137 controls and patients are lower than *NOVA1* (Supplementary Table S1). Therefore, we decided
138 to focus our efforts on understanding the role played by *NOVA1* and *RBFOX1/2/3* in triggering
139 aberrant splicing patterns in PTHS.

140 In fact, several patient-parent pairs exhibited lower expression of *RBFOX1/2/3* and *NOVA1*
141 genes in PTHS neurons. For *RBFOX1*, the fold-changes (in log base 2) of expression in the
142 comparison between patient-control pairs #1, #2, and #3 are -4.95, -2.03, and -5.1, respectively
143 (Fig. 1a-d); for patient #4, the expression is low, comparable to the levels displayed by the other
144 3 patients, although a matching parental control was unavailable in this case (Supplementary
145 Table S1). For *RBFOX2*, the fold-changes between patient-control pairs #1, #2, and #3 are -
146 2,26, -1.79, and -1,61, respectively (Fig. 1a-d and Supplementary Table S1). *RBFOX3,* the most
147 studied splicing regulator in neurons, coding for the neuronal marker NeuN, exhibits fold-
148 changes of -10.9, -2,93, and -5,14, for pairs #1, #2, and #3, respectively (Fig. 1a-d and
149 Supplementary Table S1). Finally, *NOVA1* is also significantly less expressed in PTHS samples
150 as compared to parental controls for patient-control pairs #1, #2, and #3 (Fig. 1a-d), and its
151 expression is low in patient #4 (Supplementary Table S1).

152 The impaired expression of *RBFOX1/2/3* and *NOVA1* in PTHS neurons seems to be specific to
153 this neural cell category, as PTHS neural progenitor cells do not show a diminishment in the
154 expression levels of such genes: only the expression of *RBFOX1* is lower in patient-parent pairs
155 #1 to #4, while the other genes show average fold-changes close to zero (Fig. 1e-i). Importantly,
156 the overall expression abundance of these genes is somewhat lower in neural progenitor cells
157 as compared to neurons, as judged by the metric of expression abundance TPM (transcripts per
158 million reads) (Supplementary Table S1).

159 Next, we performed independent neuronal differentiation experiments to quantify the expression
160 of *NOVA1* and *RBFOX1/2/3* in PTHS neurons via reverse transcriptase-coupled quantitative
161 PCR (RT-qPCR) in comparison with two controls of normal genotype. Similar to the
162 transcriptomic data obtained from bulk RNA-Seq experiments in 2D cultured neurons, *RBFOX1*
163 expression levels were -1.9, -1.5, and -0.9-fold lower in patients #1, #2, and #3 in the
164 comparison with control #1, or -1.8, -1.5, and -0.8-fold lower in patients #1, #2, and #3 in the
165 comparison with control #3 (Fig. 1j). Similarly, *RBFOX2* expression levels were also lower in
166 patients #1, #2, and #3 in the comparison with control #1 and in patient #2 in the comparison
167 with control #3 (Fig. 1j). On the other hand, *RBFOX3* expression was only found to be lower in
168 patient #2, and NOVA1 expression was lower in patients #1 and #2 in the comparison with
169 patient #1 (Fig. 1j).

170 In combination, the RNA-Seq and qPCR results indicate that splicing regulators in the NOVA
171 and RBFOX families have lower expression levels in neurons of certain PTHS individuals
172 cultured using traditional 2D techniques. It is noteworthy that *RBFOX1* is consistently
173 downregulated in PTHS neurons across all subjects in our cohort.

174

175 **Complex dysregulation in the expression of *RBFOX1/2/3* and *NOVA1* in Pitt-Hopkins**
176 **Syndrome cortical and subpallial organoids**

177 Because there was a difference in expression abundances for *RBFOX1/2/3* and *NOVA1*
178 between two different neural cell types in our bulk RNA-Seq data, we sought to explore the
179 expression of these genes in single-cell transcriptomic data from PTHS patient-derived cortical
180 and subpallial organoids, which are complex 3D structures able to model the events that
181 happen during early neurodevelopment in vivo and. These models are more representative of
182 the molecular characteristics of neuronal cells than traditional 2D cultures[34,37,38]. Because the
183 organoids contain a more diverse set of neurons and the interactions between different cell

184 types may influence the dysregulation of *NOVA* and *RBFOX* expression, we reasoned that the
185 organoids would more faithfully recapitulate the changes in expression occurring in vivo. In
186 organoid single-cell RNA sequencing (scRNA-Seq) libraries previously described by our
187 group[34], we could annotate six neural cell subpopulations, including progenitor cells,
188 intermediate progenitors, and neurons[34] in two distinct differentiation lineages that produce
189 either glutamatergic or GABAergic neurons (Fig. 2a). In cortical organoids (CtOs), the
190 predominant differentiation lineage leads to glutamatergic neurons (Fig. 2b), while subpallial
191 organoids (sPOs) generate mostly GABAergic cells (Fig. 2b), matching the differentiation profile
192 found in the cortex and subpallial brain regions in vivo[39,40].

193 The fractions of cells expressing *NOVA1* are substantial in control CtO and sPO subpopulations
194 (Fig. 2c,d) – 80.3% of CtO progenitor cells (Pr-Glut), 67.6% of CtO intermediate progenitor cells
195 (IP-Glut), 49.1% of CtO glutamatergic neurons (N-Glut), 96.2% of sPO progenitor cells (Pr-
196 GABA), 97.1% of sPO intermediate progenitor cells (IP-GABA), and 98.4% of sPO GABAergic
197 neurons (N-GABA) express *NOVA1* above a threshold level of 0.5. Curiously, *NOVA1*
198 expression remains stable along the differentiation trajectory from progenitor cells to neurons in
199 the glutamatergic lineage of control CtOs (Fig. 2c) (mean expression of 1.77, 1.26, and 1.3 in
200 Pr-Glut, IP-Glut, and N-Glut, respectively), whereas expression increases from progenitors to
201 neurons in the GABAergic lineage of control sPOs (Fig. 2d) (mean expression of 5.7, 5.6, and
202 12.0 in Pr-GABA, IP-GABA, and N-GABA, respectively).

203 In contrast, the numbers of cells expressing *RBFOX1* are much lower than those expressing
204 *NOVA1,* across all CtO and sPO subpopulations (Fig. 2e,f) – 1.56% of CtO progenitor cells (Pr-
205 Glut), 3.21% of CtO intermediate progenitor cells (IP-Glut), 10.7% of CtO glutamatergic neurons
206 (N-Glut), 0.6% of sPO progenitor cells (Pr-GABA), 2.0% of sPO intermediate progenitor cells
207 (IP-GABA), and 8.8% of sPO GABAergic neurons (N-GABA) express *RBFOX1* above a
208 threshold level of 0.5. *RBFOX1* expression increases dramatically along the differentiation
209 trajectory from progenitor cells to neurons in both the glutamatergic and GABAergic lineages of
210 control CtOs and sPOs, respectively (Fig. 2e,f) (mean expression of 0.04, 0.06, 0.23, 0.001,
211 0.03, and 0.2 in Pr-Glut, IP-Glut, N-Glut, Pr-GABA, IP-GABA, and N-GABA, respectively).

212 On the other hand, *RBFOX2* is expressed in a large fraction of cells in CtO and sPO
213 subpopulations (Fig. 2g,h) – 55.4% of CtO progenitor cells (Pr-Glut), 84.9% of CtO intermediate
214 progenitor cells (IP-Glut), 73.6% of CtO glutamatergic neurons (N-Glut), 28.9% of sPO
215 progenitor cells (Pr-GABA), 50.3% of sPO intermediate progenitor cells (IP-GABA), and 88.5%
216 of sPO GABAergic neurons (N-GABA) express *RBFOX2* above a threshold level of 0.5. It is

217 very clear that *RBFOX2* expression increases markedly along the differentiation trajectory from
218 progenitor cells to neurons in both the glutamatergic lineage of control CtOs (Fig. 2g) (mean
219 expression of 0.98, 2.1, and 2.6 in Pr-Glut, IP-Glut, and N-Glut, respectively) and the
220 GABAergic lineage of control sPOs (Fig. 2h) (mean expression of 0.46, 0.96, and 3.39 in Pr-
221 GABA, IP-GABA, and N-GABA, respectively).

222 Although *RBFOX3* is a widely recognized neuronal marker, it is expressed in a significant
223 fraction of progenitor cells and intermediate progenitors in CtO and sPO subpopulations (Fig.
224 2i,j) – 11.7% of CtO progenitor cells (Pr-Glut), 49.4% of CtO intermediate progenitor cells (IP-
225 Glut), and 28.5% of CtO glutamatergic neurons (N-Glut), whereas the fractions of sPO cells
226 expressing *RBFOX3* are less prominent – 0% of sPO progenitor cells (Pr-GABA), 4.4% of sPO
227 intermediate progenitor cells (IP-GABA), and 14% of sPO GABAergic neurons (N-GABA)
228 express *RBFOX2* above a threshold level of 0.5. Different than *RBFOX1* and *RBFOX2*,
229 *RBFOX3* expression peaks at the intermediate progenitor stage in the glutamatergic lineage of
230 control CtOs (Fig. 2i) (mean expression of 0.21, 1.0, and 0.53 in Pr-Glut, IP-Glut, and N-Glut,
231 respectively) and increases marginally along the GABAergic lineage of control sPOs (Fig. 2j)
232 (mean expression of 0.007, 0.08, and 0.25 in Pr-GABA, IP-GABA, and N-GABA, respectively).

233 In the comparison between PTHS and control organoids, we verified that the percentage of N-
234 Glut neurons expressing *NOVA1* is higher in PTHS CtOs in contrast to control organoids (left
235 panels in Fig. 3a) and that the percentage of *NOVA1*+ N-GABA neurons remains practically
236 unchanged in PTHS sPOs (right panels in Fig. 3a). Although the mean levels of *NOVA1*
237 expression are statistically significantly different in the comparison between PTHS and control
238 N-Glut neurons in CtOs, being higher in the patient-derived organoids, the effect size is not
239 large (2.0 expression abundance in PTHS versus 1.3 in control; Fig. 2c). Likewise, *NOVA1*
240 expression is statistically significantly higher in PTHS N-GABA neurons of sPOs, and the effect
241 size is similarly not large (14.8 expression abundance in PTHS versus 12.0 in controls; Fig. 2d).

242 Similar to *NOVA1*, the percentages of N-Glut neurons expressing *RBFOX1* and *RBFOX2* are
243 practically unchanged in PTHS CtOs in contrast to control organoids (left panels in Fig. 3b,c),
244 the percentage of N-GABA neurons expressing *RBFOX1* is slightly higher in PTHS sPOs, and
245 the percentage of N-GABA neurons expressing *RBFOX2* is lower in PTHS sPOs (right panels in
246 Fig. 3b,c). Interestingly, the percentages of cells expressing RBFOX3 are higher in both CtOs
247 and sPOs (Fig. 3d). Finally, the mean expression levels of *RBFOX1*, *RBFOX2*, or *RBFOX3* are
248 statistically significantly dysregulated in PTHS N-Glut neurons in CtOs, with small effect sizes
249 (0.19, 3.14, and 1.0 expression abundances in PTHS versus 0.23, 2.6, and 0.53 in control,

250 respectively; Fig. 2e,g,i), being higher in PTHS N-GABA neurons of sPOs, with equally small
251 effect sizes (0.74, 3.6, and 0.61 expression abundances in PTHS versus 0.2, 3.4, and 0.25 in
252 control, respectively; Fig. 2f,h,j).

253 The higher expression of *NOVA1* in PTHS CtO and sPO organoids (Fig. 2c,d) as well the higher
254 expression of *RBFOX1/2/3* in PTHS sPO organoids (Fig. 2f,h,j) contrasts with their lower
255 expression in PTHS neurons grown in traditional 2D culture (Fig. 1d). As stated above, this
256 difference may be due to the much wider diversity of cell types and connectivity found in the
257 organoids. Regardless, the bulk and single-cell RNA-Seq results described here and in the
258 previous section indicate alterations in the expression of *NOVA1* and *RBFOX1/2/3* splicing
259 regulators in Pitt-Hopkins Syndrome. These findings prompted us to further investigate possible
260 changes in the RNA splicing process in the patient-derived neural tissue.

261

262 **Aberrant splicing patterns in neurons derived from Pitt-Hopkins Syndrome patients**

263 The consistent finding of altered expression of RBFOX and NOVA splicing regulators in
264 neuronal cultures derived from PTHS patients led us to hypothesize that such dysregulated
265 expression may have led to aberrant splicing. To investigate this, we conducted a computational
266 analysis of the RNA-Seq data from neuronal cultures derived from 5 patient-control pairs using
267 the software rMATS-turbo[41], a computational tool for analyzing alternative splicing using short-
268 read RNA-seq data. In short, rMATS-turbo works by initially using fastq input files to produce
269 splicing graphs for each RNA-seq sample, followed by integration across samples and
270 replicates to quantify alternative splicing events of five basic types: skipped exon (SE),
271 alternative 5′ splice sites (A5SS), alternative 3′ splice sites (A3SS), mutually exclusive exons
272 (MXE) and retained intron (RI). rMATS-turbo adopts the widely used PSI metric to quantify
273 splicing, which represents the percentage of transcripts that include a specific exon or splice
274 site, as calculated from the read counts supporting those specific exons or splice junctions,
275 normalized by the effective lengths of the gene's transcripts[41]. Differences between PSI values
276 in the control and PTHS samples were then computed as ΔPSI, which can be used to infer the
277 existence of different splicing patterns in one or another condition (for example, in PTHS
278 neurons). Positive ΔPSI values imply that the splicing event under analysis is more prevalent in
279 the control samples, while negative ΔPSI values indicate stronger usage in the PTHS samples.

280 After filtering for PSI values (averaged across biological replicates for PTHS or control
281 conditions) smaller than 5% or greater than 95% in both samples, for a false discovery rate

282 smaller than 5% and for |ΔPSI| greater than 5%, we could detect a significant number of splicing
283 events with different PSI in the comparison between neurons derived from each PTHS patient
284 against respective control, in all five categories of splicing (Supplementary Files 1-25). As an
285 example, we could detect 3695 alternative SE events, where one exon of a gene is excluded
286 from a transcript at a higher rate in one sample versus the other, with |ΔPSI| greater than 5% in
287 the comparison between patient and parent #2 (Fig. 4a; Supplementary Table S2). These
288 events were related to transcripts from 2137 different genes (Supplementary Table S2). The
289 numbers of differential splicing events and the numbers of genes in which these events happen,
290 in the 5 patient-control pairs, can be found in Supplementary Table S2. It should be noted that
291 SE events are more numerous than the other types, and RI or A5SS events are the rarest (Fig.
292 4a). When we intersected the splicing events across the five patient-control pairs, we found 492
293 SE events that are common to all pairs (Fig. 4b), in 365 genes (Fig. 4c; Supplementary Table
294 S3). The numbers of common splicing events of the other 4 types are found in Fig. 4b,c and
295 Supplementary Table S3.

296 To better understand the scope of changes in PTHS neurons due to aberrant splicing patterns,
297 we conducted Over-Representation Analysis (ORA) followed by Gene Ontology (GO) and
298 KEGG Pathway analyses, using a list of unique genes for which rMATS-turbo detected splicing
299 events in the comparison between each PTHS patient and a control neuronal cell line. This
300 analysis was performed using the web-based WebGestalt tool[42], which results in an over-
301 representation *z*-score and enrichment *p*-value for each GO term. Interestingly, for SE events,
302 when we analyzed the set of genes from the comparison between patient #2 and its control, we
303 detected a significant enrichment of genes in Gene Ontology-Biological Processes (GO-BP) and
304 KEGG-Pathway categories linked to synapse physiology, axon projection, and cell adhesion
305 (Fig. 4d,e). GO and Pathway analyses results for other patient-control pairs yielded similar
306 results (Supplementary Figure S1), highlighting the possibly relevant role of aberrant splicing
307 patterns in the neural tissue cellular and molecular abnormalities that accompany PTHS.

308 Likewise, we conducted Gene Set Enrichment Analysis (GSEA) followed by Gene Ontology
309 (GO) and KEGG Pathway analyses, using the same list of unique genes for which rMATS-turbo
310 detected splicing events in the comparison between each PTHS patient and a control neuronal
311 cell line. For these analyses, the list of genes was ordered according to the |ΔPSI| values of the
312 corresponding splicing events from rMATS-turbo. GSEA results indicate that, among the genes
313 at the top of the list (meaning genes with splicing events that are more dissimilarly frequent in
314 the comparison between PTHS and control), there are genes related to synapse formation,

synapse physiology, and ion transport (Fig. 4f,g). It is interesting to note that the GO and Pathway enrichment analyses described above show the presence of aberrant splicing patterns in genes related to synapse formation, which may be responsible for the lower electrical activity found in PTHS organoids[34].

The finding of a substantial number of splicing events that seem to occur differentially between PTHS and control neuronal cell lines led us to hypothesize that the alterations found in the expression levels of *NOVA1* and *RBFOX1/2/3* in PTHS neurons are causally related to the aberrant splicing patterns observed.

**NOVA1 and RBFOX1/2/3 RNA binding sites are enriched in transcripts predicted to have aberrant alternative splicing patterns in PTHS**

To further study the relationship between the aberrant splicing patterns observed in PTHS neurons and the dysregulation of expression of NOVA and RBFOX splicing regulators, we investigated the occurrence of NOVA1 and RBFOX1/2/3 binding sites in the genes detected by rMATS-turbo described in the previous section. First, we extracted the genomic sequences containing the exons involved in the SE splicing events detected by rMATS-turbo, plus additional sequences of 200 bp in the neighboring introns. The resulting collection of exon sequences flanked by adjacent introns was then analyzed for RBP (RNA binding protein) binding site enrichment using the SEA program in the MEME suite[43]. Our reasoning was that, if NOVA1 or RBFOX regulators were involved in the aberrant splicing patterns observed in PTHS, then we would expect to see an enrichment in binding sites near the splicing events relative to the occurrence of RBP binding sites in a random collection of genomic sequences. Two NOVA1 recognition motifs were analyzed[44] and both were found to be significantly enriched in the genomic sequences near the SE events detected by rMATS-turbo (*E*-values of $9.2 \times 10^{-8}$ and $3.0 \times 10^{-8}$; Fisher's exact test; Fig. 5a), with a lower *E*-value than those calculated for a set of genes known to be regulated by NOVA1 (positive control; *E*-values of $2.9 \times 10^{-1}$ and $1.6 \times 10^{-6}$, respectively; Fig. 5a; Supplementary File S4). The same type of analysis was conducted for binding site motifs recognized by RBFOX1 and RBFOX2/3 splicing regulators[44], indicating that both are significantly enriched in the genomic sequences near the SE events detected by rMATS-turbo (*E*-values of $5.8 \times 10^{-4}$ and $4.1 \times 10^{-7}$, respectively; Fisher's exact test; Fig. 5a); although these values are larger than those calculated for a set of genes known to be regulated by RBFOX1/2/3 (positive control; *E*-values of $8.4 \times 10^{-8}$ and $9.2 \times 10^{-16}$, respectively; Fig. 5a;

347 Supplementary File S4), the *E*-values are still much smaller than when the enrichment analysis
348 was performed by comparing two randomly chosen sets of genomic sequences (Fig. 5a).

349 Next, we conducted an experiment to detect putative RBP binding sites across the entire
350 genomic sequence of the genes whose splicing patterns were found to be dysregulated by
351 rMATS-turbo. This analysis was performed using the RNA-binding protein motif detection
352 software FIMO in the MEME suite[45]. Examples of genes shown to have enrichment of NOVA1
353 and RBFOX1/2/3 sites are shown in Fig. 5b,c, in comparison with the lower abundance of
354 binding motifs in genes in the random set (negative control; Fig. 5d). We observed a difference
355 in site count which is statistically significant for NOVA1 (Fig. 5e) when considering the genes for
356 the top splicing pattern alterations, as measured by the absolute value of ΔPSI (larger than
357 30%).

358 These enrichment analyses suggest that NOVA1 and RBFOX splicing regulators may be
359 involved in dictating the aberrant splicing patterns observed in PTHS neurons. Direct evidence
360 for this relationship was further pursued by the analysis of published CLiP-Seq (HITS-CLiP)
361 experimental datasets, from studies that performed transcriptome-wide determination of NOVA
362 and RBFOX binding sites[1,11,46]. Considering the 1634 genes harboring aberrant SE splicing
363 events with absolute ΔPSI values larger than 0.1 (larger splicing usage either in the parent or
364 the patient), there is a substantial overlap with the hundreds of genes whose transcripts were
365 shown to be directly bound to NOVA1 (Fig. 5f,g; Supplementary File S5), including genes
366 coding for synapse organization and neuronal physiology. Likewise, there is a very substantial
367 overlap with the thousands of genes whose transcripts have been shown to be directly bound to
368 RBFOX1/2/3 splicing regulators (Fig. 5h; Supplementary File S5).

369 In combination, these results suggest that the altered splicing patterns detected in the
370 transcriptomes of PTHS neurons may be caused by the dysregulation of NOVA1 and
371 RBFOX1/2/3 splicing regulators.

372

373 **The Neurexin-3 gene (*NRXN3*) exhibits an exon skipping event that leads to a transcript**
374 **encoding a soluble protein isoform**

375 Considering that the list of skipped exon (SE) splicing events detected by rMATS-turbo is longer
376 than the lists for the other types of differential splicing events (Fig. 4a), we decided to further
377 investigate this category to better understand the types of events and genes affected by

378 aberrant splicing patterns. GO and Pathway analyses conducted on the list of genes for which
379 rMATS-turbo detected splicing events revealed the presence of genes in enriched categories
380 related to synapse formation and physiology, axonal projection, and cell adhesion (Fig. 4d-g).
381 Curiously, we detected a significant number of splicing events that occur in genes in the
382 Neurexin family, which are traditionally regarded as critical components of synapse
383 physiology[47].

384 In the list of SE splicing events in the comparison between patient-control pairs #1, #2, and #3,
385 we could detect 6, 12, and 1 events, respectively, in genes in the Neurexin family (*NRXN1,*
386 *NRXN2,* and *NRXN3*) (Supplementary File S2). Because some qPCR experiments described
387 below were conducted comparing patient #2 with the parent #1 control (due to non-availability of
388 parent #2 samples), we also conducted the parent #1/patient #2 comparison using rMATS-
389 turbo, which yielded 9 differential SE events linked to *NRXN1/2/3* genes. Finally, in the
390 comparison between parent #2 and patient #4, for whom there is no associated parent in our
391 cohort, we could also detect 2 SE events linked to *NRXN2* and *NRXN3* genes (Supplementary
392 File S2).

393 The numerous events related to *NRXN* genes are further highlighted by the presence of 2
394 events associated with *NRNX2* and *NRXN3* among the top 20 SE splicing events with positive
395 ΔPSI (higher usage in control samples) in the patient #2-parent #2 comparison (Supplementary
396 File S2). The event in the *NRXN2* gene has a ΔPSI equal to 0.681 (Fig. 6a,b), with a higher
397 splicing event usage (ranging from 0.797 to 0.86) in the parent #2 biological replicate libraries,
398 while the PSI values range from 0.117 to 0.196 in the patient #2 samples (Supplementary File
399 S2). It is interesting to note that the exon preferentially included in the transcripts of the *NRXN2*
400 gene in the control sample (SE-NRXN2-i splicing event) is very short (24 nt; Fig. 6b,e) and is
401 equivalent to exon 7 in the *NRXN2* Ensembl transcript ENST00000409571.6. In the PTHS
402 patient, exon 7 is preferentially excluded (SE-NRXN2-e splicing event (Fig. 6b). Interestingly,
403 there exists a 45-bp annotated exon in the same genomic region for the canonical *NRNX2*
404 Ensembl transcript (accession ENST00000265459.11) (Fig. 6c). The only difference between
405 these two annotated transcripts is the 21-bp difference in this exon. The exon skipping event
406 does not lead to a change in the coding sequence reading frame and results in a loss of a few
407 amino acid residues relative to the protein encoded by the canonical transcript (Fig. 6c);
408 therefore, we decided not to pursue this splicing event further.

409 On the other hand, the SE event in the *NRXN3* gene has a ΔPSI equal to 0.608 (Fig. 6d,e), with
410 splicing event inclusion levels ranging from 0.579 to 1.00 in the parent #2 biological replicate

411 libraries, while the PSI values range from 0.167 to 0.185 in patient #2 samples (Supplementary
412 File S2). For ease of reference, we will name the SE event that includes the exon as SE-
413 NRXN3-i, while the splicing configuration that does not include the exon will be called SE-
414 NRXN3-e. Moreover, we chose to call the included exon as E20.5, thus named because it is
415 included between exons 20 and 21 of the canonical *NRXN3* Ensembl transcript
416 ENST00000335750.7 (this exon is also equivalent to exon 19 in the *NRXN3* Ensembl transcript
417 ENST00000554738.5). The exon that is preferentially included in the control samples has a size
418 of 679 nt (Fig. 6e,f). This sequence is found in only 2 out of 14 *NRNX3* Ensembl transcripts with
419 recognized coding sequences (ENST00000554738.5 and ENST00000555387.1); on the other
420 hand, this is not an annotated exon in the canonical *NRXN3* Ensembl transcript
421 (ENST00000335750.7) (Fig. 6f). Additionally, this included exon has been described before in
422 the context of *NRXN3* transcripts expressed in post-mortem human cerebral cortex samples of
423 individuals harboring a SNP near a splicing site in the *NRXN3* gene (rs8019381)[48]. As will be
424 described in detail below, this included exon creates a premature stop codon and the resulting
425 protein truncating transcript variant (PTV) leads to a significantly shorter NRNX3 protein
426 isoform.

427 Interestingly, when we analyzed the sequences of *NRXN2* and *NRXN3* genes using the RNA
428 binding protein binding site motif detection software FIMO[45] (MEME suite), we could detect 4
429 and 223 NOVA1 binding sites in *NRXN2* and *NRNX3* (Fig. 6g,h), respectively. Of these, 1 site
430 falls into adjacent introns and none falls inside the skipped exon for the *NRXN2* event (Fig. 6g),
431 whereas 9 sites fall inside the flanking introns and none falls in the skipped exon for the *NRXN3*
432 splicing event (Fig. 6h). Likewise, we could detect 13 and 156 RBFOX1/2/3 binding sites in the
433 *NRXN2* and *NRNX3* genes (Fig. 6g,h), respectively. Of these, none falls into neighboring introns
434 or inside the skipped exon for the *NRXN2* event (Fig. 6g), while one falls inside the flanking
435 introns and none falls inside the skipped exon for the *NRXN3* splicing event (Fig. 6h).

436 In combination, these results argue in favor of the *NRNX2* and *NRXN3* events under analysis
437 are controlled by NOVA and RBFOX splicing regulators. To confirm the differential splicing in
438 the *NRXN3* gene that creates the SE events mentioned in the previous paragraphs, we decided
439 to conduct reverse-transcription-coupled real-time quantitative PCR (RT-qPCR) as well as
440 standard PCR amplification, using gene-specific primers able to discriminate between those
441 splicing events. To detect the SE-NRXN3-i variant, we placed the forward oligo (p5) in exon 19
442 of the canonical transcript and the reverse oligo (p4) in the splice junction between exons 20
443 and 20.5 (Fig. 7a, right panel). To detect the SE-NRNX3-e variant, we used another forward

444 oligo (p1) in exon 19, and the reverse oligo (p3) was placed in the splice junction between
445 exons 20 and 21 of the canonical *NRXN3* Ensembl transcript ENST00000335750.7. To detect
446 both variants (and the vast majority of transcripts, according to the *NRXN3* Ensembl entry), we
447 used forward and reverse oligos (p1 and p2) in exons 19 and 20, respectively (Fig. 7a, left
448 panel).

449 The mean overall levels of *NRXN3* expression are lower in PTHS individuals, as judged by
450 qPCR with the oligos to globally detect the gene's transcripts (Fig. 7b), which is in keeping with
451 the lower transcriptomic expression found in samples from two PTHS subjects in RNA
452 sequencing experiments (Fig. 7c)[34]. Additionally, the combinations of oligos to detect the SE-
453 NRXN3-i and SE-NRXN3-e splicing variants showed that both have lower expression in PTHS
454 samples (Fig. 7d). Importantly, we calculated the ratio between the expression abundances
455 detected via qPCR for the SE-NRXN3-i versus the SE-NRXN3-e events; compared with the
456 ratios found in patients #1 and #2, controls #1 and #3 have a higher ratio, indicating that the SE-
457 NRXN3-i event is proportionally more abundant than the SE-NRXN3-e event in the control
458 samples in the comparison with the ratio found in those PTHS patient neurons (Fig. 7e).
459 Likewise, these numbers allow us to conclude that the difference in abundances for NRXN3-i
460 and SE-NRXN3-e events is less prominent in patients #1 and #2 than in parents #1 and #3.
461 Curiously, patient #3 also presents an imbalance between both variants, in this case in the
462 opposite direction, which also represents an aberrant splicing pattern with potential functional
463 consequences.

464 To further investigate the SE-NRXN3-i and SE-NRXN3-e splicing events, we used oligos for
465 regular PCR amplification, designed to discriminate between variants based on amplicon size.
466 One pair of primers was placed in exons 20 and 21 (Fig. 7f, top panel); the expected band size
467 for the PCR performed on the SE-NRXN3-e event with this primer pair (p6 and p7) is 334 bp
468 (Fig. 7f), although we could rule out the possibility that other band sizes would exist for other
469 configurations of spliced sites involving exons 20 and 21. Another primer pair (p6 and p8) was
470 placed in exons 20 and 20.5, which is expected to lead to band of size 252 bp only from the SE-
471 NRXN3-i event (Fig. 7f, bottom panel). The PCR results confirm the expected amplicons for
472 these primer pair/splicing event combinations (Fig. 7g), providing experimental evidence that
473 confirms the in silico analyses executed by rMATS-turbo on our PTHS neuron RNA-Seq data,
474 which predicted the existence of the SE-NRXN3-i and SE-NRXN3-e splicing variants.

475 The PCR results showed that we could amplify most of the expected bands that would support
476 the existence of alternative splicing of exon 20.5. However, other bands were also amplified

477 (Fig. 7g), therefore, we decided to clone the amplicons in bulk in a PCR cloning vector, followed
478 by Sanger sequencing of several clones. With the primer pair positioned in exons 20 and 21, we
479 detected amplicons with an exon configuration that matches the expected joining of exons 20
480 and 21 without the intervening exon 20.5 in both the control and PTHS samples (Supplementary
481 Fig. S4a-c), and the percentage of clones supporting the existence of this exon configuration is
482 larger in PTHS than in control samples (Supplementary Fig. S4c), which concurs with the
483 results from rMATS-turbo showing that transcripts with SE-NRXN3-e events are more prevalent
484 than SE-NRXN3-i in PTHS samples. Importantly, the combination of oligos pairing with exons
485 20 and 20.5, which is expected to amplify only the SE-NRXN3-i event-related amplicons,
486 resulted in sequences that match this expected exon configuration in the control samples
487 (Supplementary Fig. S4a-c); moreover, we could only find clones matching this exon
488 configuration in control samples (Supplementary Fig. S4c), which concurs with the positive ΔPSI
489 for the SE-NRXN3-i event in the controls (Fig. 6e). It is also curious to note that, although this
490 type of experiment is not quantitative, the intensity of the 334-bp band with oligos in exons 20
491 and 21 as well as the 252-bp band with oligos in exons 20 and 20.5 are stronger in the control
492 samples (Fig. 7g), which is in keeping with the lower global expression levels for the *NRXN3*
493 gene in PTHS samples revealed by the qPCR experiments (Fig. 7b,c).

494 Although other bands were detected in these PCR experiments, some larger than the predicted
495 size for the corresponding amplicons, they did not yield sequence reads in the Sanger
496 sequencing experiments performed later. The significance of these bands remains to be
497 investigated. In combination, the PCR results presented above confirm the existence of a
498 transcript containing an SE exon in the *NRXN3* gene (SE-NRXN3-i event) identified in silico in
499 our RNA-Seq data and that this splicing event has a lower usage in PTHS neurons, as
500 compared to the transcript without exon 20.5 (SE-NRXN3-e event).

501

502 **The PTHS-depleted alternatively spliced *NRXN3* transcript leads to a prematurely**
503 **stopped soluble Neurexin-3 protein isoform**

504 Next, we investigated the Neurexin-3 protein isoforms coded for by transcripts containing the
505 SE-NRXN3-i and SE-NRXN3-e events. The alternative *NRXN3* Ensembl transcript
506 ENST00000554738.5, containing the splicing configuration with exon 20.5 (SE-NRXN3-i event),
507 which is depleted in PTHS, results in a protein isoform with 6 laminin G domains and 3 EGF-like
508 domains (Fig. 8a), as judged by protein domain prediction with the InterPro tool[49].

509 Comparatively, the canonical *NRXN3* Ensembl transcript ENST00000335750.7 encodes a
510 protein with the same number of protein domains, but shorter C-terminus (Fig. 7A). The shorter
511 protein encoded when exon 20.5 is included results from a premature stop codon in this exon
512 (Fig. 8b). Because Neurexin proteins are transmembrane (TM) cellular components involved in
513 synapse organization[47], with the TM domain positioned near the C-terminus, we reasoned that
514 the shorter Neurexin isoform encoded by the transcript including exon 20.5 might be lacking part
515 of or all the TM domain. Therefore, we conducted a prediction of TM domains using the
516 DeepTMHMM tool[50], which showed that the shorter isoform completely lacks the TM domain
517 and may represent a soluble form of Neurexin (Fig. 8c). Because the balance between a
518 transmembrane and a secreted soluble form of Neurexin may have important functional
519 implications, we checked whether the protein isoforms from the canonical and the alternative
520 *NRNX3* transcripts are predicted to be secreted, using the SignalP (v 6.0) tool[51]. Both the
521 complete Neurexin isoform including exon 20.5 and the shorter isoform encoded by the
522 transcript including exon 20.5 are predicted to be secreted, because they contain a signal
523 peptide (Fig. 8d).

524 In combination, these results point to the higher prevalence of a distinct Neurexin protein
525 isoform due to abnormal splicing patterns, potentially dictated by the dysregulated expression of
526 RBFOX and NOVA1 splicing regulators in neurons derived from Pitt-Hopkins Syndrome
527 patients.

528 **DISCUSSION**

529 RNA splicing plays a major role in producing the incredible variety of transcripts and their
530 encoded protein products found in human cells. Yet, how the dysregulation of splicing events
531 lead to disease is a topic that has only recently emerged. Over the last two decades, several
532 studies have sought to elucidate how splicing can go from an important regulatory phenomenon
533 able to regulate human gene expression and generate diverse mature mRNAs to pathogenic
534 altered processes that can contribute to disease onset and progression. Of particular interest
535 are splicing events that lead to exon skipping (ES), intron retention (IR), alternative 5' or 3'
536 splice site usage (A5SS or A3SS), or the use of mutually exclusive exon[7]. Many of these
537 alternative splicing events lead to frameshifts and premature stop codons that result in protein
538 truncation (protein truncating variants; PTV), whose role in neurodevelopmental disorders are
539 becoming increasingly clear[2].

540 Even though the several recognized types of autism are the result of a complex interplay
541 between genetic factors and environmental influences, their pathophysiology is still mostly
542 unknown, and the disease mechanisms remain to be determined in most cases. It is becoming
543 apparent that, among the many molecular features of this complex set of disorders,
544 dysregulation of splicing is a more common phenomenon than initially anticipated[2]. A theme that
545 is starting to be a topic of intense research in the field of autism molecular neurobiology is the
546 decipherment of how altered splicing events or aberrant splicing patterns contribute to the
547 cellular characteristics of the neural tissue in autistic children.

548 In this study, we determined that several genes exhibit aberrant splicing patterns in neurons
549 derived from patients with a monogenic form of autism known as Pitt-Hopkins Syndrome (Figs.
550 3 and 4). We also showed that these altered splicing events are concomitant with the
551 dysregulated expression of NOVA1 and RBFOX1/2/3 splicing regulators (Figs. 1 and 2). These
552 are well known RNA binding proteins (RBPs) that facilitate spliceosome assembly and ensure
553 accurate recognition of splice sites to lead to appropriate intron excision[52,53]. Extensive RBP-
554 RNA interaction experiments have unveiled the participation of NOVA and RBFOX family
555 members in splicing regulation, determining that they are prominently expressed in neurons and
556 crucial to neurogenesis, axon guidance, and synapse formation[2]. An example of a neuronal
557 gene regulated by alternative splicing is *DLG4* (encoding PSD-95), which may generate
558 transcripts via the inclusion of specific exons that result in the incorporation of a domain lacking
559 palmitoylation sites, thereby reducing membrane localization and enhancing excitatory synaptic
560 transmission[54].

561 Importantly, altered splicing in neurons of autistic patients impacts genes involved in
562 neurodevelopment and synaptic plasticity, including synapse-related genes known to be risk
563 factors for autism, such as *NLGN3*, *NLGN4*, *NRXN1*, *SHANK3*, and *CADPS*[55]. Our data showed
564 that one of the aberrant splicing patterns observed in PTHS affect the *NRXN2* and *NRXN3*
565 genes (Fig. 6), which encode Neurexin protein family members. It is interesting to note that
566 altered splicing has been shown to affect the *NRXN1* gene in other forms of autism: these
567 studies have pinpointed the existence of region-specific splicing of *NRXN1* in the prefrontal
568 cortex, potentially influencing synapse formation[56]. Although this splicing event also involves a
569 gene in the Neurexin family, our data show a different case in which a monogenic form of
570 autism, caused by haploinsufficient heterozygous mutations in the *TCF4* gene, lead to changes
571 in RBFOX and NOVA splicing regulators, potentially in turn leading to aberrant splicing patterns
572 involving exon skipping events (Figs. 6 and 7), thereby producing a change in Neurexin 3
573 transcripts that modifies the balance between transmembrane and secreted forms of this
574 synapse-related protein (Figs. 7 and 8).

575 Changes in NOVA and RBFOX expression have been studied by many groups over the last
576 years, showing that these events disrupt neuronal development, leading to neurodevelopmental
577 and neurodegenerative diseases[57]. Other studies further sought to understand how such
578 dysregulation leads to pathological changes in splicing patterns. For example, RBFOX
579 expression was shown to be downregulated in autism, a characteristic that leads to aberrant
580 splicing in several ASD-associated genes in the SFARI database, including *SHANK3*,
581 *CACNA1C*, and *TSC2*[1] and synapse-related genes *NRCAM* and *GRIN1*[58,59]. Moreover,
582 mutations in *NOVA2* have been associated with autism and other neurodevelopmental
583 disorders, impairing neurite outgrowth and axonal pathfinding[60,61]. Thousands of NOVA2 binding
584 sites have been located in the mouse genome, particularly in genes involved in synapse and
585 axon guidance such as *Dcc*, *Robo2*, *Slit2*, and *Epha5*[62,63]. Our study significantly expands these
586 studies by describing that NOVA1 and RBFOX1/2/3 expression are impaired in a monogenic
587 form of autism, PTHS, and by detecting thousands of genes in which splicing patterns are
588 altered (Figs. 3 and 4).

589 It is noteworthy that many genes characterized by aberrant splicing patterns in PTHS are
590 involved in synapse formation and organization (Fig. 4), which prompted us to focus on one
591 particular splicing event (exon skipping) that affects the synapse-related *NRNX3* gene (Fig. 6).
592 This finding may explain the impaired electrical activity we observed in the PTHS neural
593 tissue[34], as judged by the analysis of neurons in traditional 2D culture and in patient-derived

594 brain organoids. These findings concur with our detection of transcript variants derived from
595 alternative splicing in genes in the Neurexin family, which are known to control synapse
596 formation during normal brain development[36,47]. In particular, the *NRXN3* gene encodes two
597 major types of Neurexin, a large isoform encoded by longest isoform including most exons (α
598 Neurexin-3 type) and a short isoform derived from transcription from an internal promoter (β
599 type), both transmembrane-bound proteins[47]. Neurexins interact with Neuroligin proteins to
600 control the rate at which synapses are formed in maturing neurons[64,65].

601 Our study detected a change in the abundances of two transcripts derived from alternative
602 splicing: in neurons of normal genotype, there seems to be an alternative splicing transcript that
603 results from exon exclusion, leading to a α-type Neurexin lacking the transmembrane domain
604 which is expected to be secreted[66]; in contrast, in PTHS neurons, the abundance of this
605 secreted form is diminished relative to the transmembrane form (Fig. 8). Although the transcript
606 with exon exclusion (SE-NRXN3-e) has already been described in the literature[66,67], its
607 involvement in the context of a neurodevelopmental disorder is novel and may open new
608 avenues of investigation to understand how such splicing dysregulation leads to aberrant
609 synapses and altered electrical activity in the neural tissue of autistic children. In the context of a
610 neurodegenerative disorder (Alzheimer's disease), the ratio between the Neurexin-3
611 transmembrane and soluble isoforms was found to be reduced[67], which is exactly in opposition
612 to what we observed in the context of autism. Further studies will illuminate the functional
613 contribution of these ratio alterations in both diseases.

614 Another interesting possibility is that the changes in ratio between the transmembrane and
615 secreted isoforms of Neurexin-3 may lead to differences in the content of GABAergic and
616 glutamatergic synapses in PTHS. We raised this possibility based on the observation that
617 deletion of the *Nrxn3* gene in mice led to a decrease in GABA-mediated inhibitory responses[65].
618 Outstandingly, we observed an augmentation of GABAergic neurons in PTHS cortical
619 organoids, while these cells are mostly lacking in control cortical organoids (Fig. 2b,c,e,g,i), a
620 piece of observation that may explain why PTHS cortical organoids have lower electrical
621 activity[34].

622 In conclusion, our study described alterations in the expression of splicing regulators in PTHS
623 and how this is accompanied by aberrant splicing patterns affecting hundreds of genes involved
624 in neurodevelopment. Further studies are needed to unravel how these alterations contribute to
625 the pathology in autistic brains and to result in the identification of molecular therapeutic targets,

626 eventually leading to the development of strategies to correct abnormalities in splicing regulation
627 in autism.

628 # METHODS

629 ## Patient-derived neural progenitor cell lines and neuronal differentiation

630 Our studies are based on RNA-Seq and single-cell RNA-Seq from our previous publication
631 using neurons and brain organoids derived from patients with PTHS[34]. This study included
632 subjects from volunteering families recruited through the University of Campinas
633 (Supplementary Table 6) harboring heterozygous mutations in *TCF4* mutation. Details on the
634 patients' PTHS clinical symptoms are reported in Supplementary Table 6. Control subjects were
635 the patients' same-sex parents, with no prior history of psychiatric or genetic disorders. The
636 participation of all subjects was approved by the Human Subjects Ethics Committees of the
637 University of Campinas, under CAAE number 86136925.6.0000.5404. Written informed consent
638 was obtained from all participating families after receiving a thorough description of the study
639 and no compensation was provided to participants.

640 In our previous study, neural progenitor cells were obtained from reprogrammed iPSCs derived
641 from skin fibroblasts subjected to cellular reprogramming via over-expression of the Yamanaka
642 factors *OCT4*, *SOX2*, *KLF4*, and *MYC* (Cytotune iPS 2.0 Sendai reprogramming kit; Thermo
643 Fisher Scientific)[34]. All iPSC lines were validated via immunostaining and SNP mapping to rule
644 out the presence of unwanted chromosomal abnormalities and mutations. Results reported in
645 this paper are from experiments conducted with one P15 iPSC clone per subject. Cultures were
646 tested every two weeks for mycoplasma.

647 iPSC colonies maintained in mTeSR1 Plus medium (Stem Cell Technologies), followed by
648 addition of 10 mM SB431542 and 1 mM dorsomorphin to form embryoid bodies. After 2 weeks
649 of culturing, the embryoid bodies were plated onto Matrigel-coated plates with DMEM/F12
650 medium plus N2 and SM1 supplements (Stem Cell Technologies) and 20 ng/mL FGF-2. Neural
651 rosettes were then picked and replated onto Matrigel-coated dishes, and the resulting NPCs
652 were dissociated with Accutase and re-seeded onto plates coated with 10 µg/mL poly-ornithine
653 (Sigma-Aldrich) and 5 µg/mL laminin (Thermo Fisher Scientific).

654 For neuronal differentiation, NPCs were seeded onto plates coated with poly-ornithine and
655 laminin. At 90% confluency, medium was replaced with DMEM/F12 containing N2 and B27
656 supplements (Thermo Fisher Scientific) and 1% penicillin/streptomycin, with media changes
657 occurring every 3 to 4 days.

658

659 **RNA sequencing from NPCs and neuronal cultures and single-cell RNA sequencing data**
660 **from organoids.**

661 Our study employed bulk RNA sequencing data from NPCs and neurons derived from PTHS
662 syndrome, which have been previously published by our group[34]. Briefly, these libraries were
663 produced from NPCs of 4 subjects and 4 respective parental controls at passage 15 and from
664 differentiating neuronal cultures of 3 parent-patient pairs and an additional patient line at 2
665 months in vitro. RNA was extracted from 3 independently prepared biological replicates,
666 followed by Illumina library preparation using the stranded TruSeq kit (Illumina) and sequencing
667 on an Illumina NovaSeq 6000 S4 instrument with 150 bp paired-end reads. Transcript-level
668 expression abundance was estimated using Salmon (version 0.14.1) software[68], with selective
669 mapping ('--validateMappings') and correction for sequence-specific biases ('--seqBias'), GC-
670 content biases ('--seqBias'), and fragment position bias ('--posBias'). Reference transcripts for
671 read mapping were obtained from the GENCODE 32 basic annotation[69]. MA plots of specific
672 genes in RNA-Seq data was produced from dataframes containing the expression abundances
673 in each individual library, using the 'ggplot' package for *R* language.

674 Single-cell RNA-Seq data used in this study was also previously published by our team (REF
675 our paper), with samples from cortical organoids (CtOs) and subpallial organoids (sPOs).
676 Briefly, CtOs were produced from iPSC colonies dissociated using Accutase (Thermo Fisher
677 Scientific), followed by neural induction in mTeSR1 Plus medium (Stem Cell Technologies)
678 supplemented with 10 mM SB431542 (Stemgent) and 1 mM dorsomorphin (R&D Systems).
679 Embryoid bodies formed were transferred to Neurobasal medium (Thermo Fisher Scientific)
680 containing GlutaMAX, 1% B27 supplement (Thermo Fisher Scientific), 1% N2 NeuroPlex
681 supplement (Gemini Bio-Products), 1% NEAA (Thermo Fisher Scientific), 1%
682 penicillin/streptomycin (Thermo Fisher Scientific), 10 mM SB431542, and 1 mM dorsomorphin,
683 for 7 days. NPC proliferation medium was then used, consisting of Neurobasal medium
684 containing GlutaMAX, 1% B27, 1% NEAA, 20 ng/mL FGF-2 (Thermo Fisher Scientific) for 7
685 days, followed by 7 additional days with 20 ng/mL EGF (PeproTech). Neuronal differentiation
686 was achieved in Neurobasal medium containing 1% GlutaMAX, 1% Gem21, 1% NEAA, 10
687 ng/mL of BDNF, 10 ng/mL of GDNF, 10 ng/mL of NT-3 (all from PeproTech), 200 mM L-
688 ascorbic acid, and 1 mM dibutyryl-cAMP (Sigma-Aldrich), for 7 days. GABAergic-enriched
689 subpallial organoids (sPOs) were producing using a similar protocol, with neural induction
690 medium containing 5 μM IWP-2 (SelleckChem), from day 4 until day 10, followed by incubation
691 in proliferation medium (Neurobasal medium containing 1% GlutaMAX, 1% Gem21, 1% NEAA,

692 20 ng/mL FGF-2, and 100 nM SAG (SelleckChem) for 7 days, followed by 2 additional days in
693 the same medium supplemented with 20 ng/mL EGF (PeproTech).

694 Single cell RNA-Seq was performed on dissociated CtOs and GbOs at 8 weeks in vitro, with 15
695 organoids per library (5 organoids from each of three wells from independent experiments)
696 dissociated with forceps and enzymatic digestion with Accutase. Libraries were prepared using
697 the Chromium Single Cell 3' v3 Library kit (10× Genomics) according to the manufacturer's
698 protocol, with a target of 20,000 cells per sample on the Chromium chip and PCR-amplification
699 for a total of 10 cycles. Sequencing was performed on a NovaSeq 6000 S4 sequencer (Illumina)
700 to produce 20,000 reads per single cell. Feature count matrices for each single-cell RNA-Seq
701 library were generated separately using the 'cellranger count' command in the Cell Ranger
702 software (version 4.0.0) and the GRCh38 2020-A reference dataset of human transcripts,
703 yielding an estimated number of cells across all libraries in the range 2,944 to 6,466 per library,
704 with a mean number of reads per cell ranging from 58,501 to 169,931. Cell type subpopulations
705 were demarcated via automated annotation and refinement after manual inspection. Seurat
706 (version 3.2.2)[70] was used for producing UMAP and violin plots in this study, after normalization
707 to 10,000 counts per cell and log-transformation with the 'NormalizeData' function, finding of
708 variable features with the 'FindVariableFunctions' function, scaling with the 'ScaleData' function,
709 and linear dimensional reduction (PCA) with the 'RunPCA' function. Non-linear reduction was
710 done using the UMAP algorithm in the 'RunUMAP' function. For statistical comparisons of gene
711 expression levels, we used Mann-Whitney U test.

712

713 **rMATS-turbo analysis of splicing events**

714 rMATS-turbo analysis was conducted using control and patient RNA-seq data previously
715 published by our group (GEO accession numbers GSE159392 and GSE189121)[34]. For
716 alternative splicing analysis conducted with libraries with different read sequence lengths,
717 sequences were trimmed down to the shortest length across all libraries (101 bp) using
718 Trimmomatic software[71]. rMATS-turbo was run using the following arguments: end-pairing ('-t')
719 equal to 'paired', read length ('—readLength') equal to 101 or 151, depending on the library, as
720 well as a function to find novel splice-sites ('—novelSS'). The raw output data was then filtered
721 using rMATS-turbo specific in-built filtering script 'rmats_filtering.py'[72], with parameters set to
722 filter events with a minimum average number of 10 read counts per sample, to filter out events
723 with PSI < 0.05 or > 0.95, to filter events with FDR < 0.05, and to filter events with absolute
724 values of IncLevelDifference (ΔPSI) > 0.05. Over-representation (ORA) and Gene-Set

725 Enrichment Analysis (GSEA) were performed using the lists of genes for which SE events had
726 absolute values of ΔPSI > 0.3, for each patient-control pair, using the WebGestalt software[73]
727 and Gene Ontology - Biological Process and KEGG pathway functional databases, with the
728 'genome protein-coding' reference set. For GSEA, the list of genes was ordered according to
729 the absolute values of IncLevelDifference, with GO-Biological Process and KEGG pathway
730 functional databases.

731

732 **RBP binding site recognition and enrichment analyses**

733 For the analysis of motif enrichment for RBFOX and NOVA regulators, we recovered the
734 genomic coordinates for the list of SE events that had absolute values of ΔPSI > 0.1, for patient-
735 control pair #2. To recover the sequences of the exons involved in these SE events plus 200 bp
736 of the neighboring introns, we subtracted 199 from the value of the initial exonic genomic
737 coordinate ('exonStart_0base' column in the rMATS-turbo output table) and added 200 to the
738 value of the final exonic genomic coordinate ('exonEnd' column). Sequences were recovered in
739 fasta format using the 'faidx' function in samtools[74] from the GRCh38 human genome database.
740 RBP motif enrichment was conducted using the SEA software in the MEME suite[43], with RNA
741 motif database 'Ray2013_rbp_Homo_sapiens.dna_encoded.meme', which was modified to
742 include motifs for NOVA1, RBFOX2, and RBFOX3. A position-specific scoring matrix (PSSM)
743 for NOVA1 was obtained from the RBPmap developing team[75,76]; PSSM for RBFOX2 and
744 RBFOX3 were produced using the 'matrix2meme' script in the MEME suite using published
745 motif sequences from RNA Bind-n-Seq (RBNS) results (Supplementary Table S3 in ref[44]). SEA
746 was run using default parameters plus arguments to avoid trimming ('--notrim') and the '--
747 hofract' parameter set to 0.5. As a negative control sequence set, we started with a random
748 selection of 5,000 exons collected across all genes in the GRCh38 database using 'awk', 'sort',
749 and 'head', to produce a FASTA file with exonic and neighboring intronic (200 bp) sequences.
750 As positive control sequences for RBFOX1/2/3 and NOVA1, we selected genes whose splicing
751 is known to regulated by these proteins[2,77] and recovered all exonic plus neighboring intronic
752 (200 bp) sequences.

753 For the detection of binding sites of NOVA1 and RBFOX1/2/3 in specific genes (including
754 *NRXN3*), we first obtained the FASTA files for the target genes from Ensembl (genome
755 assembly GRCh38.p14; GCA_000001405.29) and then run the motif search using Find
756 Individual Motif Occurrences (FIMO) software in the MEME suite[45], with the '--norc' argument to
757 prevent motif scanning in the reverse complement strand. We ran this analysis for 50 genes at

758 the top of the SE event list from rMATS-turbo, ordered according to the absolute values of ΔPSI.
759 In parallel, we ran FIMO scanning for 50 random genes that were not found in the rMATS-turbo
760 output SE event list. The frequencies of RBP binding sites were then normalized to 100,000 bp,
761 followed by enrichment analysis using unpaired *t*-test in GraphPad Prism.

762

763 **PCR and sequencing of cloned bands**

764 PCR was conducted using GoTaq DNA polymerase (Promega; cat# M3001) with primers p6 (5'
765 TCTGCTGAATGTTCAAGTG 3') and p7 (5' TCTTCAGTTTGGTGCTGTC 3') for the first
766 reaction and primers p6 and p8 (5' GGAGATGGATGACAGCAC 3') for the second reaction (Fig.
767 7f,g), using the following cycling parameters: 94°C 5 min, and 38 cycles of 94°C 45 s, 58°C 45 s,
768 and 72°C 2 min. The fragments were then cloned into the pGEM-T-easy vector (Promega) and
769 Sanger sequenced using universal M13 primer. Alignment was performed using Benchling
770 alignment tool.

771

772 **Protein secondary structure analysis**

773 Protein secondary structure analysis was performed using InterProScan[49] for general domain
774 prediction using the coding amino acid sequence of *NRXN3* ENST00000335750.7 and
775 ENST00000554738.5 transcripts, representing the transmembrane and soluble isoforms,
776 respectively, using default parameters. Specific transmembrane domain prediction was
777 performed using DeepTMHMM 1.0 software[50], inputting the same sequences mentioned above.
778 The likelihood of secretion was evaluated using the SecretomeP (v 2.0) tool

779

780 **Real-time quantitative qPCR**

781 For real-time qPCR, neuronal cultures were differentiated from NPCs seeded onto 6-well plates,
782 followed by withdrawal of FGF-2 for 3 weeks, prior to RNA extraction with the RNeasy Mini Plus
783 kit (Qiagen). All samples were analyzed with 3 independent biological replicates. cDNA was
784 synthesized from total RNA using the Improm II First-Strand Reverse Transcription System
785 (Promega). Real-time quantitative PCR (RT-qPCR) was performed using regular primers for
786 SYBR Green technology, on an Applied Biosystems Real Time PCR system, with the following
787 cycling parameters: 94°C for 3 min, followed by 40 cycles of 94°C for 30 s and 68°C for 1 min.
788 Amplification and denaturation curves for all probes were analyzed to verify amplification of just

789 one amplicon. Normalization was achieved with endogenous control gene *ACTB*, and relative
790 expression was calculated using the traditional ΔΔCt method. The following qPCR primers were
791 used to detect all *NRXN3* transcript variants: p1 (5' ACAACCCGTAAGAATGCGTC 3') and p2
792 (5' CATCACTTGAACATTCAGCAG 3'). The following variant-specific qPCR primers were used:
793 p1, p3 (5' GGAAAGACTCTTATCTGTACTC 3'), p4 (5' TAGAGCTTCTGGCTGTACTC 3'), and
794 p5 (5' CGTAAGAATGCGCTCTACAG 3').

795

796 **Statistical analysis and reproducibility**

797 Data are presented as mean + SEM, unless otherwise indicated. Different types of statistical
798 test were used throughout the study, including comparisons of means between two groups
799 (PTHS against parent) in experiments that measured relative expression levels, or expression
800 abundances used two-sample Welch's t test, assuming unequal variances and
801 heteroskedasticity. For comparing mean gene expression in single cell RNA-Seq data between
802 two samples, we used the non-parametric Mann–Whitney U-test. Sample sizes are indicated in
803 the figure legends. *P*-values are reported as asterisks in the figures for significance levels
804 defined as $p<0.05$ (*), $p<0.01$ (**), or $p<0.001$ (***). Supplementary Table S8 presents extended
805 results for all statistical tests performed, including sample sizes, statistical tests employed, effect
806 sizes, statistics metrics (*H, F, t*, or *W*), along with exact *p*-values, listed according to the order of
807 appearance in figure panels throughout the study. The numbers of replicates are also indicated
808 in the figure legends and all attempts at replication were successful. When experiments involved
809 data collected from independent biological replicates, results were collected from randomly
810 chosen replicates (wells/plates of experiments involving cells in 2D culture) and blinding was
811 used for most analyses comparing patients and control samples, including immunostaining.
812 Blinding was not used when analyzing results from RNA sequencing and single cell RNA
813 sequencing experiments, due to the inherently unbiased nature of the bioinformatic approaches
814 used for quantitating gene expression and determining differential expression between
815 genotypes or cell types. Statistical analyses were performed using Prism software (GraphPad;
816 version 9.2.0). Figures were composed with Prism and Illustrator CS4 (Adobe; version 14.0.0).

817 **Materials Availability**

818 Further information and requests for resources and reagents should be directed to and will be
819 fulfilled by the corresponding author, Fabio Papes (papesf@unicamp.br), upon reasonable
820 request. When provided to others, unique reagents generated in this study will be available with
821 a completed Materials Transfer Agreement.

822

823 **Data Availability**

824 Data supporting the findings in this study are included within the Supplementary Material. RNA
825 sequencing and single cell RNA sequencing raw and processed data used in this study are
826 deposited at the Gene Expression Omnibus (GEO) of the National Center for Biotechnology
827 Information (NCBI), under accession numbers GSE159392
828 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159392), GSE159859
829 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159859), GSE159860
830 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE159860), and GSE189121
831 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE189121), which are available publicly
832 without restriction.

833

834 **Code Availability**

835 No new custom software or code has been used in this paper. Codes (R programming
836 language) used for bioinformatic analyses are all strictly based on pre-existing, regularly used
837 published codes for RNA-Seq and single-cell RNA-Seq analyses, as described above.

838 ## References


839 1. Weyn-Vanhentenryck, S. M. *et al.* HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-
840 Regulatory Network Linked to Brain Development and Autism. *Cell Rep* **6**, 1139–1152 (2014).

841 2. Jeong, J., Yoo, H. J., An, J. Y. & Jeong, S. Dysregulated RNA-binding proteins and alternative
842 splicing: Emerging roles in autism spectrum disorder. *Mol Cells* **48**, 100237 (2025).

843 3. Kjer-Hansen, P. & Weatheritt, R. J. The function of alternative splicing in the proteome: rewiring
844 protein interactomes to put old functions into new contexts. *Nature Structural & Molecular*
845 *Biology 2023 30:12* **30**, 1844–1856 (2023).

846 4. Zheng, S. Alternative splicing programming of axon formation. *Wiley Interdiscip Rev RNA* **11**,
847 e1585 (2020).

848 5. Quesnel-Vallières, M., Weatheritt, R. J., Cordes, S. P. & Blencowe, B. J. Autism spectrum disorder:
849 insights into convergent mechanisms from transcriptomics. *Nature Reviews Genetics 2018 20:1*
850 **20**, 51–63 (2018).

851 6. Kim, K. K., Adelstein, R. S. & Kawamoto, S. Identification of neuronal nuclei (NeuN) as Fox-3, a
852 new member of the Fox-1 gene family of splicing factors. *Journal of Biological Chemistry* **284**,
853 31052–31061 (2009).

854 7. Jacko, M. *et al.* Rbfox Splicing Factors Promote Neuronal Maturation and Axon Initial Segment
855 Assembly. *Neuron* **97**, 853-868.e6 (2018).

856 8. An, J. Y. *et al.* Towards a molecular characterization of autism spectrum disorders: An exome
857 sequencing and systems approach. *Transl Psychiatry* **4**, (2014).

858 9. Teplova, M. *et al.* Protein-RNA and protein-protein recognition by dual KH1/2 domains of the
859 neuronal splicing factor Nova-1. *Structure* **19**, 930–944 (2011).

860 10. Scala, M. *et al.* De novo truncating NOVA2 variants affect alternative splicing and lead to
861 heterogeneous neurodevelopmental phenotypes. *Hum Mutat* **43**, 1299–1313 (2022).

862 11. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA
863 processing. *Nature* **456**, 464–469 (2008).

864 12. O'Leary, A. *et al.* Behavioural and functional evidence revealing the role of RBFOX1 variation in
865 multiple psychiatric disorders and traits. *Mol Psychiatry* **27**, 4464 (2022).

866 13. Brockschmidt, A. *et al.* Severe mental retardation with breathing abnormalities (Pitt-Hopkins
867 syndrome) is caused by haploinsufficiency of the neuronal bHLH transcription factor TCF4. *Hum*
868 *Mol Genet* **16**, 1488–94 (2007).

869 14. Amiel, J. *et al.* Mutations in TCF4, Encoding a Class I Basic Helix-Loop-Helix Transcription Factor,
870 Are Responsible for Pitt-Hopkins Syndrome, a Severe Epileptic Encephalopathy Associated with
871 Autonomic Dysfunction. *The American Journal of Human Genetics* **80**, 988–993 (2007).

872 15. Zweier, C. *et al.* Haploinsufficiency of TCF4 Causes Syndromal Mental Retardation with
873 Intermittent Hyperventilation (Pitt-Hopkins Syndrome). *The American Journal of Human Genetics*
874 **80**, 994–1001 (2007).

875 16. Pitt, D. & Hopkins, I. A syndrome of mental retardation, wide mouth and intermittent
876 overbreathing. *Aust Paediatr J* **14**, 182–184 (1978).

877 17. Kim, H., Berens, N. C., Ochandarena, N. E. & Philpot, B. D. Region and Cell Type Distribution of
878 TCF4 in the Postnatal Mouse Brain. *Front Neuroanat* **14**, (2020).

879 18. de Pontual, L. *et al.* Mutational, functional, and expression studies of the TCF4 gene in Pitt-
880 Hopkins syndrome. *Hum Mutat* **30**, 669–76 (2009).

881 19. Jung, M. *et al.* Analysis of the expression pattern of the schizophrenia-risk and intellectual
882 disability gene TCF4 in the developing and adult brain suggests a role in development and
883 plasticity of cortical and hippocampal neurons. *Mol Autism* **9**, 1–15 (2018).

884 20. Sepp, M., Kannike, K., Eesmaa, A., Urb, M. & Timmusk, T. Functional Diversity of Human Basic
885 Helix-Loop-Helix Transcription Factor TCF4 Isoforms Generated by Alternative 5' Exon Usage and
886 Splicing. *PLoS One* **6**, e22138 (2011).

887 21. Chen, E. S. *et al.* Molecular convergence of neurodevelopmental disorders. *Am J Hum Genet* **95**,
888 490–508 (2014).

889 22. Schmidt-Edelkraut, U., Daniel, G., Hoffmann, A. & Spengler, D. Zac1 Regulates Cell Cycle Arrest in
890 Neuronal Progenitors via Tcf4. *Mol Cell Biol* **34**, 1020–1030 (2014).

891 23. Hill, M. J. *et al.* Knockdown of the schizophrenia susceptibility gene TCF4 alters gene expression
892 and proliferation of progenitor cells from the developing human neocortex. *Journal of Psychiatry*
893 *and Neuroscience* **42**, 181–188 (2017).

894 24. Forrest, M. P. *et al.* The Psychiatric Risk Gene Transcription Factor 4 (TCF4) Regulates
895 Neurodevelopmental Pathways Associated With Schizophrenia, Autism, and Intellectual
896 Disability. *Schizophr Bull* **44**, 1100–1110 (2018).

897 25. Page, S. C. *et al.* The schizophrenia-and autism-associated gene, transcription factor 4 regulates
898 the columnar distribution of layer 2/3 prefrontal pyramidal neurons in an activity-dependent
899 manner. *Mol Psychiatry* **23**, 304–315 (2018).

900 26. Fischer, B. *et al.* E-proteins orchestrate the progression of neural stem cell differentiation in the
901 postnatal forebrain. *Neural Dev* **9**, 23 (2014).

902 27. Marangi, G. & Zollino, M. Pitt-Hopkins Syndrome and Differential Diagnosis: A Molecular and
903 Clinical Challenge. *J Pediatr Genet* **4**, 168–76 (2015).

904 28. Zollino, M. *et al.* Diagnosis and management in Pitt-Hopkins syndrome: First international
905 consensus statement. *Clin Genet* **95**, 462–478 (2019).

906 29. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747
907 (2009).

908 30. Ripke, S. *et al.* Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*
909 **43**, 969–976 (2011).

910 31. Smoller, J. W. *et al.* Identification of risk loci with shared effects on five major psychiatric
911 disorders: a genome-wide analysis. *The Lancet* **381**, 1371–1379 (2013).

912 32. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the
913 genetic architecture of major depression. *Nat Genet* **50**, 668–681 (2018).

914 33. Gelernter, J. *et al.* Genome-wide association study of post-traumatic stress disorder
915 reexperiencing symptoms in >165,000 US veterans. *Nat Neurosci* **22**, 1394–1401 (2019).

916 34. Papes, F. *et al.* Transcription Factor 4 loss-of-function is associated with deficits in progenitor
917 proliferation and cortical neuron content. *Nat Commun* **13**, (2022).

918 35. Teixeira, J. R., Szeto, R. A., Carvalho, V. M. A., Muotri, A. R. & Papes, F. Transcription factor 4 and
919 its association with psychiatric disorders. *Transl Psychiatry* **11**, (2021).

920 36. Gomez, A. M., Traunmüller, L. & Scheiffele, P. Neurexins: Molecular Codes for Shaping Neuronal
921 Synapses. *Nat Rev Neurosci* **22**, 137 (2021).

922 37. Trujillo, C. A. *et al.* Complex Oscillatory Waves Emerging from Cortical Organoids Model Early
923 Human Brain Network Development. *Cell Stem Cell* **25**, 558-569.e7 (2019).

924 38. Yoon, S.-J. *et al.* Reliability of human cortical organoid generation. *Nat Methods* **16**, 75–78 (2019).

925 39. Gasiorowska, A. *et al.* The Biology and Pathobiology of Glutamatergic, Cholinergic, and
926 Dopaminergic Signaling in the Aging Brain. *Front Aging Neurosci* **13**, 654931 (2021).

927 40. Sun, L. *et al.* Transcriptomic insights into fate choice of pallial versus subpallial GABAergic
928 neurons. *Nature Communications* **16**, 1–20 (2025).

929 41. Wang, Y. *et al.* rMATS-turbo: an efficient and flexible computational tool for alternative splicing
930 analysis of large-scale RNA-seq data. *Nature Protocols 2024 19:4* **19**, 1083–1104 (2024).

931 42. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive,
932 powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res* **45**,
933 W130–W137 (2017).

934 43. Bailey, T. L. & Grant, C. E. SEA: Simple Enrichment Analysis of motifs. *bioRxiv* 2021.08.23.457422
935 (2021) doi:10.1101/2021.08.23.457422.

936 44. Dominguez, D. *et al.* Sequence, Structure, and Context Preferences of Human RNA Binding
937 Proteins. *Mol Cell* **70**, 854-867.e9 (2018).

938 45. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif.
939 *Bioinformatics* **27**, 1017–1018 (2011).

940 46. Yuan, Y. *et al.* Cell type-specific CLIP reveals that NOVA regulates cytoskeleton interactions in
941 motoneurons. *Genome Biol* **19**, 1–19 (2018).

942 47. Zhang, R., Jiang, H. X., Liu, Y. J. & He, G. Q. Structure, function, and pathology of Neurexin-3.
943 *Genes Dis* **10**, 1908–1919 (2023).

944 48. Hishimoto, A. *et al.* Neurexin 3 polymorphisms are associated with alcohol dependence and
945 altered expression of specific isoforms. *Hum Mol Genet* **16**, 2880–2891 (2007).

946 49. Blum, M. *et al.* InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res*
947 **53**, D444–D456 (2025).

948 50. Hallgren, J. *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep
949 neural networks. *bioRxiv* 2022.04.08.487609 (2022) doi:10.1101/2022.04.08.487609.

950 51. Teufel, F. *et al.* SignalP 6.0 predicts all five types of signal peptides using protein language
951 models. *Nat Biotechnol* **40**, 1023–1025 (2022).

952 52. Kumari, A., Sedehizadeh, S., Brook, J. D., Kozlowski, P. & Wojciechowska, M. Differential fates of
953 introns in gene expression due to global alternative splicing. *Hum Genet* **141**, 31–47 (2022).

954 53. Jeong, S. SR Proteins: Binders, Regulators, and Connectors of RNA. *Mol Cells* **40**, 1–9 (2017).

955 54. Han, K. & Kim, E. Synaptic adhesion molecules and PSD-95. *Prog Neurobiol* **84**, 263–283 (2008).

956 55. Quesnel-Vallières, M., Weatheritt, R. J., Cordes, S. P. & Blencowe, B. J. Autism spectrum disorder:
957 insights into convergent mechanisms from transcriptomics. *Nat Rev Genet* **20**, 51–63 (2019).

958 56. Fuccillo, M. V. *et al.* Single-Cell mRNA Profiling Reveals Cell-Type-Specific Expression of Neurexin
959 Isoforms. *Neuron* **87**, 326–340 (2015).

960 57. Parra, A. S. & Johnston, C. A. Emerging Roles of RNA-Binding Proteins in Neurodevelopment. *J*
961 *Dev Biol* **10**, (2022).

962 58. An, J. Y. *et al.* Towards a molecular characterization of autism spectrum disorders: an exome
963 sequencing and systems approach. *Transl Psychiatry* **4**, (2014).

964 59. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular
965 pathology. *Nature* **474**, 380–386 (2011).

966 60. Mattioli, F. *et al.* De Novo Frameshift Variants in the Neuronal Splicing Factor NOVA2 Result in a
967 Common C-Terminal Extension and Cause a Severe Form of Neurodevelopmental Disorder. *Am J*
968 *Hum Genet* **106**, 438–452 (2020).

969 61. Scala, M. *et al.* De novo truncating NOVA2 variants affect alternative splicing and lead to
970 heterogeneous neurodevelopmental phenotypes. *Hum Mutat* **43**, 1299–1313 (2022).

971 62. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA
972 processing. *Nature* **456**, 464–469 (2008).

973 63. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215
974 (2003).

975 64. Graf, E. R., Zhang, X., Jin, S. X., Linhoff, M. W. & Craig, A. M. Neurexins Induce Differentiation of
976 GABA and Glutamate Postsynaptic Specializations via Neuroligins. *Cell* **119**, 1013 (2004).

977 65. Südhof, T. C. Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits. *Cell*
978 **171**, 745 (2017).

979 66. Hishimoto, A. *et al.* Neurexin 3 polymorphisms are associated with alcohol dependence and
980 altered expression of specific isoforms. *Hum Mol Genet* **16**, 2880–2891 (2007).

981 67. Hishimoto, A. *et al.* Neurexin 3 transmembrane and soluble isoform expression and splicing
982 haplotype are associated with neuron inflammasome and Alzheimer's disease. *Alzheimers Res*
983 *Ther* **11**, (2019).

984 68. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-
985 aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).

986 69. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic*
987 *Acids Res* **47**, D766–D773 (2019).

988 70. Sun, A. X. *et al.* Potassium channel dysfunction in human neuronal models of Angelman
989 syndrome. *Science (1979)* **366**, 1486–1492 (2019).

990 71. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data.
991 *Bioinformatics* **30**, 2114 (2014).

992 72. Wang, Y. *et al.* rMATS-turbo: an efficient and flexible computational tool for alternative splicing
993 analysis of large-scale RNA-seq data. *Nat Protoc* **19**, 1083–1104 (2024).

994 73. Elizarraras, J. M. *et al.* WebGestalt 2024: faster gene set analysis and new support for
995 metabolomics and multi-omics. *Nucleic Acids Res* **52**, W415–W421 (2024).

996 74. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).

997 75. Paz, I., Argoetti, A., Cohen, N., Even, N. & Mandel-Gutfreund, Y. RBPmap: A Tool for Mapping and
998 Predicting the Binding Sites of RNA-Binding Proteins Considering the Motif Environment.
999 *Methods Mol Biol* **2404**, 53–65 (2022).

1000 76. Paz, I., Kosti, I., Ares, M., Cline, M. & Mandel-Gutfreund, Y. RBPmap: A web server for mapping
1001 binding sites of RNA-binding proteins. *Nucleic Acids Res* **42**, (2014).

1002 77. Piton, A. NOVA1/2 genes and alternative splicing in neurodevelopment. *Curr Opin Genet Dev* **93**,
1003 (2025).

1004 78. Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism
1005 spectrum disorders (ASDs). *Mol Autism* **4**, 36 (2013).

1006

1007

1008 **Acknowledgments**
1009 We thank Drs. Gonçalo Pereira, J. Andrés Yunes, and Maria Andreia Delbin for resources, Ana
1010 Paula Ferreira and Welbe Bragança for administrative and technical help, Dr. Elsa Molina
1011 (former director of the UCSD Stem Cell Genomics Core) for technical assistance with single cell
1012 RNA-Seq (made possible in part by CIRM Major Facilities grant #FA1-00607 to the Sanford
1013 Consortium for Regenerative Medicine). This work was supported by Sao Paulo Research
1014 Foundation (FAPESP) grants (#2020/11451-7 and #2022/12762-1), by FAPESP fellowships
1015 (#2024/18734-5 to P.K.K.F., #2024/12825-9 to B.B.H., #2025/09026-0 to J.G.B., #2024/20077-2
1016 to G.D.V., and #2020/16525-9 to I.C.V.), and by a CNPq fellowship (L.L.D.). We would also like
1017 to thank Yael Mandel-Gutfreund and Dina Alexandrovich (RBPmap team[76]) for providing the
1018 NOVA1 motif PSSM.
1019

1020 **Author contributions**

1021 Conceptualization, F.P.; Methodology, F.P., P.K.K.F., A.P.C., and M.C.; Investigation, P.K.K.F.,
1022 L.L.D., B.B.H., J.G.B., G.D.V., I.C.V., and M.H.B.; Writing – Original Draft, F.P., P.K.K.F.;
1023 Writing – Review & Editing, F.P. and P.K.K.F.; Funding Acquisition, F.P.; Resources, F.P.,
1024 M.H.B., and M.C.; Supervision, F.P.

1025

1026 **Competing interests**

1027 Authors declare no competing interests.

1028 ## Figure legends

1029 **Figure 1: Lower expression of splicing regulators NOVA1 and RBFOX1/2/3 in neurons**
1030 **derived from Pitt-Hopkins Syndrome (PTHS) patients. (a-c)** MA plots showing differential
1031 gene expression in neurons in 2D culture in the comparison between different pairs of control
1032 and PTHS subjects – pair #1 in (a), pair #2 in (b) and pair #3 in (c). See Supplementary Table
1033 S6 for patient information. Each parent-patient comparison is shown separately. Gray dots are
1034 genes that are not statistically significantly differentially expressed between control and PTHS.
1035 Colored dots are statistically significant DE genes, and the color indicates adjusted *p*-value (*s*-
1036 value). Red dots represent genes coding for splicing regulators NOVA1, RBFOX1, RBFOX2,
1037 and RBFOX3. The *x* axes represent average normalized expression across all libraries. Fold-
1038 changes (in log scale, base 2) in the *y* axes indicate the contrast between expression levels in
1039 the PTHS patient and the corresponding control (negative values indicate lower expression in
1040 PTHS neurons). **(d)** Summary of fold changes (in log scale, base 2) of *NOVA1* and *RBFOX1/2/3*
1041 gene expression for each patient-parent pair. Symbols indicate different pairs, as shown in
1042 Supplementary Table S6. **(e-h)** MA plots showing differential gene expression in neural
1043 progenitor cells in 2D culture in the comparison between different pairs of control and PTHS
1044 subjects – pair #1 in (e), pair #2 in (f), pair #3 in (g), and pair #4 in (h). **(i)** Summary of fold
1045 changes of *NOVA1* and *RBFOX1/2/3* gene expression in neural progenitor cells. Notice the lack
1046 of change in expression of some of these splicing regulators in neural progenitors. **(j,k)** Relative
1047 expression levels for *NOVA1*, *RBFOX1*, *RBFOX2*, and *RBFOX3* genes, as determined from RT-
1048 qPCR experiments, in iPSC-derived neuronal cultures. Three patients per graph, with mean
1049 expression normalized to 1 in control #1 (j) or control #2 (k). N = 2 technical replicates per
1050 sample. Bar graphs represent mean + SEM. ns, not statistically significantly different, *$p<0.05$,
1051 **$p<0.01$, ***$p<0.001$; two-sample Welch's t-test assuming unequal variances.

1052 **Figure 2: Aberrant expression of splicing regulators in PTHS cortical and subpallial**
1053 **organoids. (a)** Uniform Manifold Approximation and Projection (UMAP) bidimensional reduction
1054 of scRNA-Seq profiling of cortical (CtO) and subpallial (sPO) organoids at 8 weeks in vitro, with
1055 pooled cells from 15 organoids per library. Color code represents 6 annotated subpopulations:
1056 Pr-Glut: neural progenitors in glutamatergic lineage; IP-Glut: intermediate progenitors in
1057 glutamatergic lineage; N-Glut: glutamatergic neurons; Pr-GABA: neural progenitors in inhibitory
1058 lineage; IP-GABA: intermediate progenitors in inhibitory lineage; N-GABA: neuronal population
1059 containing GABAergic interneurons. Minority subpopulations exist but were not the focus of our
1060 study. **(b)** UMAP of annotated cells in CtO and sPO organoids, separated according to organoid
1061 type, showing prevalence of excitatory neuron lineage in CtOs and preponderance of
1062 GABAergic interneurons in sPO libraries. **(c-j)** UMAPs showing expression of *NOVA1* (c,d),
1063 *RBFOX1* (e,f), *RBFOX2* (g,h), and *RBFOX3* (i,j), in control and PTHS CtOs and sPOs. Warmer
1064 colors indicate higher expression levels for each gene, while black indicates absence of
1065 detectable expression.

1066 **Figure 3: Altered percentage of cells expressing splicing regulators in PTHS organoids**.
1067 **(a-d)** Left graphs in each group: Violin plots showing comparison of expression of *NOVA1* (a),
1068 *RBFOX1* (b), *RBFOX2* (c), and *RBFOX3* (d) between N-Glut subpopulations in the parent and
1069 PTHS CtOs (left groups) or between N-GABA subpopulations in the parent and PTHS sPOs
1070 (right groups). N = 1401 (parent) or 380 (PTHS) N-Glut cells, or N = 2661 (parent) and 988
1071 (PTHS) N-GABA cells. Right graphs in each group: changes in percentages of neuronal
1072 subtypes in the analyzed subpopulations of CtOs and sPOs (cells expressing each gene above
1073 threshold corresponding to 0.5 log-normalized expression levels).

1074 **Figure 4: Genes exhibiting aberrant splicing patterns in PTHS are involved in synapse**
1075 **formation and cell-cell communication. (a)** Pie charts representing the distribution of splicing
1076 events in 5 alternative splicing categories, in several pairs of patient and control samples. **(b)**
1077 Pie chart indicating the number of events in each alternative splicing category for the common
1078 splicing events across all patient-control pairs. **(c)** Pie chart representing the number of genes
1079 for which splicing events were detected in each alternative splicing category, in the intersection
1080 across all patient-control pairs. **(d,e)** Enrichment ratios observed in Over-Representation
1081 Analysis (ORA) of splicing events with an absolute value of DPSI larger than 0.3 in the patient-
1082 control #2 pair, subjected to either Gene Ontology – Biological Process (d) or KEGG Pathway
1083 analysis (e). Only the 10 categories with the largest enrichment ratios are shown. The color of
1084 each bar represents the false discovery rate (FDR) and the category names indicated in blue
1085 represent those linked to synapse formation and function or cell-cell communication. **(f.g)**
1086 Enrichment ratios observed in similar analyses conducted in d,e but with Gene-Set Enrichment
1087 Analysis (GSEA).

1088 **Figure 5: Enrichment of binding sites for splicing regulators NOVA1 and RBFOX1/2/3 in**
1089 **genes with altered splicing patterns in PTHS. (a)** Results from site enrichment analysis
1090 conducted on the sequences of skipped exons and neighboring introns for splicing events
1091 detected by rMATS-turbo in PTHS neurons, for RNA-binding proteins (RBP) NOVA1 and
1092 RBFOX1/2/3, with two motifs for NOVA1, one motif for RBFOX1 and one common motif for
1093 RBFOX2 and RBFOX3 (motif logos in the middle column). The statistical results from Fisher's
1094 exact test (*E*-value) in the comparison between the test dataset (sequences from splicing events
1095 in PTHS) and genomic sequences from a random set of genes are found in boldface text. For
1096 contrast, the *E*-values for the comparison between a positive control dataset (sequences of
1097 genes known to be regulated by the corresponding RBP) and the random control dataset are
1098 given in parentheses, and the *E*-values for the comparison between a negative control dataset
1099 (sequences of genes not known to be regulated by the corresponding RBP) and the random
1100 control dataset are given between <> symbols. **(b,c)** Examples of genomic sequence structure
1101 for two genes in our aberrant splicing dataset detected by rMATS-turbo, including the *SLC4A10*
1102 gene, shown to be enriched for NOVA1 binding sites (b) and the *CAMTA1* gene, shown to be
1103 enriched for RBFOX1/2/3 binding sites (c). The exons are shown as blue rectangles, connected
1104 by lines representing the introns. The binding sites are shown as black bars above the gene
1105 representation, and the chromosome numbers and genomic coordinate are also displayed. **(d)**
1106 Representation of NOVA1 and RBFOX1/2/3 binding sites in a random gene (MANBAL) not
1107 present in the aberrant splicing event list from PTHS neurons. **(e)** Violin plots representing the
1108 statistical comparison of the numbers of NOVA1 (left graph) and RBFOX1/2/3 (right graph)
1109 sites, in 100,000 nt, in the genes with the highest positive values of ∆PSI from the rMATS-turbo
1110 analysis of PTHS neurons (n=11 genes) compared with a random set of genes not present in
1111 the splicing event list in PTHS (n=12 genes). The median/interquartile range are shown as
1112 horizontal bars inside the violins. Welch's unpaired t-test assuming unequal variances; * $p<0.05$;
1113 ns, not statistically significantly different. **(f-h)** Venn diagrams representing the comparison
1114 between the number of genes with differential splicing events in PTHS neurons, as detected by
1115 rMATS-turbo in RNA-Seq data (absolute values of ∆PSI larger than 0.1), and the number of
1116 genes whose mRNA was shown in the literature to bind NOVA1 (f; ref [46]), NOVA1 (g; ref [63]), and
1117 RBFOX1 (h; ref [1]), in HITS-CliP experiments. The numbers of genes in each sector are shown
1118 inside the Venn diagrams, and some genes involved in neurodevelopment are highlighted in
1119 each intersection. Genes highlighted in blue are those present in the SFARI database of autism-
1120 related genes[78].

1121 **Figure 6: Aberrant splicing in Neurexin genes in PTHS. (a)** MA plot representing the
1122 differential splicing event prevalence, in the comparison between the PTHS patient-derived
1123 neurons and control neurons cultivated in vitro. The *x*-axis represents the expression
1124 abundance for each gene (represented as dots), in transcripts per million (TPM), while the *y*-
1125 axis shows the difference between the PSI (percent spliced in) for the patient cells and the
1126 control neurons (ΔPSI). Each dot represents a gene. Only a fraction of the genes is shown, to
1127 highlight the region of the graph where we observed a splicing event in the *NRXN2* gene
1128 chosen for further analysis. Blue dots indicate splicing events with a statistically significantly
1129 higher usage in control neurons, involving the number of genes displayed in parenthesis. The
1130 red dots indicate splicing events with higher usage in patient neurons. Gray dots are splicing
1131 events with a non-statistically significant difference in usage between patient and control cells.
1132 **(b)** Sashimi plots representing the splicing patterns for the NRXN2 splicing event highlighted in
1133 (a), depicting the location of exons 6, 7, and 8 in chromosome 11 (genomic coordinates below
1134 the plots). The vertical bars indicate the coverage of RNA-Seq reads in each exon or supporting
1135 the existence of each splice junction, in a *y*-axis scale shown as RPKM (reads per kilobase per
1136 million total reads). The cusp lines connecting the exons above or below the vertical bars
1137 indicate reads that support the existence of each corresponding exon-exon junction, with full
1138 lines representing junctions that include SE (skipped exon) 7, while dashed lines display
1139 functions that exclude said exon. The cusp thickness is proportional to the number of reads
1140 supporting the existence of each splice junction. The Sashimi plot on top (purple) represents the
1141 splicing pattern in control neurons, and the plot at the bottom (red) indicates the pattern in PTHS
1142 neurons, with the PSI values indicated above each graph. **(c)** Schematic representation of a
1143 fraction of the genomic sequence of NRXN2 gene, showing exons 6, 7, and 8. Three transcripts
1144 are shown: one represents the transcript that excludes exon 7 (SE-NRXN2-e), as detected by
1145 our rMATS-turbo analysis of PTHS neurons; the second represents the transcript including exon
1146 7 (SE-NRNX2-i), which also corresponds to a transcript in the Ensembl database (ENST entry
1147 number shown); the transcript at the bottom is another Ensembl transcript that includes a
1148 slightly longer version of exon 7. The amino acid translations for each transcript are shown
1149 below the DNA sequences. **(d)** Similar to (a), but for an event involving the *NRXN3* gene,
1150 chosen for further analyses. **(e)** Similar to (b), but for the event in (d), involving exons 20 and 21
1151 of the canonical transcript in the Ensembl database, and an additional exon not annotated in the
1152 canonical transcript (exon 20.5). **(f)** Similar to (c), showing two transcripts of the *NRXN3* gene,
1153 either including (SE-NRXN3-i) or excluding (SE-NRXN3-e) exon 20.5. **(g,h)** Genomic sequence
1154 structure for NRNX2 (g) and NRNX3 (h), showing the location of NOVA1 and RBFOX1/2/3

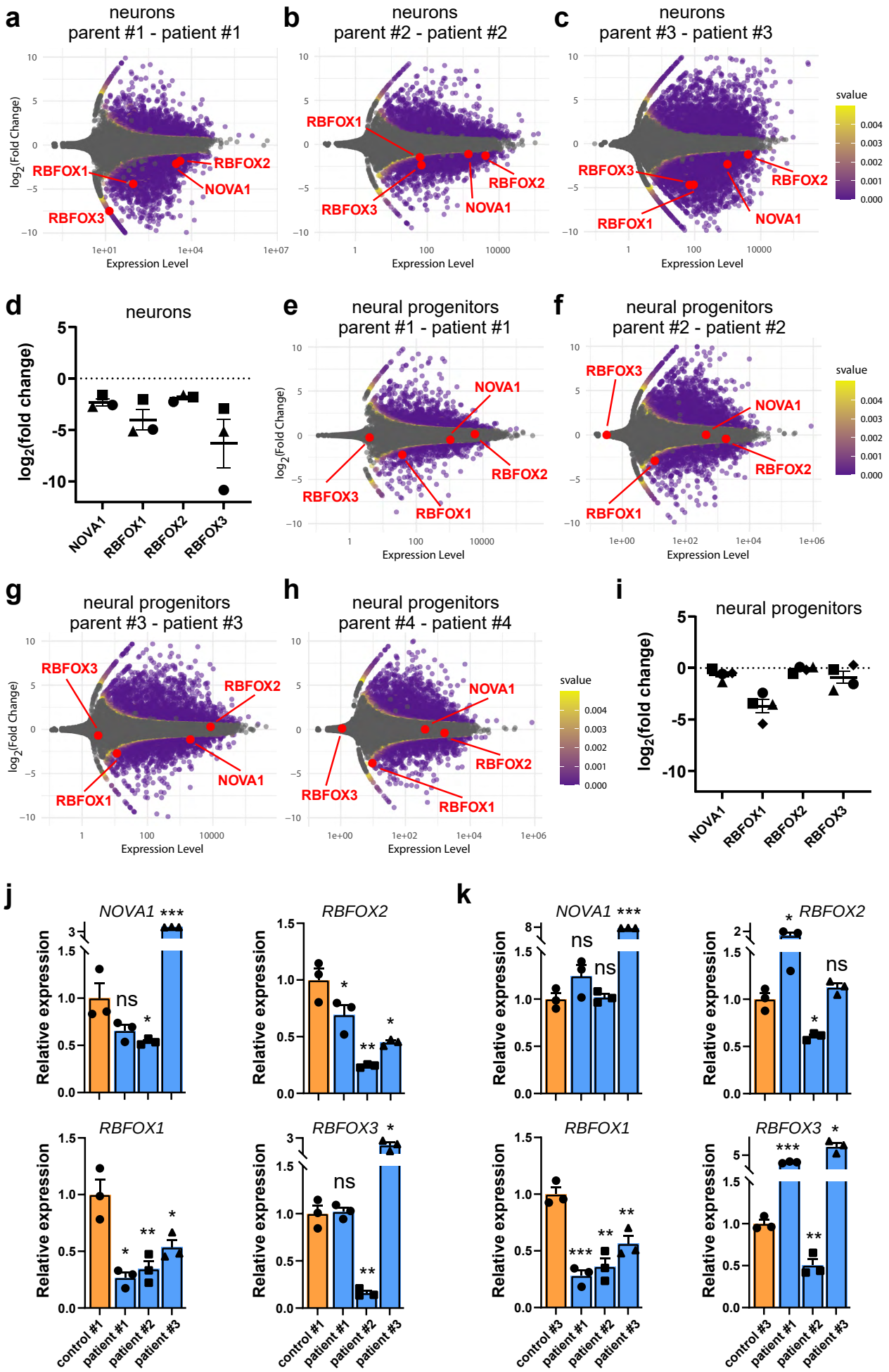
1155 binding sites. The exons are shown as blue rectangles (exon numbers below the gene
1156 structure), connected by lines representing the introns. The binding sites are shown as black
1157 bars above the gene representation, and the chromosome numbers and genomic coordinate
1158 are also displayed. The region containing the exons involved in the SE events are blown up
1159 below. The models at the bottom schematically depict the exons (to scale) and the splicing
1160 patterns in the SE-NRXN2/3-i and SE-NRXN2/3-e transcripts.

1161 **Figure 7: Changes in the abundances of transcript variants via exon skipping at the**
1162 **Neurexin-3 gene *NRXN3* in PTHS neurons in vitro. (a)** Schematic representation of the
1163 location of PCR primers in exons of the *NRXN3* gene, used in real-time qPCR experiments.
1164 Some oligos were designed to detect all transcript variants (left), because they detect a
1165 combination of exons present in both the transcript including (SE-NRXN3-i) and excluding (SE-
1166 NRXN3-e) exon 20.5. Other PCR oligos (right) were designed as variant-specific primers,
1167 because they recognize exon-exon splice junctions typical of either the SE-NRXN3-I or SE-
1168 NRXN3-e transcripts. **(b)** Real-time qPCR results showing the expression of *NRXN3* with the
1169 combination of oligos p1 and p2, which recognize both transcript variants. Each graph
1170 compares the expression in three patient lines (blue bars) and control line #1 (orange bar on the
1171 left graph) or control line #3 (orange bar on the right graph). Error bars are S.E.M. Each symbol
1172 represents a patient line, according to Supplementary Table S6. The expression is each graph
1173 is normalized to 1 in the control line, and the statistical comparisons (Welch's unpaired t-test
1174 assuming unequal variances) are between each patient line and the control shown in each
1175 graph. * $p<0.05$; ** $p<0.01$; ns, non-statistically significantly different. N=3 biological replicates
1176 per group. **(c)** Left: Bar plots representing the expression abundances of NRXN3 in control
1177 (orange) and PTHS (blue) neuronal cell lines, as judged by transcripts per million (TPM) in
1178 RNA-Seq data. Each symbol represents a patient or control. The control group has 3 subjects,
1179 while the PTHS group has 4 subjects; in three cases, the PTHS patient lines are matched with
1180 the corresponding parental control lines, and a fourth PTHS neuronal line is not matched to
1181 controls. Error bars are S.E.M. Notice the tendency for lower mean expression in the PTHS
1182 group, but the difference fell below the statistical significance threshold level, due to the low
1183 expression values in patient-parent pair #1 (circle symbols). Right: Direct comparison between
1184 the TPM expression abundances in three patient-parent pairs, represented as fold-change in
1185 the $\log_2$ scale. Error bars are S.E.M. **(d)** qPCR results with variant-specific oligos to detect the
1186 SE-NRXN3-e transcript (p1 and p3) or the SE-NRXN3-i (p4 and p5). The cell lines and
1187 comparisons are the same as in (c). Welch's unpaired t-test assuming unequal variances; *
1188 $p<0.05$; ** $p<0.01$. **(e)** Ratio between the relative expression of SE-NRXN3-i and SE-NRXN3-e
1189 transcript abundances, produced by contrasting the RQ (relative quantitation) levels with oligos
1190 p4+p5 with those obtained with primers p1+p3. In each graph, the mean ratio was normalized to
1191 1 in either control #1 (left) or control #3 (right) neuronal cell lines. N=3 independent biological
1192 replicates. Notice the change in the balance between the two transcript variants in two patient
1193 lines, with a reduction in the abundance of the transcript containing exon 20.5, as indicated by
1194 the mean ratio below 1 (dashed line). (f) Schematic representation of the location of primers for

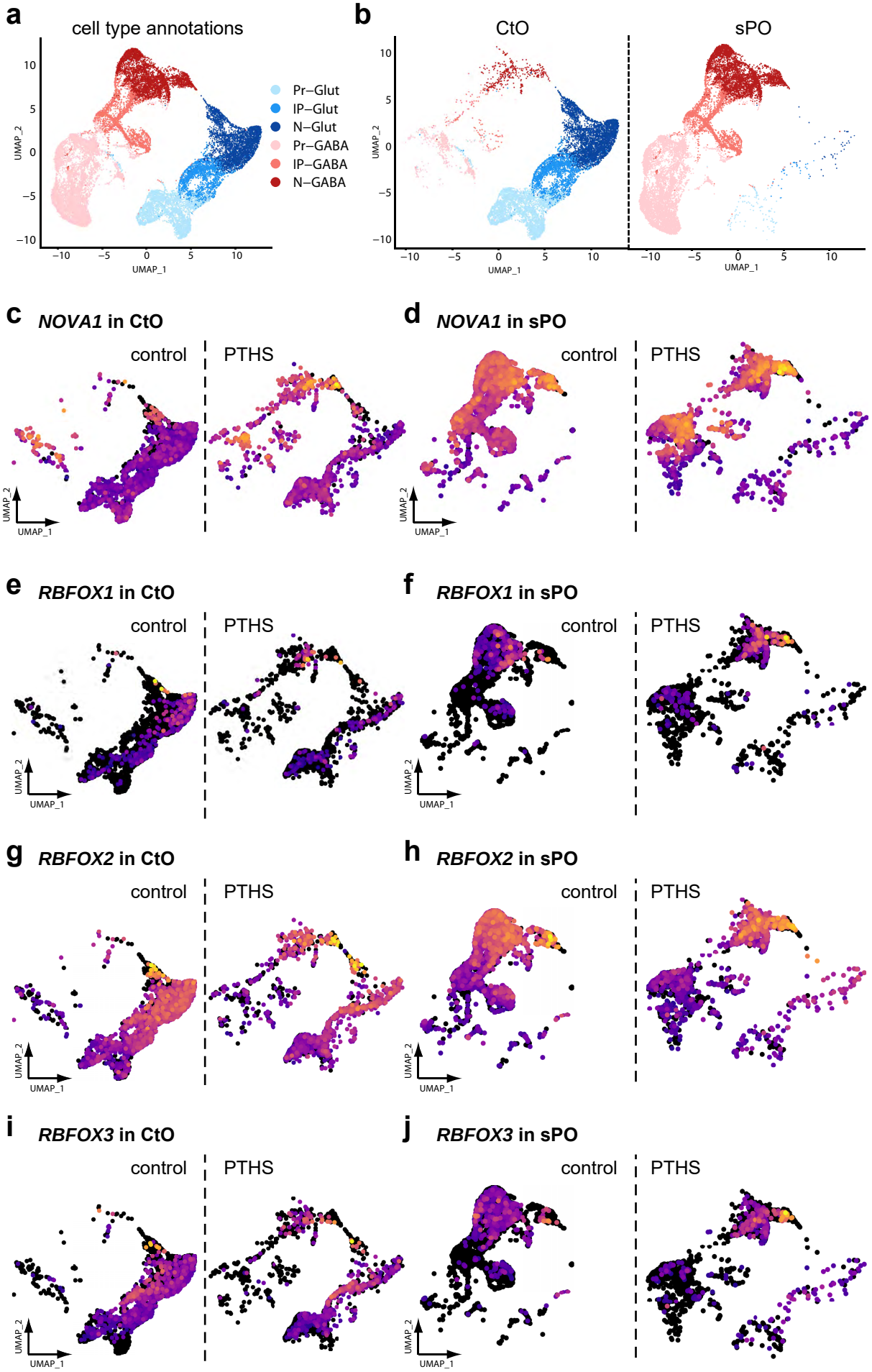
1195 regular PCR amplification of specific *NRXN3* gene transcripts: oligos p6+p7 were designed to
1196 amplify just the SE-NRNX3-e variant, as p7 recognizes exon 21; oligos p6+p8 on the other hand
1197 were designed to amplify just the SE-NRNX3-i variant, as p8 recognizes the skipped exon 20.5
1198 (yellow). (g) Image of the results of gel electrophoresis showing amplification of several bands
1199 with the p6+p7 and p6+p8 primer combinations, with samples from patient line #2 and control
1200 line #1. L: Kasvi 1 kb DNA ladder. The open triangle indicates the 252-bp band expected from
1201 the SE-NRXN3-i transcript, while the filled triangle shows the 334-bp band expected from the
1202 SE-NRXN3-e transcript. Notice the almost complete absence of the band from the SE-NRXN3-i
1203 transcript in PTHS samples.

1204 **Figure 8: Evidence that the transcript variants of *NRXN3* encode transmembrane and**
1205 **secreted Neurexin-3 protein isoforms. (a)** Location of protein domains laminin G and EGF in
1206 the Neurexin-3 protein isoforms encoded by the SE-NRNX3-e and SE-NRNX3-i variants, as
1207 detected by InterProScan, showing the shorter length of the protein predicted to be produced
1208 from the variant included exon 20.5, due to the presence of a premature stop codon**. (b)**
1209 Schematic representation showing the coding sequences in transcripts SE-NRNX3-e and SE-
1210 NRNX3-i (hatched pattern), showing a detail of the premature stop codon introduced when exon
1211 20.5 is spliced into the transcript. Exon sizes are shown to scale, except for exon 21 in the
1212 bottom transcript, which is shown partially. **(c)** Prediction of protein topology using the
1213 DeepTMHMM tool on the proteins encoded by the SE-NRNX3-e (top) and SE-NRNX3-i (bottom)
1214 transcripts. The protein topology in each case is shown above the graph, color-coded according
1215 to the prediction of extracellular, intracellular, or transmembrane domains. The *y*-axis in the
1216 graph represents the probability of each portion of the protein being extracellular (orange),
1217 transmembrane (purple), or intracellular (blue). **(d)** Prediction of signal peptide and probability of
1218 secretion with the SignalP tool, on the first 70 amino acid residues in the protein sequences
1219 predicted from the SE-NRNX3-e and SE-NRNX3-i transcripts. Each line indicates the probability
1220 of the existence of a signal peptide (sigP) having features of N-terminal, central, or C-terminal
1221 portions of the signal peptide, while the dashed line is the remainder of the probability.

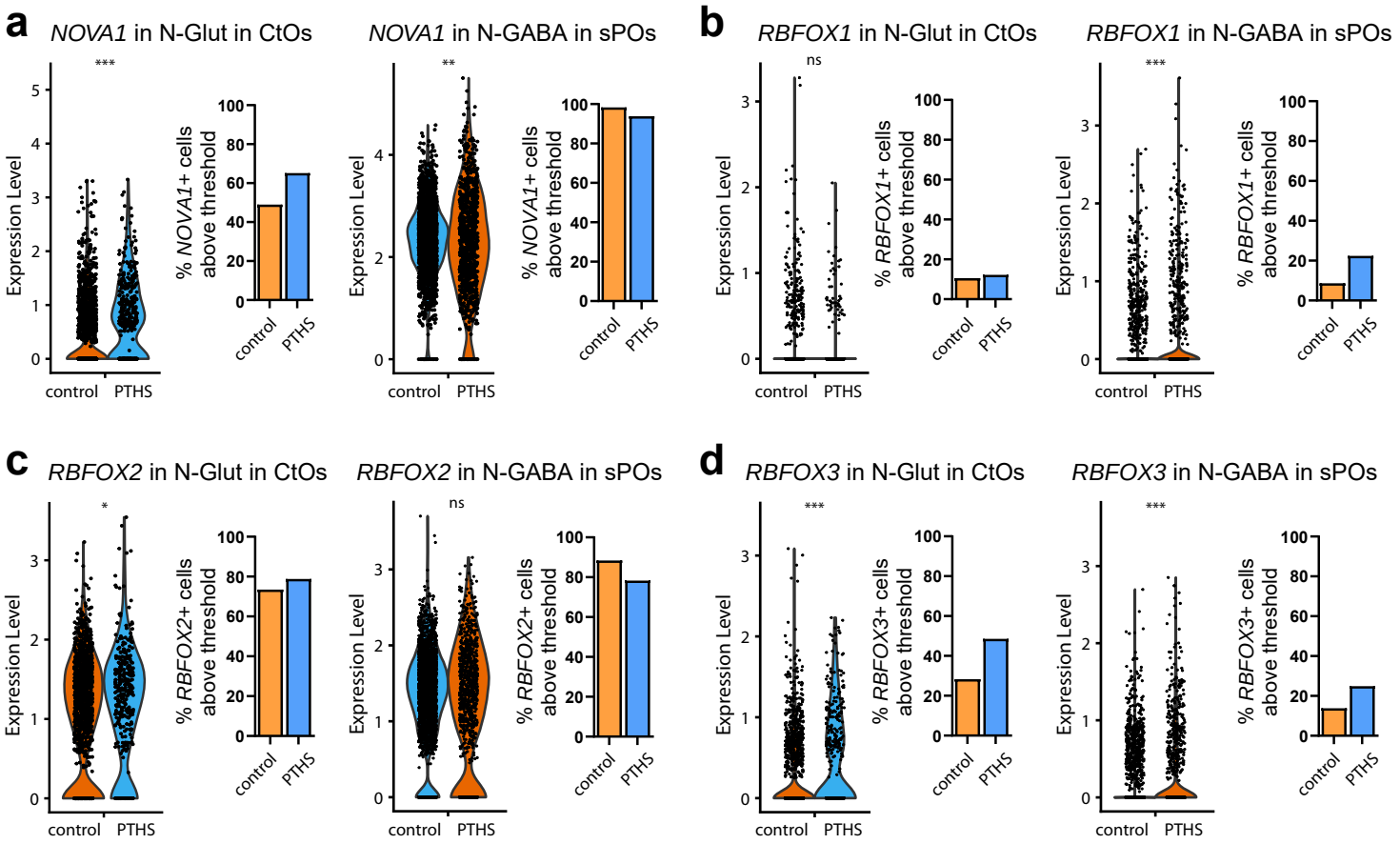
# Figure 1


a
neurons
parent #1 - patient #1
b
neurons
parent #2 - patient #2
c
neurons
parent #3 - patient #3
svalue
0.004
0.003
0.002
0.001
0.000
log2(Fold Change)
Expression Level
RBFOX1
RBFOX2
RBFOX3
NOVA1
d
neurons
log2(fold change)
NOVA1
RBFOX1
RBFOX2
RBFOX3
e
neural progenitors
parent #1 - patient #1
f
neural progenitors
parent #2 - patient #2
g
neural progenitors
parent #3 - patient #3
h
neural progenitors
parent #4 - patient #4
i
neural progenitors
j
NOVA1 ***
RBFOX2
RBFOX1
RBFOX3 *
Relative expression
ns
control #1
patient #1
patient #2
patient #3
k
NOVA1 ***
RBFOX2
RBFOX1
RBFOX3 *
control #3
patient #1
patient #2
patient #3

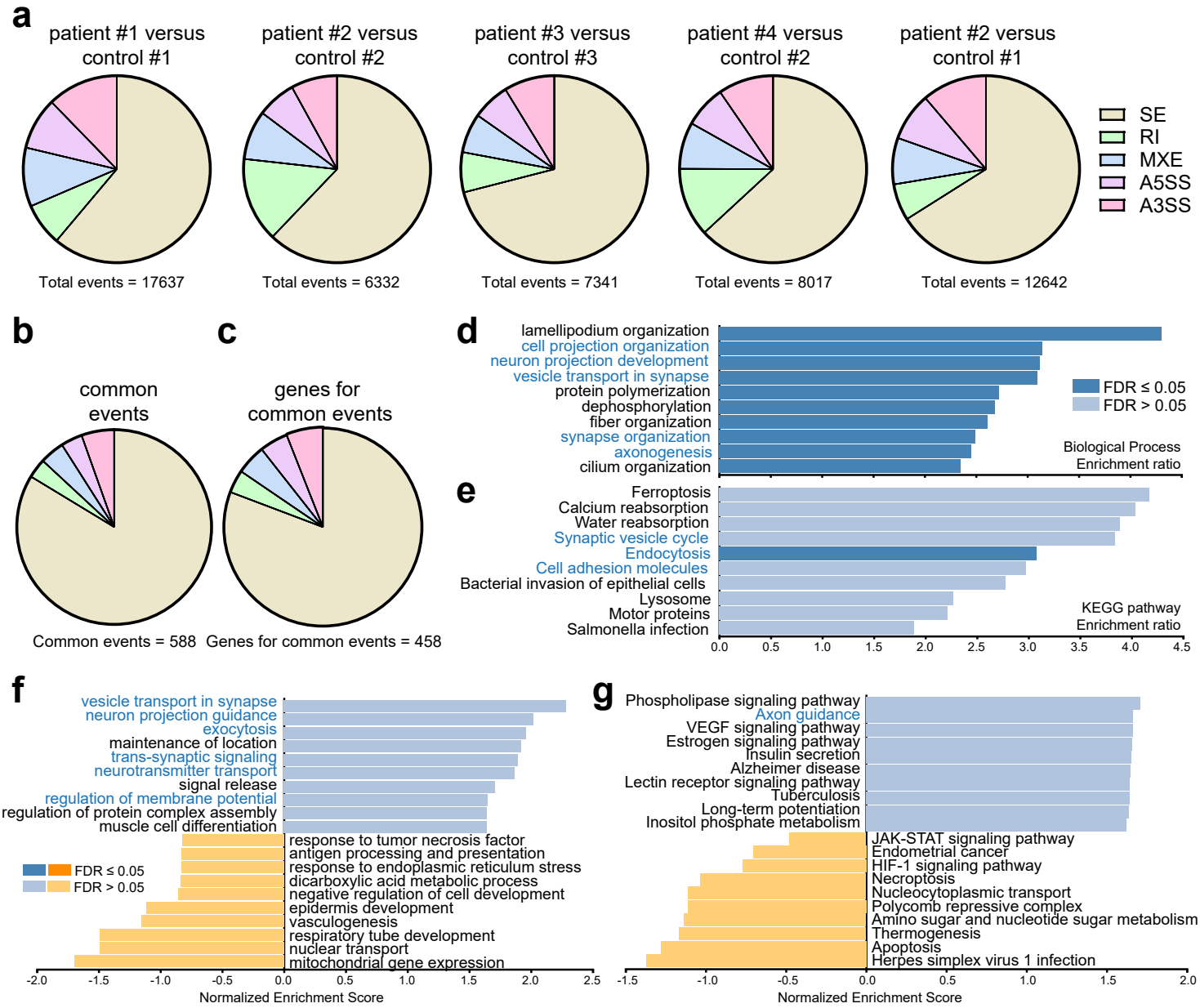
# Figure 2


a
cell type annotations
Pr–Glut
IP–Glut
N–Glut
Pr–GABA
IP–GABA
N–GABA
UMAP_1
UMAP_2
b
CtO
sPO
c NOVA1 in CtO
control
PTHS
d NOVA1 in sPO
e RBFOX1 in CtO
f RBFOX1 in sPO
g RBFOX2 in CtO
h RBFOX2 in sPO
i RBFOX3 in CtO
j RBFOX3 in sPO

# Figure 3


a
NOVA1 in N-Glut in CtOs
NOVA1 in N-GABA in sPOs
b
RBFOX1 in N-Glut in CtOs
RBFOX1 in N-GABA in sPOs
c
RBFOX2 in N-Glut in CtOs
RBFOX2 in N-GABA in sPOs
d
RBFOX3 in N-Glut in CtOs
RBFOX3 in N-GABA in sPOs
Expression Level
% NOVA1+ cells above threshold
% RBFOX1+ cells above threshold
% RBFOX2+ cells above threshold
% RBFOX3+ cells above threshold
control
PTHS
***
**
ns
*

# Figure 4

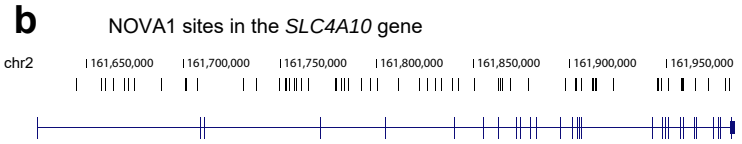

a
patient #1 versus control #1
patient #2 versus control #2
patient #3 versus control #3
patient #4 versus control #2
patient #2 versus control #1
SE
RI
MXE
A5SS
A3SS
Total events = 17637
Total events = 6332
Total events = 7341
Total events = 8017
Total events = 12642
b
common events
Common events = 588
c
genes for common events
Genes for common events = 458
d
lamellipodium organization
cell projection organization
neuron projection development
vesicle transport in synapse
protein polymerization
dephosphorylation
fiber organization
synapse organization
axonogenesis
cilium organization
FDR ≤ 0.05
FDR > 0.05
Biological Process
Enrichment ratio
e
Ferroptosis
Calcium reabsorption
Water reabsorption
Synaptic vesicle cycle
Endocytosis
Cell adhesion molecules
Bacterial invasion of epithelial cells
Lysosome
Motor proteins
Salmonella infection
KEGG pathway
Enrichment ratio
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5
f
vesicle transport in synapse
neuron projection guidance
exocytosis
maintenance of location
trans-synaptic signaling
neurotransmitter transport
signal release
regulation of membrane potential
regulation of protein complex assembly
muscle cell differentiation
response to tumor necrosis factor
antigen processing and presentation
response to endoplasmic reticulum stress
dicarboxylic acid metabolic process
negative regulation of cell development
epidermis development
vasculogenesis
respiratory tube development
nuclear transport
mitochondrial gene expression
FDR ≤ 0.05
FDR > 0.05
-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0 2.5
Normalized Enrichment Score
g
Phospholipase signaling pathway
Axon guidance
VEGF signaling pathway
Estrogen signaling pathway
Insulin secretion
Alzheimer disease
Lectin receptor signaling pathway
Tuberculosis
Long-term potentiation
Inositol phosphate metabolism
JAK-STAT signaling pathway
Endometrial cancer
HIF-1 signaling pathway
Necroptosis
Nucleocytoplasmic transport
Polycomb repressive complex
Amino sugar and nucleotide sugar metabolism
Thermogenesis
Apoptosis
Herpes simplex virus 1 infection
-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0
Normalized Enrichment Score

# Figure 5

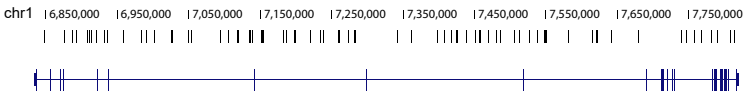
**a**

| RBP | motif logo | site enrichment *E*-value test against random set (pos. ctrl against random set) <neg. ctrl against random set> |
|---|---|---|
| NOVA1 motif #1 | | **9.2 x E-08** (2.9 x E-01) <2.2 x E-00> |
| NOVA1 motif #2 | | **3.0 x E-08** (1.6 x E-06) <5.9 x E-01> |
| RBFOX1 | | **5.8 x E-04** (8.4 x E-08) <3.2 x E-00> |
| RBFOX2/3 | | **4.1 x E-07** (9.2 x E-16) <2.4 x E-00> |

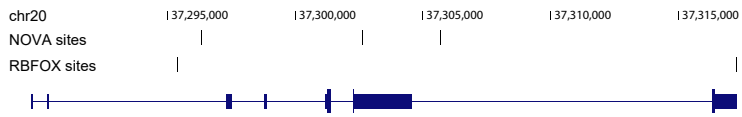
**b**



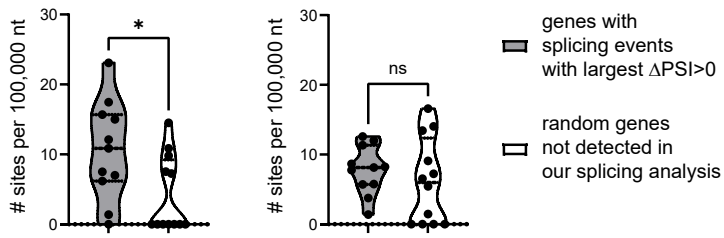

**c**



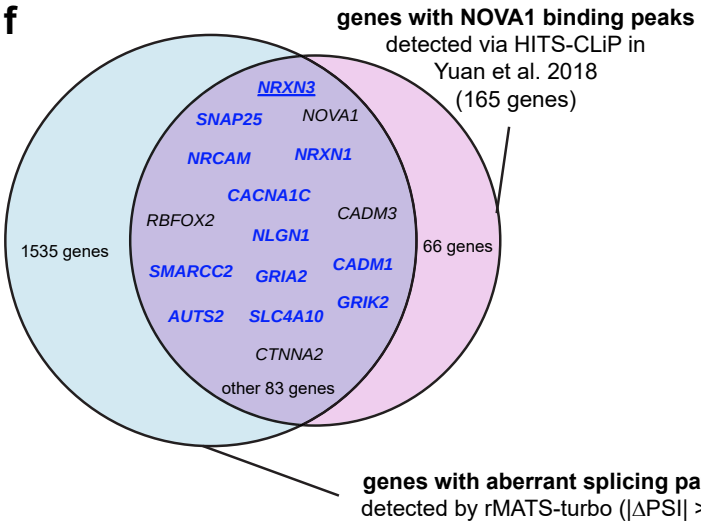

**d**



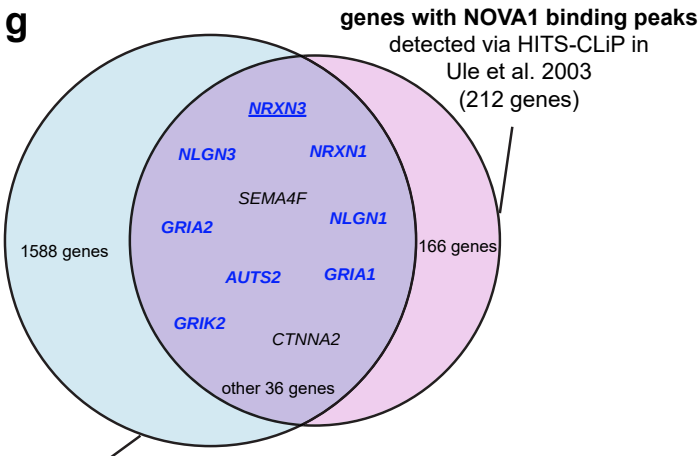

**e**




**f**

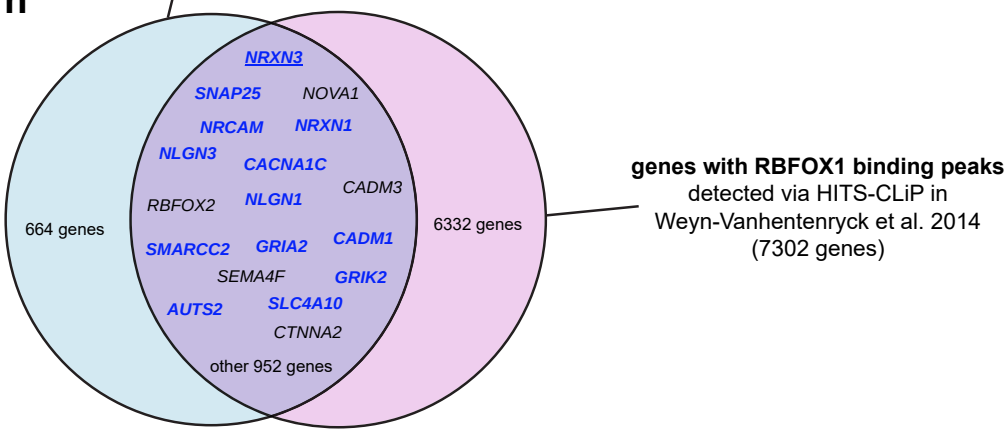



**g**

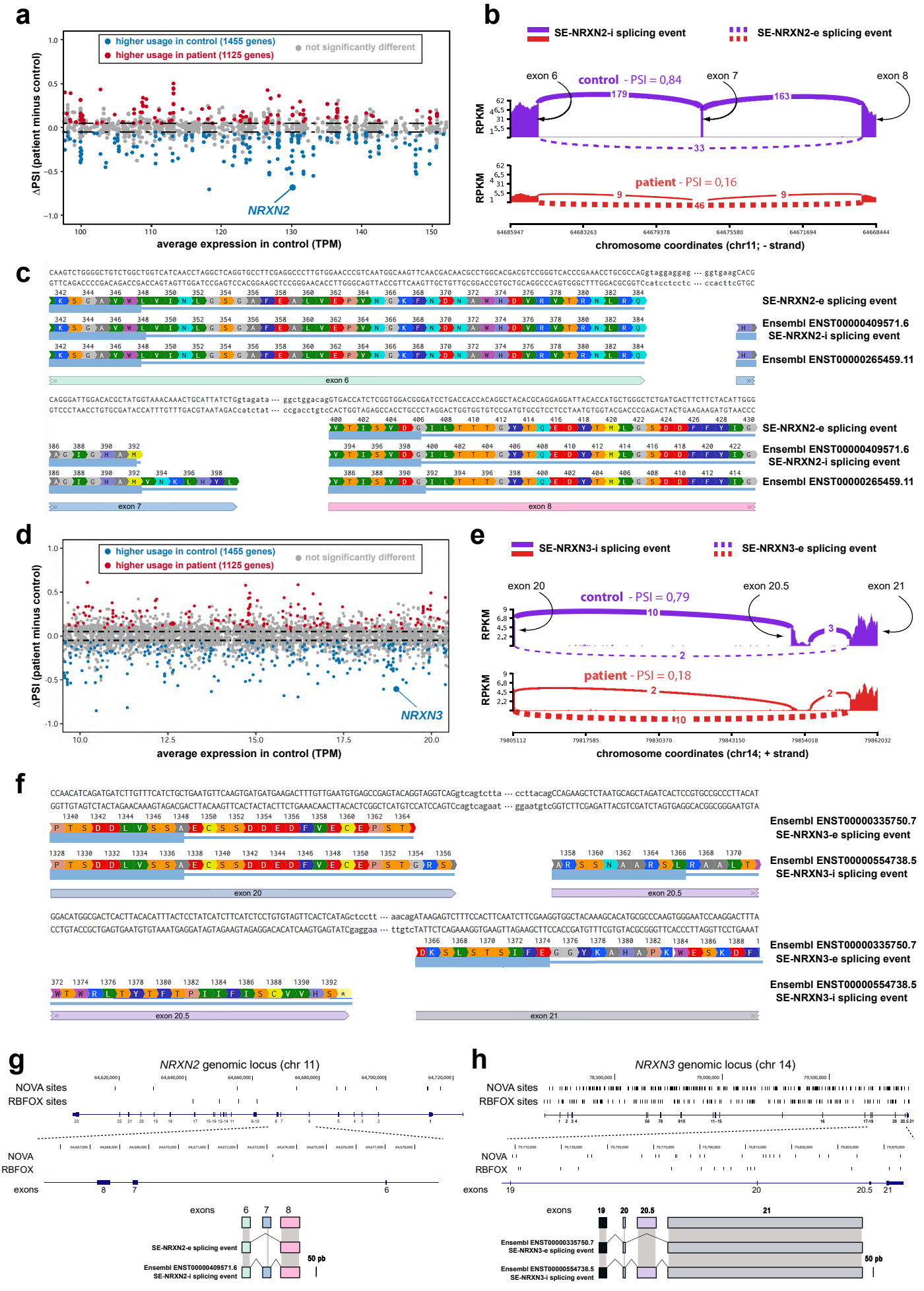



**genes with aberrant splicing pattern** detected by rMATS-turbo (|ΔPSI| > 0.1) (1634 genes)

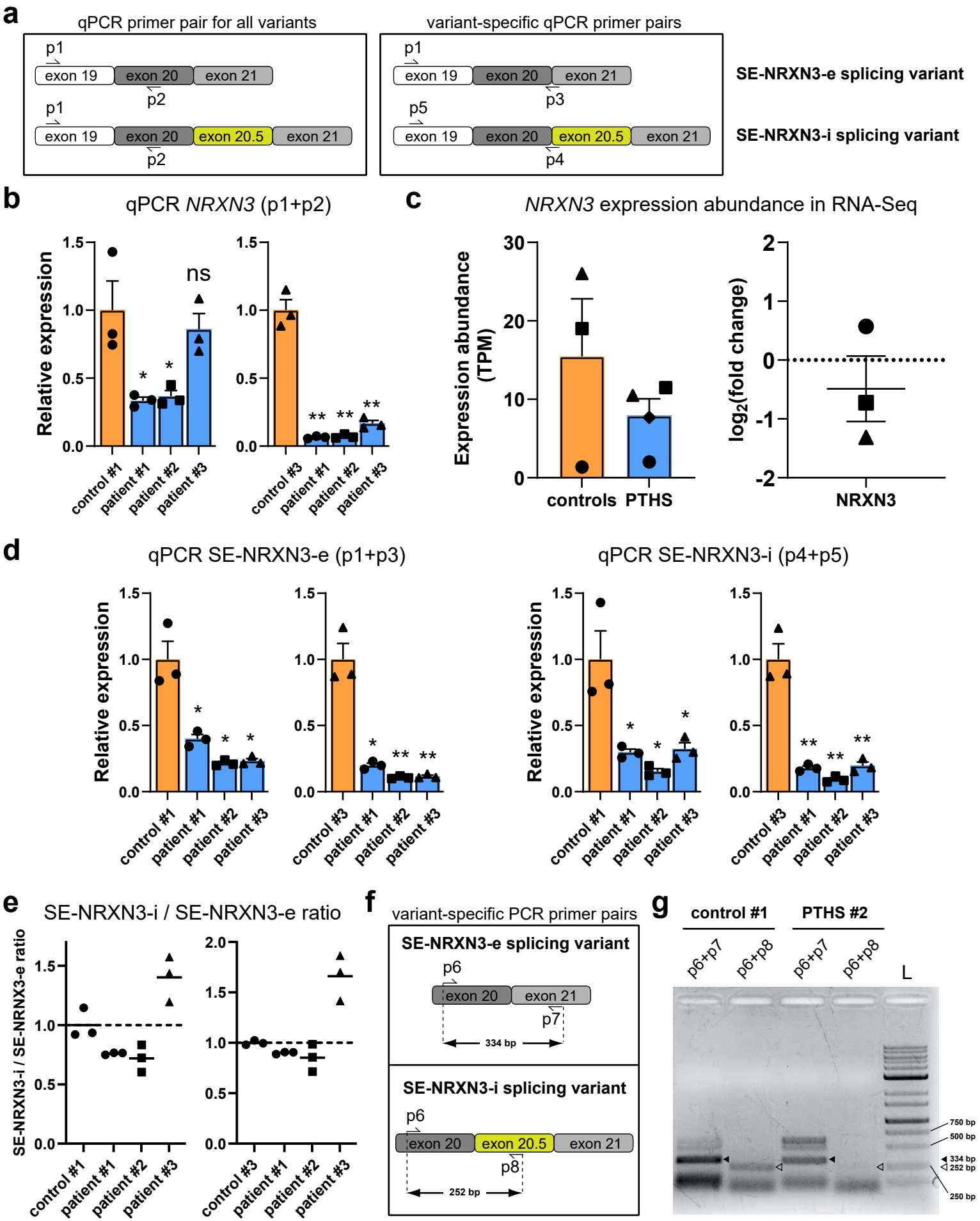
**h**



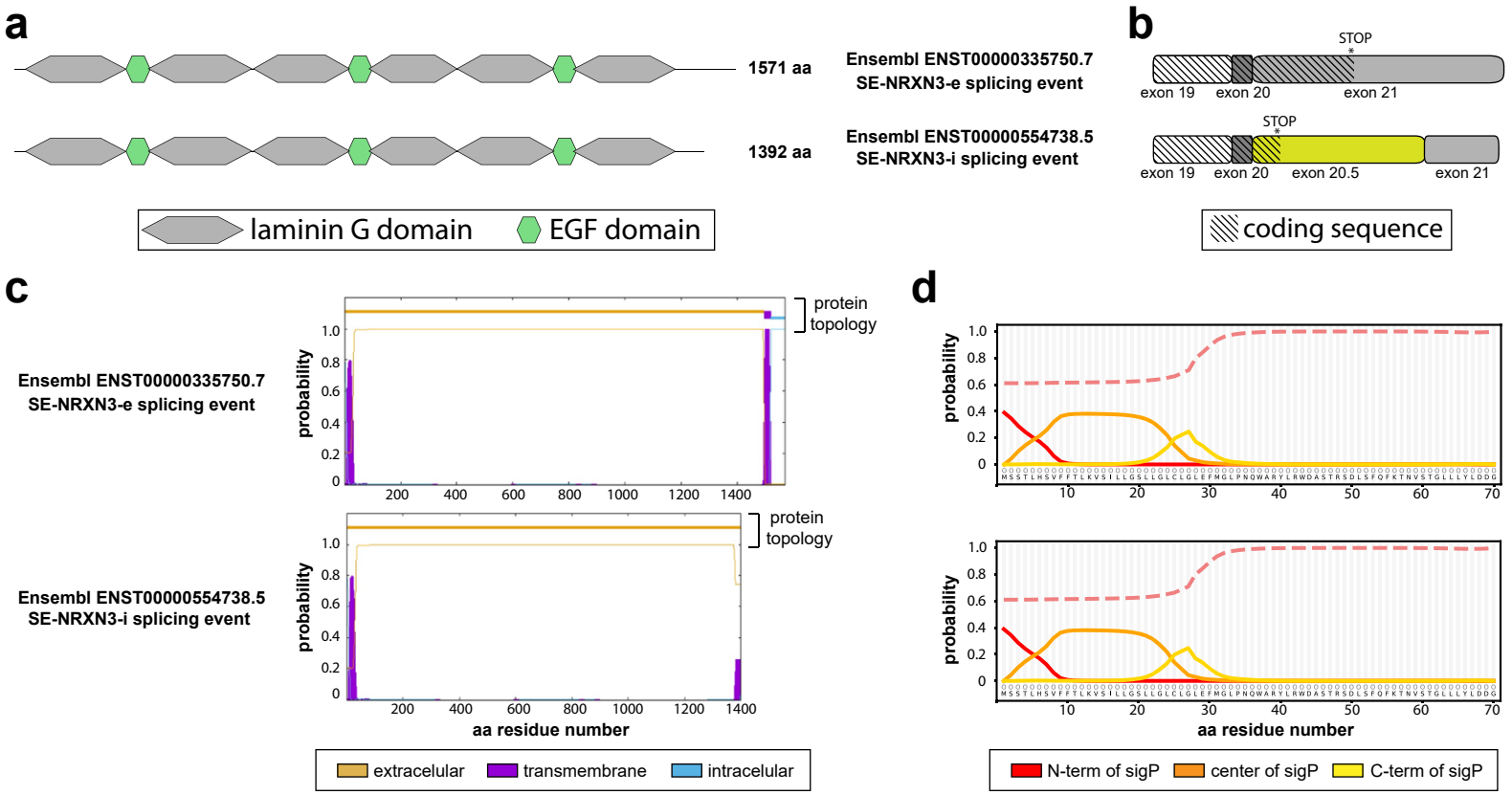
# Figure 6


a
higher usage in control (1455 genes)
higher usage in patient (1125 genes)
not significantly different
ΔPSI (patient minus control)
NRXN2
average expression in control (TPM)
b
SE-NRXN2-i splicing event
SE-NRXN2-e splicing event
exon 6
control - PSI = 0,84
exon 7
exon 8
RPKM
patient - PSI = 0,16
chromosome coordinates (chr11; - strand)
c
SE-NRXN2-e splicing event
Ensembl ENST00000409571.6
SE-NRXN2-i splicing event
Ensembl ENST00000265459.11
exon 6
exon 7
exon 8
d
higher usage in control (1455 genes)
higher usage in patient (1125 genes)
not significantly different
ΔPSI (patient minus control)
NRXN3
average expression in control (TPM)
e
SE-NRXN3-i splicing event
SE-NRXN3-e splicing event
exon 20
control - PSI = 0,79
exon 20.5
exon 21
RPKM
patient - PSI = 0,18
chromosome coordinates (chr14; + strand)
f
Ensembl ENST00000335750.7
SE-NRXN3-e splicing event
Ensembl ENST00000554738.5
SE-NRXN3-i splicing event
exon 20
exon 20.5
exon 21
g
NRXN2 genomic locus (chr 11)
NOVA sites
RBFOX sites
NOVA
RBFOX
exons
SE-NRXN2-e splicing event
Ensembl ENST00000409571.6
SE-NRXN2-i splicing event
50 pb
h
NRXN3 genomic locus (chr 14)
NOVA sites
RBFOX sites
NOVA
RBFOX
exons
Ensembl ENST00000335750.7
SE-NRXN3-e splicing event
Ensembl ENST00000554738.5
SE-NRXN3-i splicing event
50 pb

# Figure 7


a
qPCR primer pair for all variants
variant-specific qPCR primer pairs
p1
exon 19
exon 20
exon 21
p2
exon 20.5
p5
p3
p4
SE-NRXN3-e splicing variant
SE-NRXN3-i splicing variant
b
qPCR NRXN3 (p1+p2)
Relative expression
ns
control #1
patient #1
patient #2
patient #3
control #3
c
NRXN3 expression abundance in RNA-Seq
Expression abundance (TPM)
controls
PTHS
log2(fold change)
NRXN3
d
qPCR SE-NRXN3-e (p1+p3)
qPCR SE-NRXN3-i (p4+p5)
e
SE-NRXN3-i / SE-NRXN3-e ratio
f
variant-specific PCR primer pairs
SE-NRXN3-e splicing variant
p6
p7
334 bp
SE-NRXN3-i splicing variant
p8
252 bp
g
control #1
PTHS #2
p6+p7
p6+p8
L
750 bp
500 bp
334 bp
252 bp
250 bp

# Figure 8


a
1571 aa
Ensembl ENST00000335750.7
SE-NRXN3-e splicing event
1392 aa
Ensembl ENST00000554738.5
SE-NRXN3-i splicing event
laminin G domain
EGF domain
b
STOP
exon 19
exon 20
exon 21
STOP
exon 19
exon 20
exon 20.5
exon 21
coding sequence
c
Ensembl ENST00000335750.7
SE-NRXN3-e splicing event
Ensembl ENST00000554738.5
SE-NRXN3-i splicing event
probability
protein topology
aa residue number
extracelular
transmembrane
intracelular
d
probability
aa residue number
N-term of sigP
center of sigP
C-term of sigP