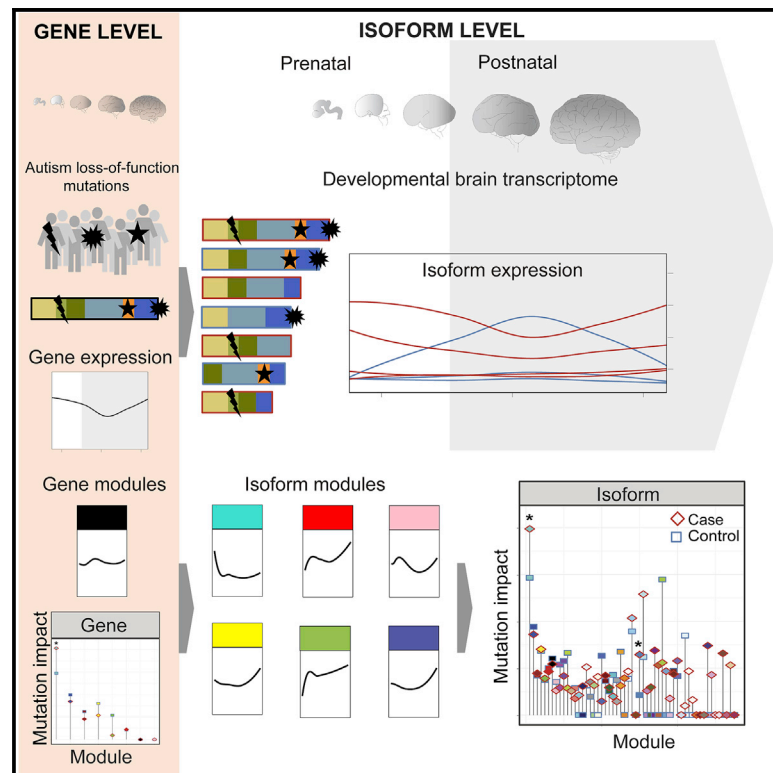


Full-length isoform transcriptome of the developing human brain provides further insights into autism

Graphical abstract



Authors

Kevin K. Chau, Pan Zhang,
Jorge Urresti, ..., Roser Corominas,
Guan Ning Lin, Lilia M. Iakoucheva

Correspondence

lilyak@ucsd.edu

In brief

Alternative splicing plays an important role in brain development. Chau et al. analyze the full-length isoform transcriptome of the developing brain and identify differentially expressed isoforms and isoform co-expression modules undetectable by gene-level analyses. They demonstrate that neurodevelopmental disease mutations affect different splicing isoforms, highlighting the importance of the isoform transcriptome for future studies.

Highlights

- Full-length isoform transcriptome reveals pathways undetectable by gene-level analyses
- Splicing isoforms affected by autism mutations exhibit higher prenatal expression
- Isoform modules with splicing and synaptic functions are enriched in autism mutations
- Functional effect of mutations should be investigated at the full-length isoform level



Article

Full-length isoform transcriptome of the developing human brain provides further insights into autism

Kevin K. Chau,^{1,2} Pan Zhang,^{1,3} Jorge Urresti,¹ Megha Amar,¹ Akula Bala Pramod,¹ Jiaye Chen,¹ Amy Thomas,¹ Roser Corominas,^{1,4} Guan Ning Lin,^{1,5} and Lilia M. Iakoucheva^{1,6,*}

¹Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA

²Department of Biology, University of California, San Diego, La Jolla, CA, USA

³Present address: Department of Psychiatry, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

⁴Present address: Department of Genetics, Microbiology and Statistics, Universitat de Barcelona, Barcelona, Spain

⁵Present address: School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China

⁶Lead contact

*Correspondence: lilyak@ucsd.edu

<https://doi.org/10.1016/j.celrep.2021.109631>

SUMMARY

Alternative splicing plays an important role in brain development, but its global contribution to human neurodevelopmental diseases (NDDs) requires further investigation. Here we examine the relationships between splicing isoform expression in the brain and *de novo* loss-of-function mutations from individuals with NDDs. We analyze the full-length isoform transcriptome of the developing human brain and observe differentially expressed isoforms and isoform co-expression modules undetectable by gene-level analyses. These isoforms are enriched in loss-of-function mutations and microexons, are co-expressed with a unique set of partners, and have higher prenatal expression. We experimentally test the effect of splice-site mutations and demonstrate exon skipping in five NDD risk genes, including *SCN2A*, *DYRK1A*, and *BTRC*. Our results suggest that the splice site mutation in *BTRC* reduces translational efficiency, likely affecting Wnt signaling through impaired degradation of β -catenin. We propose that functional effects of mutations should be investigated at the isoform- rather than gene-level resolution.

INTRODUCTION

More than 95% of multi-exon human genes undergo alternative splicing (AS) and/or use alternative promoters to increase transcriptomic and proteomic diversity, with an estimated average of five to seven isoforms transcribed per gene (Pan et al., 2008; Steijger et al., 2013; Wang et al., 2008). AS is highly specific, and expression of isoforms is often restricted to certain organs, tissues, or cell types (Barbosa-Morais et al., 2012; Sapkota et al., 2019; Shalek et al., 2013; Trapnell et al., 2010). In addition, many isoforms are expressed only during specific developmental periods (Kalsotra and Cooper, 2011). Alternatively spliced isoforms encoded by the same gene can also be expressed at different levels in the same tissue or during the same developmental period (Wang et al., 2008).

The developing human brain has one of the highest frequencies of AS events (Calarco et al., 2011; Melé et al., 2015; Raj and Blencowe, 2015; Yeo et al., 2004). Many of the processes that occur during neural development, including cell fate determination, neuronal migration, axon guidance, and synaptogenesis, are controlled by differentially expressed alternatively spliced isoforms (Grabowski, 2011; Kim et al., 2013; Li

et al., 2007). Several recent studies, including one by us, began to investigate isoform-level transcriptome dysregulation in psychiatric diseases (Gandal et al., 2018; Li et al., 2018; Parikshak et al., 2016). However, spatiotemporal full-length isoform transcriptome of the developing human brain remains relatively unexplored.

Integration of the brain gene transcriptome with genetic data from exome and whole-genome sequencing studies have provided important insights into neurodevelopmental diseases (NDDs) (Li et al., 2018; Lin et al., 2015; Parikshak et al., 2013; Satterstrom et al., 2020; Willsey et al., 2013). Most of the recent work in this area has been focused on understanding the effect of mutations at gene-level resolution, whereas the isoform-specific effect of loss-of-function (LoF) mutations in the context of brain development has not yet been fully investigated.

It is important to map LoF mutations to transcripts because protein isoforms encoded by different transcripts have drastically different protein interaction capabilities. As we have demonstrated previously, the majority of the isoforms encoded by the same gene share less than a half of their interacting partners in the human interactome network (Yang et al., 2016). This observation points to striking functional differences between



splicing isoforms that are not accounted for by the majority of the existing gene-level studies. In addition, our recent work demonstrated that isoform-level networks of autism risk genes and copy number variants provide better resolution and depth around disease proteins (Corominas et al., 2014).

To better understand how NDD risk mutations dysregulate normal brain development, we analyzed the temporal isoform transcriptome of the developing human brain using the BrainSpan RNA-seq dataset from the PsychEncode Consortium (Li et al., 2018), summarized to full-length isoforms, as described previously (Gandal et al., 2018). We identified hundreds of differentially expressed isoforms (DEIs) and dozens of isoform co-expression modules at brain developmental periods starting from fetal to adult. Compared with the gene-level transcriptome, the full-length isoform transcriptome provides more meaningful insights and paints a more complete picture of neurodevelopmental processes. Importantly, many DEIs and isoform co-expression modules were undetectable by gene-level analyses. Mapping autism spectrum disorder (ASD) risk mutations to DEIs revealed that ASD LoF-affected isoforms have higher prenatal expression, more frequently carry microexons, and are preferentially involved in key neuronal processes compared with non-affected isoforms. Furthermore, isoform co-expression modules with splicing-related and synaptic functions were enriched in LoF-affected isoforms, implicating these functions in NDDs. Finally, we experimentally tested the effect of several splice-site LoF mutations and demonstrated that they cause exon skipping to produce novel isoforms with altered biological properties. Our study makes a strong case for investigation of disease mutations at full-length isoform- rather than gene-level resolution.

RESULTS

Construction, quality control, and validation of the full-length isoform transcriptome of the developing human brain

To investigate global patterns of isoform expression across brain development, we analyzed the temporal full-length isoform transcriptome of the developing human brain (Figure S1). We used the BrainSpan RNA-seq dataset from the PsychEncode Consortium (Li et al., 2018), summarized to full-length isoforms, as described previously (Gandal et al., 2018). After rigorous quality control, which included sample outlier detection through weighted gene co-expression network analysis (WGCNA) (Oldham et al., 2012) and detection of confounding variables with surrogate variable analyses (Leek and Storey, 2007; Figures S2 and S3; STAR Methods), we obtained expression profiles for 100,754 unique full-length isoforms corresponding to 26,307 brain-expressed human genes, resulting in ~ 3.8 isoforms/gene.

To experimentally validate the quality of the BrainSpan isoform transcriptome, we used quantitative RT-PCR (qRT-PCR) to estimate the relative expression difference of 44 unique isoforms of 32 genes between two independent RNA samples that were age-, sex-, and brain region-matched to the samples from BrainSpan (STAR Methods). The relative qRT-PCR isoform expression values in independent samples of the frontal lobe of a 22-week-old fetus and the cerebral cortex of a 27-year-old adult were compared with the values obtained from BrainSpan. We

observed a positive correlation ($R^2 = 0.46$) between experimental and BrainSpan-derived values for these isoforms despite using independent samples for validation (Table S1; Figure S4A; STAR Methods). This correlation was comparable with our previous validation values from other PsychEncode datasets ($R^2 = 0.48$ for ASD and $R^2 = 0.51$ for SCZ [schizophrenia]) (Gandal et al., 2018).

Differential isoform expression reveals distinct signals relative to differential gene expression

We recently demonstrated that isoform-level changes capture larger disease effects than gene-level changes in the context of three major psychiatric disorders (Gandal et al., 2018). Here we investigated the role of full-length isoform expression in the context of normal brain development. We performed differential expression analysis among all pairs of adjacent developmental periods as well as between pooled prenatal (P02–P07) and pooled postnatal (P08–P13) (PrePost) samples, yielding sets of differentially expressed genes (DEGs) and DEIs (STAR Methods; Tables S2 and S3). We observed the largest number of DEGs and DEIs in the P06/P07 (late mid-fetal/late fetal) and P07/P08 (late fetal/neonatal) developmental periods, supporting critical brain remodeling right before and after birth (Figure 1A). In P06/P07, 8.3% of genes and 20.3% of isoforms (converted to gene identifiers) were differentially expressed, whereas in P07/P08, 13.2% of genes and 20.4% of isoforms (converted to gene identifiers) were differentially expressed (Table S4). Overall, 48.4% of genes and 64.9% of isoforms (converted to gene identifiers) were differentially expressed between PrePost periods. These results indicate that the expression levels of over half of all isoforms change significantly between the prenatal and postnatal periods, suggesting profound transcriptomic remodeling during brain development.

In addition to the greater fraction of DEIs between adjacent and PrePost periods, we also observed significantly increased effect sizes (absolute \log_2 fold changes) among DEIs compared with DEGs, overall and in nearly every developmental period (Figure 1B). This suggests that the levels of differential expression were more pronounced at the full-length isoform level relative to the gene level, consistent with previous results obtained from postmortem brains of individuals with NDDs (Gandal et al., 2018). Thus, the full-length isoform transcriptome is likely to provide additional information about brain development that is missed by the gene transcriptome.

To better understand the biological basis of brain transcriptome differences at the gene and isoform levels, we performed enrichment analyses of unique non-overlapping DEGs and DEIs (lightly shaded subsets from Figure 1A) using cell type and literature-curated gene lists (Figure 1C). We used a published gene-level single-cell RNA-seq dataset (Zhong et al., 2018) for cell type enrichment and NDD-related gene lists for gene enrichment analyses in each period as well as in the PrePost dataset (STAR Methods). Overall, DEGs captured weaker enrichment signals than DEIs, potentially because of smaller DEG dataset sizes. Among cell types, DEIs were enriched significantly in excitatory neuron markers, especially in the prenatal to early childhood developmental periods (Fisher's exact test, max Bonferroni-adjusted $p < 1E-09$, odds ratio [OR] = 2.39–3.29, min. 95% confidence interval [CI] = 2.07, max 95% CI = 3.98 for P02/P03–P09/P10)

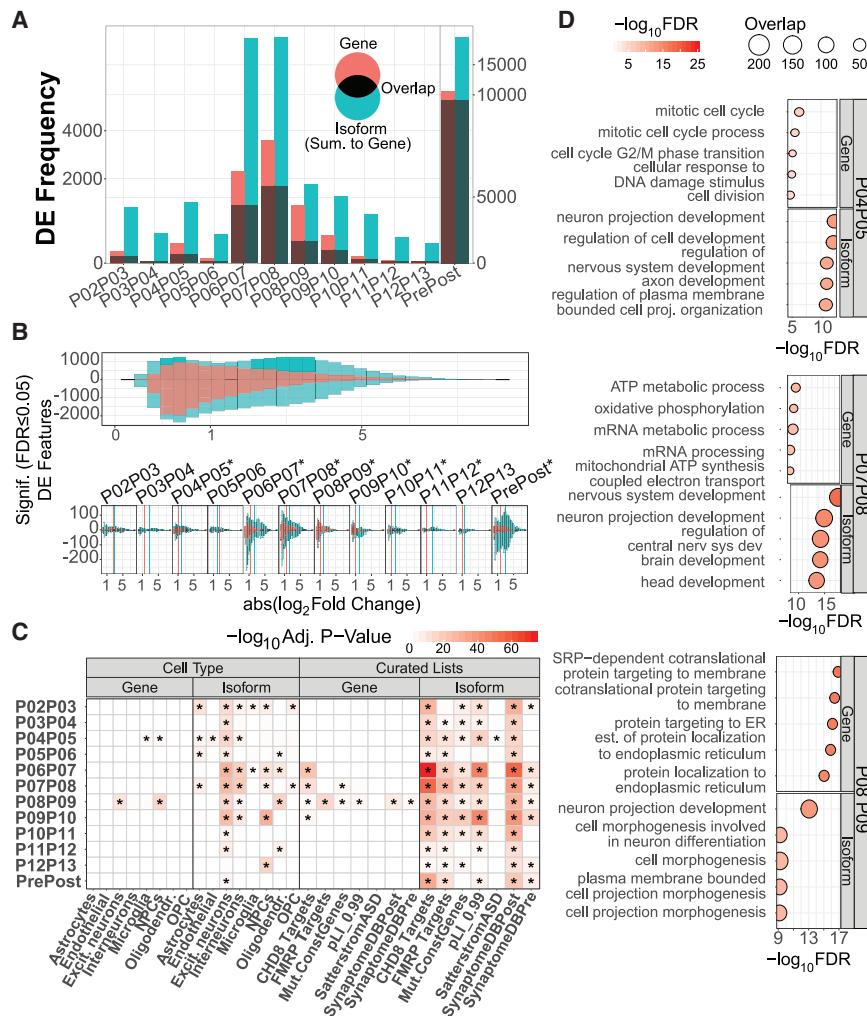


Figure 1. Differential gene and isoform expression analyses

(A) Number of significant DEGs and DEIs in the adjacent brain developmental periods and in pre-natal versus postnatal periods. Isoform identifiers were summarized to gene identifiers for simplicity of comparison. Shaded areas represent identifiers shared between gene and isoform datasets, whereas unshaded bars represent genes (red) or isoforms (turquoise) unique to each dataset.

(B) Effect size (absolute \log_2 fold change) distribution of DEGs (red) and DEIs (turquoise) of combined data (top) and per developmental period (bottom). Average absolute effect sizes for genes and isoforms are marked by corresponding colored vertical lines, and differences were tested using two-sample t tests ($^*FDR < 0.05$).

(C) Enrichment of cell types and literature-curated gene sets among genes and isoforms unique to each dataset (unshaded sets from a panel). Fisher's exact test was used to calculate p values.

(D) Gene Ontology (GO) enrichment of DEGs and DEIs unique to each dataset (unshaded sets from a panel). Three adjacent periods are shown as examples (P04/P05, P07/P08, and P08/P09). DEIs are enriched in nervous system-related processes.

(Figure 1C, left panel). The DEIs from almost all periods were also enriched in post-synaptically expressed genes as well as FMRP (Fragile X Mental Retardation Protein) and CHD8 targets, with the most significant enrichment during P06/P07 (late mid-fetal/late fetal). Interestingly, the DEIs from only P04/P05 (early mid-fetal) were enriched in autism risk genes (Satterstrom et al., 2020) (Fisher's exact test, Bonferroni-adjusted $p = 0.005$, OR = 3.88, 95% CI = 2.11–3.68) (Figure 1C, right panel), and this signal was not observed at the gene level. The mid-to-late fetal developmental period has been identified previously as critical for ASD pathogenesis (Parikhshak et al., 2013; Willsey et al., 2013; Lin et al., 2015).

The Gene Ontology (GO) enrichment analyses in P04/P05, P07/P08, and P08/P09 demonstrated stronger enrichment of DEI in neurodevelopmental processes compared with DEGs (Figure 1D; Tables S5 and S6). For example, “neuron projection development,” “brain development,” and “nervous system development” were enriched among DEIs but not among DEGs. In contrast, DEGs were enriched in basic biological function-related processes, such as “mitotic cell cycle,” “metabolic processes,”

“protein targeting,” and “localization.” This suggests that the full-length isoform transcriptome provides better biological insights into brain development than the gene transcriptome.

DEIs affected by autism LoF mutations have higher prenatal expression

To improve our understanding of the effect of NDD mutations on brain development, we mapped rare *de novo* LoF variants identified in the largest autism exome sequencing study (Satterstrom et al., 2020) to the full-length isoform transcriptome. 12,111 ASD case variants and 3,588 control variants were processed through Ensembl's Variant Effect Predictor (VEP) and filtered for consequences likely to result in LoF of the affected gene or isoform (STAR Methods; Table S7). In total, 1,132 ASD case and 262 control variants fit this criterion, affecting 4,050 isoforms from 1,189 genes. At the isoform level, 3,128 isoforms were affected by ASD case variants (ASD LoF), 848 isoforms by control variants (control LoF), and 74 isoforms by both. We also defined a dataset of isoforms that were not affected by ASD variants (non-affected by ASD LoF) as an internal control.

In every prenatal developmental period as well as in the pooled prenatal dataset, expression of the ASD LoF-affected isoforms was found to be significantly higher than control LoF-affected isoforms or non-affected by ASD LoF isoforms (Mann-Whitney test, Benjamini-Hochberg-adjusted $p \leq 0.05$) (Figure 2A). This suggests that the potential decrease or loss of expression of

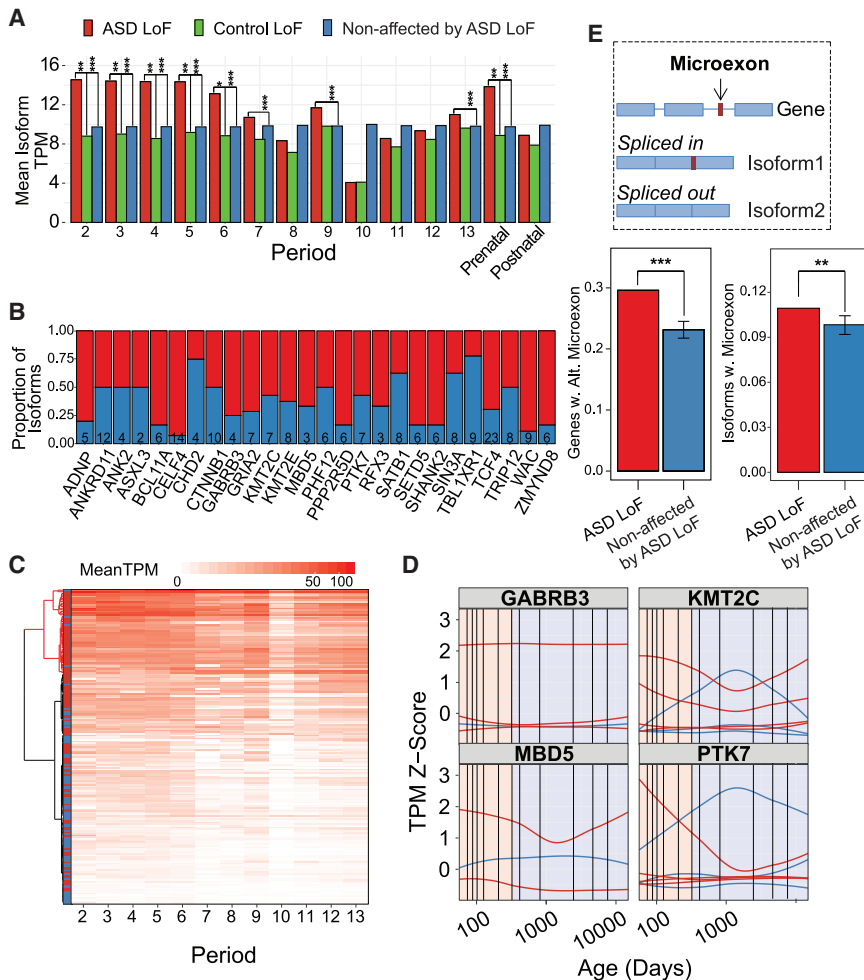


Figure 2. Analyses of isoforms affected by rare *de novo* ASD LoF variants

(A) Mean expression of isoforms affected by case rare *de novo* ASD LoF variants (affected by ASD LoF) is significantly higher in prenatal periods compared with those affected by control LoF mutations or to non-affected isoforms.

(B) Proportion of protein-coding isoforms of high-risk ASD genes from Satterstrom et al. (2020), uniquely differentially expressed at isoform level, affected (red) or not affected (blue) by rare *de novo* ASD LoF variants.

(C) Ward hierarchical clustering of isoforms from (B) based on average expression values across developmental periods.

(D) Expression profiles of affected and non-affected isoforms of four ASD risk genes across development, demonstrating higher prenatal expression of some affected isoforms.

(E) Schematic of alternatively regulated microexons (top panel), proportion of all brain-expressed genes with alternatively regulated microexons (bottom left), and proportion of all brain-expressed isoforms with alternatively regulated microexons (bottom right).

* $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

these highly expressed isoforms in the normal prenatal human brain as a result of LoF mutation may contribute to ASD pathogenesis.

We then selected genes with DEIs between adjacent development periods for which at least one isoform is ASD LoF-affected and at least one other isoform is non-affected by ASD LoF; 26 genes of 102 Satterstrom genes satisfied this criterion (Figure 2B). Hierarchical clustering of the isoforms from these genes based on expression values identified a prenatally expressed cluster consisting largely of the ASD LoF-affected isoforms (Figure 2C). A higher fraction of LoF-affected isoforms carried microexons (i.e., short exons 3–27 bp in length) compared with non-affected isoforms (permutation test, $n = 1,000$ permutations, $p = 0.04$) (Figure 2D), recapitulating previous findings at the gene level and in agreement with the important role of microexons in autism (Irimia et al., 2014; Li et al., 2015). The affected and non-affected isoforms of some genes (*KMT2C*, *MBD5*, and *PTK7*) had opposite developmental trajectories, whereas for other genes (*GABRB3*), the affected isoforms were highly expressed throughout brain development (Figure 2E). It is likely that a LoF mutation that affects a highly prenatally expressed isoform can severely

help us to better understand their functional effects in the context of brain development.

Isoform co-expression modules capture distinct trajectories of brain development

To understand how brain development is regulated at the full-length isoform level, we carried out WGCNA and isoform co-expression network analysis (Langfelder and Horvath, 2008; STAR Methods). This analysis identified modules of genes and isoforms with highly correlated expression profiles across all BrainSpan samples. We identified a total of 8 gene and 55 isoform co-expression modules by analyzing the gene and isoform transcriptomes (Tables S8 and S9). Gene and isoform networks followed scale-free topology and had similarly low soft-thresholding betas (two for gene and three for isoform) selected to allow for module detection (Figures S4C and S4D).

The hierarchical clustering of modules by eigengenes demonstrated that each gene co-expression module closely clustered with a corresponding isoform co-expression module (Figure 3A). Further characterization of these gene/isoform module pairs via GO annotations showed overlapping functions and pathways (Tables S10 and S11). For example, gene module gM2 and

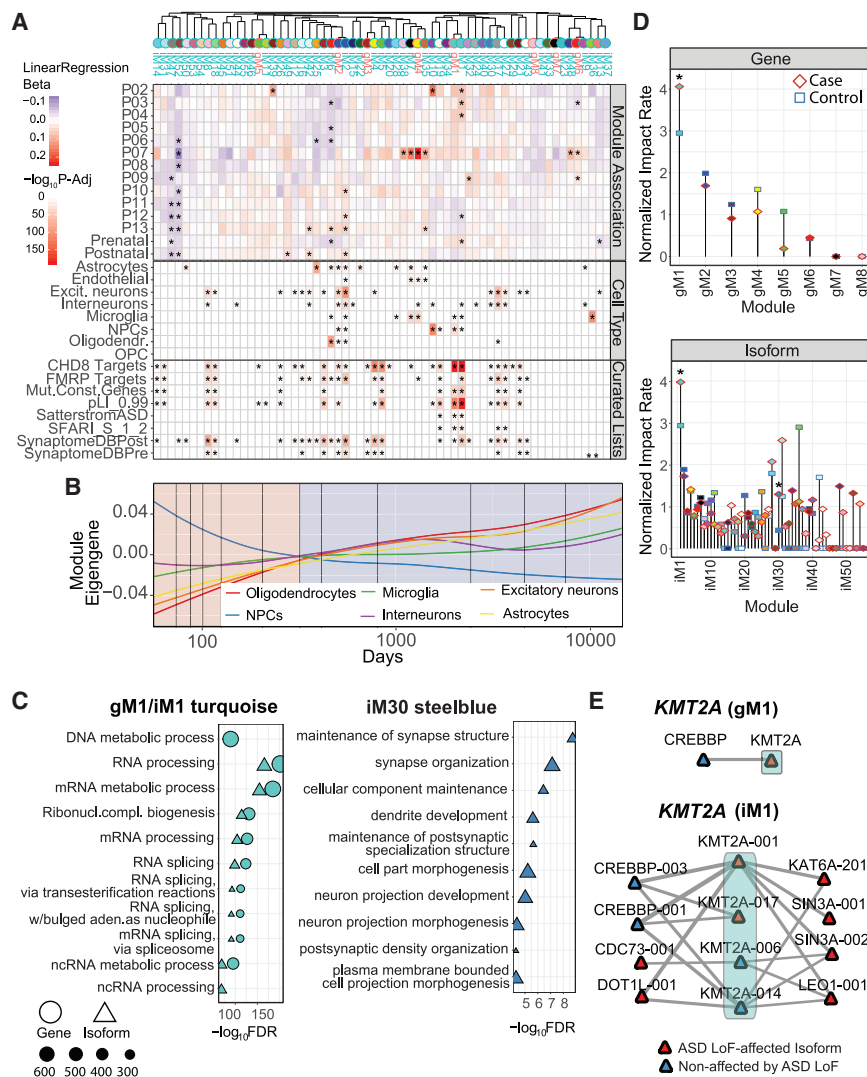


Figure 3. Gene and isoform co-expression analyses

(A) Association of gene and isoform co-expression modules clustered by module eigengene with developmental periods (top). Linear regression beta coefficients were calculated using linear mixed-effects models. Module enrichment in cell type and literature curated gene sets (bottom) was calculated using Fisher's exact test.

(B) Module eigengene expression profiles across brain development for modules most significantly associated with each cell type: astrocytes, iM25; oligodendrocytes, iM6; microglia, iM36; NPCs, iM10; excitatory neurons, iM2; interneurons, iM17.

(C) GO functional enrichment analyses of gM1/iM1 and iM30 modules significantly affected by case ASD LoF mutations.

(D) Gene (top panel) and isoform (bottom panel) co-expression modules affected by case and control ASD LoF mutations. Normalized effect rate per module is shown. Significance was calculated by permutation test (1,000 permutations, *FDR \leq 0.05).

(E) Gene-level and isoform-level co-expressed protein interaction networks for the *KMT2A* gene from the gM1 and iM1 turquoise modules. Only edges in the top 10% of expression PCCs that are also supported by gene-level protein interactions are retained.

co-expression modules with unique developmental trajectories and biological functions that were not detected by gene co-expression analyses.

To relate each co-expression module with brain developmental periods, we calculated module-period associations using linear mixed effects models (*STAR Methods*). We found modules that were significantly associated with several developmental periods (Figure 3A, top panel); iM1 was significantly associated

isoform module iM2 were enriched for GO terms related to synaptic transmission. This indicates that the isoform co-expression network recapitulates functional aspects of the gene co-expression network.

This hierarchical clustering of modules by eigengenes also revealed isoform modules that are completely unique relative to the gene-level modules (Figure 3A). We further performed a Pearson correlation analysis between the isoform modules and the gene modules by their respective eigengene expression and identified isoform modules with low Pearson correlation coefficients (PCCs) with any gene module (Figure S4E). This set of isoform modules with unique expression trajectories includes those related to chemical synaptic transmission and neurogenesis (iM18, PCC = 0.13); kinase activity (iM51, PCC = 0.27); cell adhesion, neuron fate, and cerebral cortex regionalization (iM40, PCC = 0.34); behavior, cognition, learning, and memory (iM13, PCC = 0.38); and chromatin, neuron organization, and migration (iM31, PCC = 0.3) as examples. These results identify

with prenatal periods P02 (false discovery rate [FDR]-adjusted $p = 0.009$), P03 (FDR-adjusted $p = 0.003$), and P04 (FDR-adjusted $p = 0.008$), whereas iM10 and iM39 were associated with P02 (FDR-adjusted $p = 6.59E-04$ and FDR-adjusted $p = 0.026$, respectively). Functional GO analyses of these modules demonstrated that iM1 was enriched in splicing functions and iM10 in mitosis and cell cycle-related processes, whereas iM39 was enriched in embryonic development; all functions were related to early fetal brain development (Table S11). The developmental trajectory of the iM10 module has a clear peak at early developmental periods that levels down after birth, and both modules have unique GO functions along with intermediate to low correlation with gene modules (Figure S4E).

Several other modules (gM4, iM35, iM7, and iM38) were strongly associated with late fetal period P07 (gM4: FDR-adjusted $p = 1.78E-09$; iM35: FDR-adjusted $p = 8.23E-04$; iM7: FDR-adjusted $p = 3.83E-04$; iM38: FDR-adjusted $p = 0.009$). These modules were enriched for angiogenesis and

extracellular matrix organization GO functions (Table S11) and had high PCC and overlapping GO functions with gene modules (Figure S4E).

The analysis of cell-type markers extracted from single-cell sequencing studies (STAR Methods) identified modules that were significantly enriched in specific cell types (Figure 3A, center panel). For example, iM10, which was associated with a very early P02 period, was also enriched in neuroprogenitors (NPCs), the cells that give rise to other neuronal cell populations and are often found very early in brain development. Likewise, iM2 was primarily associated with postnatal periods and strongly enriched in excitatory neurons, a mature neuronal population. Interestingly, the cluster of modules that was strongly associated with the late fetal P07 period (gM4, iM35, iM7, and iM38), was enriched in microglia, or innate immune cells of the brain, that peak around late mid-fetal to late fetal development. Furthermore, isoform module eigengene trajectories captured the appropriate signals from each cell type, with NPCs steadily decreasing and neuronal cell types increasing from prenatal to postnatal brain development (Figure 3B). To ensure that the observed cell-type-specific signatures are driven by isoform expression rather than cellular composition, we performed decomposition analyses of bulk BrainSpan RNA-seq data with fetal brain single-cell RNA-seq from two previous studies (Polio-udakis et al., 2019; Zhong et al., 2018). We observed that cell type composition only explained a relatively small portion of total variance compared with sample age (i.e., developmental period) (Figures S5A and S5B).

Analysis of curated gene lists in the context of co-expression modules identified gM1/iM1 as being enriched in ASD risk genes and CHD8 target and functionally constrained and mutation-intolerant (pLI > 0.99) genes (Figure 3A, bottom panel). The same modules were significantly associated with prenatal periods and were enriched in RNA processing and splicing GO functions (Figure 3C, left panel). Another module enriched in ASD risk genes was iM19, annotated with chromatin and histone-related GO functions. This is consistent with previous observations regarding enrichment of chromatin modifier genes among ASD risk genes (De Rubeis et al., 2014). Analyses of isoform co-expression modules broaden our knowledge of the developing human brain at the full-length isoform transcriptome level.

LoF-affected co-expression modules point to dysregulation of RNA splicing and synaptic organization

We next investigated enrichment of rare *de novo* ASD variants from cases and controls (Satterstrom et al., 2020), and identified co-expression modules that were significantly affected by LoF case but not control mutations (Table S12; STAR Methods). We observed three modules significantly affected by case ASD variants: one gene module (gM1) and two isoform modules (iM1 and iM30) (Figure 3D). Unsurprisingly, gM1 and iM1 clustered together and were enriched in similar GO functions related to RNA processing and splicing, including non-coding RNA splicing (Figure 3C). This agrees with the already demonstrated crucial role of splicing dysregulation in ASD (Gandal et al., 2018; Parikshak et al., 2016). Functional enrichment of isoform co-expression module iM30 pointed to dysregulation of synapse

organization, dendrite development, and neuronal projection pathways (Figure 3C), which are strongly implicated in ASD (Iakoucheva et al., 2019; Pinto et al., 2014). The highest PCC of iM30 with any gene module is 0.62, and its developmental trajectory and GO functions are distinct compared with gene modules (Figure S4E). Thus, isoform modules reflect processes implicated previously in ASD and point to specific isoforms (rather than genes) that can contribute to this dysregulation.

To demonstrate how isoform co-expression modules could be useful for future studies, we built isoform co-expressed protein-protein interaction (PPI) networks for the gM1 and iM1 modules (Figure S5C). The isoform network was focused on ASD risk genes that had at least one isoform affected by an LoF mutation, and the edges that had gene-level PPI information (because of scarcity of isoform-level PPIs) were filtered for the top 10% PCC (STAR Methods). Clearly, gM1 had fewer connections than iM1, and iM1 highlighted some interesting isoform co-expressed PPIs that were not detectable from the gene-level network. For example, 9 genes from this module (*ARID1B*, *CHD8*, *KMD5B*, *KMT2A*, *MED13L*, *PCM1*, *PHF12*, *POGZ*, and *TCF4*) had at least one ASD LoF-affected isoform and at least one that was not affected by mutation. These isoforms were co-expressed and interacted with different partner isoforms. For example, ASD LoF-affected and non-affected isoforms of the *KMT2A* gene had shared as well as unique protein interaction partners (Figure 3E). This could lead to different networks being disrupted as a result of ASD mutation, and these networks were not observed at the gene level, with only one *KMT2A* partner (*CREBBP*) in the gene network. Another interesting observation from co-expressed PPI networks was that LoF-affected isoforms tended to have higher correlation with corresponding partners than non-affected isoforms (Mann-Whitney test, $p = 1.53E-05$), suggesting potentially greater functional effects on networks.

De novo splice-site mutations of NDD risk genes cause exon skipping, partial intron retention, or have no effect on isoforms

One type of LoF mutation is mutations that affect splice sites directly. Here we experimentally investigated the effect of *de novo* splice site mutations identified by exome sequencing studies in four NDD risk genes (*DYRK1A*, *SCN2A*, *DLG2*, and *CELF2*) to better understand their functional effects. All highly prenatally expressed isoforms of these genes were found in iM1. We used an exon trapping assay (STAR Methods) to test the following *de novo* splice site mutations: *SCN2A* (chromosome 2 [chr2]:166187838, A:G, acceptor site; Fromer et al., 2014), *DYRK1A* (chr21: 38865466, G:A, donor site; O’Roak et al., 2012), *DLG2* (chr11: 83194295, G:A, donor site; Fromer et al., 2014), and *CELF2* (chr10: 11356223, T:C, donor site; Xu et al., 2011). The mutation in *SCN2A* caused out-of-frame exon skipping and potential inclusion of 30 new amino acids into the translated protein before ending with a premature stop codon that most likely will result in nonsense-mediated decay (NMD) (Figure 4A). In contrast, the mutation in *DYRK1A* caused in-frame exon skipping, potentially producing a different variant of the same protein, and, thus, is expected to have milder or perhaps no functional effects (Figure 4B). In the case of *DLG2*, the mutation affected a splice site adjacent to exon 5, which is

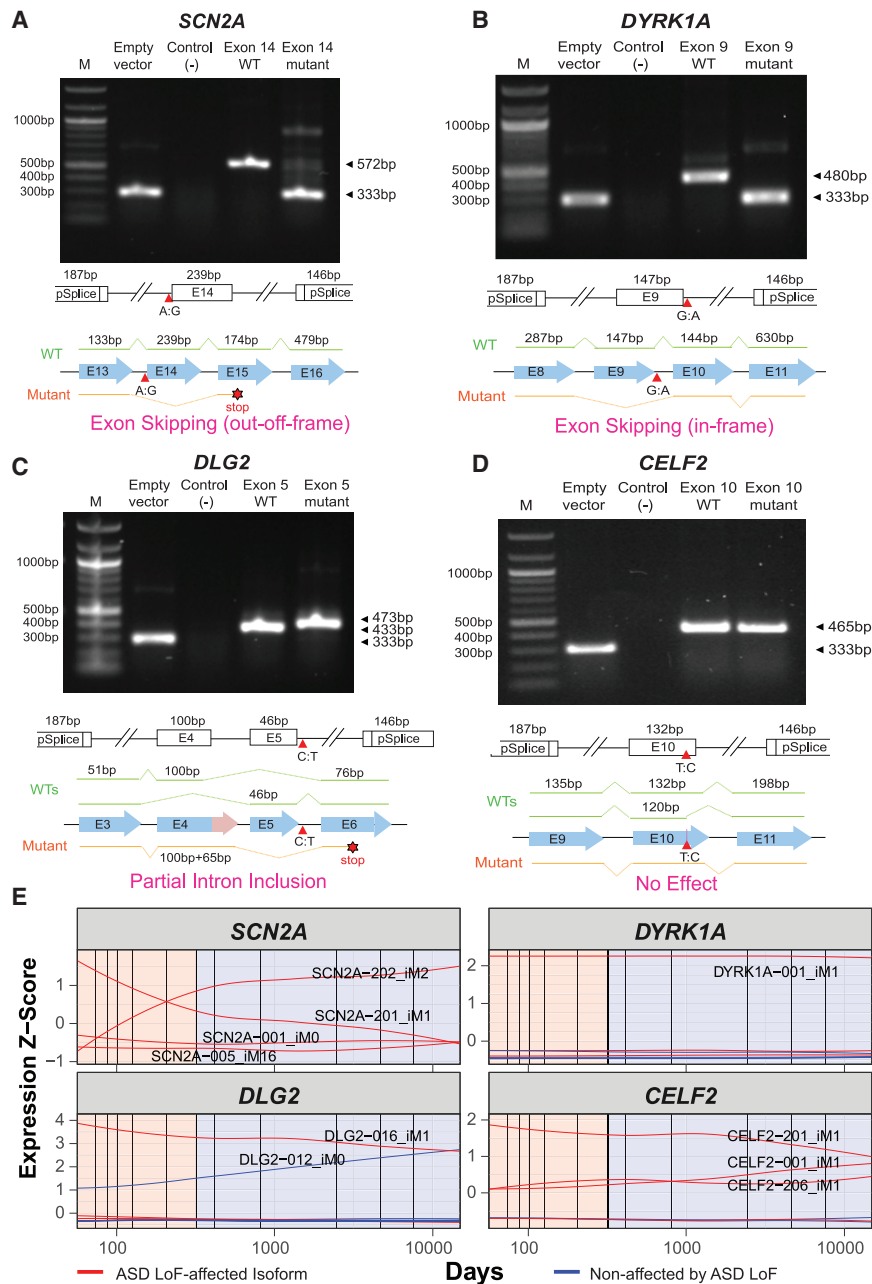


Figure 4. Functional effect of the *de novo* splice site mutations from individuals with NDDs

(A–D) Minigene assays demonstrate the effect of splice-site mutations in four genes: *SCN2A* (A), *DYRK1A* (B), *DLG2* (C), and *CELF2* (D). A schematic of the cloned minigenes, the expected splicing patterns, and the effects of the mutations are shown below the gel image. Numbers denote base pairs. M, molecular marker; E, exon.

(E) Expression profiles across brain development of the brain-expressed isoforms transcribed by these four genes, annotated with module memberships; highly overlapping expression profiles are unlabeled for readability.

isoform generated from the WT minigene included the isoform carrying a longer exon with mutation (Figure 4D). Thus, after introducing the mutation, no difference between WT and mutant constructs was observed. This is not surprising, given the fact that the splice site mutation behaves like an exonic missense mutation in the isoform predominantly expressed from our construct. These results suggest that mutations can affect different isoforms of the same gene by different mechanisms; i.e., splice site mutation in one isoform could represent a missense mutation in another isoform.

Further analysis of expression profiles of the brain-expressed isoforms transcribed by these genes (Figure 4E) suggested that highly prenatally expressed isoforms (*SCN2A-201*, *DYRK1A-001*, *DLG2-016*, and *CELF2-201*) were most likely targets for the “pathogenic” effect of mutations. Furthermore, given distinct co-expressed PPI partners of affected versus non-affected isoforms (Figure S5D) in most cases, the effect of mutation would be propagated onto different networks, affecting different signaling pathways. Our experiments showcased different scenarios of the effect of splice site mutations and

alternatively spliced in the wild-type (WT) isoforms (Figure 4C). We constructed a minigene that includes exon 5 together with the preceding exon 4 and observed that exon 5 was constitutively spliced out from our construct independent of the presence of mutation. However, the mutation caused partial (i.e., 65-bp) intron inclusion downstream of exon 4. At the translational level, this mutation would likely result in a truncated protein one residue after the end of exon 4 because of a premature stop codon. Finally, the *CELF2* mutation affected an alternative splice site that also mapped to an exonic region of another alternatively spliced isoform. When cloned into the exon-trapping vector, the

confirmed the need to investigate their functional effects at isoform- rather than gene-level resolution.

The splice site mutation in *BTRC* reduces its translational efficiency

As shown above, mutations can affect different protein interaction networks depending on the splicing isoform they affect and whether the affected isoform is expressed at specific periods of brain development. Next we investigated in greater detail how specific mutations mapped to different isoforms may disrupt downstream signaling pathways. For this, we selected three

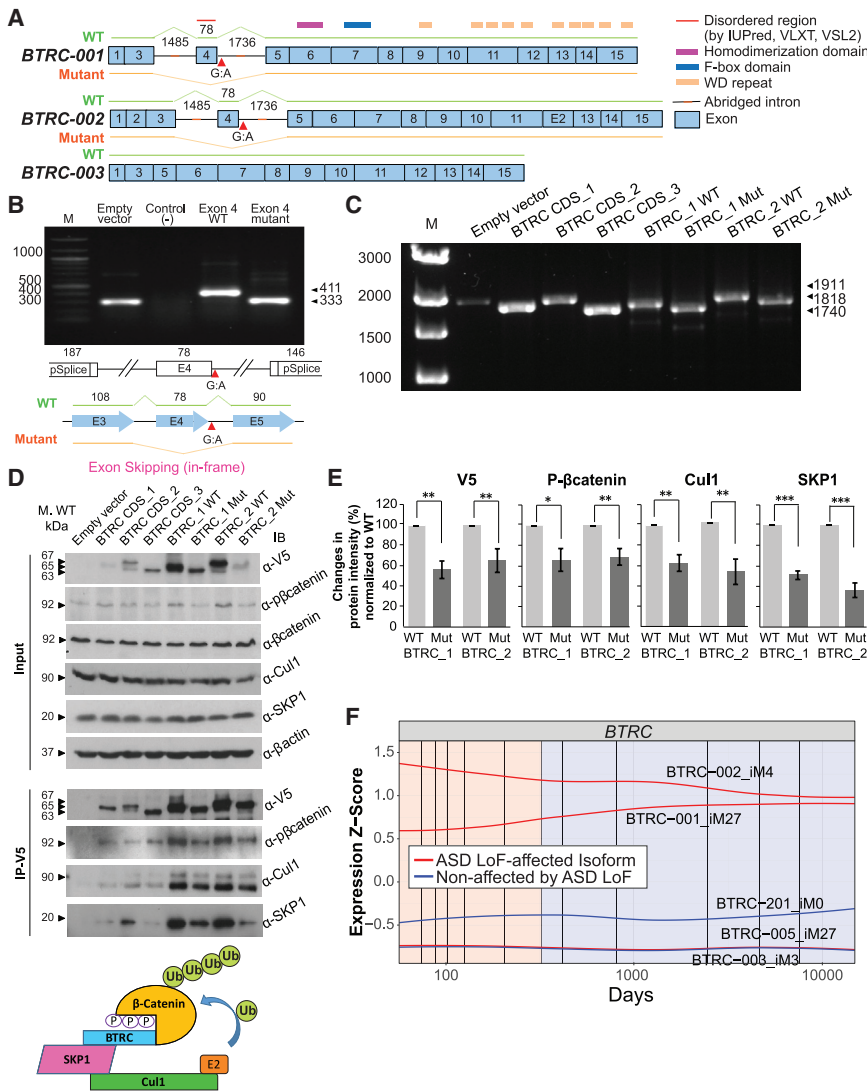


Figure 5. The *de novo* autism splice-site mutation causes exon skipping in *BTRC* isoforms and reduces their translational efficiency

(A) The exon structure of three splicing isoforms of the *BTRC* gene showing positions of the cloned abridged introns and the splice-site mutation; numbers denote base pairs.

(B) Minigene assays demonstrate exon 4 skipping as a result of the splice-site mutation. The assays show the results of RT-PCR performed using total RNA from HeLa cells transfected with *BTRC* minigene constructs; numbers denote base pairs.

(C) Splicing assays with full-length constructs carrying abridged introns, confirm exon skipping observed in the minigene assays.

(D) Immunoblot (IB) of whole-cell lysates of HeLa cells transfected with different *BTRC* minigene constructs and an empty vector, as indicated. Membranes were probed to observe *BTRC* overexpression and investigate expression of β -catenin, Cul1, and SKP1. β -Actin was used as a loading control. IP was performed with the antibody recognizing the V5 tag, and proteins were detected by IB with β -catenin, Cul1, SKP1, and V5 antibodies. The splice-site mutation causes reduced translational efficiency of the *BTRC*_1Mut and *BTRC*_2Mut mutant isoforms compared with their WT counterparts. A schematic of the Skp1-Cul1-*BTRC* ubiquitin protein ligase complex is shown at the bottom.

(E) Quantification of protein pull-down with V5 IP using ImageJ software. The band intensity values were normalized to WT expression levels. Error bars represent 95% CI based on 3 independent experiments. On average, a 40% reduction of *BTRC* protein expression is observed as a result of a mutation. Consequently, a reduction of the corresponding *BTRC* binding partners (β -catenin, Cul1, and SKP1) is also observed.

(F) Expression profiles of brain-expressed *BTRC* isoforms show higher expression of ASD-affected *BTRC*-001 and *BTRC*-002.

Numbers denote base pairs (A–C) or kilodaltons (D). * $p < 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

full-length isoforms (*BTRC*-001, *BTRC*-002, and *BTRC*-003) of an ASD risk gene, *BTRC* (beta-transducin repeat containing, also known as β -*TrCP* or *FBXW1A*) (Ruzzo et al., 2019), based on their availability from our previous study (Yang et al., 2016; Figure 5A). Two *de novo* mutations, one missense (chr10:103285935,G-A; Ruzzo et al., 2019) and one splice site (chr10: 103221816, G:A, donor site; De Rubeis et al., 2014), were identified in individuals with ASD and none in controls, making *BTRC* one of 69 high-confidence ASD risk genes with genome-wide significance ($0.05 < \text{FDR} \leq 0.1$; Ruzzo et al., 2019). We demonstrated that the splice site *BTRC* mutation caused in-frame exon 4 (78 bp) skipping in the exon trapping assay (Figure 5B). To further test the effect of this mutation on different *BTRC* transcripts, we generated additional constructs by inserting abridged introns surrounding exon 4 into the coding sequences (CDSs) of two isoforms, *BTRC*-001 and *BTRC*-002 (Figure 5C; STAR Methods). The third isoform, *BTRC*-003, did not carry exon 4, and its structure and size are identical

BTRC-001 after exon 4 is skipped. We also generated the mutant constructs *BTRC*-001Mut and *BTRC*-002Mut carrying the mutation in the abridged intron (Figure 5A). RT-PCR following exon trapping assays on the full-length CDS as well as WT and mutant constructs with abridged introns confirmed the correct sizes of all constructs and validated the exon skipping event because of splice site mutation (Figure 5C). The expression levels of the mutant constructs were comparable with the WT. Furthermore, western blot confirmed the expected sizes of the protein products from the WT and mutant constructs (Figure 5D). The splice site mutation significantly reduced the amount of protein produced from mutant isoforms, suggesting their decreased translational efficiency (Figure 5E). Higher amounts of protein products from all constructs with abridged introns compared with CDSs were consistent with previous observations of increased translational efficiency of RNAs produced by splicing compared with their intron-less counterparts (Diem et al., 2007). Further, *BTRC*-001 and *BTRC*-002 are highly

expressed relative to the non-affected *BTRC*-003 (Figure 5F) and are co-expressed and interact with non-overlapping sets of protein partners in the co-expressed PPI networks (Figure S5E). This suggests that mutations in different isoforms could affect different cellular networks.

Next we investigated binding properties of all isoforms using co-immunoprecipitation (coIP) (Figure 5D). The *BTRC* gene encodes a protein of the F box family and is a component of the SCF (Skp1-Cul1-F box protein) E3 ubiquitin-protein ligase complex. One of the well-known substrates of this complex is β -catenin (CTNNB1). The SCF complex ubiquitinates and regulates degradation of β -catenin, an essential component of the *Wnt* signaling pathway (Winston et al., 1999). *Wnt* plays key roles in cell patterning, proliferation, polarity, and differentiation during embryonic development of the nervous system (Ciani and Salinas, 2005), and it has been implicated consistently in ASD (Urresti et al., 2021; Iakoucheva et al., 2019; Kwan et al., 2016). β -Catenin and Cul1 carry *de novo* mutations identified in individuals with NDD (Satterstrom et al., 2020).

The interaction of *BTRC* with its partners Cul1, Skp1, and β -catenin demonstrated reduced binding with mutant *BTRC*, potentially suggesting a shortage of the SCF ligase complexes (Figures 5D and 5E). In agreement with previous observations, we found that *BTRC* only binds to the phosphorylated form of β -catenin (Winston et al., 1999). This suggests that the amount of protein complex is strongly dependent on the availability of *BTRC* protein, which is reduced significantly because of splice-site mutation. Thus, our results indicate that the *BTRC* splice-site mutation causes exon skipping in *BTRC* isoforms and reduces the translational efficiency of the resulting protein product. This, in turn, decreases the amount of SCF protein ligase complexes available for β -catenin ubiquitination. We hypothesize that this may lead to impaired degradation of β -catenin, its cellular accumulation, and upregulation of *Wnt* signaling as a result of this ASD risk mutation. Further studies in neuronal cells are needed to test this hypothesis.

DISCUSSION

Recent large-scale whole-exome and whole-genome sequencing studies have greatly facilitated discovery of the genetic causes of neurodevelopmental disorders. One of the bottlenecks in translating these findings into molecular mechanisms is our limited understanding of the transcriptional and translational programs governing brain development. The brain is one of the most complex human organs with the highest number of alternatively spliced events (Melé et al., 2015; Raj and Blencowe, 2015). Thus, knowledge regarding its splicing repertoire is crucial for future translational studies of brain diseases.

Several earlier studies have addressed the effects of NDD mutations at exon-level resolution. For example, Uddin et al. (2014) identified highly expressed critical exons with *de novo* ASD mutations that were enriched in cases compared with controls. A more recent study correlated exons affected by ASD LoF mutations with phenotypes of affected individuals, and found that those with mutations in the same exon had more similar phenotypes (Chiang et al., 2021). However, none of the previous

studies investigated the effects of LoF mutations in the context of the full-length isoform transcriptome.

We demonstrated previously that integration of genetic data with isoform-level co-expression and protein interaction networks is crucial for improving our understanding of the molecular mechanisms of neurodevelopmental disorders (Corominas et al., 2014; Gandal et al., 2018; Lin et al., 2015, 2017). The importance of isoform-level networks was further emphasized by our earlier observation that protein products encoded by different splicing isoforms of the same gene share less than half of their interaction partners (Yang et al., 2016). These studies underscore the importance of the brain isoform transcriptome for future studies of NDDs. Here we analyzed the full-length isoform transcriptome of the developing brain and demonstrated its utility for investigating LoF mutations implicated in autism.

We demonstrated that brain differential isoform expression analysis identified a fairly large set of DEIs that were not detected by the gene-level analysis. Furthermore, DEIs captured more relevant functions than DEGs in the context of brain development. Processes such as neuron projection development, axon development, and head, brain, and nervous system development were primarily supported by DEIs uniquely identified only through analysis of the isoform transcriptome.

By mapping LoF mutations from autism cases and controls onto the full-length isoform transcriptome, we found that ASD LoF-affected isoforms had significantly higher prenatal expression than non-affected isoforms or isoforms affected by control mutations. The expression trajectories of affected and non-affected isoforms across brain development were remarkably different for some of the autism risk genes. For example, two LoF-affected isoforms of *KMT2C* histone lysine methyltransferase, a high-confidence ASD risk gene (De Rubeis et al., 2014; Iossifov et al., 2015; O’Roak et al., 2011) were highly expressed prenatally and had opposing temporal trajectories compared with non-affected isoforms that were highly expressed postnatally (Figure 2D). Similar developmental trajectories were observed for some of the LoF-affected isoforms of *PTK7* and *MBD5* genes. This demonstrates that future studies of these and other genes with similar properties should focus on affected isoforms with high prenatal expression rather than on all available isoforms because they may be more relevant to brain development.

In general, we consistently gain additional information from the full-length isoform transcriptome across various types of analyses. At the level of co-expression, isoform co-expression modules provide important insights into neurodevelopment and on how it may be disrupted by autism mutations. Importantly, many isoform modules, especially those with a low PCC with gene modules, have unique and distinct developmental trajectories and biological functions that are not identified through gene-level analyses (Figure S4E). Many of these distinct isoform modules were annotated with important and ASD-relevant functions, such as translation (iM21), protein/histone demethylation (iM47), chromatin organization and neuron migration (iM31), and cell adhesion, neuron fate, and cerebral cortex regionalization (iM40). Other isoform modules that have a high PCC with gene modules could also provide important insights. The isoform module iM1 was significantly enriched in isoforms affected by case LoF mutations (Figure 3D). Functionally,

it was enriched in RNA splicing and processing pathways that have been implicated previously in ASD (Parikshak et al., 2016). iM1 was significantly associated with prenatal developmental periods; enriched in interneurons, microglia, and NPCs; and enriched in CHD8 target genes and mutation-intolerant genes and was also one of a few modules enriched in ASD risk genes. Given all of these lines of evidence, it clearly is a very important module for further investigation.

Using isoforms from the iM1 module, we experimentally investigated the functional effects of splice site-disrupting LoF mutations in five genes. The results demonstrated exon skipping or disruption of normal splicing patterns but not in all cases. A more detailed analysis at the isoform level suggested that not all isoforms were affected by mutations. For example, at least one known isoform of the *BTRC* gene did not carry an exon with mutation and therefore was not expected to be affected by it. We next demonstrated that a *BTRC* mutation decreased translational efficiency of the affected isoforms because a lower amount of the resulting protein was observed (Figure 5D). This, in turn, led to reduced interaction between *BTRC* and its protein partners, potentially disrupting *Wnt* signaling (Figure 5E). Because β -catenin is a substrate of the *BTRC-Cul1-Skp1* ubiquitin ligase complex, shortage of this complex may lead to impaired ubiquitination and degradation of β -catenin and its neuronal accumulation. Interestingly, transgenic mice overexpressing β -catenin have enlarged forebrains, arrest of neuronal migration, and dramatic disorganization of the layering of the cerebral cortex (Chenn and Walsh, 2002). It would be interesting to investigate whether individuals carrying the *de novo BTRC* splice-site mutation have similar brain abnormalities.

Typically, mutations affecting essential splice sites are automatically classified as LoF mutations when considering gene-level analyses. Here we demonstrate that this is not always the case and that splice-site mutations affecting one isoform of a gene may serve as a missense mutation in another isoform that carries a longer exon spanning the splice site, like in the case of *CELF2* (Figure 4D). Thus, depending on where, when, at what level, and which isoform of the gene is expressed, the functional effect of the same mutation may differ dramatically. In addition, the mutation could also be “silent” when the isoform is highly expressed but does not carry an exon affected by a specific mutation. This suggests that the effects of mutations should be investigated at isoform- rather than gene-level resolution, and the expression levels of splicing isoforms in disease-relevant tissues should be taken into consideration to better guide hypotheses regarding potential mechanisms of disease and future treatments.

Although our study provided further insights into functional effects of NDD mutations at the level of the full-length isoform transcriptome, it has several limitations that warrant discussion. The BrainSpan developmental brain transcriptome was sequenced using short-read sequencing technologies; therefore, full-length isoform assignment, especially for low-abundance transcripts, is less reliable compared to the long-read sequencing. Long-read RNA-seq technologies could offer large improvements in the accuracy of full-length isoform identification and discovery. When it becomes less cost prohibitive, creating a long-read-sequenced isoform transcriptome of the developing human brain would be invaluable for the scientific community. Our study directly

demonstrates the value of using full-length transcriptome data for future human disease studies and for investigating functional effects of disease-associated mutations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Brain transcriptome data
 - Cell lines
- METHOD DETAILS
 - Processing of RNA-Seq gene and transcript BrainSpan data
 - Validation of isoform expression with RT-qPCR
 - Differential gene and isoform expression analysis
 - Cell type and literature curated gene sets enrichment analyses
 - Gene Ontology (GO) functional enrichment analyses
 - Rare *de novo* ASD loss-of-function variants
 - Weighted gene/isoform co-expression network analyses (WGCNA)
 - Co-expression module characterization
 - Decomposition of bulk BrainSpan data with fetal brain single cell RNA-seq
 - Variant impact analysis for co-expression modules
 - Integration of protein-protein interaction networks with co-expression modules
 - Minigenes cloning
 - Exon trapping and RT-PCR
 - Co-Immunoprecipitation and Western Blot
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Expression validation by RT-qPCR
 - Variant impact analysis
 - Co-expression module characterization

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2021.109631>.

ACKNOWLEDGMENTS

The study was supported by grants R01MH109885 (to L.M.I.), R01MH105524 (to L.M.I.), and R01 MH108528 (to L.M.I.) and in part by the Simons Foundation for Autism Research (#345469 to L.M.I.). We thank Dr. Marc Vidal and Dr. David Hill for providing clones of the *BTRC* isoforms. We also thank Dr. Hyun-Jun Nam and Bom Han from the lakoucheva lab for helping with genetic data curation. We thank the PsychENCODE Consortium for generating the data. Data generation as part of the PsychENCODE Consortium was supported by U01MH103339, U01MH103365, U01MH103392, U01MH103340, U01MH103346, R01MH105472, R01MH094714, R01MH105898, R21MH102791, R21MH105881, R21MH103877, and P50MH106934 awarded to Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford

(Duke), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (USC), Mark Gerstein (Yale), Daniel Geschwind (UCLA), Thomas M. Hyde (LIBD), Andrew Jaffe (LIBD), James A. Knowles (USC), Chunyu Liu (UIC), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Nenad Sestan (Yale), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (UCSF), Patrick Sullivan (UNC), Flora Vaccarino (Yale), Sherman Weissman (Yale), Kevin White (UChicago), and Peter Zandi (JHU).

AUTHOR CONTRIBUTIONS

L.M.I., G.N.L., and R.C. conceived and conceptualized the study. K.K.C., P.Z., A.B.P., and J.C. performed bioinformatics analyses. J.U., M.A., and A.T. performed experiments. L.M.I. supervised the work and acquired funding. K.K.C. and L.M.I. wrote the paper with input from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 1, 2020

Revised: April 23, 2021

Accepted: August 9, 2021

Published: August 31, 2021

REFERENCES

Barbosa-Morais, N.L., Irimia, M., Pan, Q., Xiong, H.Y., Gueroussov, S., Lee, L.J., Slobodeniuc, V., Kutter, C., Watt, S., Colak, R., et al. (2012). The evolutionary landscape of alternative splicing in vertebrate species. *Science* 338, 1587–1593.

Calarco, J.A., Zhen, M., and Blencowe, B.J. (2011). Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts. *RNA* 17, 775–791.

Chatr-Aryamontri, A., Oughtred, R., Boucher, L., Rust, J., Chang, C., Kolas, N.K., O'Donnell, L., Oster, S., Theesfeld, C., Sellam, A., et al. (2017). The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* 45 (D1), D369–D379.

Chenn, A., and Walsh, C.A. (2002). Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science* 297, 365–369.

Chiang, A.H., Chang, J., Wang, J., and Vitkup, D. (2021). Exons as units of phenotypic impact for truncating mutations in autism. *Mol. Psychiatry* 26, 1685–1695.

Ciani, L., and Salinas, P.C. (2005). WNTs in the vertebrate nervous system: from patterning to neuronal connectivity. *Nat. Rev. Neurosci.* 6, 351–362.

Corominas, R., Yang, X., Lin, G.N., Kang, S., Shen, Y., Ghamsari, L., Broly, M., Rodriguez, M., Tam, S., Trigg, S.A., et al. (2014). Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. *Nat. Commun.* 5, 3650.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247–261.

Das, J., and Yu, H. (2012). HINT: High-quality protein interactomes and their applications in understanding human disease. *BMC Syst. Biol.* 6, 92.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al.; DDD Study; Homozygosity Mapping Collaborative for Autism; UK10K Consortium (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209–215.

Diem, M.D., Chan, C.C., Younis, I., and Dreyfuss, G. (2007). PYM binds the cytoplasmic exon-junction complex and ribosomes to enhance translation of spliced mRNAs. *Nat. Struct. Mol. Biol.* 14, 1173–1179.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., et al. (2014).

De novo mutations in schizophrenia implicate synaptic networks. *Nature* 506, 179–184.

Gandal, M.J., Zhang, P., Hadjimirchev, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al.; PsychENCODE Consortium (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* 362, eaat8127.

Grabowski, P. (2011). Alternative splicing takes shape during neuronal development. *Curr. Opin. Genet. Dev.* 21, 388–394.

Hoffman, G.E., and Schadt, E.E. (2016). variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* 17, 483.

Huttlin, E.L., Ting, L., Bruckner, R.J., Gebreab, F., Gygi, M.P., Szpyt, J., Tam, S., Zarraga, G., Colby, G., Baltier, K., et al. (2015). The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162, 425–440.

Iakoucheva, L.M., Muotri, A.R., and Sebat, J. (2019). Getting to the Cores of Autism. *Cell* 178, 1287–1298.

Iossifov, I., Levy, D., Allen, J., Ye, K., Ronemus, M., Lee, Y.H., Yamrom, B., and Wigler, M. (2015). Low load for disruptive mutations in autism genes and their biased transmission. *Proc. Natl. Acad. Sci. USA* 112, E5600–E5607.

Irimia, M., Weatheritt, R.J., Ellis, J.D., Parikhshak, N.N., Gonatopoulos-Pournatzis, T., Babor, M., Quesnel-Vallières, M., Tapial, J., Raj, B., O'Hanlon, D., et al. (2014). A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 159, 1511–1523.

Jew, B., Alvarez, M., Rahmani, E., Miao, Z., Ko, A., Garske, K.M., Sul, J.H., Pietiläinen, K.H., Pajukanta, P., and Halperin, E. (2020). Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* 11, 1971.

Kalsotra, A., and Cooper, T.A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* 12, 715–729.

Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. (2009). Human Protein Reference Database–2009 update. *Nucleic Acids Res.* 37, D767–D772.

Kim, K.K., Nam, J., Mukoyama, Y.S., and Kawamoto, S. (2013). Rbfox3-regulated alternative splicing of Numb promotes neuronal differentiation during development. *J. Cell Biol.* 200, 443–458.

Kishore, S., Khanna, A., and Stamm, S. (2008). Rapid generation of splicing reporters with pSpliceExpress. *Gene* 427, 104–110.

Kwan, V., Unda, B.K., and Singh, K.K. (2016). Wnt signaling networks in autism spectrum disorder and intellectual disability. *J. Neurodev. Disord.* 8, 45.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.

Leek, J.T. (2014). svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res.* 42, e161.

Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 3, 1724–1735.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291.

Li, Q., Lee, J.A., and Black, D.L. (2007). Neuronal regulation of alternative pre-mRNA splicing. *Nat. Rev. Neurosci.* 8, 819–831.

Li, Y.I., Sanchez-Pulido, L., Haerty, W., and Ponting, C.P. (2015). RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.* 25, 1–13.

Li, T., Wernersson, R., Hansen, R.B., Horn, H., Mercer, J., Slodkowitz, G., Workman, C.T., Riggins, O., Rapacki, K., Stærfeldt, H.H., et al. (2017). A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat. Methods* 14, 61–64.

Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y., et al.; BrainSpan Consortium; PsychENCODE Consortium; PsychENCODE Developmental

- Subgroup (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362, eaat7615.
- Lin, G.N., Corominas, R., Lemmens, I., Yang, X., Tavernier, J., Hill, D.E., Vidal, M., Sebat, J., and Iakoucheva, L.M. (2015). Spatiotemporal 16p11.2 protein network implicates cortical late mid-fetal brain development and KCTD13-Cul3-RhoA pathway in psychiatric diseases. *Neuron* 85, 742–754.
- Lin, G.N., Corominas, R., Nam, H.J., Urresti, J., and Iakoucheva, L.M. (2017). Comprehensive Analyses of Tissue-Specific Networks with Implications to Psychiatric Diseases. *Methods Mol. Biol.* 1613, 371–402.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 122.
- Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al.; GTEx Consortium (2015). Human genomics. The human transcriptome across tissues and individuals. *Science* 348, 660–665.
- O’Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C., et al. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* 43, 585–589.
- O’Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485, 246–250.
- Oldham, M.C., Langfelder, P., and Horvath, S. (2012). Network methods for describing sample relationships in genomic datasets: application to Huntington’s disease. *BMC Syst. Biol.* 6, 63.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40, 1413–1415.
- Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 155, 1008–1021.
- Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandai, M.J., Hartl, C., Leppa, V., Ubieta, L.T., Huang, J., et al. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 540, 423–427.
- Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., et al. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* 94, 677–694.
- Pirooznia, M., Wang, T., Avramopoulos, D., Valle, D., Thomas, G., Haganir, R.L., Goes, F.S., Potash, J.B., and Zandi, P.P. (2012). SynaptomeDB: an ontology-based knowledgebase for synaptic genes. *Bioinformatics* 28, 897–899.
- Polioudakis, D., de la Torre-Ubieta, L., Langerman, J., Elkins, A.G., Shi, X., Stein, J.L., Vuong, C.K., Nichterwitz, S., Gevorgian, M., Opland, C.K., et al. (2019). A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* 103, 785–801.e8.
- Raj, B., and Blencowe, B.J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron* 87, 14–27.
- Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198. <https://doi.org/10.1093/nar/gkz369>.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Ruepp, A., Waegele, B., Lechner, M., Brauner, B., Dungen-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C., and Mewes, H.W. (2010). CORUM: the comprehensive resource of mammalian protein complexes—2009. *Nucleic Acids Res.* 38, D497–D501.
- Ruzzo, E.K., Pérez-Cano, L., Jung, J.Y., Wang, L.K., Kashef-Haghighi, D., Hartl, C., Singh, C., Xu, J., Hoekstra, J.N., Leventhal, O., et al. (2019). Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* 178, 850–866.e26.
- Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., et al. (2014). A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* 46, 944–950.
- Sapkota, D., Lake, A.M., Yang, W., Yang, C., Wesseling, H., Guise, A., Uncu, C., Dalal, J.S., Kraft, A.W., Lee, J.M., et al. (2019). Cell-Type-Specific Profiling of Alternative Translation Identifies Regulated Protein Isoform Variation in the Mouse Brain. *Cell Rep.* 26, 594–607.e7.
- Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R., Grove, J., Klei, L., et al.; Autism Sequencing Consortium; iPSYCH-Broad Consortium (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* 180, 568–584.e23.
- Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaubblomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240.
- Steijger, T., Abril, J.F., Engström, P.G., Kokocinski, F., Hubbard, T.J., Guigó, R., Harrow, J., and Bertone, P.; RGASP Consortium (2013). Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* 10, 1177–1184.
- Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45 (D7), D362–D368.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515.
- Uddin, M., Tammimies, K., Pellecchia, G., Alipanahi, B., Hu, P., Wang, Z., Pinto, D., Lau, L., Nalpathamkalam, T., Marshall, C.R., et al. (2014). Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. *Nat. Genet.* 46, 742–747.
- Urresti, J., Zhang, P., Moran Losada, P., Yu, N.-K., Negraes, P.D., Trujillo, C.A., Antaki, D., Amar, M., Chau, K., Pramod, A.B., et al. (2021). Cortical organoids model early brain development disrupted by 16p11.2 Copy Number Variants in autism. *Mol Psychiatry*, In press. <https://doi.org/10.1038/s41380-021-01243-6>.
- Wang, E.T., Sandberg, R., Luo, S., Khrebukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476.
- Wilkinson, B., Grepo, N., Thompson, B.L., Kim, J., Wang, K., Evgrafov, O.V., Lu, W., Knowles, J.A., and Campbell, D.B. (2015). The autism-associated gene chromodomain helicase DNA-binding protein 8 (CHD8) regulates noncoding RNAs and autism-related genes. *Transl. Psychiatry* 5, e568.
- Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155, 997–1007.
- Winston, J.T., Strack, P., Beer-Romero, P., Chu, C.Y., Elledge, S.J., and Harper, J.W. (1999). The SCFbeta-TRCP-ubiquitin ligase complex associates specifically with phosphorylated destruction motifs in IkappaBalpha and beta-catenin and stimulates IkappaBalpha ubiquitination in vitro. *Genes Dev.* 13, 270–283.
- Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., and Karayiorgou, M. (2011). Exome sequencing supports

a de novo mutational paradigm for schizophrenia. *Nat. Genet.* 43, 864–868.

Yang, X., Coulombe-Huntington, J., Kang, S., Sheynkman, G.M., Hao, T., Richardson, A., Sun, S., Yang, F., Shen, Y.A., Murray, R.R., et al. (2016). Widespread Expansion of Protein Interaction Capabilities by Alternative Splicing. *Cell* 164, 805–817.

Yeo, G., Holste, D., Kreiman, G., and Burge, C.B. (2004). Variation in alternative splicing across human tissues. *Genome Biol.* 5, R74.

Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N., et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* 555, 524–528.

Zuberi, K., Franz, M., Rodriguez, H., Montojo, J., Lopes, C.T., Bader, G.D., and Morris, Q. (2013). GeneMANIA prediction server 2013 update. *Nucleic Acids Res.* 41, W115–22.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-V5-tag magnetic beads	MBL International Cat# M167-9	RRID: AB_10795284
anti-V5 antibody	Thermo Fisher Scientific Cat# R960-25	RRID: AB_2556564
anti- β -catenin	Abcam Cat# ab22656	RRID: AB_447227
anti-p- β -catenin	Cell Signaling Technology Cat# 9561	RRID: AB_331729
anti-Cul1	Abcam Cat# 75812	RRID: AB_1310104
anti-SKP1	Cell Signaling Technology Cat#2156	RRID: AB_2270271
anti- β -actin	Thermo Scientific Cat# MA5-15739	RRID: AB_10979409
Biological samples		
Human fetal brain total RNA	AMSBIO	CAT #636526
Human adult brain total RNA	AMSBIO	CAT #540143
Critical commercial assays		
QIAquick Gel Extraction Kit	QIAGEN	CAT #28704
Gateway BP Clonase II Enzyme mix	Invitrogen	CAT# 11789020
Gateway LR Clonase Enzyme Mix	Invitrogen	CAT# 11791019
RNAeasy Mini Kit	QIAGEN	CAT# 74104
Pierce-ECL Western Blotting Substrate Kit	Thermo Scientific	CAT# 32209
SuperScript III First Strand system	Invitrogen	CAT# 18080051
Deposited data		
BrainSpan gene/isoform expression data	PsychEncode Consortium	syn12080241, syn8241756
Processed summary-level BrainSpan data	PsychEncode Consortium	syn8466658
Original analysis code	GitHub	10.5281/zenodo.5091071
Single-Cell RNA-seq data	Polioudakis et al., 2019 ; Zhong et al., 2018	0.1016/j.neuron.2019.06.011; 10.1038/nature25980
Large exome sequencing variant data	Satterstrom et al., 2020	10.1101/484113
Mutational constraint genes	Samocha et al., 2014	10.1038/ng.3050
FMRP target genes	Darnell et al., 2011	10.1016/j.cell.2011.06.013
High risk ASD genes	Satterstrom et al., 2020	10.1016/j.cell.2019.12.036
CHD8 target genes	Wilkinson et al., 2015	10.1038/tp.2015.62
SynaptomeDB	Pirooznia et al., 2012	10.1093/bioinformatics/bts040
Mutation-intolerant genes	Lek et al., 2016	10.1038/nature19057
SFARI	https://gene.sfari.org/	N/A
Bioplex	Huttlin et al., 2015	10.1016/j.cell.2015.06.043
HPRD	Keshava Prasad et al., 2009	10.1093/nar/gkn892
InWeb	Li et al., 2017	10.1038/nmeth.4083
HINT	Das and Yu, 2012	10.1186/1752-0509-6-92
BioGRID	Chatr-Aryamontri et al., 2017	10.1093/nar/gkw1102
GeneMANIA	Zuberi et al., 2013	10.1093/bioinformatics/btu671
STRING	Szklarczyk et al., 2017	10.1093/nar/gkw937
CORUM	Ruepp et al., 2010	10.1093/nar/gkp914
Experimental models: Cell lines		
HeLa cells	ATCC	CAT #CRL-12401

(Continued on next page)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
All oligonucleotides used in the study are listed in Table S1 and in the STAR Methods section	N/A	N/A
Recombinant DNA		
pDESTSplice plasmid	Addgene	CAT #32484
pDON223.1 plasmid	Invitrogen	10.1101/gr.2505604
Software and algorithms		
R v3.6.3	The R Foundation	https://cran.r-project.org/
Surrogate Variable Analysis	Leek and Storey, 2007	https://bioconductor.org/packages/release/bioc/html/sva.html
limma v3.44.3	Ritchie et al., 2015	https://bioconductor.org/packages/release/bioc/html/limma.html
WGCNA: Weighted Correlation Network Analysis v1.69	Langfelder and Horvath, 2008	https://cran.r-project.org/web/packages/WGCNA/index.html
Variante Effect Predictor v96	McLaren et. al. 2016	https://github.com/Ensembl/ensembl-vep
gProfiler	Raudvere et al., 2019	https://biit.cs.ut.ee/gprofiler/
Bisque	Jew et al., 2020	10.1038/s41467-020-15816-6
variancePartition	Hoffman and Schadt, 2016	10.1186/s12859-016-1323-z
Custom code	N/A	10.5281/zenodo.5091071

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Lilia M lakoucheva (lilyak@ucsd.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Raw isoform-level RNA sequencing (RNA-seq) BrainSpan data are available at PsychENCODE Capstone Data Collection, <https://doi.org/10.7303/syn12080241>. The processed summary-level BrainSpan data are available at <http://Resource.PsychENCODE.org>. The code used for isoform RNA-seq data analysis generated during this study is available from GitHub (https://github.com/lakouchevaLab/Isoform_BrainSpan), <https://doi.org/10.5281/zenodo.5091071>. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Brain transcriptome data

RNA-Seq datasets quantified at the gene and isoform levels were downloaded from PsychENCODE Knowledge Portal, PEC Capstone Collection, Synapse ID: syn8466658 (<https://www.synapse.org/#!Synapse:syn12080241>). RNA-Seq from post-mortem brain tissue of 57 donors aged between 8 weeks post-conception to 40 years, across a number of different brain regions, for a total of 606 samples, has been carried out as previously described ([Li et al., 2018](#); [Figure S1B](#)).

Cell lines

HeLa cells were cultured using DMEM media (GIBCO) supplemented with 10% FBS (GIBCO) and 1x penicillin/streptomycin (GIBCO). Cells were maintained in 10 cm dishes at 37°C and 5% CO₂ until 60%–80% confluency reached. When confluent, cells were sub-cultured in 1:10 – 1:20 dilution. The cells from passage 3–20 were used for the lipofection experiments.

METHOD DETAILS

R version 3.6.0 was used throughout this analysis. Downstream bioinformatics analysis is outlined in [Figure S1A](#).

Processing of RNA-Seq gene and transcript BrainSpan data

RNA-seq data processing was performed as described ([Gandal et al., 2018](#)). Briefly, FASTQs were trimmed for adaptor sequence and low base call quality (Phred score < 30 at ends) using cutadapt (v1.12). Trimmed reads were then aligned to the GRCH37.p13 (hg19) reference genome via STAR (2.4.2a) using comprehensive gene annotations from Gencode (v19). BAM files were produced in both genomic and transcriptome coordinates and sorted using samtools (v1.3). Gene and isoform-level quantifications were calculated using RSEM (v1.2.29). Quality control metrics were calculated using RNA-SeQC (v1.1.8), featureCounts (v1.5.1), PicardTools (v1.128), and samtools (v1.3.1). Subsequently, TPM matrices for both gene and transcript datasets were filtered for TPM \geq 0.1 in at least 25% of samples, yielding a total of 100,754 isoforms corresponding to 26,307 genes.

Sample connectivity analysis was performed to detect sample outliers as previously described ([Oldham et al., 2012](#)). In brief, bi-weight mid-correlation was calculated among sample expression vectors in both filtered datasets. These values were converted into connectivity Z-scores. 55 samples were identified as having sample connectivity Z-scores \leq -2, and were removed from downstream analysis, resulting in 551 final samples.

Surrogate variable analysis (SVA) was performed to remove latent batch effects in the data, taking into consideration age, brain region, sex, ethnicity, and study site ([Leek, 2014](#); [Leek and Storey, 2007](#)). The number of surrogate variables was chosen to minimize apparent batch effects while avoiding overfitting based on evidence from principal components analysis and relative log expression ([Figures S2 and S3](#)). 16 surrogate variables were found to be sufficient for downstream analysis of both gene and transcript data.

Validation of isoform expression with RT-qPCR

RT-qPCR was used to estimate the relative expression difference of 44 unique isoforms of 32 genes between two independent RNA samples that were age, sex and brain region-matched to the samples from the BrainSpan; these results were then compared to the computationally-assigned BrainSpan values ([Table S1](#); [Figure S4A](#)). RNA from a frontal lobe tissue sample of a 22 weeks old male (fetal brain), and RNA from cerebral cortex tissue sample of a 27 years old male (adult brain) (AMSBIO, UK), corresponding to P06 (late mid-fetal) and P13 (young adult) in the BrainSpan data, was used. The isoforms were selected to encompass all possible ranges of differential expression - from high differential expression (high \log_2 FC, e.g., > 4), to low or no differential expression (low \log_2 FC, e.g., < 1) between the same isoforms from two samples from different time points. To further discriminate between isoform expression and absolute gene expression, isoforms were selected based on the presence of unique exons, for which the primers were designed ([Table S1](#)). The majority of selected isoforms were expressed with > 1 TPM values to ensure that they were detectable by the RT-qPCR. However, we also included 12 isoform pairs with at least one isoform being lowly expressed (< 1 TPM) to ensure that validation covers both highly and lowly expressed isoforms. This resulted in a total of 44 unique isoforms from 32 genes. Primers were designed using unique exon-exon junctions, specific for each of the selected isoforms. 3 μ g of RNA using SuperScript II Kit (Invitrogen) were reverse transcribed to cDNA, following manufacturer's instructions. Then, the cDNA was diluted ten times to use as a template for the RT-qPCR reaction. SYBR Green II Master Mix (Invitrogen) was used for the RT-qPCR reaction, performed in a CFX Connect 96X Thermal Cycler, using standard parameters for SYBR Green. Relative expression between each isoform in the two samples was calculated by normalizing each expression value against two housekeeping genes (*RPL28* and *MRSP36*) as control using QIAGEN control primers, and the $\Delta\Delta t$ method was applied using the CFX Manager Software. Comparison of these relative expression values against the BrainSpan computational expression values resulted in positive correlation ([Figure S4A](#)).

Differential gene and isoform expression analysis

Differential gene and isoform expression analysis was performed using the *limma* (v3.40.6) R package ([Ritchie et al., 2015](#)). Relevant covariates (brain region, sex, ethnicity, and study site) and surrogate variables were included in the linear model as fixed effects. The *duplicateCorrelation* function was used to fit the donor identifier as a random effect to account for the nested expression measurements due to multiple brain regions derived from the same donor. Genes and isoforms with an absolute fold change of \geq 1.5 and FDR-adjusted p value of \leq 0.05 between adjacent developmental periods, or between prenatal and postnatal periods (PrePost) were defined as significantly differentially expressed.

Cell type and literature curated gene sets enrichment analyses

Fisher's exact tests were performed on gene lists and isoform lists (converted to gene identifiers) against curated gene lists: mutational constraint genes (Mut. Const. Genes) ([Samocha et al., 2014](#)), FMRP target genes ([Darnell et al., 2011](#)), high-risk ASD genes (Satterstrom ASD) ([Satterstrom et al., 2020](#)), CHD8 target genes ([Wilkinson et al., 2015](#)), synaptic genes (Synaptome DB) ([Pirooznia et al., 2012](#)), genes intolerant to mutations (Pli_0.99) ([Lek et al., 2016](#)), syndromic and rank 1 and 2 ASD risk genes (SFARI_S_1_2) (<https://gene.sfari.org/>). Cell types were extracted from a recent gene-level single cell sequencing study ([Zhong et al., 2018](#)).

Gene Ontology (GO) functional enrichment analyses

Functional enrichment analysis was performed using the *gprofiler2* v0.1.5 R package (Raudvere et al., 2019). Ensembl gene or isoform (converted to gene) identifiers were used to test for enrichment in two Gene Ontology categories, Biological Processes (BP) and Molecular Functions (MF). Enrichment p values were Benjamini-Hochberg corrected for multiple hypothesis testing, and overly general terms (i.e., terms with more than 1,000 members) were filtered out.

Rare *de novo* ASD loss-of-function variants

Rare *de novo* variant data was downloaded from Satterstrom et al. (2020), and was processed using Ensembl's Variant Effect Predictor v96 (McLaren et al., 2016) using human genome version GRCh37 to annotate variants for predicted functional consequences. Loss-of-function (LoF) variants were defined as those impacting essential splice donor/acceptor sites, frameshift insertions or deletions, predicted start losses, and predicted stop gains.

Weighted gene/isoform co-expression network analyses (WGCNA)

Co-expression networks were constructed using the *WGCNA* (v1.68) R package (Langfelder and Horvath, 2008). Relevant covariates (brain region, sex, ethnicity, and study site) and surrogate variables were first regressed out of both gene and isoform expression datasets using linear mixed effects models. Each transformed expression matrix was then tested for scale-free topology to estimate a soft-thresholding power, *beta*. We used *beta* = 2 for gene co-expression and *beta* = 3 for isoform co-expression networks, and signed networks were constructed blockwise using a single block for the gene network and three blocks for the isoform network with *deepSplit* = 2 and *minModuleSize* = 20 for module detection in both networks. Both networks followed scale-free topology (Figures S4C and S4D).

Co-expression module characterization

Module eigengene and developmental period association analysis was performed using linear mixed effects models, considering fixed effects (age, brain region, sex, ethnicity, and study site) and random donor effects to account for multiple brain region samples per donor. Module enrichment analysis was performed using Fisher's exact tests against curated gene lists; isoform identifiers within modules were converted to gene identifiers for this purpose. Gene Ontology functional enrichment analysis for modules was performed using *gprofiler2* with queries ordered by module membership (kME) rank.

Decomposition of bulk BrainSpan data with fetal brain single cell RNA-seq

Bisque (Jew et al., 2020) was used to decompose bulk RNA-seq BrainSpan data to estimate the cell type proportions. Single-cell RNA-seq data of human fetal prefrontal cortex from two studies (Polioudakis et al., 2019; Zhong et al., 2018) was used to compute the reference profile. BrainSpan data was compared to the reference profile to estimate the cell type proportion in each sample. Variance partition (Hoffman and Schadt, 2016) was used to investigate cell type proportions (Figures S5A and S5B).

Variant impact analysis for co-expression modules

We quantified the impact of rare *de novo* LoF case and control variants on co-expression modules by mapping variants to the members of each module. To account for gene and isoform sizes in each module, we calculated the non-overlapping length, in base pairs, of the members of each module and normalized the module impact rates by these genomic coverages. Further, to control for total mutation rate, we scaled the impact rate by the total number of variants. We further added a scaling factor of 1,000,000 to make the numbers more manageable, in a similar fashion to a TPM calculation. Differences in the impact rates between case and control variants for each module were tested using permutation test with 1,000 iterations of module member resampling (controlling for length and GC content, $\pm 10\%$ for each attribute) (Table S12). Modules impacted by significantly more case mutations were identified.

Integration of protein-protein interaction networks with co-expression modules

Gene-level PPI network data was manually curated and filtered for physical and co-complex interactions extracted from Bioplex (Huttlin et al., 2015), HPRD (Keshava Prasad et al., 2009), Inweb (Li et al., 2017), HINT (Das and Yu, 2012), BioGRID (Chatr-Aryamontri et al., 2017), GeneMANIA (Zuberi et al., 2013), STRING (Szklarczyk et al., 2017), and CORUM (Ruepp et al., 2010). To build co-expressed PPI networks, gene and isoform modules were first filtered for connections (i.e., edges) supported by the gene-level PPIs; isoform edges were retained if corresponding gene edges were supported by PPIs. Subsequently, networks were filtered to only retain edges supported by the top 10% of co-expression PCCs between genes or isoforms. Genes or isoforms without any connections were removed from the networks.

Minigenes cloning

The following genes impacted by rare *de novo* splice site mutations identified in patients with NDDs were selected for the experiments: *SNC2A* (chr2:166187838, A:G, acceptor site) (Fromer et al., 2014); *DYRK1A* (chr21: 38865466, G:A, donor site) (O'Roak et al., 2012), *CELF2* (chr10: 11356223, T:C, donor site) (Xu et al., 2011), *DLG2* (chr11: 83194295, G:A, donor site) (Fromer et al.,

2014) and *BTRC* (chr10: 103221816, G:A, donor site) (De Rubeis et al., 2014). The exons of these genes that are likely impacted by splice site mutations, together with the ~1kb of their flanking intronic sequence, were cloned. The constructs were cloned into pDESTSplice exon trapping expression vector (Kishore et al., 2008). The site-directed mutagenesis by two-step stich PCR was performed to introduce the mutation affecting the splice site.

The minigenes were generated by PCR-amplifying the desired sequences from genomic DNA (Clontech). Primers were designed for each minigene, and attB sites were added at the 5' end of the primers. The sequences of the primers were as follows: (1) *SCN2A*; Fw: GGAAGCTATGTTTAGCCAGGATACATTTGG, Rv: CCAGATGATGTCCCCTCCCTACATAGTCC; (2) *DYRK1A*: Fw: GTTGGGAA AATTTCCCCTATTTAAGC, Rv: CCCAGAGGCTTAATAAAGTATGGACC; (3) *CELF2*: Fw: GGAGTTGGAATGACAGACGTTCCACA TGC, Rv: CCGCTGTGGGCTGAGGATCAGTTTCC; (4) *DLG2*: Fw: GAGGTTCCAGAGACATTCAATTCCC, Rv: CTTGATGCTGTCCAGA TAATGC; (5) *BTRC*: Fw: GGGCTCAGAATGACACAGTACG, Rv: GAACCTGCGTTTCTGTTTTTGCC. After PCR amplification, amplicons were loaded in a 1% low EEO agarose gel (G-BioSciences) and purified using the QIAquick Gel Extraction Kit (QIAGEN) following manufacturer's instructions. Purified amplicons were subcloned into the pDON223.1 expression vector using the BP-Gateway System (Invitrogen). At least six different clones for each minigene were sequenced to verify correct sequences of the minigenes. The clone with the desired sequence and highest DNA concentration was used for subcloning into the pDESTSplice expression vector (Addgene) using the LR-Gateway System (Invitrogen).

Exon trapping and RT-PCR

HeLa cells were seeded at 2×10^5 cells per well in 6-well plates (Falcon). After 24h, cells were transfected using Lypofectamine 3000 (Invitrogen) following manufacturer's instructions, and then harvested after additional 24h. RNA was purified using the RNAeasy Mini Kit (QIAGEN) following manufacturer's instructions. Two μg of RNA was used to generate cDNA using the SuperScript III First Strand kit (Invitrogen), and PCR was carried out. In the case of exon trapping assays, we used primers specific for the rat insulin exons constitutively present in the pDESTSplice vector: Fw: CCTGCTGGCCCTGCTCA, Rv: TAGTTGCAGT AGTTCTCCAGTTGG. In the case of the *BTRC* RT-PCR, we used primers specific for 5' and 3' sequences of the *BTRC* gene. Amplicons were loaded into the agarose gel (G-BioSciences) and visualized using Gel-Doc XR+ Imaging System (Bio-Rad).

Co-Immunoprecipitation and Western Blot

HeLa cells were harvested and rinsed once with ice-cold 1xPBS, pH 7.2, and lysed in immunoprecipitation lysis buffer (20 mM Tris, pH 7.4, 140 mM NaCl, 10% glycerol, and 1% Triton X-100) supplemented with 1xEDTA-free complete protease inhibitor mixture (Roche) and phosphatase inhibitor cocktails-II, III (Sigma Aldrich). The cells were centrifuged at $16,000 \times g$ at 4°C for 30min, and the supernatants were collected. Protein concentration was quantified by modified Lowry assay (DC protein assay; Bio-Rad). The cell lysates were resolved by SDS-PAGE and transferred onto PVDF Immobilon-P membranes (Millipore). After blocking with 5% nonfat dry milk in TBS containing 0.1% Tween 20 for 1hr at room temperature, membranes were probed overnight with the appropriate primary antibodies. They were then incubated for 1h with the species-specific peroxidase-conjugated secondary antibody. Membranes were developed using the Pierce-ECL Western Blotting Substrate Kit (Thermo Scientific).

For immunoprecipitation experiments, samples were lysed and quantified as described above. Then, 3 mg of total protein was diluted with immunoprecipitation buffer to achieve a concentration of 3 mg/ml. A total of $30 \mu\text{l}$ of anti-V5-magnetic beads-coupled antibody (MBL) was added to each sample and incubated for 4h at 4°C in tube rotator. Beads were then washed twice with immunoprecipitation buffer and three more times with ice cold 1xPBS. The proteins were then eluted with $40 \mu\text{l}$ of 2xLaemli buffer. After a short spin, supernatants were carefully removed, and SDS-PAGE was performed. The following primary antibodies were used: anti-V5 (1:1000; Invitrogen), anti- β -catenin (1:1000; Abcam), anti-p- β -catenin (1:1000; Cell Signaling), anti-Cul1 (1:1000; Abcam), anti-SKP1 (1:1000; Cell Signaling), and anti- β -actin (1:10000; Thermo Scientific).

QUANTIFICATION AND STATISTICAL ANALYSIS

In all Figures, the “***” represents statistically significant result, with significance defined as $p < 0.05$ unless otherwise stated. All *p-values* underwent FDR multiple hypothesis correction where appropriate unless otherwise stated. Gene Ontology enrichment analyses were performed using *gProfiler*, with overly general terms (terms with more than 1,000 members) filtered out; enrichment analyses of curated gene lists were performed with Fisher's exact tests. Additional statistical analyses details including *p* values could be found in Figure legends and in the Results section of the manuscript.

Expression validation by RT-qPCR

Expression quantification was validated through RT-qPCR of 44 unique isoforms from 32 genes for two age-, sex-, and region-matched brain samples. Relative qPCR expression was compared against the BrainSpan expression level for each of one 22-week-old male frontal lobe fetal sample and one 27-year-old male cerebral cortex adult sample. Expressions of each isoform were compared using Pearson correlation analysis. Results can be found in Table S1 and Figure S4A.

Variant impact analysis

Co-expression modules were analyzed for their rate of impact by LoF variants. This metric (Variants per Million, VPM) was calculated for each module as the number of LoF variants impacting the module members divided by the non-overlapping length of member genes/isoforms, scaled by the total number of LoF variants times 1,000,000. Statistical analysis of these module VPMs was performed using permutation tests for each module, comparing VPM of case variants against VPM of control variants, with $n = 1,000$ permutations. Genes and isoforms were sampled in each permutation controlling for length and GC content, $\pm 10\%$ for each attribute. Results of this analysis can be found in [Figure 3D](#) and [Table S12](#).

Co-expression module characterization

Co-expression modules were characterized by their association with specific developmental periods. This was done using linear mixed effects models, fixing age, brain region, sex, ethnicity, and study site and designating random donor effects to account for multiple samples per donor in the dataset. Modules were further characterized through module enrichment analysis by Fisher's exact tests against curated gene lists and Gene Ontology enrichment with *gProfiler*, with modules ordered by their kME rank. Results can be found in [Figure 3](#) and [Tables S10](#) and [S11](#).