Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms

Ashok Patowary[1-4,+], Pan Zhang[1-4,+], Connor Jops[5-7,+], Celine K. Vuong[1-3], Xinzhu Ge[8], Kangcheng Hou[9], Minsoo Kim[1,2,4], Naihua Gong[5], Michael Margolis[4], Daniel Vo[5-7], Xusheng Wang[10-11], Chunyu Liu[12-13], Bogdan Pasaniuc[4,9,14-16], Jingyi Jessica Li[4,8,9,16-17], Michael J. Gandal[1-7,*], Luis de la Torre-Ubieta[1-3,*]

Affiliations

[1] Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.
[2] Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA 90095, USA.
[3] Intellectual and Developmental Disabilities Research Center, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA 90095, USA.
[4] Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.
[5] Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.
[6] Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.
[7] Lifespan Brain Institute at Penn Med and the Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.
[8] Department of Statistics, University of California, Los Angeles, CA 90095, USA.
[9] Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA 90095, USA.
[10] Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, Tennessee 38103, USA
[11] Center for Proteomics and Metabolomics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA.
[12] Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY 13210, USA.
[13] Center for Medical Genetics & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University; Changsha, Hunan, 410008, China
[14] Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles; Los Angeles, CA 90095, USA.
[15] Institute for Precision Health, University of California, Los Angeles; Los Angeles, CA 90095, USA.
[16] Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles; Los Angeles, CA 90095, USA
[17] Department of Biostatistics, University of California, Los Angeles; Los Angeles, CA 90095, USA

+ These authors contributed equally to this work
* Corresponding authors: ldelatorreubieta@mednet.ucla.edu, michael.gandal@pennmedicine.upenn.edu

Abstract:

RNA splicing is highly prevalent in the brain and has strong links to neuropsychiatric disorders, yet the role of cell-type-specific splicing or transcript-isoform diversity during human brain development has not been systematically investigated. Here, we leveraged single-molecule long-read sequencing to deeply profile the full-length transcriptome of the germinal zone (GZ) and cortical plate (CP) regions of the developing human neocortex at tissue and single-cell resolution. We identified 214,516 unique isoforms, of which 72.6% are novel (unannotated in Gencode-v33), and uncovered a substantial contribution of transcript-isoform diversity, regulated by RNA binding proteins, in defining cellular identity in the developing neocortex. We leveraged this comprehensive isoform-centric gene annotation to re-prioritize thousands of rare de novo risk variants and elucidate genetic risk mechanisms for neuropsychiatric disorders.

One-Sentence Summary:

A cell-specific atlas of gene isoform expression helps shape our understanding of brain development and disease.

Structured Abstract

INTRODUCTION
The development of the human brain is regulated by precise molecular and genetic mechanisms driving spatio-temporal and cell-type-specific transcript expression programs. Alternative splicing, a major mechanism increasing transcript diversity, is highly prevalent in the human brain, influences many aspects of brain development, and has strong links to neuropsychiatric disorders. Despite this, the cell-type-specific transcript-isoform diversity of the developing human brain has not been systematically investigated.

RATIONALE
Understanding splicing patterns and isoform diversity across the developing neocortex has translational relevance and can elucidate genetic risk mechanisms in neurodevelopmental disorders. However, short-read sequencing, the prevalent technology for transcriptome profiling, is not well suited to capturing alternative splicing and isoform diversity. To address this, we employed third-generation long-read sequencing, which enables capture and sequencing of complete individual RNA molecules, to deeply profile the full-length transcriptome of the germinal zone (GZ) and cortical plate (CP) regions of the developing human neocortex at tissue and single-cell resolution.

RESULTS
We profiled microdissected GZ and CP regions of post-conception week (PCW) 15-17 human neocortex in bulk and at single-cell resolution across six subjects using high-fidelity long-read sequencing (PacBio IsoSeq). We identified 214,516 unique isoforms, of which 72.6% were novel (unannotated in Gencode), and >7,000 novel exons, expanding the proteome by 92,422 putative proteoforms. We uncovered thousands of isoform switches during cortical neurogenesis

predicted to impact RNA regulatory domains or protein structure and implicating previously uncharacterized RNA-binding proteins in cellular identity and neuropsychiatric disease. At the single-cell level, early-stage excitatory neurons exhibited the greatest isoform diversity, and isoform-centric single-cell clustering led to the identification of previously uncharacterized cell states. We systematically assessed the contribution of transcriptomic features, and localized cell and spatio-temporal transcript expression signatures across neuropsychiatric disorders, revealing predominant enrichments in dynamic isoform expression and utilization patterns and that the number and complexity of isoforms per gene is strongly predictive of disease. Leveraging this resource, we re-prioritized thousands of rare de novo risk variants associated with autism spectrum disorders (ASD), intellectual disability (ID), and neurodevelopmental disorders (NDDs), more broadly, to potentially more severe consequences and revealed a larger proportion of cryptic splice variants with the expanded transcriptome annotation provided in this study.

CONCLUSION

Our study offers a comprehensive landscape of isoform diversity in the human neocortex during development. This extensive cataloging of novel isoforms and splicing events sheds light on the underlying mechanisms of neurodevelopmental disorders and presents an opportunity to explore rare genetic variants linked to these conditions. The implications of our findings extend beyond fundamental neuroscience, as they provide crucial insights into the molecular basis of developmental brain disorders and pave the way for targeted therapeutic interventions. To facilitate exploration of this dataset we developed an online portal (https://sciso.gandallab.org/).

Main Text

Human brain development is a tightly coordinated process under precise molecular and genetic control. Transcriptomics has provided substantial insights into the cellular and molecular processes regulating human brain development (*1*, *2*), including characterization of the underlying heterogeneous cell-types, states, and lineages through single-cell RNA-sequencing (scRNA-Seq) (*3–5*) as well as their developmental trajectories (*6*), gene regulatory networks, and cell-to-cell variability (*7*). However, the technological limitations of short-read scRNA-Seq have largely prevented systematic characterization of the full complexity of cell-type-specific alternative splicing (AS) and resulting isoform diversity present during neurodevelopment (*8*).

AS is a fundamental form of tissue-specific gene regulation present in >90% of multi-exon genes (*9*). RNA transcript diversity results from the combinatorial effects of alternative transcription start sites (TSS), regulated exon inclusion/exclusion or intron retention, and distinct transcript termination sites due to alternative polyadenylation (APA) (*10*). In individual cases, synaptic genes such as *Nrxn1* have thousands of unique isoforms (*11*). Human brain-expressed genes, which are longer and contain the most exons, undergo the greatest degree of splicing compared with other tissues and species - a mechanism contributing to the vast proteomic, phenotypic, and evolutionary complexity of the human brain (*12*, *13*). Indeed, AS plays an important role in synaptogenesis, synapse specification (*14*) and brain development more broadly (*15*). Substantial cell-type-specificity of splicing and isoform diversity has been observed in the mouse brain, even among closely related neurons, and often with precise temporal regulation (*16–20*).

AS has been implicated as a critical mechanism linking genetic variation and neuropsychiatric disease (*2*, *21–24*). For example, recent work identified marked splicing and isoform expression dysregulation in the brains of individuals with ASD and schizophrenia (SCZ) - a signal substantially more widespread than gene expression changes with greater enrichment for genetic risk (*25–27*). Common variants associated with ASD and SCZ from genome-wide association studies (GWAS) showed substantial enrichment among splicing-QTLs in the developing and adult human brain (*2*, *24*). Likewise, rare de novo variants identified in ASD and ID are enriched for predicted splice-altering consequences (*23*). These genetic signals also exhibited convergence during mid-fetal brain development, highlighting the clinical relevance of this critical developmental period characterized by rapid increases in neurogenesis (*28*, *29*). Finally, a comprehensive understanding of splicing and isoform complexity can have direct therapeutic relevance, as demonstrated recently for spinal muscular atrophy (*30*).

The advent of third-generation, long-read sequencing technologies has enabled accurate, in-depth characterization of the full-length alternatively-spliced transcriptome at scale (*10*, *31*). Coupled with single-cell barcoding, recent work has begun to catalog the isoform-centric transcriptome with single-cell resolution (*20*, *32*). Here, we leverage this approach to deeply profile the major cell-types of the developing human neocortex at mid-gestation. We uncover >150k previously unannotated transcript-isoforms and thousands of spliced exons, many within high-confidence NDD risk genes. To functionally annotate these isoforms, we integrated proteomics, characterized regional patterns of isoform-switching during corticogenesis, embedded isoforms within coexpression networks, and mapped transcripts to 16 distinct cell-type clusters. Altogether, results highlight the tremendous complexity of transcript-isoform diversity during

neurodevelopment, which we leverage to uncover mechanisms of cell-fate specification and genetic risk mechanisms for NDDs.

Results

Full-length transcriptome of the developing human brain

We performed high-depth, high-fidelity long-read sequencing (PacBio HiFi IsoSeq) to comprehensively profile the full-length polyadenylated transcriptome of the developing human neocortex across six donors at mid-gestation (15-17 PCW) (**Fig. 1A**). To interrogate patterns of differential transcript expression and usage during peak neurogenesis, samples were first microdissected into the neural progenitor-enriched germinal zone (GZ) and the neuron-enriched cortical plate (CP). Following sequencing and comprehensive quality control (**Fig. 1B-C)**, we generated >33 million high-quality (>Q20) circular consensus sequence (CCS) reads across samples (**fig. S1-2).** Using minimap2 (*33*), >99% of full-length reads were confidently aligned to the reference genome, a marked improvement over the ~85% mapping rate of short-read RNA-seq (**fig. S2A-B)**. Using the TALON pipeline (*34*), 214,516 unique transcripts were found in the bulk tissue transcriptome (**data S1**), corresponding to 24,554 genes; of these, >175k isoforms from 17,299 genes were expressed at >0.1 transcripts per million (TPM) in ≥50% of the samples (**Fig. 1D-E)**. Isoform-level expression exhibited reproducibility across technical and biological replicates (**Fig. 1B** and **fig. S2C**), with clear separation of CP and GZ samples (**Fig. 1C**). The median read length was 2.99 kb (range: 80-14200 bp), consistent with the high RNA quality and expected distribution of mammalian mRNA transcripts (**Fig. 1F).**

We next generated a high-quality, custom reference annotation of the developing neocortex transcriptome (**table S1**), merging data across samples. Compared with Gencode v33, only 65,006 (30.3%) of observed isoforms matched existing transcripts (**Fig. 1D-E**). We further classified isoforms based on their splice junction match to Gencode, strand specificity, and 5' and/or 3' overlap with known transcripts (*35*) (**fig. 1D**; **table S1**). As a class, novel transcripts were more lowly expressed than known transcripts **(Fig. 1G**), although several were individually highly expressed, including *WASF1*, *CYFIP2*, *MAP1B*, *NEFL*, and *SMARCA4* (**fig. S3A-E**).

Most isoforms not found in Gencode were classified as incomplete splice match (ISM; **fig 1D-E**). Although ISM transcripts are often disregarded as artifacts of RNA degradation or internal priming, gene body coverage did not show evidence of 3' bias (**Fig. 1F; fig. S2A**). Further, novel transcripts (including ISMs) were significantly longer (P<2e-16, Kruskal-Wallis; **Fig. 1H**) and contained more exons (P<2e-16, Kruskal-Wallis; **Fig. 1I)** compared with known transcripts. Consequently, we retained ISM transcripts with additional supporting evidence (*36*), and note that many exhibited functional roles as "hub" isoforms in network analyses detailed below.

Finally, given this large number of isoforms detected, we sought to characterize their patterns of usage and potential functional relevance to NDDs. For multi-isoform genes, we calculated the proportion of gene expression attributable to the top (dominant) isoform, stratified by total gene abundance (**Fig. 1J**). Whereas lowly expressed genes exhibited a relatively even distribution, for

highly expressed genes, the dominant isoform tended to capture the majority of expressed reads. Nevertheless, the number of detected isoforms per gene was strongly predictive of NDD risk gene status, even accounting for gene length and total gene expression (P=5.2e-03, logistic regression; **Fig. 1K**). This association remained significant even when restricting only to novel transcripts (OR 1.8, P=5.5e-04, **fig. S3F**).

Expanded transcriptomic and proteomic complexity in the developing human brain

To contextualize this expanded transcriptomic complexity in the developing brain, we next integrated several orthogonal genomic annotations (*37–43*) and predicted potential downstream protein coding consequences (**Fig. 2)**. Altogether, ~80% of novel isoform TSS's were supported by proximal, annotated CAGE and/or ATAC-Seq peaks. At the 3' end, ~91% of novel transcripts were supported by nearby polyA sites/motifs–a rate higher than for known transcripts (**Fig. 2A-B**). Of 38,115 splice junctions not observed in Gencode, 74% were validated by Intropolis (**Fig. 2C**). Finally, 53% of novel transcripts were validated when combining the latest version of Gencode (v43) with multiple independent long-read datasets, despite the majority representing non-neural tissues (**fig. S3G**). Together, these results provided broad orthogonal support for novel isoforms discovered in this study.

In total, compared with Gencode, our data extended by ~27Mb the transcribed portion of the human genome. Of this, 3.85 Mb comprised >7000 previously unannotated spliced exons, spanning >3500 unique genes (**Fig. 2D**). Nearly 80% of these exons were supported by canonical splice junctions **(table S2**) and we successfully validated several using RT-PCR (**Fig. 2E**; **table S2)**. Additionally, we identified 319 multi-exonic genes not matching any existing gene model (**table S1**), of which 256 were antisense to existing genes and 63 were located within intergenic regions. For example, we identified an 18-exon gene with 4 unique splice isoforms antisense to the ASD risk gene *DPYSL2* (**fig. S3H**).

We next sought to determine the functional relevance of this expanded compendium of novel exons and transcripts at the protein-coding level (**Fig. 2F)**. Of the novel transcripts observed, 92,422 exhibited protein coding potential, with at least one complete putative open reading frame (ORF). Integration of human brain proteomics data (*44*) provided peptide-level support for 35,467 novel transcripts with unique proteoforms (**table S2**). For example, a novel, alternatively spliced 30 bp microexon found in the NDD risk gene *NF1* was predicted to add an additional 10 amino acids (AA) to the known protein (**Fig. 2A**). Searching against mass-spectrometry proteomics data, we identified spectra (n=15) that confidently supported the existence of this 10 AA sequence (**Fig. 2G**). Notably, isoforms containing this microexon were only detected in neuronally-enriched CP samples (**fig. S4A**). Extending these analyses beyond a single gene locus, we observed broad peptide-level support for novel transcripts across all classification categories (**Fig. 2H**).

Isoform Switching during Neurogenesis

We next sought to contrast gene and isoform usage across GZ/CP samples to identify genes exhibiting isoform-switching during neurogenesis and/or neuronal maturation. Whereas differential gene expression (DGE) signatures of neuronal maturation have been extensively characterized in the developing human brain at bulk tissue and single-cell resolution (*3*, *4*, *37*), AS and isoform usage have not been investigated transcriptome-wide during this critical developmental period.

Consistent with previous work (*3*), many genes exhibited significant DGE patterns between GZ and CP (4,475 of 24,554 genes, at FDR <0.05; **table S3**). Likewise, of genes with multiple expressed isoforms, a large proportion exhibited differential transcript usage (DTU) across GZ and CP (2,679 of 10,809 genes at FDR <0.05; **Fig. 3A**; **table S3**), with the majority (57%) of significantly switching isoforms (5,630 isoforms at FDR<0.05) coming from novel transcripts (**Fig. 3B**). Although there was a significant overlap among DGE and DTU genes (1,010 genes, OR=2, $P<10^{-46}$, Fisher's exact test; **Fig. 3A**), 1,669 genes exhibited isoform switching without changes in overall expression. The majority of GZ/CP isoform switching events had observed and/or predicted functional consequences (**Fig. 3C)**, summarized in **table S3**. Of 5,630 isoform switching events, 3,204 (57%) were predicted to alter the ORF length, with longer ORFs predicted among CP-upregulated isoforms (FDR-corrected P<0.0001). However, CP-upregulated isoforms were also more likely to exhibit predicted nonsense-mediated decay (NMD) sensitivity (243 vs 100 isoforms; FDR-corrected $P<10^{-15}$).

To further interrogate how the 3' untranslated region (UTR) of transcripts differed between regions, we performed a complementary, read-depth based analysis of distal polyadenylation (polyA) usage (**Fig. 3D**). For the 9,896 transcripts with multiple annotated polyA sites, we used DaPars2 to compute a distal polyA usage index (DPUI) - the fraction of total reads mapping to the longer (distal) 3'UTR (*45*). In total, 1,013 transcripts exhibited significant differences in DPUI between GZ and CP (repeated measures ANOVA; FDR-corrected P < 0.05; **Fig. 3D; table S3**). DPUIs were increased in CP versus GZ for the majority (772) of transcripts, and DPUI was on average greater in the CP (two-sample paired T-test, p < 2.2e-16), indicating overall increased 3'UTR lengths with neuronal maturation. Pathway enrichment among genes exhibiting GZ/CP DPUI differences was notable for RNA/mRNA binding as well as cytoplasmic stress granule gene ontologies. Finally, we found RNA-binding proteins (RBPs) (*46*) were overrepresented among transcripts exhibiting significant DPUI changes (453/3527 RBP genes vs 560/6369 non-RBP genes; one-tailed Fisher's Exact Test, p = 2.06e-10). These findings indicated that 3'UTR lengthening in the CP is prevalent among genes encoding RBPs, suggesting that tight regulation of RBP activity is important for the transition from neural progenitor to neuron.

We next conducted pathway analyses for all genes exhibiting significant (DTU, one-sided Fisher's exact test, P<0.05, **table S3A**) isoform switches across regions. Enriched biological pathways included dendrite morphogenesis, cadherin binding, and the chromatin modifier nBAF and SWI/SNF complexes, known to harbor convergent genetic risk for neuropsychiatric disorders (**Fig. 3E**; (*47*, *48*)). For example, among the top genes was *SMARCC2*, a high-confidence ASD risk gene encoding a subunit of the SWI/SNF chromatin remodeling complex (**Fig. 3F**). Although overall expression of *SMARCC2* did not differ between CP and GZ, two newly identified isoforms of *SMARCC2* showed preferential usage in GZ and exhibited exon skipping compared with known transcripts. Other notable isoform-switching genes included

several known splicing regulators and RBPs (*SRRM1*, *SRRM4*, *CELF1*, *PTBP2*, *ELAVL2*, *ELAVL4*, *RBFOX2*), chromatin modifiers (*KMT2E*, *KMT5B*, *SMARCA1*, *SMARCD3*, *SMARCE1*), transcription factors (*FOXP2*), regulators of synaptic transmission (*GRIA3*, *VAMP1*, *GAD1*), and synaptic adhesion molecules (*NLGN4X*, *NRXN1*). Isoform-switching genes were broadly expressed across cell-types, with particular enrichment for excitatory neuron lineages (**Fig. 3G**).

Putative RNA-binding protein regulators of isoform switching events

RBPs are a diverse class of proteins regulating the processing and fate of target mRNAs. Through regulation of AS, RBPs alter isoform and consequently protein diversity and play important roles in mammalian neural development and function (*17*, *49*). To identify potential RBP regulators of isoform switching events in the developing human brain, we assessed DGE, DTE, and DTU genes for enrichment for RBP targets identified through AS changes or direct binding by cross-linking immunoprecipitation (CLIP) using two datasets – 1) targets of RBPs that regulate AS during neural development and/or maturation curated from previous work ("brain-enriched"; black bar, **Fig. 3H and fig. S4B-C**), and 2) targets defined by systematic profiling of RBPs in the ENCODE repository (gray bar; **Fig. 3H and fig. S4B-C**).

Examining isoform diversity, we found that DTU genes significantly overlapped with targets of known regulators of AS in the brain (one-sided Fisher's exact test, FDR-corrected $P < 0.05$, **Fig. 3H;** (*49*)). Splicing targets of brain-enriched RBPs are more enriched in DTU genes than DTE/DGE genes – consistent with known roles of these RBPs as regulators of AS in the brain (*49*). Correspondingly, gene expression targets, rather than AS targets, were more enriched in DGE and DTE than DTU genes (**fig. S4B**). We also observed that gene expression targets of RBPs that are more highly expressed in neurons than progenitors (*SRRM3/4*; (*49*)) are enriched in genes that increase in expression over neural development, while those that are targets of progenitor-associated RBPs (*SAM68*, *PTBP1*; (*49*)) show the opposite trend (**fig. S4C**).

To identify RBPs important for but not previously studied during the developmental transition from progenitor to neuron, we performed enrichment analysis of DTU, DTE, and DGE genes compared to a comprehensive set of RBP targets compiled from ENCODE (**Fig. 3H**; (*50*)). Similar to our results comparing target genes of brain-enriched RBPs, we found that ENCODE RBP targets were more enriched among DTU genes compared to DTE/DGE. Within DTU genes, ENCODE RBPs targeting introns, 5'/3' splice sites, and 3' splice sites were more enriched than those targeting the 3'UTR/CDS or 3'UTRs (**Fig. 3H**; dark green, magenta, and aqua boxes vs. green and dark blue boxes). This result matches the overall trend for RBP target enrichment in DTU over DGE, the latter of which is often mechanistically regulated through 3' UTR binding.

Of the ENCODE RBPs showing significant target overlap with DTU genes (one-sided Fisher's exact test, FDR-corrected $P < 0.05$, **Fig. 3H**), several have been increasingly recognized to play critical roles in neural development and disease. These included RBPs known to regulate RNA metabolism and splicing such as *LIN28B*, *EFTUD2*, *KHSRP*, and *DGCR8* (*51–53*). We observed a strong enrichment between DTU genes and targets of *DDX3X*, an X-linked RNA helicase where de novo mutations lead to sexually dimorphic ID and ASD (*54*, *55*). We also found strong

enrichment of DTU genes among RBP targets with known roles in RNA metabolism, but which have not yet been studied in the context of neural development. For example, targets bound by components of the exosome (the RNA degradation system) and those involved in rRNA biogenesis, such as *EXOSC5*, *UTP18*, and *SUPV3L1* (*56–58*), and the nuclear matrix protein *SAFB*, implicated in heterochromatin regulation (*59*), are enriched in DTU genes. Altogether, these results indicate that while many of the GZ/CP isoform switching events are likely regulated by brain-enriched RBPs through AS, many more switching events are expected to be produced through diverse mechanisms regulated by RBPs previously not known to function in neural development.

Network Context of Developmental Isoform Regulation

Given the large number of isoform switching events, we next leveraged weighted gene correlation network analysis (WGCNA) to place these results within a systems-level context during human brain development (*60*, *61*). We separately built unsupervised co-expression networks for gene (geneExpr), isoform expression (isoExpr), and transcript usage quantifications (isoUsage), the proportion of each gene's total abundance attributable to a given isoform. For each network, genes or isoforms were assigned to modules based on shared patterns of covariation across samples (**fig. S5**; **table S4**), enabling in silico deconvolution of cell-type isoform usage as well as functional inference through pathway analysis (**fig. S5**).

We next compared properties across the three networks to better understand which factors drove differential transcript/gene expression and transcript usage. We observed that while both geneExpr and isoExpr networks were strongly driven by cell-type identity, isoExpr modules were enriched for specific progenitor and neuron subtypes (**Fig. 4A** and **fig. S6**). Multiple modules across isoExpr/isoUsage networks were enriched for RNA processing, cytoskeletal function and chromatin regulation pathways (**fig. S7A-D**). We further identified several disease-associated modules including isoExpr.M11, where a novel isoform of the ASD risk gene *DDX3X* is a hub transcript (**fig. S7E**). Together, these results show that isoform-level transcript expression further refines the resolution of cell type-specific modules and that isoform regulation is important for neurogenesis.

In contrast to geneExpr and isoExpr networks, the isoUsage network did not display strong cell type-specific enrichments (**Fig. 4A** and **fig. S6)**, but instead was better defined by RBP isoform usage patterns and showed higher enrichment for RBP targets (**Fig. 4B** and **fig. S6B**). A detailed examination of isoUsage modules revealed expected enrichment patterns with modules exhibiting GZ/CP specificity enriched for targets of established progenitor- or neuronal-enriched RBPs such as *PTBP1*, *SRRM4*, *PTBP2*, and *RBFOX1/2/3* (**Fig. 4C**). However, many modules were also significantly enriched for targets of RBPs less studied in brain development, such as *SAFB*, *UTP18* and *SRSF9* (one-sided Fisher's exact test, FDR-corrected $P<0.05$; (*50*)). Below, we focus on two module pairs with RBP isoforms in the top 30 hub transcripts and concomitant enrichment for their targets, and highlight example genes for which DTU informs neurodevelopmental processes.

IsoUsage.M1/M2 showed reciprocal GZ/CP specificity and concordant pathway enrichments for progenitor cell function and neuronal morphogenesis pathways, respectively (**Fig. 4C-E**). We also observed enrichment for neuronal markers in isoUsage.M1, reflecting the presence of neurons in the intermediate zone separating GZ and CP. Across isoUsage.M1/M2, we identified two novel NIC hub isoforms of the RBP *ELAVL2* (**Fig. 4C**) and concomitant enrichment of their targets (**Fig. 4F**). Inclusion of exon 2 in the CP-enriched isoUsage.M2 isoform (TALONT000708623) alters the translational start, adding 29 amino acids to the RNA-recognition motif (RRM) and potentially altering *ELAVL2's* RNA metabolism function. Additionally, we found that in both modules, the *ELAVL2* hub isoforms contained an alternative 5' TSS (**fig. S7F**) which may serve a regulatory function, as recent work demonstrated translational regulation of another ELAVL family member, *ELAVL4*, at its alternative 5' UTRs (*62*). Consistent with GZ/CP DTU of BAF complex proteins (**Fig. 3F)**, we identified three hub isoforms of *SMARCE1* (**Fig. 4G**), differing in the first four exons encoding an intrinsically disordered region (IDR). The isoUsage.M1 isoform encodes the full-length protein, whereas the two isoUsage.M2 isoforms lack either all or part of the first IDR, a low complexity protein domain (*63*) which mediates binding with higher-order complexes involved in chromatin remodeling and RNA splicing (*64*, *65*). Loss of all or part of the IDR in CP-enriched isoUsage.M2 suggests that *SMARCE1* may change its interaction with the BAF complex and/or associate with other protein complexes during neurogenesis.

In a second module pair, isoUsage.M3/M8, we observed complementary GZ/CP specificity and pathway enrichments, (**Fig. 4H**) and identified multiple RBP hub transcripts (*ELAVL1*, *RBFOX2*, *CELF2*) (**Fig. 4I-J**) with concomitant enrichment of their targets (**Fig. 4K**). While isoUsage.M3 was enriched in mRNA metabolism and RNA splicing processes, isoUsage.M8 was enriched in cytoskeletal function and cell projection organization (**Fig. 4I**), processes important in neuronal migration and maturation (*66*, *67*) and which match its CP enrichment (**Fig. 4J**). The M3 *ELAVL1* isoform encodes a short transcript containing only one of three RRMs, potentially affecting protein function (**Fig. 4L**). *ELAVL1* is also known to bind the 3'UTRs of target transcripts, including its own, to increase their stability (*68*) Both M3/M8 *ELAVL1* isoforms contain a primate/human-conserved novel 3'UTR with a ~200bp intron spliced out, potentially affecting auto-regulation (**Fig. 4L** and **fig. S7G**). Altogether, these networks refine our understanding of the specific "RNA regulons" active in the developing brain (*46*).

Isoform Expression at Single-Cell Resolution

Specific patterns of gene expression shape the differentiation and function of neural cells. Although gene expression in the developing neocortex has been extensively profiled at the single-cell level, isoform expression has yet to be systematically characterized. To gain single-cell resolution, we leveraged the recently developed ScISOrSeq (*20*) approach to profile >7,000 single cells across an additional 3 unique donor samples derived from microdissected GZ and CP regions (**Fig. 1A**; **fig. S8**). Barcoded full-length single-cell cDNA libraries generated using Drop-seq, with incorporated UMIs as published (*3*), were used as input to generate >26.4M high-quality PacBio CCS reads. To obtain cell-type specificity, cell-barcodes were matched to the high-depth, short-read sequencing dataset previously published on the same libraries (*3*). Of 7,189 individual single cell full-length transcriptomes, 4,281 had matching barcodes from short-

read sequencing (**figs. S8 and S9A; table S5A**). All subsequent analyses were performed on this matched subset of the scIso-Seq data.

Following strict quality control and downstream processing (*36*) (**figs. S8** and **S9A-B**), we detected on average 530 unique transcripts per cell, mapping in aggregate to 18,541 genes and 138,497 unique isoforms (**fig. S9C, data S2**). We observed high concordance between pseudo-bulk short-read and long-read based gene expression ($R = 0.9$, $p<2.2e^{-16}$) and detection ($R = 0.92$, $p<2.2e^{-16}$) (**fig. S9D** and **E**), and high inter-donor reproducibility ($R = 0.84\text{-}0.87$, $p<2.2e^{-16}$) (**fig. S9F)**, demonstrating the robustness of the data. Similar to the bulk tissue transcriptome, the majority of detected isoforms (71.7%) were previously unannotated (**fig. S9G**). We found broad support for these isoforms in bulk tissue IsoSeq and in independent long-read datasets, with >80% matching both 5' and 3' end termini (**fig. S9H)**, >75% containing CAGE or ATAC-seq peaks near the TSS, >85% supported by nearby polyA sites or motifs (**fig. S9I)**, and ~87% of splice junctions detected in bulk IsoSeq (**fig. S9J**). Altogether, 67,183 isoforms detected by scIso-Seq (49% of total) fully match isoforms detected in bulk tissue and 40% match independent datasets (**fig. S9K)**, including those containing 2,452 out of 5,165 previously unannotated spliced-in exons (**table S5B**), a handful of which were also validated by RT-PCR (**fig. S9L**).

We next connected single cells with their specific cellular identities. Previous unsupervised graph-based clustering in Seurat (*69*) using high-depth short-read RNA-Seq identified 16 transcriptionally distinct cell-type clusters in the developing human neocortex (**Fig. 1A**) (*3*). Through barcode matching, we detected cells from all 16 clusters **(Fig. 5A**), allowing us to construct cell-type-specific isoform expression profiles (**table S5** and **Fig. 5B)**. Comparing isoform expression diversity across cell types, we observed that excitatory neuron clusters, in particular those corresponding to newly-born migrating (ExN) and maturing neurons (ExM), harbored the largest number of isoforms (**Fig. 5C**)**.** This was not due to differences in sequencing depth across clusters (**fig. S9M**) or gene detection (**fig. S10A**). These same cell types exhibited the greatest diversity of unannotated expressed isoforms (**Fig. 5C**), highlighting a role for these transcripts in early neuronal maturation processes.

Selective isoform expression across different cell types has been reported in the adult brain for a few genes (*19*, *70*, *71*) with potential implications for neuropsychiatric disorders (*25*). Thus, we next sought to systematically characterize isoform expression and utilization across cells in the developing neocortex. We first characterized patterns of DTE between each cell-type cluster versus all other clusters (*36*) (**table S5C**) to identify transcripts that define cellular identities and can serve as molecular markers (**Fig. 5B**). As expected, the majority of these transcripts belonged to genes previously identified as canonical markers of the respective cell-types, including *HES1* (RG), *CRYAB* (vRG), *HOPX* (oRG), *EOMES* (IP), *LMO3* (ExDp), and *SATB2* (ExM-U), among others (*3*). Of 1,040 transcripts enriched in specific cell-types, 257 (24.7%) corresponded to isoforms newly identified in this study (**table S5C**). Among the top enriched transcripts for each cluster, we identify novel isoforms of *NRG1*, *LMO3*, *NEFL*, and *SYT4* enriched in oRG, ExDp, ExM-U, and ExM cell classes, respectively, all of which have established roles in brain development and function (**Fig. 5B, table S5C;** (*72–77*)).

We next conducted a pairwise DTE analysis to identify transcripts changing across specific cell-type transitions, detecting 409 transcripts corresponding to 147 genes (**table S5D,** $P_{nominal}<0.05$). Focusing on genes with multiple isoforms showing dynamic expression between progenitors and neurons, we observed isoforms of *PFN2*, which functions in actin polymerization dynamics and morphogenesis (*78*), with opposing expression patterns between progenitors and neurons (**Fig. 5D**). Similarly, we identified progenitor and neuron-specific isoforms of *RTN4*, a canonical regulator of axon growth and neuronal migration (*79*, *80*) (**Fig. 5D**). Together, these examples highlight changes in isoform expression across cell-types and developmentally-relevant transitions with putative consequences to the structure or stability of their encoded protein products.

Given the degree of isoform switching observed between GZ and CP (**Fig. 3**), we sought to quantify similar events across individual cell types (*36*). We identified 1,695 genes where the proportion of expressed isoforms for a given gene differed across at least two cell types (single-cell DTU; **Fig. 5E** and **table S5E**). These instances represent switches in isoform utilization across cell types that may be missed by traditional DGE analyses. Of the 2,284 specific transcripts exhibiting DTU across these 1695 genes, 48.5% showed proportional differences in progenitors and 43% in neurons, with an average number of 221 DTU transcripts per cell-type with roughly similar distribution across these cell classes. DTU genes were enriched in regulation of mRNA splicing (*CELF2*, *CIRBP*, *HNRNPA2B1*), cell division, regulation of synapse maturation (*NRXN1*, *YWHAZ*), and GO categories related to cytoskeleton dynamics and vesicle transport (*MAP1B*, *ANXA6*, *TPM1*, *GOPC*) (**Fig. 5E** and **fig. S10B**). Consistent with these results and a role for isoform switching in cell identity, GZ/CP DTU transcripts primarily clustered by expression across progenitors, neurons, or support cells (**fig. S10C**).

Additional Cell Types Uncovered from Isoform-level Clustering

Given the broad changes in isoform diversity and expression observed across cells, we leveraged these data to expand current cell-type classification catalogs. Re-clustering cells based on isoform expression yielded 15 highly-stable clusters largely mapping to many of the same cell classes as defined by gene-based clustering (*36*) (**Fig. 5 F-H** and **fig. S10D**). However, progenitors transitioning into neurons and early-born excitatory neurons were split into additional clusters, providing higher-resolution cell maturation stages than those observed by traditional gene-based clustering. In particular, newborn migrating neurons (ExN) split into three clusters (ExN1-3), encompassing cells previously annotated to IP, ExN, and ExM clusters, and two additional new clusters, vRG-ExN and ExN-ExM, representing cells in states on either side of a maturity spectrum centered around ExN cells (**Fig. 5 F-G)**. This was supported by pseudotime lineage inference analysis, whereby vRG-ExN cells represented a path of direct neurogenesis from vRG cells distinct from another path through oRG cells, and where ExN-ExM cells preceded the most mature neuronal clusters (**Fig. 5I**). To better understand the molecular programs and markers of these new cell states, we repeated DTE analyses (**table S5C**). Across ExN clusters, isoforms of *ENC1, ANKRD18CP*, and *RASD1* were enriched in ExN1, ExN2, and ExN3 cells, respectively (**fig. S10E-F**). Moreover, vRG-ExN cells were defined by a majority of transcripts involved in mitochondrial metabolism, consistent with a recent report of the role of mitochondria in regulating neuronal maturation (*81*). Overall, the increased resolution in ExN and ExM cells obtained from isoform-based clustering matches the observed increase in isoform

diversification in those cells (**Fig. 5C**) and supports a role for this mechanism in the early processes of neurogenesis.

Isoform-Centric Localization of Convergent Risk Gene Mechanisms

We next performed enrichment analyses to localize rare-variant association signals from large-scale whole exome and genome sequencing studies of neurodevelopmental and psychiatric disorders, including ASD (*82*), NDD (*82*, *83*), SCZ (*84*), bipolar disorder (BIP) (*85*), and epilepsy (*86*). Risk genes for NDD, ASD, and DDD (Developmental Disabilities) had significantly more isoforms (log2 scale; OR's 1.14-1.4; q's<3e-4) and exons (log2 scale; ORs 1.19-1.64; q's<0.01) compared to non-disease genes (**Fig. 6A**, **fig. S11A, table S6A;** FDR-corrected p-values from logistic regression, correcting for gene length and coding length). These associations were not observed for epilepsy (ORs 1.14, 1.22; NS), BIP (ORs 1.12, 1.22; NS), or SCZ (ORs 1.4, 1.56; NS). Disease-associated genes showed significant overlap with those exhibiting DGE, DTE, and DTU during cortical neurogenesis (logistic regression, FDR-corrected P<0.05; **Fig. 6A** and **fig. S11B**). Overall, these associations were observed mainly for genes and isoforms upregulated in the CP (DGE.up, DTE.up), indicative of neuronal expression, and were shared across NDD, non-syndromic ASD, DDD, and epilepsy, but not other diseases. In particular, NDD genes were enriched for those exhibiting DTE or DTU, but not changing in overall gene expression (DTU.not.DGE, DTE.not.DGE). Consistently, at the single-cell level we observed that NDD genes were significantly enriched in DTU (logistic regression, FDR-corrected P<0.05; **Fig. 6B,** "SingleCell-DTU"), and many NDD and ASD genes showed DTU across cell types in the developing neocortex (**Fig. 6C; fig. S12A)**. NDD, DDD, and ASD gene isoforms were primarily enriched in excitatory neurons (ExM-U, ExDp, or ExM) (**Fig. 6B; fig. S12B)**. However, isoforms of NDD genes and those causing syndromic forms of ASD were also enriched in mitotic progenitors and radial glia (**Fig. 6B; fig S12B**), as expected from the broader phenotypic spectrum in these disorders. Finally, NDD genes were enriched in ExN1, the newly-defined cell state based on isoform-level quantifications (**Fig. 6B** and **Fig. 5**).

We next leveraged our gene/isoform co-variation networks to localize disease gene convergence at the molecular level. Disease gene signal mainly coalesced among isoExpr modules (51.9% at $p_{nominal}$<0.05), followed by isoUsage (26.8%) and geneExpr (21.3%) modules (**fig. S13)**. NDD, DDD and non-syndromic ASD shared overlapping molecular signatures with modules mainly enriched for neuronal markers (**fig. S13**) and isoExpr modules regulating chromatin/histone modification (isoExpr.M7; **Fig. 6A and fig. S7B**) and RNA metabolism/splicing (isoExpr.M11/ M10/M28; **table S6A and fig. S7C-E**). Notably, isoExpr.M11 contained a hub isoform of the ASD-associated RNA helicase *DDX3X* with concomitant enrichment for its targets (**fig. S7E**). NDD and ASD genes were also enriched across excitatory neuronal modules regulating cytoskeleton, synaptic vesicles, and neurite morphogenesis (isoExpr.M30/M24, **table S6A** and **fig. S7D**) and ribosomal RNA processing and chromatin (isoUsage.M29, **table S6A and fig. S7C**). These findings support a major role for isoform expression and diversification in neuropsychiatric disease mechanisms during development, consistent with recent findings (*25*), regulating chromatin remodeling, cytoskeletal dynamics and RNA processing.

Reprioritization of De Novo Variants in Individuals with NDDs

Finally, to move from population to individual genetic risk mechanisms, we used our atlas to reinterpret de novo, noncoding genetic variants identified in large-scale sequencing studies of ASD (*87–89*) and intellectual disability/developmental disorders (ID/DD) (*89*). We reasoned that some variants previously disregarded as noncoding may actually fall within the >27MB with newfound transcriptional activity in our data or disrupt newly identified splice junctions. To test this hypothesis, we complemented the Gencode v33 annotation with our newly identified protein-coding isoforms and re-annotated the set of compiled genetic variants (N_{total}=272,187; N_{case}=145,880; $N_{control}$=126,307) using VEP. Altogether, this new annotation framework uncovered more severe consequences for 1.24% of all variants (**Fig. 6D**, **table S6B**). For example, we observed a novel *AKT3* isoform (TALONT000820688) with an alternative last exon extending the coding sequence (**Fig. 6E; right**) and expressed at comparable levels as known *AKT3* isoforms (**fig. S14**). This extended coding region overlapped a de novo mutation (DNM) from the ASD cohort (*90*), previously classified as a benign intronic variant but now predicted to cause a missense mutation in the novel AA sequence. In another example, a reported DNM from the ID/DD cohort (*91*) was predicted to cause a missense mutation in the *KLC1* protein based on known isoforms from Gencode (v33). The *KLC1* gene encodes a member of the kinesin light chain family, involved in microtubule cargo transport. We observed a *KLC1* isoform (TALONT000423578) with a novel TSS, supported by overlapping CAGE peaks (**fig. S15A**). This isoform was predicted to code for a protein with a novel start codon. Given the structure of this novel isoform, the DNM would lead to the loss of the start codon--a potentially more severe consequence. The protein encoded by TALONT000423578 had an alternative carboxy termini not observed in Gencode (v33) and strongly supported by our proteomics data (**fig. S15B**).

Cryptic splicing variation, in which splice-disrupting variants fall outside of the essential GT and AG dinucleotide motif, is a major recently uncovered mechanism underlying genetic risk for neurodevelopmental disorders, as exemplified by SpliceAI (*23*), a state-of-the-art deep neural network trained using pre-mRNA sequence to predict cryptic splice mutations. To determine whether our isoform-centric transcriptome could improve prediction of NDD-associated cryptic splice variants, we retrained SpliceAI with these annotations and overlapped its predictions with the compiled set of DNMs. With the addition of novel transcripts, a significantly larger proportion of variants were predicted to be cryptic splice variants (**Fig. 6F**, **table S6C** 1.44% with Gencode and novel transcripts, 1.37% with only Gencode transcripts, $p = 0.0028$, binomial test). By way of example, an ID/DD associated DNM was predicted to alter the splicing of a novel *AKT3* isoform (TALONT000820468), while it had no effect on the splicing of known *AKT3* isoforms (**Fig. 6E**). TALONT000820468 is the most highly expressed *AKT3* isoform detected (**fig. S14**). This variant caused the loss of the nearby splice acceptor and the retention of the last intron. The retained intron leads to shortened coding sequence and eliminates part of the protein kinase C-terminal domain (**Fig. 6E**). *AKT3* is a key regulator of the PI3K-AKT-mTOR pathway in the nervous system (*92*), and dysregulation of *AKT3* is associated with NDDs (*93*). Our findings highlight multiple de novo mutations may contribute to NDDs by affecting specific *AKT3* isoforms. More broadly, these results demonstrate that a more complete catalog of mid-gestation-brain-expressed full-length isoforms provides more granular molecular insight into the genetic risk mechanisms underlying neurodevelopmental disorders.

Discussion

Here, we have provided a detailed view of the full-length, alternatively-spliced transcriptome in the developing human neocortex at midgestation, with regional and cell-type specificity. Although splicing and isoform regulation are known to be critical for proper neural development (*15*), and are strongly implicated in NDD risk (*21*, *23*, *25*), technical challenges have made it difficult to delineate the path from genetic mutation to functional isoform changes, in part due to reliance on short-read sequencing as well as incomplete genomic annotations. Using high-depth long-read sequencing, we identified 149,510 previously unannotated transcript isoforms in the developing human brain, extending by >27MB the transcriptionally active content of the genome and expanding the proteomic diversity of the human brain. The majority of these unannotated transcripts were validated across independent datasets and data modalities; the remaining fraction may be unique to the specific donors, developmental periods, and/or regions profiled here or, alternatively, represent false positives inherent to challenges of isoform annotation. Examining the functional consequences of novel isoforms and profiling their expression across a wider range of time periods, cell-types, and distinct genetic backgrounds, facilitated by decreasing sequencing costs, will be critical next steps for future studies to further refine these results. However, our analyses here provide considerable support for their functional importance during neurogenesis and in neurodevelopmental disorders.

Assessing differential transcript-isoform activity across the developing cortex, we found wide-ranging changes in isoform expression and usage implicating chromatin remodeling via the BAF complex, and cytoskeletal dynamics important for neuronal morphogenesis. Isoform switching during corticogenesis implicated known neuronal splicing regulators, as well as RBPs previously not studied in context of brain development, including KHSRP, SUPV3L1, and SRSF9. APA analysis of GZ/CP differential DTU genes supported previous work showing that 3'UTR lengthening in differentiated cells is conserved throughout many cell lineages including neurons (*94*, *95*). Though there are many individual examples of APA in neuronal RBP genes (*94*, *96*, *97*), our analysis found that APA of RBP genes occurs on a large scale during human corticogenesis, indicating that tight regulation of RBP activity is important for neurogenesis.

In network analyses, we found that while geneExpr and isoExpr networks are cell-type driven, the isoUsage network was defined by RBP regulatory dynamics. IsoUsage modules contained hub isoforms with putative structural and regulatory differences including those predicted to alter protein domains of the NDD risk gene *SMARCE1*, and encoding primate and human-specific novel 3'UTR sequence of the RBP *ELAVL1*. While many developmental processes are conserved across vertebrates in early neural development, primate- and human-specific differences are important to understand given the unique cell types found in the expanded cortices of these species (*98*, *99*). Correspondingly, isoform-level single-cell transcriptomics demonstrated differential isoform expression and usage across cell types and enabled the identification of additional cell states in newborn excitatory neurons (ExN 1-3) as well as states encompassing the transition from progenitor to neuron and neuronal maturation (vRG-ExN and ExN-ExM). Together, these results increased the catalog of isoforms expressed during corticogenesis and strongly implicated splicing and RBP regulation of isoform expression and usage in neurogenesis.

The data generated in this study can inform current and future genome-wide disease risk association studies. We show that genes associated with NDD, ASD, and DDD exhibit increased isoform diversity, and NDD and ASD rare variants are enriched for isoform expression and usage changes during corticogenesis. Finally, we used our isoform-centric atlas to re-annotate and re-prioritize thousands of de novo ASD/IDD rare variants. The large number of previously unannotated transcripts identified here suggests that the functional consequences of many variants may have been missed using previous incomplete annotations. In sum, our results have broad implications for understanding cell fate specification in the developing human brain and for comprehensive interpretation of the genetic risk mechanisms underlying developmental brain disorders.

Materials and methods summary

Detailed materials and methods can be found in the supplementary materials. Human mid-gestation cortical tissue samples were obtained from the UCLA Gene and Cell Therapy Core in accordance with the IRB and the UCLA Office of Human Research Protection regulations, with full informed consent from the parent donors. GZ and CP regions of 15-17 PCW cortices were microdissected and processed for bulk IsoSeq or scIsoSeq. For bulk IsoSeq, PacBio Sequel IIe platform was used to generate ~38.5 million high-quality, full-length reads, which were filtered, aligned to the human reference genome, and analyzed using TALON, TranscriptClean, and Gencode v33 to identify and quantify known and novel genes/isoforms. Differential gene expression, transcript usage, and alternative polyadenylation site usage analyses were conducted using DESeq2, IsoformSwitchAnalyzeR, and DaPars2, respectively. Pathway enrichment was performed using gProfileR, while RBP target gene sets were analyzed for overlap with gene lists or module-associated genes/isoforms from bulk isoform sequencing or network analyses using Fisher's exact test and FDR-correction. TransDecoder, CPAT, and Comet were used to predict ORFs, assess coding potential, and validate novel proteins, respectively. WGCNA was used to generate gene and isoform expression, and isoform utilization modules, followed by over-representation and GO analysis. scIsoSeq libraries were prepared from captured single-cell full-length cDNA (*3*), and sequenced on Sequel I/II platforms. Reads were mapped to the human genome reference using minimap2 and isoforms were called with TALON. Seurat was used for clustering and differential expression analysis and Monocle3 for lineage trajectory analysis. Single-cell differential transcript usage analysis was conducted using binomial regression and p-value correction. RT-PCR was used to validate novel exons. Enrichment analyses were performed to associate rare-variant signals from neuropsychiatric disorders with transcriptome features, gene/isoform modules, or differentially expressed genes/isoforms using logistic regression controlling for gene and coding length, and transcript expression. ASD/IDD de novo variants(*87–89*) were annotated using Ensembl VEP or SpliceAI, with two rounds of annotation involving GENCODE v.33 GTF file and predicted ORFs from newly identified transcripts.

References and notes

1. J. H. Lui, T. J. Nowakowski, A. A. Pollen, A. Javaherian, A. R. Kriegstein, M. C. Oldham, Radial glia require PDGFD-PDGFRβ signalling in human but not mouse neocortex. *Nature*. **515**, 264–268 (2014).

2. R. L. Walker, G. Ramaswami, C. Hartl, N. Mancuso, M. J. Gandal, L. de la Torre-Ubieta, B. Pasaniuc, J. L. Stein, D. H. Geschwind, Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell*. **179**, 750–771.e22 (2019).

3. D. Polioudakis, L. de la Torre-Ubieta, J. Langerman, A. G. Elkins, X. Shi, J. L. Stein, C. K. Vuong, S. Nichterwitz, M. Gevorgian, C. K. Opland, D. Lu, W. Connell, E. K. Ruzzo, J. K. Lowe, T. Hadzic, F. I. Hinz, S. Sabri, W. E. Lowry, M. B. Gerstein, K. Plath, D. H. Geschwind, A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron*. **103**, 785–801.e8 (2019).

4. T. J. Nowakowski, A. Bhaduri, A. A. Pollen, B. Alvarado, M. A. Mostajo-Radji, E. Di Lullo, M. Haeussler, C. Sandoval-Espinosa, S. J. Liu, D. Velmeshev, J. R. Ounadjela, J. Shuga, X. Wang, D. A. Lim, J. A. West, A. A. Leyrat, W. J. Kent, A. R. Kriegstein, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science*. **358**, 1318–1323 (2017).

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

6. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature* (2018), doi:10.1038/s41586-018-0414-6.

7. A.-E. Saliba, A. J. Westermann, S. A. Gorski, J. Vogel, Single-cell RNA-seq: advances and future challenges. *Nucleic Acids Res.* **42**, 8845–8860 (2014).

8. J. Westoby, P. Artemov, M. Hemberg, A. Ferguson-Smith, Obstacles to detecting isoforms using full-length scRNA-seq data. *Genome Biol.* **21**, 74 (2020).

9. J. Merkin, C. Russell, P. Chen, C. B. Burge, Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science*. **338**, 1593–1599 (2012).

10. C. Foord, J. Hsu, J. Jarroux, W. Hu, N. Belchikov, S. Pollard, Y. He, A. Joglekar, H. U. Tilgner, The variables on RNA molecules: concert or cacophony? Answers in long-read sequencing. *Nat. Methods*. **20**, 20–24 (2023).

11. B. Treutlein, O. Gokce, S. R. Quake, T. C. Südhof, Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E1291–E1299 (2014).

12. M. Melé, P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segrè, S. Djebali, A. Niarchou, GTEx Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie, R. Guigó, Human genomics. The human transcriptome across tissues and individuals.

Science. **348**, 660–665 (2015).

13. B. Raj, B. J. Blencowe, Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron*. **87**, 14–27 (2015).

14. S. L. Zipursky, W. B. Grueber, The molecular basis of self-avoidance. *Annu. Rev. Neurosci.* **36**, 547–568 (2013).

15. X. Zhang, M. H. Chen, X. Wu, A. Kodani, J. Fan, R. Doan, M. Ozawa, J. Ma, N. Yoshida, J. F. Reiter, D. L. Black, P. V. Kharchenko, P. A. Sharp, C. A. Walsh, Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell*. **166**, 1147–1162.e15 (2016).

16. E. Furlanis, L. Traunmüller, G. Fucile, P. Scheiffele, Landscape of ribosome-engaged transcript isoforms reveals extensive neuronal-cell-class-specific alternative splicing programs. *Nat. Neurosci.* **22**, 1709–1717 (2019).

17. S. M. Weyn-Vanhentenryck, H. Feng, D. Ustianenko, R. Duffié, Q. Yan, M. Jacko, J. C. Martinez, M. Goodwin, X. Zhang, U. Hengst, S. Lomvardas, M. S. Swanson, C. Zhang, Precise temporal regulation of alternative splicing during neural development. *Nat. Commun.* **9**, 2189 (2018).

18. Y. Song, O. B. Botvinnik, M. T. Lovci, B. Kakaradov, P. Liu, J. L. Xu, G. W. Yeo, Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Mol. Cell*. **67**, 148–161.e5 (2017).

19. A. S. Booeshaghi, Z. Yao, C. van Velthoven, K. Smith, B. Tasic, H. Zeng, L. Pachter, Isoform cell-type specificity in the mouse primary motor cortex. *Nature*. **598**, 195–199 (2021).

20. I. Gupta, P. G. Collier, B. Haase, A. Mahfouz, A. Joglekar, T. Floyd, F. Koopmans, B. Barres, A. B. Smit, S. A. Sloan, W. Luo, O. Fedrigo, M. E. Ross, H. U. Tilgner, Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* (2018), doi:10.1038/nbt.4259.

21. Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, J. K. Pritchard, RNA splicing is a primary link between genetic variation and disease. *Science*. **352**, 600–604 (2016).

22. T. Raj, Y. I. Li, G. Wong, J. Humphrey, M. Wang, S. Ramdhani, Y.-C. Wang, B. Ng, I. Gupta, V. Haroutunian, E. E. Schadt, T. Young-Pearse, S. Mostafavi, B. Zhang, P. Sklar, D. A. Bennett, P. L. De Jager, Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat. Genet.* **50**, 1584–1592 (2018).

23. K. Jaganathan, S. Kyriazopoulou Panagiotopoulou, J. F. McRae, S. F. Darbandi, D. Knowles, Y. I. Li, J. A. Kosmicki, J. Arbelaez, W. Cui, G. B. Schwartz, E. D. Chow, E. Kanterakis, H. Gao, A. Kia, S. Batzoglou, S. J. Sanders, K. K.-H. Farh, Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. **176**, 535–548.e24 (2019).

24. A. Takata, N. Matsumoto, T. Kato, Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun.* **8**, 14519 (2017).

25. M. J. Gandal, P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, H. Won, H. van Bakel, M. Varghese, Y. Wang, A. W. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. Moran Losada, Z. Khan, J. Mleczko, Y. Xia, R. Dai, D. Wang, Y. T. Yang, M. Xu, K. Fish, P. R. Hof, J. Warrell, D. Fitzgerald, K. White, A. E. Jaffe, PsychENCODE Consortium, M. A. Peters, M. Gerstein, C. Liu, L.

M. Iakoucheva, D. Pinto, D. H. Geschwind, Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. **362** (2018), doi:10.1126/science.aat8127.

26. N. Akula, S. Marenco, K. Johnson, N. Feng, K. Zhu, A. Schulmann, W. Corona, X. Jiang, J. Cross, B. England, A. Nathan, S. Detera-Wadleigh, Q. Xu, P. K. Auluck, K. An, R. Kramer, J. Apud, B. T. Harris, C. Harker Rhodes, B. K. Lipska, F. J. McMahon, Deep transcriptome sequencing of subgenual anterior cingulate cortex reveals cross-diagnostic and diagnosis-specific RNA expression changes in major psychiatric disorders. *Neuropsychopharmacology* (2021), doi:10.1038/s41386-020-00949-5.

27. M. J. Gandal, J. R. Haney, B. Wamsley, C. X. Yap, S. Parhami, P. S. Emani, N. Chang, G. T. Chen, G. D. Hoftman, D. de Alba, G. Ramaswami, C. L. Hartl, A. Bhattacharya, C. Luo, T. Jin, D. Wang, R. Kawaguchi, D. Quintero, J. Ou, Y. E. Wu, N. N. Parikshak, V. Swarup, T. G. Belgard, M. Gerstein, B. Pasaniuc, D. H. Geschwind, Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature*. **611**, 532–539 (2022).

28. A. J. Willsey, S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin, N. Sestan, M. W. State, Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. **155**, 997–1007 (2013).

29. N. N. Parikshak, R. Luo, A. Zhang, H. Won, J. K. Lowe, V. Chandran, S. Horvath, D. H. Geschwind, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. **155**, 1008–1021 (2013).

30. E. J. Lopez Soto, M. J. Gandal, T. Gonatopoulos-Pournatzis, E. A. Heller, D. Luo, S. Zheng, Mechanisms of Neuronal Alternative Splicing and Strategies for Therapeutic Interventions. *J. Neurosci*. **39**, 8193–8199 (2019).

31. R. Stark, M. Grzelak, J. Hadfield, RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).

32. A. Joglekar, A. Prjibelski, A. Mahfouz, P. Collier, S. Lin, A. K. Schlusche, J. Marrocco, S. R. Williams, B. Haase, A. Hayes, J. G. Chew, N. I. Weisenfeld, M. Y. Wong, A. N. Stein, S. A. Hardwick, T. Hunt, Q. Wang, C. Dieterich, Z. Bent, O. Fedrigo, S. A. Sloan, D. Risso, E. D. Jarvis, P. Flicek, W. Luo, G. S. Pitt, A. Frankish, A. B. Smit, M. E. Ross, H. U. Tilgner, A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nat. Commun*. **12**, 463 (2021).

33. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. **34**, 3094–3100 (2018).

34. D. Wyman, G. Balderrama-Gutierrez, F. Reese, S. Jiang, S. Rahmanian, S. Forner, D. Matheos, W. Zeng, B. Williams, D. Trout, W. England, S.-H. Chu, R. C. Spitale, A. J. Tenner, B. J. Wold, A. Mortazavi, A technology-agnostic long-read analysis pipeline for transcriptome discovery and quantification. *bioRxiv* (2020), p. 672931.

35. M. Tardaguila, L. de la Fuente, C. Marti, C. Pereira, F. J. Pardo-Palacios, H. Del Risco, M. Ferrell, M. Mellado, M. Macchietto, K. Verheggen, M. Edelmann, I. Ezkurdia, J. Vazquez, M. Tress, A. Mortazavi, L. Martens, S. Rodriguez-Navarro, V. Moreno-Manzano, A. Conesa, SQANTI: extensive

characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res.* **28**, 396–411 (2018).

36. Materials and methods are available as supplementary materials.

37. L. de la Torre-Ubieta, J. L. Stein, H. Won, C. K. Opland, D. Liang, D. Lu, D. H. Geschwind, The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell*. **172**, 289–304.e18 (2018).

38. A. E. Trevino, F. Müller, J. Andersen, L. Sundaram, A. Kathiria, A. Shcherbina, K. Farh, H. Y. Chang, A. M. Pa□ca, A. Kundaje, S. P. Pa□ca, W. J. Greenleaf, Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell*. **184**, 5053–5069.e23 (2021).

39. R. S. Ziffra, C. N. Kim, J. M. Ross, A. Wilfert, T. N. Turner, M. Haeussler, A. M. Casella, P. F. Przytycki, K. C. Keough, D. Shin, D. Bogdanoff, A. Kreimer, K. S. Pollard, S. A. Ament, E. E. Eichler, N. Ahituv, T. J. Nowakowski, Single-cell epigenomics reveals mechanisms of human cortical development. *Nature*. **598**, 205–213 (2021).

40. S. Noguchi, T. Arakawa, S. Fukuda, M. Furuno, A. Hasegawa, F. Hori, S. Ishikawa-Kato, K. Kaida, A. Kaiho, M. Kanamori-Katayama, T. Kawashima, M. Kojima, A. Kubosaki, R.-I. Manabe, M. Murata, S. Nagao-Sato, K. Nakazato, N. Ninomiya, H. Nishiyori-Sueki, S. Noma, E. Saijyo, A. Saka, M. Sakai, C. Simon, N. Suzuki, M. Tagami, S. Watanabe, S. Yoshida, P. Arner, R. A. Axton, M. Babina, J. K. Baillie, T. C. Barnett, A. G. Beckhouse, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. J. Carlisle, H. C. Clevers, C. A. Davis, M. Detmar, T. Dohi, A. S. B. Edge, M. Edinger, A. Ehrlund, K. Ekwall, M. Endoh, H. Enomoto, A. Eslami, M. Fagiolini, L. Fairbairn, M. C. Farach-Carson, G. J. Faulkner, C. Ferrai, M. E. Fisher, L. M. Forrester, R. Fujita, J.-I. Furusawa, T. B. Geijtenbeek, T. Gingeras, D. Goldowitz, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, Y. Hasegawa, M. Herlyn, P. Heutink, K. J. Hitchens, D. A. Hume, T. Ikawa, Y. Ishizu, C. Kai, H. Kawamoto, Y. I. Kawamura, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. Klein, S. P. Klinken, A. J. Knox, S. Kojima, H. Koseki, S. Koyasu, W. Lee, A. Lennartsson, A. Mackay-Sim, N. Mejhert, Y. Mizuno, H. Morikawa, M. Morimoto, K. Moro, K. J. Morris, H. Motohashi, C. L. Mummery, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, T. Nozaki, S. Ogishima, N. Ohkura, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, R. Passier, M. Patrikakis, A. Pombo, S. Pradhan-Bhatt, X.-Y. Qin, M. Rehli, P. Rizzu, S. Roy, A. Sajantila, S. Sakaguchi, H. Sato, H. Satoh, S. Savvi, A. Saxena, C. Schmidl, C. Schneider, G. G. Schulze-Tanzil, A. Schwegmann, G. Sheng, J. W. Shin, D. Sugiyama, T. Sugiyama, K. M. Summers, N. Takahashi, J. Takai, H. Tanaka, H. Tatsukawa, A. Tomoiu, H. Toyoda, M. van de Wetering, L. M. van den Berg, R. Verardo, D. Vijayan, C. A. Wells, L. N. Winteringham, E. Wolvetang, Y. Yamaguchi, M. Yamamoto, C. Yanagi-Mizuochi, M. Yoneda, Y. Yonekura, P. G. Zhang, S. Zucchelli, I. Abugessaisa, E. Arner, J. Harshbarger, A. Kondo, T. Lassmann, M. Lizio, S. Sahin, T. Sengstag, J. Severin, H. Shimoji, M. Suzuki, H. Suzuki, J. Kawai, N. Kondo, M. Itoh, C. O. Daub, T. Kasukawa, H. Kawaji, P. Carninci, A. R. R. Forrest, Y. Hayashizaki, FANTOM5 CAGE profiles of human and mouse samples. *Sci Data*. **4**, 170112 (2017).

41. C. J. Herrmann, R. Schmidt, A. Kanitz, P. Artimo, A. J. Gruber, M. Zavolan, PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3′ end sequencing. *Nucleic Acids Res.* **48**, D174–D179 (2019).

42. L. M. Pardo, P. Rizzu, M. Francescatto, M. Vitezic, G. G. R. Leday, J. S. Sanchez, A. Khamis, H. Takahashi, W. D. J. van de Berg, Y. A. Medvedeva, M. A. van de Wiel, C. O. Daub, P. Carninci, P. Heutink, Regional differences in gene expression and promoter usage in aged human brains.

Neurobiol. Aging. **34**, 1825–1836 (2013).

43. A. Nellore, A. E. Jaffe, J.-P. Fortin, J. Alquicira-Hernández, L. Collado-Torres, S. Wang, R. A. Phillips III, N. Karbhari, K. D. Hansen, B. Langmead, J. T. Leek, Human splicing diversity and the extent of unannotated splice junctions across human RNA-seq samples on the Sequence Read Archive. *Genome Biol.* **17**, 266 (2016).

44. B. C. Carlyle, R. R. Kitchen, J. E. Kanyo, E. Z. Voss, M. Pletikos, A. M. M. Sousa, T. T. Lam, M. B. Gerstein, N. Sestan, A. C. Nairn, A multiregional proteomic survey of the postnatal human brain. *Nat. Neurosci.* **20**, 1787–1795 (2017).

45. L. Li, K.-L. Huang, Y. Gao, Y. Cui, G. Wang, N. D. Elrod, Y. Li, Y. E. Chen, P. Ji, F. Peng, W. K. Russell, E. J. Wagner, W. Li, An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.* **53**, 994–1005 (2021).

46. F. Gebauer, T. Schwarzl, J. Valcárcel, M. W. Hentze, RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* (2020), doi:10.1038/s41576-020-00302-y.

47. G. Sokpor, Y. Xie, J. Rosenbusch, T. Tuoc, Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders. *Front. Mol. Neurosci.* **10**, 243 (2017).

48. M. E. Rowland, J. M. Jajarmi, T. S. M. Osborne, A. V. Ciernia, Insights Into the Emerging Role of Baf53b in Autism Spectrum Disorder. *Front. Mol. Neurosci.* **15**, 805158 (2022).

49. C. K. Vuong, D. L. Black, S. Zheng, The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* **17**, 265–281 (2016).

50. E. L. Van Nostrand, P. Freese, G. A. Pratt, X. Wang, X. Wei, R. Xiao, S. M. Blue, J.-Y. Chen, N. A. L. Cody, D. Dominguez, S. Olson, B. Sundararaman, L. Zhan, C. Bazile, L. P. B. Bouvrette, J. Bergalet, M. O. Duff, K. E. Garcia, C. Gelboin-Burkhart, M. Hochman, N. J. Lambert, H. Li, M. P. McGurk, T. B. Nguyen, T. Palden, I. Rabano, S. Sathe, R. Stanton, A. Su, R. Wang, B. A. Yee, B. Zhou, A. L. Louie, S. Aigner, X.-D. Fu, E. Lécuyer, C. B. Burge, B. R. Graveley, G. W. Yeo, A large-scale binding and functional map of human RNA-binding proteins. *Nature*. **583**, 711–719 (2020).

51. M. Yang, S.-L. Yang, S. Herrlinger, C. Liang, M. Dzieciatkowska, K. C. Hansen, R. Desai, A. Nagy, L. Niswander, E. G. Moss, J.-F. Chen, Lin28 promotes the proliferative capacity of neural progenitor cells in brain development. *Development*. **142**, 1616–1627 (2015).

52. L. Lei, S.-Y. Yan, R. Yang, J.-Y. Chen, Y. Li, Y. Bu, N. Chang, Q. Zhou, X. Zhu, C.-Y. Li, J.-W. Xiong, Spliceosomal protein eftud2 mutation leads to p53-dependent apoptosis in zebrafish neural progenitors. *Nucleic Acids Res.* **45**, 3422–3436 (2017).

53. C. W. Bird, A. S. Gardiner, F. Bolognani, D. C. Tanner, C.-Y. Chen, W.-J. Lin, S. Yoo, J. L. Twiss, N. Perrone-Bizzozero, KSRP modulation of GAP-43 mRNA stability restricts axonal outgrowth in embryonic hippocampal neurons. *PLoS One*. **8**, e79255 (2013).

54. A. L. Lennox, M. L. Hoye, R. Jiang, B. L. Johnson-Kerner, L. A. Suit, S. Venkataramanan, C. J. Sheehan, F. C. Alsina, B. Fregeau, K. A. Aldinger, C. Moey, I. Lobach, A. Afenjar, D. Babovic-Vuksanovic, S. Bézieau, P. R. Blackburn, J. Bunt, L. Burglen, P. M. Campeau, P. Charles, B. H. Y. Chung, B. Cogné, C. Curry, M. D. D'Agostino, N. Di Donato, L. Faivre, D. Héron, A. M. Innes, B. Isidor, B. Keren, A. Kimball, E. W. Klee, P. Kuentz, S. Küry, D. Martin-Coignard, G. Mirzaa, C.

Mignot, N. Miyake, N. Matsumoto, A. Fujita, C. Nava, M. Nizon, D. Rodriguez, L. S. Blok, C. Thauvin-Robinet, J. Thevenon, M. Vincent, A. Ziegler, W. Dobyns, L. J. Richards, A. J. Barkovich, S. N. Floor, D. L. Silver, E. H. Sherr, Pathogenic DDX3X Mutations Impair RNA Metabolism and Neurogenesis during Fetal Cortical Development. *Neuron*. **106**, 404–420.e8 (2020).

55. M. L. Hoye, L. Calviello, A. J. Poff, N.-E. Ejimogu, C. R. Newman, M. D. Montgomery, J. Ou, S. N. Floor, D. L. Silver, Aberrant cortical development is driven by impaired cell cycle and translational control in a DDX3X syndrome model. *Elife*. **11** (2022), doi:10.7554/eLife.78203.

56. D. L. Makino, F. Halbach, E. Conti, The RNA exosome and proteasome: common principles of degradation control. *Nat. Rev. Mol. Cell Biol.* **14**, 654–660 (2013).

57. P. Clemente, A. Pajak, I. Laine, R. Wibom, A. Wedell, C. Freyer, A. Wredenberg, SUV3 helicase is required for correct processing of mitochondrial transcripts. *Nucleic Acids Res.* **43**, 7398–7413 (2015).

58. M. Thoms, E. Thomson, J. Baßler, M. Gnädig, S. Griesel, E. Hurt, The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell*. **162**, 1029–1038 (2015).

59. X. Huo, L. Ji, Y. Zhang, P. Lv, X. Cao, Q. Wang, Z. Yan, S. Dong, D. Du, F. Zhang, G. Wei, Y. Liu, B. Wen, The Nuclear Matrix Protein SAFB Cooperates with Major Satellite RNAs to Stabilize Heterochromatin Architecture Partially through Phase Separation. *Mol. Cell*. **77**, 368–383.e7 (2020).

60. N. N. Parikshak, M. J. Gandal, D. H. Geschwind, Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16**, 441–458 (2015).

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

62. T. Popovitchenko, Y. Park, N. F. Page, X. Luo, Z. Krsnik, Y. Liu, I. Salamon, J. D. Stephenson, M. L. Kraushar, N. L. Volk, S. M. Patel, H. R. S. Wijeratne, D. Li, K. S. Suthar, A. Wach, M. Sun, S. J. Arnold, W. Akamatsu, H. Okano, L. Paillard, H. Zhang, S. Buyske, I. Kostovic, S. De Rubeis, R. P. Hart, M.-R. Rasin, Translational derepression of Elavl4 isoforms at their alternative 5' UTRs determines neuronal development. *Nat. Commun.* **11**, 1674 (2020).

63. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).

64. C. A. Musselman, T. G. Kutateladze, Characterization of functional disordered regions within chromatin-associated proteins. *iScience*. **24**, 102070 (2021).

65. S. Calabretta, S. Richard, Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends Biochem. Sci.* **40**, 662–672 (2015).

66. R. Ayala, T. Shu, L.-H. Tsai, Trekking across the brain: the journey of neuronal migration. *Cell*. **128**, 29–43 (2007).

67. M. Valiente, O. Marín, Neuronal migration mechanisms in development and disease. *Curr. Opin. Neurobiol.* **20**, 68–78 (2010).

68. G. Ince-Dunn, H. J. Okano, K. B. Jensen, W.-Y. Park, R. Zhong, J. Ule, A. Mele, J. J. Fak, C. Yang, C. Zhang, J. Yoo, M. Herre, H. Okano, J. L. Noebels, R. B. Darnell, Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability.

Neuron. **75**, 1067–1080 (2012).

69. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*. **177**, 1888–1902.e21 (2019).

70. S. A. Hardwick, W. Hu, A. Joglekar, L. Fan, P. G. Collier, C. Foord, J. Balacco, S. Lanjewar, M. M. Sampson, F. Koopmans, A. D. Prjibelski, A. Mikheenko, N. Belchikov, J. Jarroux, A. B. Lucas, M. Palkovits, W. Luo, T. A. Milner, L. C. Ndhlovu, A. B. Smit, J. Q. Trojanowski, V. M. Y. Lee, O. Fedrigo, S. A. Sloan, D. Tombácz, M. E. Ross, E. Jarvis, Z. Boldogkői, L. Gan, H. U. Tilgner, Single-nuclei isoform RNA sequencing unlocks barcoded exon connectivity in frozen brain tissue. *Nat. Biotechnol.* (2022), doi:10.1038/s41587-022-01231-3.

71. C. R. Palmer, C. S. Liu, W. J. Romanow, M.-H. Lee, J. Chun, Altered cell and RNA isoform diversity in aging Down syndrome brains. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2114326118.

72. R. S. Schmid, B. McGrath, B. E. Berechid, B. Boyles, M. Marchionni, N. Sestan, E. S. Anton, Neuregulin 1-erbB2 signaling is required for the establishment of radial glia and their transformation into astrocytes in cerebral cortex. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 4251–4256 (2003).

73. Y. Liu, B. D. Ford, M. A. Mann, G. D. Fischbach, Neuregulin-1 increases the proliferation of neuronal progenitors from embryonic neural stem cells. *Dev. Biol.* **283**, 437–445 (2005).

74. A. R. Ypsilanti, K. Pattabiraman, R. Catta-Preta, O. Golonzhka, S. Lindtner, K. Tang, I. R. Jones, A. Abnousi, I. Juric, M. Hu, Y. Shen, D. E. Dickel, A. Visel, L. A. Pennachio, M. Hawrylycz, C. L. Thompson, H. Zeng, I. Barozzi, A. S. Nord, J. L. Rubenstein, Transcriptional network orchestrating regional patterning of cortical progenitors. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2024795118.

75. J. A. Miller, S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, A. Ebbert, Z. L. Riley, J. J. Royall, K. Aiona, J. M. Arnold, C. Bennet, D. Bertagnolli, K. Brouner, S. Butler, S. Caldejon, A. Carey, C. Cuhaciyan, R. A. Dalley, N. Dee, T. A. Dolbeare, B. A. C. Facer, D. Feng, T. P. Fliss, G. Gee, J. Goldy, L. Gourley, B. W. Gregor, G. Gu, R. E. Howard, J. M. Jochim, C. L. Kuan, C. Lau, C.-K. Lee, F. Lee, T. A. Lemon, P. Lesnar, B. McMurray, N. Mastan, N. Mosqueda, T. Naluai-Cecchini, N.-K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P. D. Parker, S. E. Parry, A. Stevens, M. Pletikos, M. Reding, K. Roll, D. Sandman, M. Sarreal, S. Shapouri, N. V. Shapovalova, E. H. Shen, N. Sjoquist, C. R. Slaughterbeck, M. Smith, A. J. Sodt, D. Williams, L. Zöllei, B. Fischl, M. B. Gerstein, D. H. Geschwind, I. A. Glass, M. J. Hawrylycz, R. F. Hevner, H. Huang, A. R. Jones, J. A. Knowles, P. Levitt, J. W. Phillips, N. Sestan, P. Wohnoutka, C. Dang, A. Bernard, J. G. Hohmann, E. S. Lein, Transcriptional landscape of the prenatal human brain. *Nature*. **508**, 199–206 (2014).

76. M. T. K. Kirkcaldie, S. T. Dwyer, The third wave: Intermediate filaments in the maturing nervous system. *Mol. Cell. Neurosci.* **84**, 68–76 (2017).

77. C. Dean, H. Liu, F. M. Dunning, P. Y. Chang, M. B. Jackson, E. R. Chapman, Synaptotagmin-IV modulates synaptic function and long-term potentiation by regulating BDNF release. *Nat. Neurosci.* **12**, 767–776 (2009).

78. K. Murk, M. Ornaghi, J. Schiweck, Profilin Isoforms in Health and Disease - All the Same but Different. *Front Cell Dev Biol*. **9**, 681122 (2021).

79. J. Yuasa-Kawada, M. Kinoshita-Kawada, Y. Tsuboi, J. Y. Wu, Neuronal guidance genes in health and diseases. *Protein Cell*, wac030 (2022).

80. V. Chiurchiù, M. Maccarrone, A. Orlacchio, The role of reticulons in neurodegenerative diseases. *Neuromolecular Med.* **16**, 3–15 (2014).

81. R. Iwata, P. Casimir, E. Erkol, L. Boubakar, M. Planque, I. M. Gallego López, M. Ditkowska, V. Gaspariunaite, S. Beckers, D. Remans, K. Vints, A. Vandekeere, S. Poovathingal, M. Bird, I. Vlaeminck, E. Creemers, K. Wierda, N. Corthout, P. Vermeersch, S. Carpentier, K. Davie, M. Mazzone, N. V. Gounko, S. Aerts, B. Ghesquière, S.-M. Fendt, P. Vanderhaeghen, Mitochondria metabolism sets the species-specific tempo of neuronal development. *Science*. **379**, eabn4705 (2023).

82. J. M. Fu, F. K. Satterstrom, M. Peng, H. Brand, R. L. Collins, S. Dong, B. Wamsley, L. Klei, L. Wang, S. P. Hao, C. R. Stevens, C. Cusick, M. Babadi, E. Banks, B. Collins, S. Dodge, S. B. Gabriel, L. Gauthier, S. K. Lee, L. Liang, A. Ljungdahl, B. Mahjani, L. Sloofman, A. N. Smirnov, M. Barbosa, C. Betancur, A. Brusco, B. H. Y. Chung, E. H. Cook, M. L. Cuccaro, E. Domenici, G. B. Ferrero, J. J. Gargus, G. E. Herman, I. Hertz-Picciotto, P. Maciel, D. S. Manoach, M. R. Passos-Bueno, A. M. Persico, A. Renieri, J. S. Sutcliffe, F. Tassone, E. Trabetti, G. Campos, S. Cardaropoli, D. Carli, M. C. Y. Chan, C. Fallerini, E. Giorgio, A. C. Girardi, E. Hansen-Kiss, S. L. Lee, C. Lintas, Y. Ludena, R. Nguyen, L. Pavinato, M. Pericak-Vance, I. N. Pessah, R. J. Schmidt, M. Smith, C. I. S. Costa, S. Trajkova, J. Y. T. Wang, M. H. C. Yu, Autism Sequencing Consortium (ASC), Broad Institute Center for Common Disease Genomics (Broad-CCDG), iPSYCH-BROAD Consortium, D. J. Cutler, S. De Rubeis, J. D. Buxbaum, M. J. Daly, B. Devlin, K. Roeder, S. J. Sanders, M. E. Talkowski, Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022).

83. J. Kaplanis, K. E. Samocha, L. Wiel, Z. Zhang, K. J. Arvai, R. Y. Eberhardt, G. Gallone, S. H. Lelieveld, H. C. Martin, J. F. McRae, P. J. Short, R. I. Torene, E. de Boer, P. Danecek, E. J. Gardner, N. Huang, J. Lord, I. Martincorena, R. Pfundt, M. R. F. Reijnders, A. Yeung, H. G. Yntema, S. Borras, C. Clark, J. Dean, Z. Miedzybrodzka, A. Ross, S. Tennant, T. Dabir, D. Donnelly, M. Humphreys, A. Magee, V. McConnell, S. McKee, S. McNerlan, P. J. Morrison, G. Rea, F. Stewart, T. Cole, N. Cooper, L. Cooper-Charles, H. Cox, L. Islam, J. Jarvis, R. Keelagher, D. Lim, D. McMullan, J. Morton, S. Naik, M. O'Driscoll, K.-R. Ong, D. Osio, N. Ragge, S. Turton, J. Vogt, D. Williams, S. Bodek, A. Donaldson, A. Hills, K. Low, R. Newbury-Ecob, A. M. Norman, E. Roberts, I. Scurr, S. Smithson, M. Tooley, S. Abbs, R. Armstrong, C. Dunn, S. Holden, S.-M. Park, J. Paterson, L. Raymond, E. Reid, R. Sandford, I. Simonic, M. Tischkowitz, G. Woods, L. Bradley, J. Comerford, A. Green, S. Lynch, S. McQuaid, B. Mullaney, J. Berg, D. Goudie, E. Mavrak, J. McLean, C. McWilliam, E. Reavey, T. Azam, E. Cleary, A. Jackson, W. Lam, A. Lampe, D. Moore, M. Porteous, E. Baple, J. Baptista, C. Brewer, B. Castle, E. Kivuva, M. Owens, J. Rankin, C. Shaw-Smith, C. Turner, P. Turnpenny, C. Tysoe, T. Bradley, R. Davidson, C. Gardiner, S. Joss, E. Kinning, C. Longman, R. McGowan, V. Murday, D. Pilz, E. Tobias, M. Whiteford, N. Williams, A. Barnicoat, E. Clement, F. Faravelli, J. Hurst, L. Jenkins, W. Jones, V. K. A. Kumar, M. Lees, S. Loughlin, A. Male, D. Morrogh, E. Rosser, R. Scott, L. Wilson, A. Beleza, C. Deshpande, F. Flinter, M. Holder, M. Irving, L. Izatt, D. Josifova, S. Mohammed, A. Molenda, L. Robert, W. Roworth, D. Ruddy, M. Ryten, S. Yau, C. Bennett, M. Blyth, J. Campbell, A. Coates, A. Dobbie, S. Hewitt, E. Hobson, E. Jackson, R. Jewell, A. Kraus, K. Prescott, E. Sheridan, J. Thomson, K. Bradshaw, A. Dixit, J. Eason, R. Haines, R. Harrison, S. Mutch, A. Sarkar, C. Searle, N. Shannon, A. Sharif, M. Suri, P. Vasudevan, N. Canham, I. Ellis, L. Greenhalgh, E. Howard, V. Stinton, A. Swale, A. Weber, S. Banka, C. Breen, T. Briggs, E. Burkitt-Wright, K. Chandler, J. Clayton-Smith, D. Donnai, S. Douzgou, L. Gaunt, E. Jones, B. Kerr, C. Langley, K. Metcalfe, A. Smith, R. Wright, D. Bourn, J.

Burn, R. Fisher, S. Hellens, A. Henderson, T. Montgomery, M. Splitt, V. Straub, M. Wright, S. Zwolinski, Z. Allen, B. Bernhard, A. Brady, C. Brooks, L. Busby, V. Clowes, N. Ghali, S. Holder, R. Ibitoye, E. Wakeling, E. Blair, J. Carmichael, D. Cilliers, S. Clasper, R. Gibbons, U. Kini, T. Lester, A. Nemeth, J. Poulton, S. Price, D. Shears, H. Stewart, A. Wilkie, S. Albaba, D. Baker, M. Balasubramanian, D. Johnson, M. Parker, O. Quarrell, A. Stewart, J. Willoughby, C. Crosby, F. Elmslie, T. Homfray, H. Jin, N. Lahiri, S. Mansour, K. Marks, M. McEntagart, A. Saggar, K. Tatton-Brown, R. Butler, A. Clarke, S. Corrin, A. Fry, A. Kamath, E. McCann, H. Mugalaasi, C. Pottinger, A. Procter, J. Sampson, F. Sansbury, V. Varghese, D. Baralle, A. Callaway, E. J. Cassidy, S. Daniels, A. Douglas, N. Foulds, D. Hunt, M. Kharbanda, K. Lachlan, C. Mercer, L. Side, I. K. Temple, D. Wellesley, L. E. L. M. Vissers, J. Juusola, C. F. Wright, H. G. Brunner, H. V. Firth, D. R. FitzPatrick, J. C. Barrett, M. E. Hurles, C. Gilissen, K. Retterer, Deciphering Developmental Disorders Study, Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* (2020), doi:10.1038/s41586-020-2832-5.

84. T. Singh, T. Poterba, D. Curtis, H. Akil, M. Al Eissa, J. D. Barchas, N. Bass, T. B. Bigdeli, G. Breen, E. J. Bromet, P. F. Buckley, W. E. Bunney, J. Bybjerg-Grauholm, W. F. Byerley, S. B. Chapman, W. J. Chen, C. Churchhouse, N. Craddock, C. M. Cusick, L. DeLisi, S. Dodge, M. A. Escamilla, S. Eskelinen, A. H. Fanous, S. V. Faraone, A. Fiorentino, L. Francioli, S. B. Gabriel, D. Gage, S. A. Gagliano Taliun, A. Ganna, G. Genovese, D. C. Glahn, J. Grove, M.-H. Hall, E. Hämäläinen, H. O. Heyne, M. Holi, D. M. Hougaard, D. P. Howrigan, H. Huang, H.-G. Hwu, R. S. Kahn, H. M. Kang, K. J. Karczewski, G. Kirov, J. A. Knowles, F. S. Lee, D. S. Lehrer, F. Lescai, D. Malaspina, S. R. Marder, S. A. McCarroll, A. M. McIntosh, H. Medeiros, L. Milani, C. P. Morley, D. W. Morris, P. B. Mortensen, R. M. Myers, M. Nordentoft, N. L. O'Brien, A. M. Olivares, D. Ongur, W. H. Ouwehand, D. S. Palmer, T. Paunio, D. Quested, M. H. Rapaport, E. Rees, B. Rollins, F. K. Satterstrom, A. Schatzberg, E. Scolnick, L. J. Scott, S. I. Sharp, P. Sklar, J. W. Smoller, J. L. Sobell, M. Solomonson, E. A. Stahl, C. R. Stevens, J. Suvisaari, G. Tiao, S. J. Watson, N. A. Watts, D. H. Blackwood, A. D. Børglum, B. M. Cohen, A. P. Corvin, T. Esko, N. B. Freimer, S. J. Glatt, C. M. Hultman, A. McQuillin, A. Palotie, C. N. Pato, M. T. Pato, A. E. Pulver, D. St Clair, M. T. Tsuang, M. P. Vawter, J. T. Walters, T. M. Werge, R. A. Ophoff, P. F. Sullivan, M. J. Owen, M. Boehnke, M. C. O'Donovan, B. M. Neale, M. J. Daly, Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*. **604**, 509–516 (2022).

85. D. S. Palmer, D. P. Howrigan, S. B. Chapman, R. Adolfsson, N. Bass, D. Blackwood, M. P. M. Boks, C.-Y. Chen, C. Churchhouse, A. P. Corvin, N. Craddock, D. Curtis, A. Di Florio, F. Dickerson, N. B. Freimer, F. S. Goes, X. Jia, I. Jones, L. Jones, L. Jonsson, R. S. Kahn, M. Landén, A. E. Locke, A. M. McIntosh, A. McQuillin, D. W. Morris, M. C. O'Donovan, R. A. Ophoff, M. J. Owen, N. L. Pedersen, D. Posthuma, A. Reif, N. Risch, C. Schaefer, L. Scott, T. Singh, J. W. Smoller, M. Solomonson, D. S. Clair, E. A. Stahl, A. Vreeker, J. T. R. Walters, W. Wang, N. A. Watts, R. Yolken, P. P. Zandi, B. M. Neale, Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nat. Genet.* **54**, 541–547 (2022).

86. Epi25 Collaborative, Ultra-rare genetic variation in the epilepsies: A whole-exome sequencing study of 17,606 individuals. *Am. J. Hum. Genet.* **105**, 267–282 (2019).

87. D. M. Werling, H. Brand, J.-Y. An, M. R. Stone, L. Zhu, J. T. Glessner, R. L. Collins, S. Dong, R. M. Layer, E. Markenscoff-Papadimitriou, A. Farrell, G. B. Schwartz, H. Z. Wang, B. B. Currall, X. Zhao, J. Dea, C. Duhn, C. A. Erdman, M. C. Gilson, R. Yadav, R. E. Handsaker, S. Kashin, L. Klei, J. D. Mandell, T. J. Nowakowski, Y. Liu, S. Pochareddy, L. Smith, M. F. Walker, M. J. Waterman, X. He, A. R. Kriegstein, J. L. Rubenstein, N. Sestan, S. A. McCarroll, B. M. Neale, H. Coon, A. J. Willsey, J. D. Buxbaum, M. J. Daly, M. W. State, A. R. Quinlan, G. T. Marth, K. Roeder, B. Devlin, M. E. Talkowski, S. J. Sanders, An analytical framework for whole-genome sequence association

studies and its implications for autism spectrum disorder. *Nat. Genet.* **50**, 727–736 (2018).

88. J.-Y. An, K. Lin, L. Zhu, D. M. Werling, S. Dong, H. Brand, H. Z. Wang, X. Zhao, G. B. Schwartz, R. L. Collins, B. B. Currall, C. Dastmalchi, J. Dea, C. Duhn, M. C. Gilson, L. Klei, L. Liang, E. Markenscoff-Papadimitriou, S. Pochareddy, N. Ahituv, J. D. Buxbaum, H. Coon, M. J. Daly, Y. S. Kim, G. T. Marth, B. M. Neale, A. R. Quinlan, J. L. Rubenstein, N. Sestan, M. W. State, A. J. Willsey, M. E. Talkowski, B. Devlin, K. Roeder, S. J. Sanders, Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science*. **362** (2018), doi:10.1126/science.aat6576.

89. B. B. Cummings, K. J. Karczewski, J. A. Kosmicki, E. G. Seaby, N. A. Watts, M. Singer-Berk, J. M. Mudge, J. Karjalainen, F. K. Satterstrom, A. H. O'Donnell-Luria, T. Poterba, C. Seed, M. Solomonson, J. Alföldi, Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, M. J. Daly, D. G. MacArthur, Transcript expression-aware annotation improves rare variant interpretation. *Nature*. **581**, 452–458 (2020).

90. F. K. Satterstrom, J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J.-Y. An, M. Peng, R. Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B. Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero, C. Dias, Autism Sequencing Consortium, iPSYCH-Broad Consortium, C. Betancur, E. H. Cook, L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Børglum, M. W. State, A. E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly, J. D. Buxbaum, Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*. **180**, 568–584.e23 (2020).

91. S. H. Lelieveld, M. R. F. Reijnders, R. Pfundt, H. G. Yntema, E.-J. Kamsteeg, P. de Vries, B. B. A. de Vries, M. H. Willemsen, T. Kleefstra, K. Löhner, M. Vreeburg, S. J. C. Stevens, I. van der Burgt, E. M. H. F. Bongers, A. P. A. Stegmann, P. Rump, T. Rinne, M. R. Nelen, J. A. Veltman, L. E. L. M. Vissers, H. G. Brunner, C. Gilissen, Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **19**, 1194–1196 (2016).

92. R. M. Easton, H. Cho, K. Roovers, D. W. Shineman, M. Mizrahi, M. S. Forman, V. M.-Y. Lee, M. Szabolcs, R. de Jong, T. Oltersdorf, T. Ludwig, A. Efstratiadis, M. J. Birnbaum, Role for Akt3/protein kinase Bgamma in attainment of normal brain size. *Mol. Cell. Biol.* **25**, 1869–1878 (2005).

93. L. Wang, K. Zhou, Z. Fu, D. Yu, H. Huang, X. Zang, X. Mo, Brain Development and Akt Signaling: the Crossroads of Signaling Pathway and Neurodevelopmental Diseases. *J. Mol. Neurosci.* **61**, 379–384 (2017).

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

95. P. J. Shepard, E.-A. Choi, J. Lu, L. A. Flanagan, K. J. Hertel, Y. Shi, Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA*. **17**, 761–772 (2011).

96. J. J. An, K. Gharami, G.-Y. Liao, N. H. Woo, A. G. Lau, F. Vanevski, E. R. Torre, K. R. Jones, Y. Feng, B. Lu, B. Xu, Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell*. **134**, 175–187 (2008).

97. B. Lackford, C. Yao, G. M. Charles, L. Weng, X. Zheng, E.-A. Choi, X. Xie, J. Wan, Y. Xing, J. M. Freudenberg, P. Yang, R. Jothi, G. Hu, Y. Shi, Fip1 regulates mRNA alternative polyadenylation to

promote stem cell self-renewal. *EMBO J.* **33**, 878–889 (2014).

98. J. H. Lui, D. V. Hansen, A. R. Kriegstein, Development and evolution of the human neocortex. *Cell*. **146**, 18–36 (2011).

99. E. Taverna, M. Götz, W. B. Huttner, The cell biology of neurogenesis: toward an understanding of the development and evolution of the neocortex. *Annu. Rev. Cell Dev. Biol.* **30**, 465–502 (2014).

100. I. Abugessaisa, S. Noguchi, A. Hasegawa, A. Kondo, H. Kawaji, P. Carninci, T. Kasukawa, refTSS: A Reference Data Set for Human and Mouse Transcription Start Sites. *J. Mol. Biol.* **431**, 2407–2422 (2019).

101. X. Adiconis, A. L. Haber, S. K. Simmons, A. Levy Moonshine, Z. Ji, M. A. Busby, X. Shi, J. Jacques, M. A. Lancaster, J. Q. Pan, A. Regev, J. Z. Levin, Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nat. Methods*. **15**, 505–511 (2018).

102. R. E. Workman, A. D. Tang, P. S. Tang, M. Jain, J. R. Tyson, R. Razaghi, P. C. Zuzarte, T. Gilpatrick, A. Payne, J. Quick, N. Sadowski, N. Holmes, J. G. de Jesus, K. L. Jones, C. M. Soulette, T. P. Snutch, N. Loman, B. Paten, M. Loose, J. T. Simpson, H. E. Olsen, A. N. Brooks, M. Akeson, W. Timp, Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nat. Methods*. **16**, 1297–1305 (2019).

103. Y. Gao, F. Wang, R. Wang, E. Kutschera, Y. Xu, S. Xie, Y. Wang, K. E. Kadash-Edmondson, L. Lin, Y. Xing, ESPRESSO: Robust discovery and quantification of transcript isoforms from error-prone long-read RNA-seq data. *Sci Adv*. **9**, eabq5072 (2023).

104. S. K. Leung, A. R. Jeffries, I. Castanho, B. T. Jordan, K. Moore, J. P. Davies, E. L. Dempster, N. J. Bray, P. O'Neill, E. Tseng, Z. Ahmed, D. A. Collier, E. D. Jeffery, S. Prabhakar, L. Schalkwyk, C. Jops, M. J. Gandal, G. M. Sheynkman, E. Hannon, J. Mill, Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Rep.* **37**, 110022 (2021).

105. D. A. Glinos, G. Garborcauskas, P. Hoffman, N. Ehsan, L. Jiang, A. Gokden, X. Dai, F. Aguet, K. L. Brown, K. Garimella, T. Bowers, M. Costello, K. Ardlie, R. Jian, N. R. Tucker, P. T. Ellinor, E. D. Harrington, H. Tang, M. Snyder, S. Juul, P. Mohammadi, D. G. MacArthur, T. Lappalainen, B. B. Cummings, Transcriptome variation in human tissues revealed by long-read sequencing. *Nature*. **608**, 353–359 (2022).

106. A. Varabyou, M. J. Sommer, B. Erdogdu, I. Shinder, I. Minkin, K.-H. Chao, S. Park, J. Heinz, C. Pockrandt, A. Shumate, N. Rincon, D. Puiu, M. Steinegger, S. L. Salzberg, M. Pertea, CHESS 3: an improved, comprehensive catalog of human genes and transcripts based on large-scale expression data, phylogenetic analysis, and protein structure. *bioRxiv* (2022), p. 2022.12.21.521274.

107. G. Pertea, M. Pertea, GFF Utilities: GffRead and GffCompare. *F1000Res.* **9** (2020), doi:10.12688/f1000research.23297.2.

108. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

109. Y. Li, X. Rao, W. W. Mattox, C. I. Amos, B. Liu, RNA-Seq Analysis of Differential Splice Junction Usage and Intron Retentions by DEXSeq. *PLoS One*. **10**, e0136653 (2015).

110. K. Vitting-Seerup, A. Sandelin, IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics*. **35**, 4469–4471 (2019).

111. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *Gigascience*. **10** (2021), doi:10.1093/gigascience/giab008.

112. B. Sundararaman, L. Zhan, S. M. Blue, R. Stanton, K. Elkins, S. Olson, X. Wei, E. L. Van Nostrand, G. A. Pratt, S. C. Huelga, B. M. Smalec, X. Wang, E. L. Hong, J. M. Davidson, E. Lécuyer, B. R. Graveley, G. W. Yeo, Resources for the Comprehensive Discovery of Functional RNA Elements. *Mol. Cell*. **61**, 903–913 (2016).

113. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics*. **10**, 421 (2009).

114. S. C. Potter, A. Luciani, S. R. Eddy, Y. Park, R. Lopez, R. D. Finn, HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).

115. L. Wang, H. J. Park, S. Dasari, S. Wang, J.-P. Kocher, W. Li, CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* **41**, e74–e74 (2013).

116. M. The, M. J. MacCoss, W. S. Noble, L. Käll, Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J. Am. Soc. Mass Spectrom.* **27**, 1719–1727 (2016).

117. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).

118. A. T. L. Lun, D. J. McCarthy, J. C. Marioni, A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res.* **5**, 2122 (2016).

119. X. Qiu, Q. Mao, Y. Tang, L. Wang, R. Chawla, H. A. Pliner, C. Trapnell, Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods*. **14**, 979–982 (2017).

Acknowledgments

We thank members of the Gandal, de la Torre-Ubieta, Li and Pasaniuc laboratories for helpful discussions and critical reading of the manuscript. We thank Dr. Geschwind for kindly providing bulk RNA and single-cell cDNA libraries used as input for long-read sequencing library preparation. Pac-Bio bulk tissue long-read library generation and sequencing were performed at the UC Davis Genomics Core and brain tissue was obtained from the UCLA CFAR (5P30 AI028697).

Funding: This work was supported by the Simons Foundation Bridge to Independence Award (MJG), the National Institute of Mental Health (R01MH121521 to MJG; R01MH124018 to LTU, T32MH073526 to MK), and the UCLA Medical Scientist Training Program (T32GM008042 to MK).

Data were generated as part of the PsychENCODE Consortium, supported by: U01DA048279, U01MH103339, U01MH103340, U01MH103346, U01MH103365, U01MH103392, U01MH116438, U01MH116441, U01MH116442, U01MH116488, U01MH116489, U01MH116492, U01MH122590, U01MH122591, U01MH122592, U01MH122849, U01MH122678, U01MH122681, U01MH116487, U01MH122509, R01MH094714, R01MH105472, R01MH105898, R01MH109677, R01MH109715, R01MH110905, R01MH110920, R01MH110921, R01MH110926, R01MH110927, R01MH110928, R01MH111721, R01MH117291, R01MH117292, R01MH117293, R21MH102791, R21MH103877, R21MH105853, R21MH105881, R21MH109956, R56MH114899, R56MH114901, R56MH114911, R01MH125516, R01MH126459, R01MH129301, R01MH126393, R01MH121521, R01MH116529, R01MH129817, R01MH117406, and P50MH106934 awarded to: Alexej Abyzov, Nadav Ahituv, Schahram Akbarian, Kristin Brennand, Andrew Chess, Gregory Cooper, Gregory Crawford, Stella Dracheva, Peggy Farnham, Michael Gandal, Mark Gerstein, Daniel Geschwind, Fernando Goes, Joachim F. Hallmayer, Vahram Haroutunian, Thomas M. Hyde, Andrew Jaffe, Peng Jin, Manolis Kellis, Joel Kleinman, James A. Knowles, Arnold Kriegstein, Chunyu Liu, Christopher E. Mason, Keri Martinowich, Eran Mukamel, Richard Myers, Charles Nemeroff, Mette Peters, Dalila Pinto, Katherine Pollard, Kerry Ressler, Panos Roussos, Stephan Sanders, Nenad Sestan, Pamela Sklar, Michael P. Snyder, Matthew State, Jason Stein, Patrick Sullivan, Alexander E. Urban, Flora Vaccarino, Stephen Warren, Daniel Weinberger, Sherman Weissman, Zhiping Weng, Kevin White, A. Jeremy Willsey, Hyejung Won, and Peter Zandi.

Author Contributions: A.P., M.J.G and L.T.U. conceived and designed the study. C.V. and L.T.U. collected and processed the tissue specimens and dissected the samples for processing. A.P. and L.T.U. generated the library for single cell IsoSeq. A.P. processed the single-cell raw data and performed the validation experiments. C.J. and A.P. analyzed bulk tissue IsoSeq data, C.J and M.J.G. analyzed bulk tissue isoform switching, P.Z., X.W. and C.L. performed the proteomic analysis, N.G. performed the alternative polyadenylation analysis, C.V. performed the RBP associated analysis, C.V., L.T.U. and M.J.G. conducted network analyses, A.P., C.J., M.M. performed single cell IsoSeq analysis, M.K., D.V., A.P. performed single cell isoform switch analysis. X.G., J.J.L. perform single cell DTU analysis, A.P and L.T.U analyzed the DTU results,

M.J.G, L.T.U., P.Z., K.H., C.J., A.P., B.P. performed the disease enrichment analysis. A.P., C.V., M.J.G., L.T.U. interpreted the data; A.P., C.V., P.Z., M.J.G., L.T.U. wrote the manuscript.

Competing Interests: M.J.G. receives grant funding from Mitsubishi Tanabe Pharma America that is unrelated to this current project. All other authors declare that they have no competing interests.

Data and materials availability:
Controlled-access bulk and single-cell IsoSeq data is available at https://doi.org/10.7303/syn4921369 and https://assets.nemoarchive.org/dat-rhocguc to investigators subject to approval by the NIMH Repository and Genomics Resources (NRGR).

A UCSC track hub browser containing mid-gestation neocortex isoforms is available at: https://genome.ucsc.edu/cgi-bin/hgTracks?hubUrl=https://raw.githubusercontent.com/ashokpatowary/Dev_Brain_IsoSeq/main/hub.txt&genome=hg19&position=lastDbPos

Code availability: Code used to process the data and generate all figures in this manuscript is available at https://github.com/gandallab/Dev_Brain_IsoSeq and from Zenodo at https://zenodo.org/badge/latestdoi/614018591

Supplementary Materials

Materials and Methods

Figs. S1 to S15

Tables S1 to S8

Data S1, S2

References 100 - 119

Figure legends

Summary Figure Legend:
An isoform-centric transcriptome of the developing human neocortex informs mechanisms of neuropsychiatric disease. We provide a systematic characterization of transcript-isoform diversity in the developing human neocortex at tissue and single-cell resolution using PacBio IsoSeq. We identify thousands of novel isoforms with specific regional (GZ/CP) and cell type expression coalescing into networks driven by cell type identity and RNA binding protein regulation. This resource reveals substantial contributions of isoform switching to cellular identity and elucidates novel genetic risk mechanisms for neurodevelopmental and neuropsychiatric disorders, including a re-annotation of thousands of de novo rare variants with potential clinical implications.

Fig. 1. The full-length cell-type specific transcriptome of the developing human neocortex at mid-gestation.

(**A**) Experimental design for isoform-centric profiling of the developing human brain transcriptome at bulk and single-cell resolution. Briefly, microdissected samples from the progenitor-enriched germinal zone (GZ) and neuronally-enriched cortical plate (CP) were profiled from 6 unique donors at mid-gestation. Full-length cDNA libraries were generated from homogenate tissue as well as from dissociated, barcoded single-cells with incorporation of UMIs. Single molecule long-read sequencing (PacBio) was used to quantify transcript-isoforms and integrated with matched short-read scRNA-Seq. Isoform expression quantifications demonstrated (**B**) strong biological reproducibility across unique donors and (**C**) region-specific clustering via principal component analysis. (**D**) Transcript isoforms identified by IsoSeq were compared against the Gencode v33 reference. Novel transcripts were further classified by their splice junction matching to annotated Gencode isoforms as described by TALON: ISM (incomplete splice match); NIC (novel in catalog); NNC (novel not in catalog). The 'other' category denotes isoforms belonging to antisense, genomic, and intergenic classes. (**E**) Number of isoforms identified based on classes described in D. (**F**) Heatmap shows uniform patterns of relative read-depth coverage across genes, arranged by length. Low coverage in dark blue, high coverage in yellow. (**G**) Abundance of the isoforms by each class as described in D. The Prefix ISM signal observed at Min observed counts = ~10^5 largely corresponds to a highly expressed isoform of the MAP1B gene. Compared with known isoforms, novel transcripts identified here were (**H**) significantly longer ($P<2e^{-16}$, Kruskal-Wallis) and (**I**) contained a significantly greater number of exons ($P<2e^{-16}$, Kruskal-Wallis). (**J**) Proportion of the dominant isoforms for each gene by gene expression percentile. For highly expressed genes, the dominant isoform contributed the most to the gene expression. (**K**) Genes ranked by the number of unique transcript isoforms detected. NDD risk genes (red) (*82*) had significantly more detected isoforms, controlling for total expression, gene length, and coding length (OR 1.56, P=3.6e-03, logistic regression).

Fig. 2. Expanded Transcriptomic and Proteomic Complexity

(**A**) The *NF1* gene locus with 20 previously unidentified brain-expressed isoforms. Tracks from top to bottom include: CAGE clusters, 3′seq clusters, ATAC-Seq peaks, Gencode isoforms that were not detected in our data, Gencode isoforms detected in our data, and novel isoforms identified from our data. A novel microexon is highlighted. (**B)** External validation of isoforms by independent datasets: 5' end validation was performed by presence of peak from CAGE (FANTOM5 and fetal brain cortex) and mid-gestation cortex ATAC-Seq (*37–39*); 3' end validation was performed by presence of polyA motif or peak from polyAsite database. Percentage of transcript with support at either end is highlighted. (**C)** The vast majority of splice junctions identified by IsoSeq are supported by the external Intropolis splice junction database. (**D)** Length distribution of the >7000 novel spliced exons uncovered in this study. T=True, F= False. (**E)** Validation of novel exons via RT-PCR. Expected product size is shown in parentheses along the name of the gene with the exon. Each exon was amplified with the primer sets shown in the schematic. (**F)** Characterization of novel protein-coding transcripts. Long-read sequencing identified a total of 214,516 transcripts, 149,510 of which were not found in Gencode v33. 92,422 of these novel transcripts were predicted to code for protein sequences, and 35,467 predicted ORFs were further confirmed by MS/MS proteomics data. (**G)** Representative mass spectrum of peptide HGLGTASALDWWPK, which confirms the translation of the identified *NF1* microexon. Matched b, y, a, and immonium ions were highlighted. (**H)** Number of total transcripts, transcripts with ORF, transcripts with novel ORF compared to UniProt human protein sequences, and transcripts with ORFs validated by MS/MS proteomics, plotted per isoform structural category.

Fig. 3. The landscape of isoform switching during human corticogenesis.
(**A**) Long-read RNA-Seq data from GZ and CP samples were contrasted for patterns of differential gene expression (DGE), differential transcript expression (DTE), and differential transcript usage (DTU). Venn diagram is shown depicting the overlap for genes exhibiting significant DGE, DTE, and DTU (FDR-corrected P<0.05). (**B**) A volcano plot depicts isoform switching across GZ and CP. The x-axis depicts the difference in isoform fraction (dIF) for a given transcript in CP versus GZ. Inset, most regionally variable DTU isoforms are not present in Gencode. (**C**) Functional consequences of isoform switch events between GZ and CP are shown. For example, CP-upregulated isoforms were more significantly likely to have gained (than lost) an exon (2,031 vs 1,260 isoforms**,** FDR-corrected $P<10^{-40}$). (**D)** Analysis of distal polyA site usage index (DPUI) for known transcripts between GZ and CP samples. On average, CP transcripts have higher DPUIs indicative of longer 3' UTRs. **(E)** Pathway enrichments for genes exhibiting cross-region DTU are notable for dendrite morphogenesis and SWI/SNF complex genes, among others. (**F**) An example of isoform switching observed within the ASD risk gene *SMARCC2*. Although total gene expression was not different between GZ and CP, significant switching was observed among DTU isoforms with two isoforms exhibiting preferential usage in GZ (* P<0.05, *** P<0.001). (**G**) Regionally variable genes were enriched for cell-type specific marker genes from scRNA-Seq. vRG = ventricular radial glia, oRG = Outer Radial Glia, PgG2M = Cycling progenitors (G2/M phase), PgS = Cycling progenitors (S phase), IP = Intermediate Progenitors, ExN = Migrating excitatory, ExM = Maturing excitatory, ExM-U = Maturing excitatory upper enriched, ExDp1 = Excitatory deep layer 1, ExDp2 = Excitatory deep layer 2, InMGE = Interneuron MGE, InCGE = Interneuron CGE, OPC = Oligodendrocytes Precursor Cells, End = Endothelial, Per = Pericyte, Mic = Microglia. (**H**) Genes containing DTU isoforms were also highly enriched for targets of known brain-enriched RBPs (black bar) and for targets of RBPs profiled in the ENCODE database (gray bar). Targets exhibiting alternative splicing (AS) or gene expression (GEX), or direct binding (e/CLIP) are indicated by Data Type. See also **fig. S4C**.

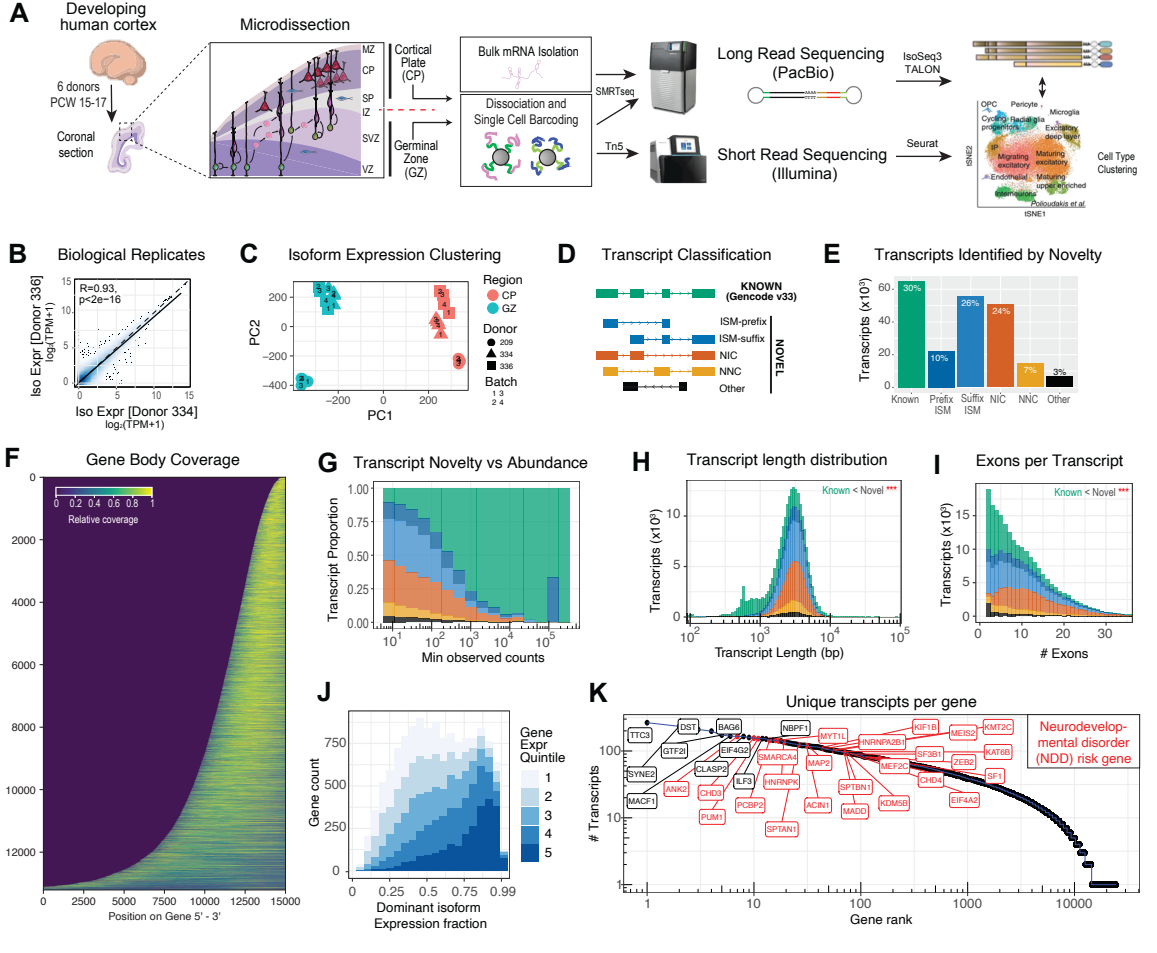
Fig. 4. Network-based contextualization of isoform usage
(A) The isoUsage network shows more and stronger enrichments for RBP targets compared to geneExpr and isoExpr. Top, density plot of cell type enrichments for the three networks; bottom, density plot of RBP enrichments. **(B)** The isoUsage network is driven by RNA-binding protein isoform usage. Dendrogram of the isoUsage network with isoforms (isoUsage and isoExpr) or genes (geneExpr) organized by their presence in isoUsage modules plotted below. Isoforms of known RBPs (*112*) are plotted below. **(C)** Module plots highlighting hub isoforms in isoUsage.M1 and M2. *SMARCE1* hub isoforms inform different cellular processes associated with progenitors and neurons. **(D)** Gene ontology for iso.Usage.M1 and M2. **(E)** Cell type marker enrichment for iso.Usage.M1 and M2. **(F)** RBP target enrichments for M1 and M2, including targets of the nELAVL RBPs, which include *ELAVL1*. **(G)** Transcript models of *SMARCE1* hub isoforms. Box highlights exon3 (M1, turquoise) which encodes part of the IDR and the shifted reading frame driven by an alternative translational start (M2, blue). M2 SMARCE1 isoforms lack either all or a portion of this IDR – in ENST00000647508, exclusion of exon 4 truncates the IDR, while in ENST00000643318 a downstream translational start in combination with exon 4 exclusion entirely removes the protein domain. **(H)** Module plots for iso.Usage.M3 and M8. **(I)** Gene ontology for M3 and M8. **(J)** Cell type enrichments for M3 and M8. **(K)** RBP target enrichment for these modules include targets of *RBFOX2*, *CELF2* and *ELAVL2* (an nELAVL) which are hub isoforms in M3 and M8. **(L)** Transcript models for *ELAVL1* hub isoforms – arrow highlights the primate/human-conserved sequence missing from these 3'UTRs.

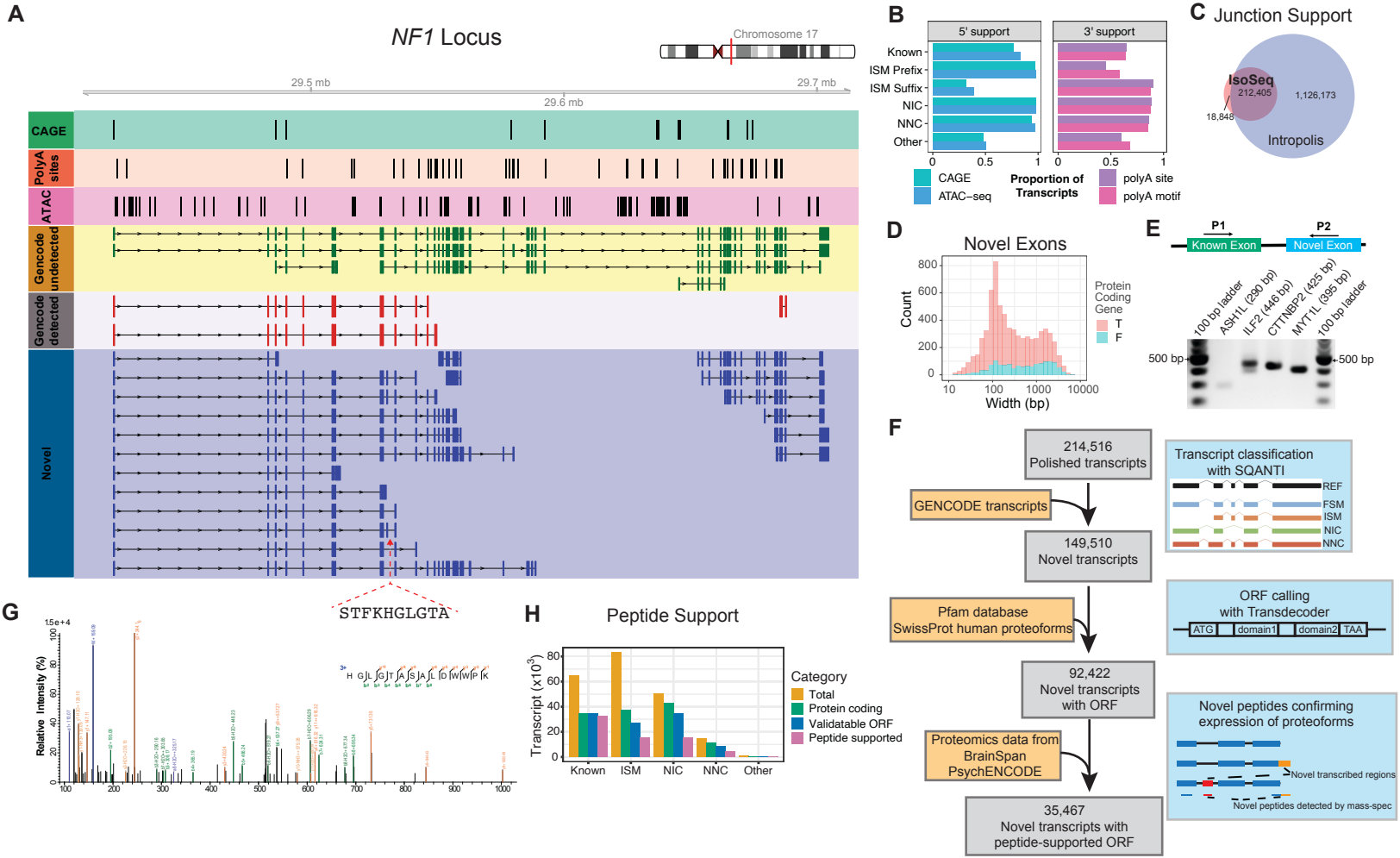
Fig. 5. Cell type-specific isoform diversity in the developing human cortex
(**A**) Uniform manifold approximation and projection (UMAP) plot of 4,281 cells detected by both 3' end short-read sequencing and by scIso-Seq. Each dot represents a single cell, colored by its corresponding cluster. UMAP position of the cells is calculated based on isoform expression while cluster labels are as previously defined ((*3*)). vRG = ventricular radial glia, oRG = Outer Radial Glia, PgG2M = Cycling progenitors (G2/M phase), PgS = Cycling progenitors (S phase), IP = Intermediate Progenitors, ExN = Migrating excitatory, ExM = Maturing excitatory, ExM-U = Maturing excitatory upper enriched, ExDp1 = Excitatory deep layer 1, ExDp2 = Excitatory deep layer 2, InMGE = Interneuron MGE, InCGE = Interneuron CGE, OPC = Oligodendrocytes Precursor Cells, End = Endothelial, Per = Pericyte, Mic = Microglia. (**B**) Heatmap showing differentially expressed isoforms across cell types defined by gene-based clustering. Novel isoforms are shown in red. (**C**) Distribution of isoforms across cell types shows greater diversity of isoforms in newborn migrating (ExN) and maturing excitatory neurons (ExM) compared to other cell types in mid-gestation human cortex. (**D**) Isoforms of PFN2 and RTN4 differentially expressed across cell types along with their predicted functional consequences. Isoform ENST00000239940,

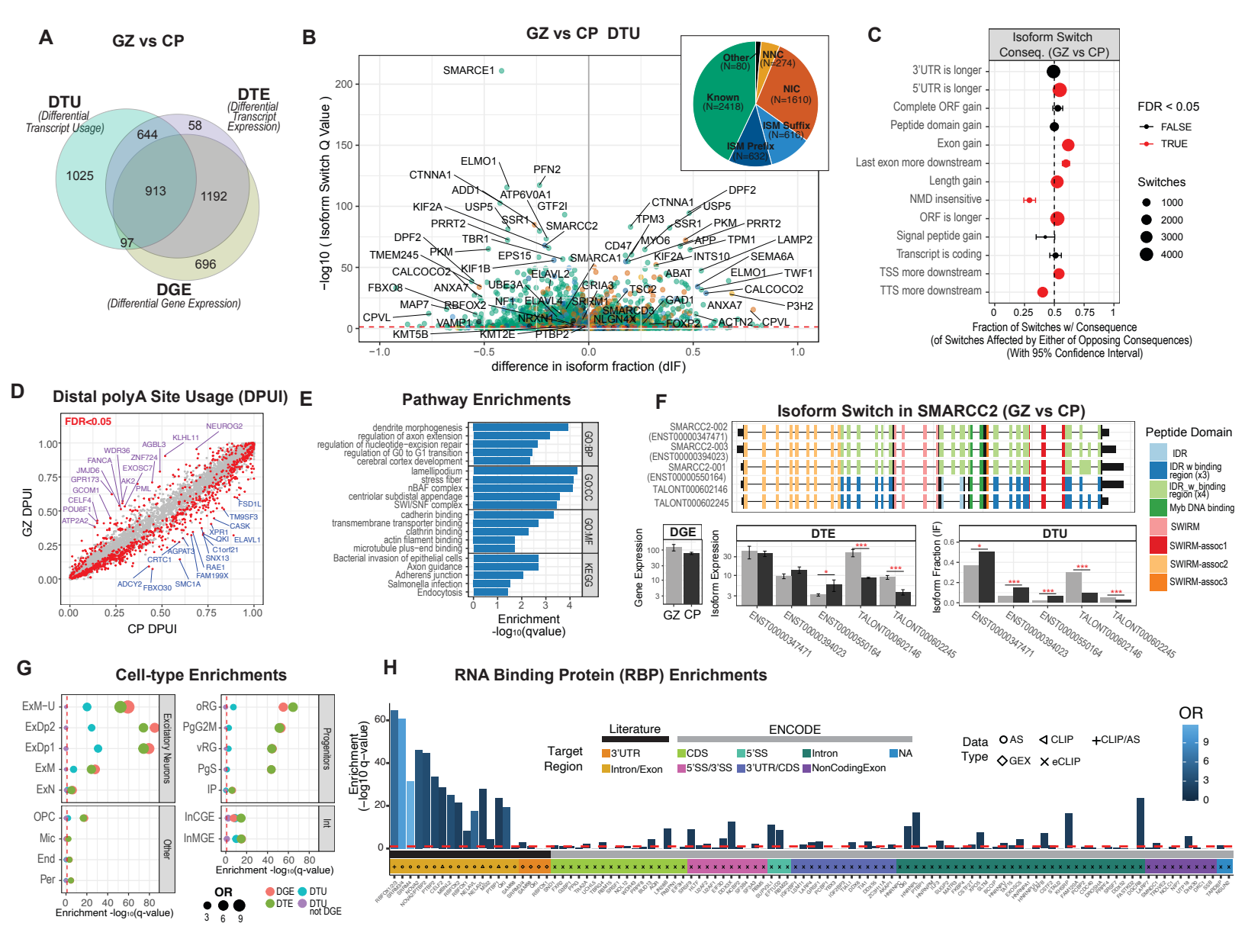
predominantly expressed in neurons, is predicted to encode IDR protein domains not found in the progenitor-enriched ENST00000452853 isoform. The novel isoform TALONT000502136 is enriched in neurons, whereas the progenitor enriched ENST00000317610 isoform is longer and contains multiple reticulon protein domains. (**E**) Heatmap showing a subset of isoforms with differential usage across cell types (DTU). (**F**) River plot showing mapping of cells from gene-based (left) to isoform-based (right) clustering. Each line represents a single cell. (**G**) UMAP of cells clustered based on isoform expression as measured by scIsoSeq. Additional stages of excitatory neuron maturation can be defined using isoform-level data. (**H**) Heatmap showing differentially expressed isoforms across cell types defined by isoform-based clustering. Novel isoforms are shown in red. (**I**) Cell lineage trajectory analysis (Monocle3) shows direct neurogenesis through vRG-ExN cells and indirect path through IP cells.

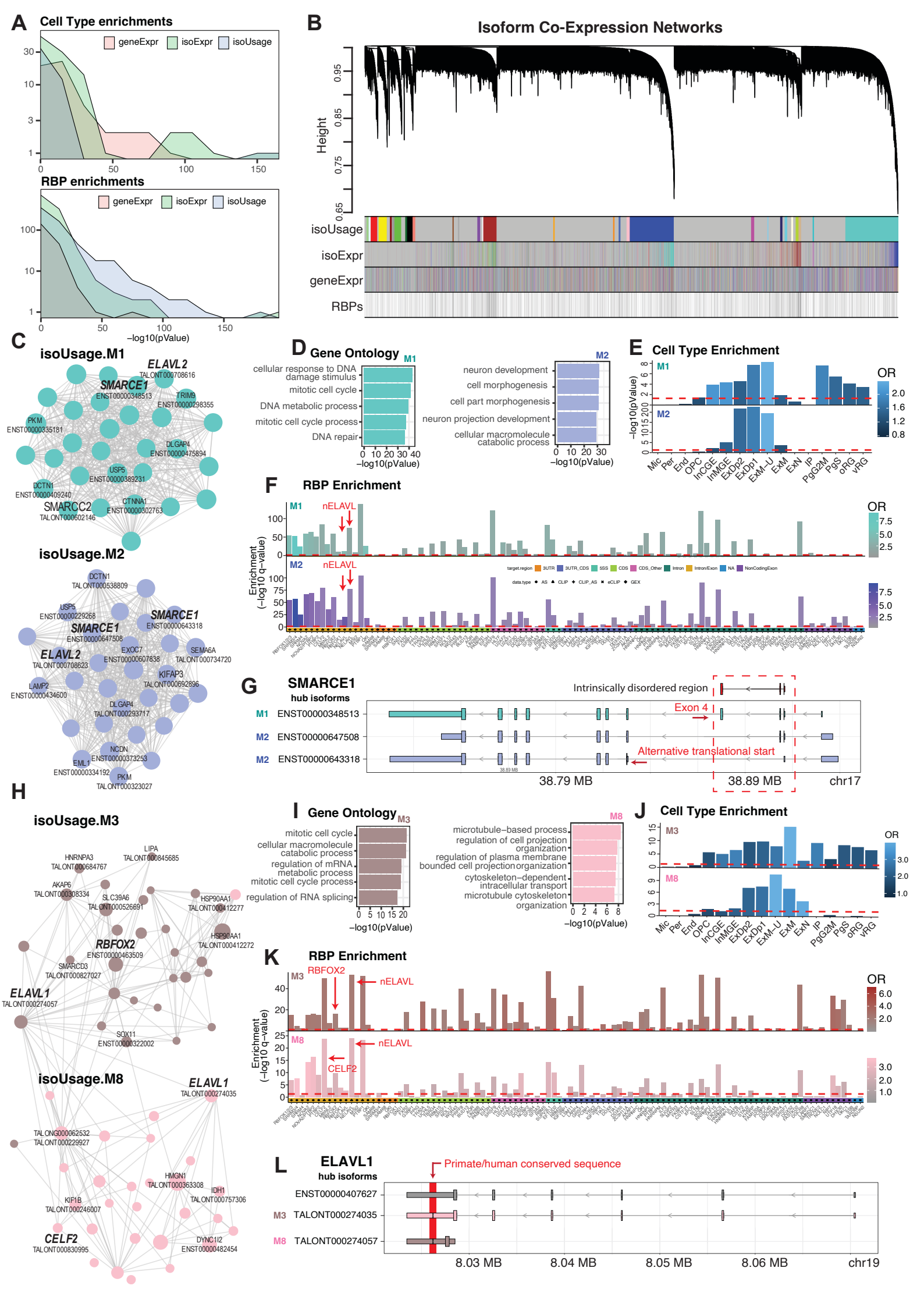
Fig. 6. Isoform-centric contextualization of neurogenetic risk mechanisms

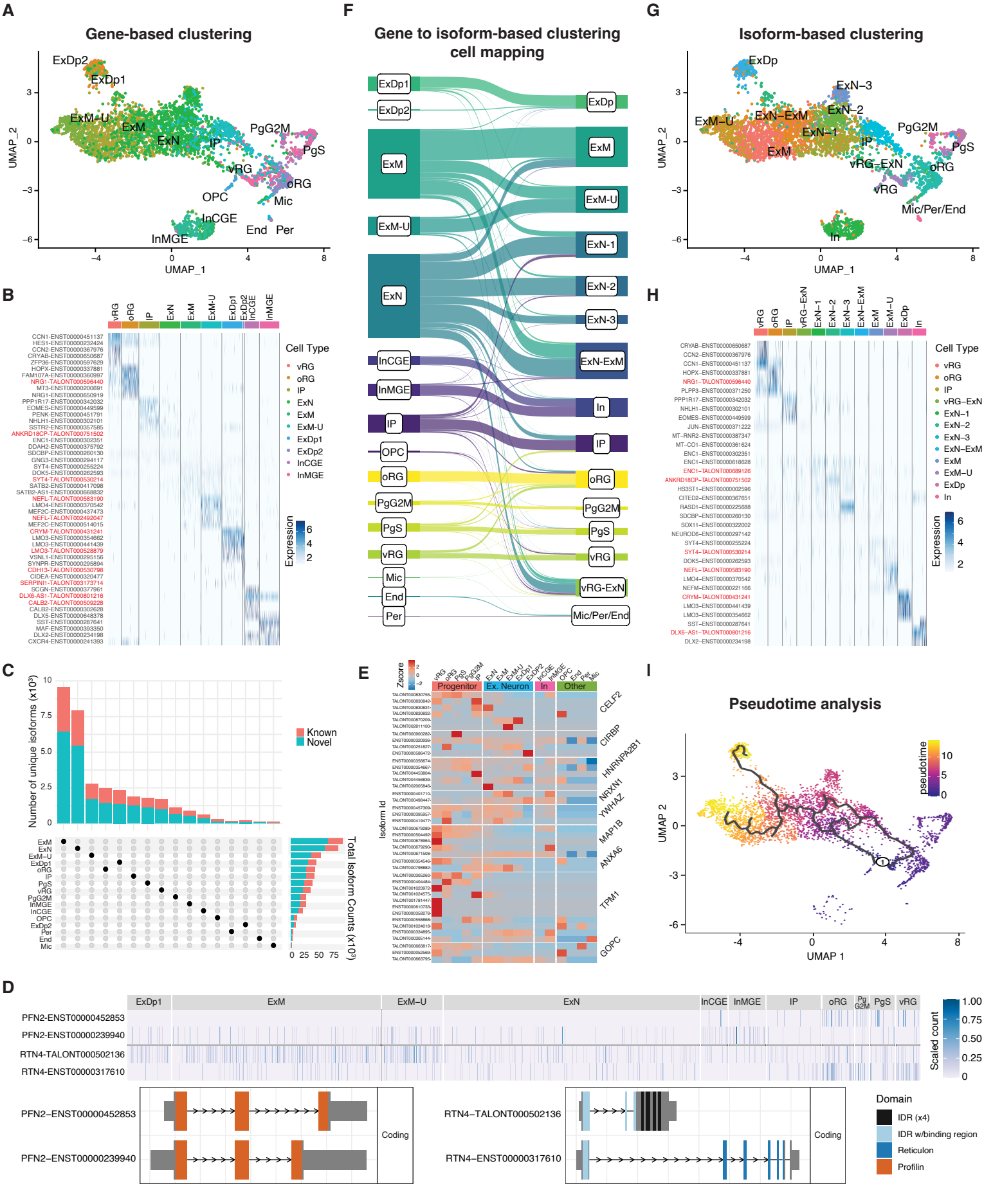
(A) Enrichment of transcriptomic features, differential expression analyses across cortical regions and cell types or isoform expression and usage networks with neuropsychiatric disorders. Red line indicates the FDR-corrected significance threshold. (**B**) Cell type enrichments indicate differential isoform expression and utilization in NDD and ASD. (**C**) Heatmap of isoforms from several NDD and ASD risk genes showing differential usage across the cell types of the developing cortex. Novel isoforms are labeled in red. (**D**) Number of variants that were reassigned to a more severe consequence after taking into account newly identified isoforms in this study. The size of the dots represents the number of variants in each category. The color of the dots indicates the source of the variants, i.e. DNM from case or control. Colored bars along the axes indicate both the severity of the consequences on a continuous scale, and the severity categories on a discrete scale, as defined by VEP. Reassignment to a different severity category may be more impactful than reassignment within the same category. (**E**) The *AKT3* gene locus with representative canonical isoforms and two novel isoforms identified from this study. Red vertical lines indicate the position of case DNMs that affect this locus. The affected regions are highlighted in the lower panels. Lower right panel, a DNM falls in the intron region in canonical protein isoforms, while leads to missense mutation in a novel protein isoform. Lower left, a DNM causes the loss of nearby splice acceptor and intron retention only in a novel protein isoform. The retained intron leads to shortened coding sequence and eliminates part of the protein kinase C-terminal domain. (**F**) Proportion of de novo mutations predicted to cryptically affect splicing, with or without the annotation of newly identified isoforms from this study.

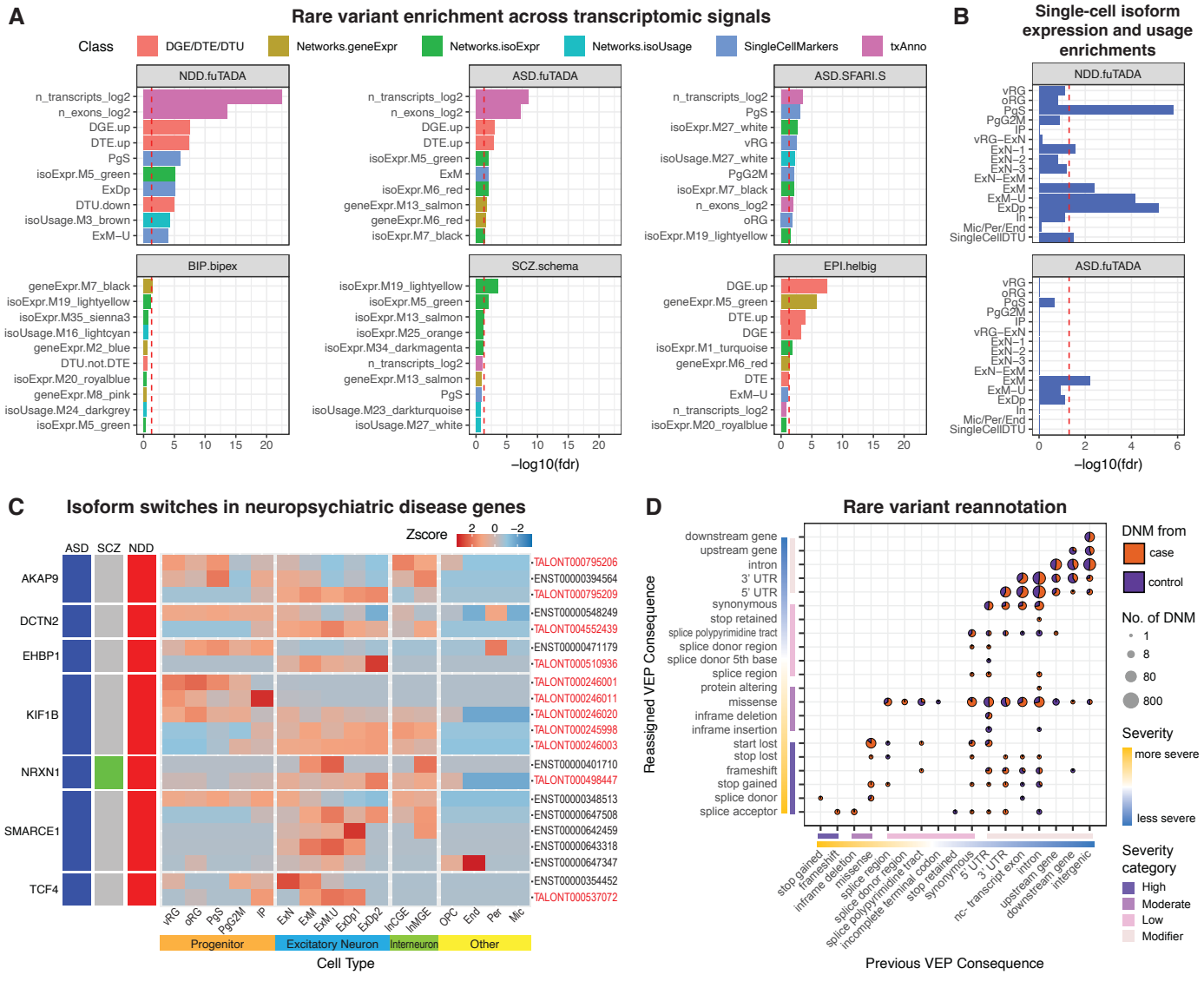

A
Developing human cortex
6 donors PCW 15-17
Coronal section
Microdissection
MZ
CP
SP
IZ
SVZ
VZ
Cortical Plate (CP)
Germinal Zone (GZ)
Bulk mRNA Isolation
Dissociation and Single Cell Barcoding
SMRTseq
Tn5
Long Read Sequencing (PacBio)
Short Read Sequencing (Illumina)
IsoSeq3 TALON
Seurat
Cell Type Clustering
tSNE1
tSNE2
Polioudakis et al.
B
Biological Replicates
R=0.93, p<2e−16
Iso Expr [Donor 336] log2(TPM+1)
Iso Expr [Donor 334] log2(TPM+1)
C
Isoform Expression Clustering
PC1
PC2
Region
CP
GZ
Donor
209
334
336
Batch
1 3
2 4
D
Transcript Classification
KNOWN (Gencode v33)
ISM-prefix
ISM-suffix
NIC
NNC
Other
NOVEL
E
Transcripts Identified by Novelty
Transcripts (x10³)
30%
10%
26%
24%
7%
3%
Known
Prefix ISM
Suffix ISM
NIC
NNC
Other
F
Gene Body Coverage
Relative coverage
Position on Gene 5' - 3'
G
Transcript Novelty vs Abundance
Transcript Proportion
Min observed counts
H
Transcript length distribution
Known < Novel ***
Transcripts (x10³)
Transcript Length (bp)
I
Exons per Transcript
Known < Novel ***
Transcripts (x10³)
Exons
J
Gene count
Dominant isoform Expression fraction
Gene Expr Quintile
1
2
3
4
5
K
Unique transcripts per gene
Transcripts
Gene rank
Neurodevelopmental disorder (NDD) risk gene
TTC3
DST
BAG6
NBPF1
MYT1L
HNRNPA2B1
KIF1B
MEIS2
KMT2C
GTF2I
EIF4G2
SMARCA4
SF3B1
KAT6B
SYNE2
CLASP2
MAP2
MEF2C
ZEB2
ANK2
ILF3
HNRNPK
SPTBN1
CHD4
SF1
MACF1
CHD3
PCBP2
ACIN1
MADD
KDM5B
EIF4A2
PUM1
SPTAN1

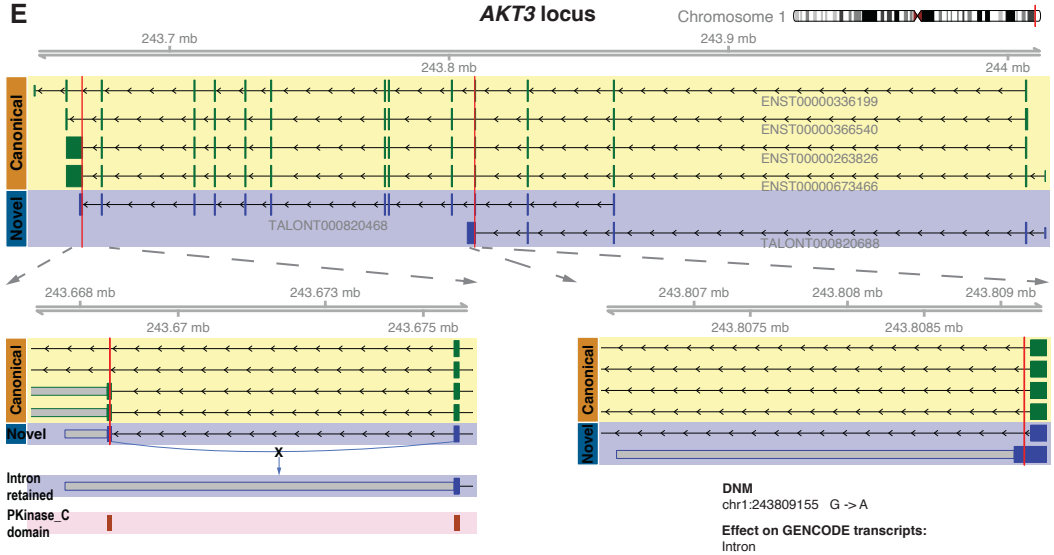

A
NF1 Locus
Chromosome 17
29.5 mb
29.6 mb
29.7 mb
CAGE
PolyA sites
ATAC
Gencode undetected
Gencode detected
Novel
B
5' support
3' support
Known
ISM Prefix
ISM Suffix
NIC
NNC
Other
CAGE
ATAC–seq
Proportion of Transcripts
polyA site
polyA motif
C Junction Support
IsoSeq
212,405
1,126,173
18,848
Intropolis
D
Novel Exons
Count
Width (bp)
Protein Coding Gene
T
F
E
P1
P2
Known Exon
Novel Exon
100 bp ladder
ASH1L (290 bp)
ILF2 (446 bp)
CTTNBP2 (425 bp)
MYT1L (395 bp)
100 bp ladder
500 bp
500 bp
F
214,516 Polished transcripts
GENCODE transcripts
149,510 Novel transcripts
Transcript classification with SQANTI
REF
FSM
ISM
NIC
NNC
Pfam database SwissProt human proteoforms
ORF calling with Transdecoder
ATG
domain1
domain2
TAA
92,422 Novel transcripts with ORF
Proteomics data from BrainSpan PsychENCODE
Novel peptides confirming expression of proteoforms
Novel transcribed regions
Novel peptides detected by mass-spec
35,467 Novel transcripts with peptide-supported ORF
G
STFKHGLGTA
Relative Intensity (%)
H
Peptide Support
Transcript (x10^3)
Known
ISM
NIC
NNC
Other
Category
Total
Protein coding
Validatable ORF
Peptide supported

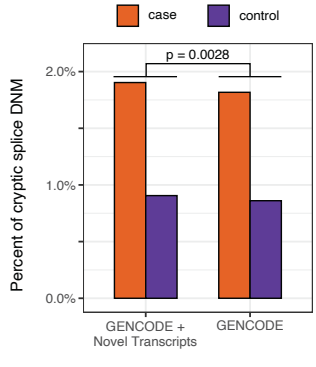

A
GZ vs CP
DTU
(Differential Transcript Usage)
DTE
(Differential Transcript Expression)
DGE
(Differential Gene Expression)
1025
644
58
913
1192
97
696
B
GZ vs CP DTU
−log10 (Isoform Switch Q Value)
difference in isoform fraction (dIF)
SMARCE1
Other (N=80)
NNC (N=274)
NIC (N=1610)
Known (N=2418)
ISM Suffix (N=616)
ISM Prefix (N=632)
C
Isoform Switch Conseq. (GZ vs CP)
3'UTR is longer
5'UTR is longer
Complete ORF gain
Peptide domain gain
Exon gain
Last exon more downstream
Length gain
NMD insensitive
ORF is longer
Signal peptide gain
Transcript is coding
TSS more downstream
TTS more downstream
FDR < 0.05
FALSE
TRUE
Switches
1000
2000
3000
4000
Fraction of Switches w/ Consequence
(of Switches Affected by Either of Opposing Consequences)
(With 95% Confidence Interval)
D
Distal polyA Site Usage (DPUI)
FDR<0.05
GZ DPUI
CP DPUI
E
Pathway Enrichments
dendrite morphogenesis
regulation of axon extension
regulation of nucleotide−excision repair
regulation of G0 to G1 transition
cerebral cortex development
lamellipodium
stress fiber
nBAF complex
centriolar subdistal appendage
SWI/SNF complex
cadherin binding
transmembrane transporter binding
clathrin binding
actin filament binding
microtubule plus−end binding
Bacterial invasion of epithelial cells
Axon guidance
Adherens junction
Salmonella infection
Endocytosis
GO:BP
GO:CC
GO:MF
KEGG
Enrichment
-log10(qvalue)
F
Isoform Switch in SMARCC2 (GZ vs CP)
SMARCC2-002 (ENST00000347471)
SMARCC2-003 (ENST00000394023)
SMARCC2-001 (ENST00000550164)
TALONT000602146
TALONT000602245
Peptide Domain
IDR
IDR w binding region (x3)
IDR_w_binding region (x4)
Myb DNA binding
SWIRM
SWIRM-assoc1
SWIRM-assoc2
SWIRM-assoc3
DGE
DTE
DTU
Gene Expression
Isoform Expression
Isoform Fraction (IF)
GZ CP
G
Cell-type Enrichments
Excitatory Neurons
Progenitors
Other
Int
Enrichment -log10(q-value)
OR
DGE
DTE
DTU
not DGE
H
RNA Binding Protein (RBP) Enrichments
Enrichment (−log10 q−value)
Literature
ENCODE
Target Region
3'UTR
CDS
5'SS
Intron
NA
Intron/Exon
5'SS/3'SS
3'UTR/CDS
NonCodingExon
Data Type
AS
CLIP
CLIP/AS
GEX
eCLIP
OR

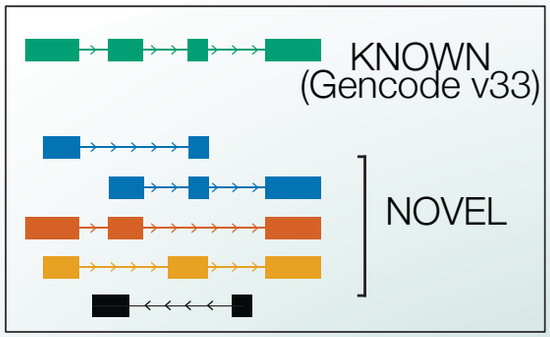

A
Cell Type enrichments
geneExpr
isoExpr
isoUsage
RBP enrichments
−log10(pValue)
B
Isoform Co-Expression Networks
Height
isoUsage
isoExpr
geneExpr
RBPs
C
isoUsage.M1
ELAVL2
SMARCE1
TRIM9
PKM
DLGAP4
USP5
DCTN1
SMARCC2
CTNNA1
isoUsage.M2
DCTN1
USP5
SMARCE1
ELAVL2
EXOC7
SEMA6A
KIFAP3
LAMP2
DLGAP4
NCDN
EML1
PKM
D
Gene Ontology
M1
cellular response to DNA damage stimulus
mitotic cell cycle
DNA metabolic process
mitotic cell cycle process
DNA repair
M2
neuron development
cell morphogenesis
cell part morphogenesis
neuron projection development
cellular macromolecule catabolic process
−log10(pValue)
E
Cell Type Enrichment
M1
M2
OR
−log10(pValue)
Mic
Per
End
OPC
InCGE
InMGE
ExDp2
ExDp1
ExM-U
ExM
ExN
IP
PgG2M
PgS
oRG
vRG
F
RBP Enrichment
M1
nELAVL
M2
nELAVL
OR
Enrichment (−log10 q−value)
target.region
3UTR
3UTR_CDS
5SS
CDS
CDS_Other
Intron
Intron/Exon
NA
NonCodingExon
data.type
AS
CLIP
CLIP_AS
eCLIP
GEX
G
SMARCE1
hub isoforms
M1 ENST00000348513
M2 ENST00000647508
M2 ENST00000643318
Intrinsically disordered region
Exon 4
Alternative translational start
38.79 MB
38.89 MB
chr17
H
isoUsage.M3
LIPA
HNRNPA3
AKAP6
SLC39A6
HSP90AA1
RBFOX2
SMARCD3
ELAVL1
SOX11
isoUsage.M8
ELAVL1
HMGN1
IDH1
KIF1B
CELF2
DYNC1I2
I
Gene Ontology
M3
mitotic cell cycle
cellular macromolecule catabolic process
regulation of mRNA metabolic process
mitotic cell cycle process
regulation of RNA splicing
M8
microtubule−based process
regulation of cell projection organization
regulation of plasma membrane bounded cell projection organization
cytoskeleton−dependent intracellular transport
microtubule cytoskeleton organization
−log10(pValue)
J
Cell Type Enrichment
M3
M8
OR
Mic
Per
End
OPC
InCGE
InMGE
ExDp2
ExDp1
ExM-U
ExM
ExN
IP
PgG2M
PgS
oRG
vRG
K
RBP Enrichment
M3
RBFOX2
nELAVL
M8
nELAVL
CELF2
OR
Enrichment (−log10 q−value)
L
ELAVL1
hub isoforms
Primate/human conserved sequence
ENST00000407627
M3 TALONT000274035
M8 TALONT000274057
8.03 MB
8.04 MB
8.05 MB
8.06 MB
chr19

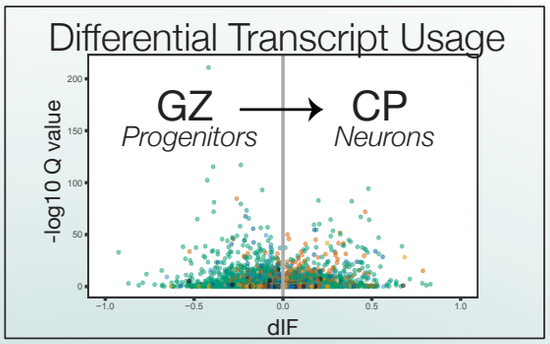

A
Gene-based clustering
ExDp2
ExDp1
ExM-U
ExM
ExN
IP
PgG2M
PgS
vRG
oRG
OPC
Mic
InCGE
InMGE
End
Per
UMAP_2
UMAP_1
F
Gene to isoform-based clustering cell mapping
ExDp1
ExDp2
ExM
ExM-U
ExN
InCGE
InMGE
IP
OPC
oRG
PgG2M
PgS
vRG
Mic
End
Per
ExDp
ExM
ExM-U
ExN-1
ExN-2
ExN-3
ExN-ExM
In
IP
oRG
PgG2M
PgS
vRG
vRG-ExN
Mic/Per/End
G
Isoform-based clustering
ExDp
ExN-3
ExN-2
ExN-ExM
ExM-U
ExN-1
PgG2M
IP
PgS
ExM
vRG-ExN
oRG
vRG
Mic/Per/End
In
B
vRG
oRG
IP
ExN
ExM
ExM-U
ExDp1
ExDp2
InCGE
InMGE
Cell Type
vRG
oRG
IP
ExN
ExM
ExM-U
ExDp1
ExDp2
InCGE
InMGE
Expression
H
vRG
oRG
IP
vRG–ExN
ExN–1
ExN–2
ExN–3
ExN–ExM
ExM
ExM–U
ExDp
In
Cell Type
vRG
oRG
IP
vRG–ExN
ExN–1
ExN–2
ExN–3
ExN–ExM
ExM
ExM–U
ExDp
In
Expression
C
Number of unique isoforms (x10³)
Known
Novel
Total Isoform Counts (x10³)
E
Zscore
Progenitor
Ex. Neuron
In
Other
Isoform Id
CELF2
CIRBP
HNRNPA2B1
NRXN1
YWHAZ
MAP1B
ANXA6
TPM1
GOPC
I
Pseudotime analysis
pseudotime
UMAP 2
UMAP 1
D
ExDp1
ExM
ExM–U
ExN
InCGE
InMGE
IP
oRG
PgG2M
PgS
vRG
PFN2-ENST00000452853
PFN2-ENST00000239940
RTN4-TALONT000502136
RTN4-ENST00000317610
Scaled count
PFN2–ENST00000452853
PFN2–ENST00000239940
Coding
RTN4–TALONT000502136
RTN4–ENST00000317610
Coding
Domain
IDR (x4)
IDR w/binding region
Reticulon
Profilin

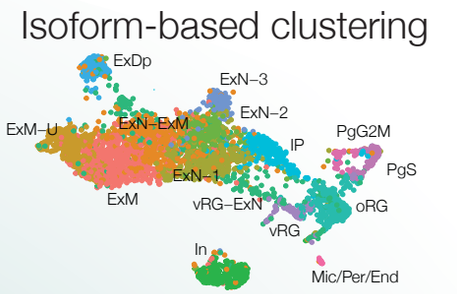

A
Rare variant enrichment across transcriptomic signals
Class
DGE/DTE/DTU
Networks.geneExpr
Networks.isoExpr
Networks.isoUsage
SingleCellMarkers
txAnno
NDD.fuTADA
ASD.fuTADA
ASD.SFARI.S
BIP.bipex
SCZ.schema
EPI.helbig
–log10(fdr)
B
Single-cell isoform expression and usage enrichments
NDD.fuTADA
ASD.fuTADA
–log10(fdr)
C
Isoform switches in neuropsychiatric disease genes
Zscore
ASD
SCZ
NDD
AKAP9
DCTN2
EHBP1
KIF1B
NRXN1
SMARCE1
TCF4
Progenitor
Excitatory Neuron
Interneuron
Other
Cell Type
D
Rare variant reannotation
Reassigned VEP Consequence
Previous VEP Consequence
DNM from
case
control
No. of DNM
Severity
more severe
less severe
Severity category
High
Moderate
Low
Modifier


E
AKT3 locus
Chromosome 1
Canonical
Novel
ENST00000336199
ENST00000366540
ENST00000263826
ENST00000673466
TALONT000820468
TALONT000820688
Intron retained
PKinase_C domain
DNM
chr1:243668598 G -> A
Effect on GENCODE transcripts:
Missense
Effect on novel transcript:
Splice acceptor loss
DNM
chr1:243809155 G -> A
Effect on GENCODE transcripts:
Intron
Effect on novel transcript:
Missense Gtc -> Atc V -> I


F
Rare variants predicted to cryptically affect splicing
case
control
p = 0.0028
Percent of cryptic splice DNM
GENCODE + Novel Transcripts
GENCODE







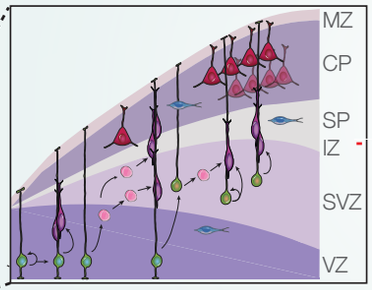
Isoform Discovery ***Spatio Temporal Dynamics*** ***Cell Type Specificity***

Isoform-centric transcriptome of developing human cortex

PCW 15-17
Cortex

6 donors

Microdissection

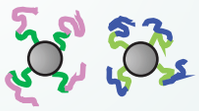


CP

GZ

Bulk mRNA

Single Cells



SMRTseq

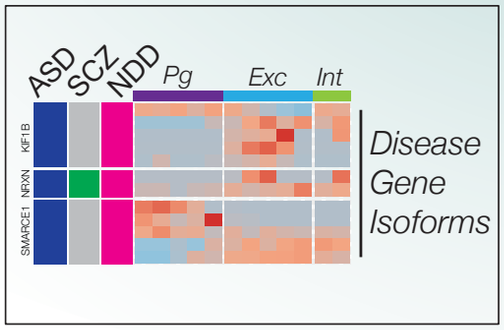
Long Read Sequencing

Drop-Seq

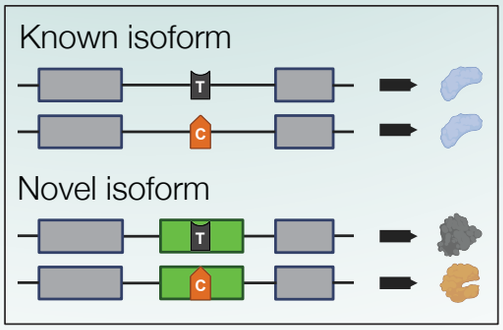
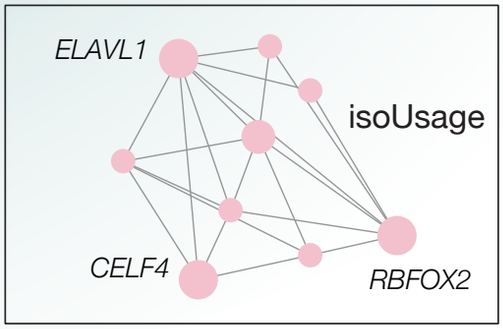
Short Read Sequencing

Neuropsychiatric Disease Mechanisms

Disease Enrichment



Variant Reannotation



RBPs & Isoform Networks