RESEARCH ARTICLE SUMMARY

PSYCHENCODE2

Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms

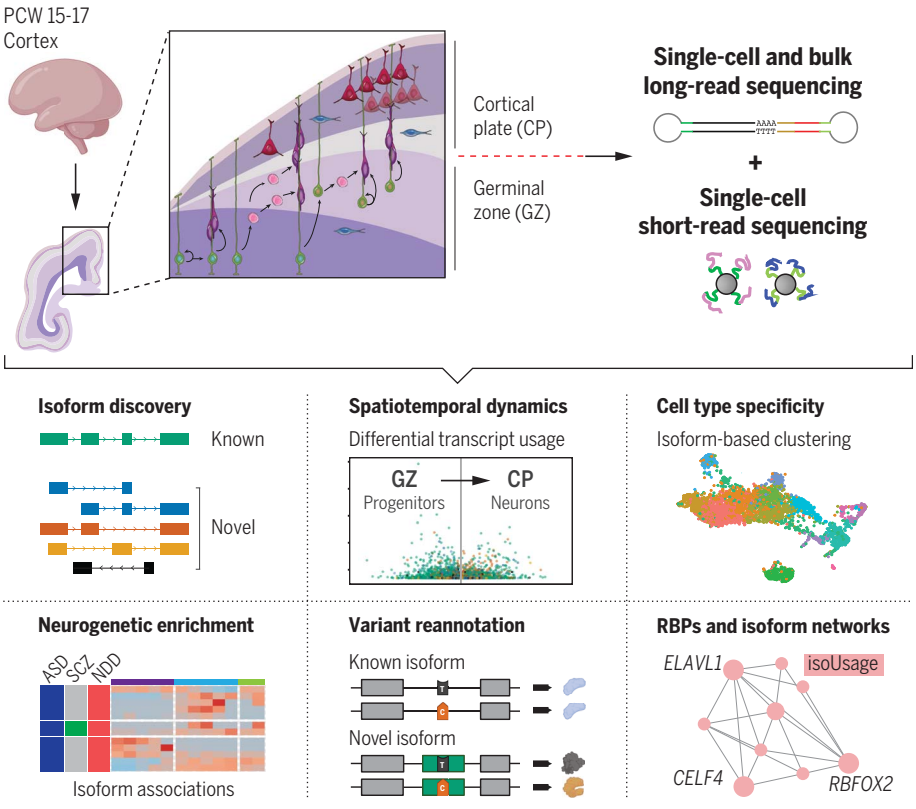
Ashok Patowary†, Pan Zhang†, Connor Jops†, Celine K. Vuong, Xinzhou Ge, Kangcheng Hou, Minsoo Kim, Naihua Gong, Michael Margolis, Daniel Vo, Xusheng Wang, Chunyu Liu, Bogdan Pasaniuc, Jingyi Jessica Li, Michael J. Gandal*, Luis de la Torre-Ubieta*

INTRODUCTION: The development of the human brain is regulated by precise molecular mechanisms driving spatiotemporal 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: Short-read sequencing, the prevalent technology for transcriptome profiling, is not well suited to capturing alternative splicing and isoform diversity. To address this, we used third-generation long-read sequencing, which enables the capture of complete R molecules, to profile the full-length transc tome 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 postconception week (PCW) 15 to 17 human neocortex across six subjects using high-fidelity long-read sequencing. We identified 214,516 distinct isoforms, of which 72.6% were novel (not previously annotated in Gencode v33), and >7000 novel exons, expanding the proteome by 92,422 putative proteoforms. We uncovered thousands of isoform switches during cortical neurogenesis predicted to affect RNA regulatory domains or protein structure and implicating previously uncharacterized RNA binding proteins (RBPs) 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 spatiotemporal transcript expression signatures across neuropsychiatric disorders. This revealed predominant enrichments in dynamic isoform expression and utilization patterns and that the number and complexity of isoforms per gene were strongly predictive of disease. Leveraging this resource, we reprioritized thousands of rare de novo risk variants associated with autism spectrum disorders (ASDs), intellectual disability, and neurodevelopmental disorders (NDDs) to potentially more-severe consequences and revealed a larger proportion of cryptic splice variants than previously reported.

CONCLUSION: Our study offers a comprehensive landscape of isoform diversity in the human neocortex during development. This extensive cataloging of isoforms and splicing events sheds light on the underlying mechanisms of NDDs and presents an opportunity to explore rare genetic variants linked to these conditions. Our findings also 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/). ■



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 long-read Iso-Seq. The bracket pointing out from a schematic summarizing data generation indicates major analyses performed in this study. We identified thousands of isoforms with specific regional (GZ or CP) and cell type expression coalescing into networks driven by cell type identity and RBP regulation. This resource reveals substantial contributions of isoform switching to cellular identity and elucidates genetic risk mechanisms for neurodevelopmental and neuropsychiatric disorders, including a reannotation of thousands of de novo rare variants with potential clinical implications. SCZ, schizophrenia. [Figure created with BioRender]

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ldelatorreubieta@mednet.ucla.edu (L.d.l.T.-U.); michael.gandal@pennmedicine.upenn.edu (M.J.G.)
†These authors contributed equally to this work
Cite this article as A. Patowary *et al.*, *Science* **384**, eadh7688 (2024). DOI: 10.1126/science.adh7688

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.adh7688

RESEARCH ARTICLE

PSYCHENCODE2

Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms

Ashok Patowary[1,2,3,4]†, Pan Zhang[1,2,3,4]†, Connor Jops[5,6,7]†, Celine K. Vuong[1,2,3], Xinzhou Ge[8], Kangcheng Hou[9], Minsoo Kim[1,2,4], Naihua Gong[5], Michael Margolis[4], Daniel Vo[5,6,7], Xusheng Wang[10,11], Chunyu Liu[12,13], Bogdan Pasaniuc[4,9,14,15,16], Jingyi Jessica Li[4,8,9,16,17], Michael J. Gandal[1,2,3,4,5,6,7]*, Luis de la Torre-Ubieta[1,2,3]*

RNA splicing is highly prevalent in the brain and has strong links to neuropsychiatric disorders; yet, the role of cell type–specific splicing and transcript-isoform diversity during human brain development has not been systematically investigated. In this work, we leveraged single-molecule long-read sequencing to deeply profile the full-length transcriptome of the germinal zone and cortical plate regions of the developing human neocortex at tissue and single-cell resolution. We identified 214,516 distinct isoforms, of which 72.6% were novel (not previously annotated in Gencode version 33), 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 reprioritize thousands of rare de novo risk variants and elucidate genetic risk mechanisms for neuropsychiatric disorders.

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 that regulate neurodevelopment (*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 multiexon genes (*9*). RNA transcript diversity results from the combinatorial effects of alternative transcription start sites (TSSs), alternative splicing, and distinct transcript termination sites due to alternative polyadenylation (APA) (*10*). In individual cases, such as for the synaptic gene *Nrxn1*, thousands of distinct isoforms have been identified (*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*). 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 have been observed in the mouse brain, even among closely related neurons, and often with precise temporal regulation (*16*–*20*).

AS has been implicated as an important mechanism linking genetic variation and neuropsychiatric disease (*2*, *21*–*25*). For example, recent work has identified marked splicing and isoform expression dysregulation in the brains of individuals with autism spectrum disorder (ASD) and schizophrenia (SCZ)—a signal substantially more widespread than gene expression changes with greater enrichment for genetic risk (*26*–*28*). Common variants associated with ASD and SCZ from genome-wide association studies (GWASs) exhibit substantial enrichment among splicing quantitative trait loci (sQTLs) in the developing and adult human brain (*2*, *24*). Likewise, rare de novo variants associated with ASD and intellectual disability (ID) are enriched for predicted splice-altering consequences (*23*). These genetic signals also exhibit convergence during midfetal brain development, which highlights the clinical relevance of this critical developmental period characterized by rapid increases in neurogenesis (*29*, *30*). Finally, a comprehensive understanding of splicing and isoform complexity can have direct therapeutic relevance, as has been demonstrated recently for spinal muscular atrophy (*31*).

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*, *32*). Coupled with single-cell barcoding, recent work has begun to catalog the isoform-centric transcriptome with single-cell resolution (*20*, *33*). In this work, we leverage this approach to deeply profile the major cell types of the developing human neocortex at midgestation. We uncover >150,000 previously unannotated transcript-isoforms and thousands of spliced exons, many within high-confidence neurodevelopmental disorder (NDD) risk genes. To functionally annotate these isoforms, we integrate proteomics, characterize regional patterns of isoform switching during corticogenesis, embed isoforms within coexpression networks, and map 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.

Full-length transcriptome of the developing human brain

We performed high-depth, high-fidelity long-read sequencing (PacBio HiFi Iso-Seq) to comprehensively profile the full-length polyadenylated transcriptome of the developing human neocortex across six donors at midgestation [postconception week (PCW) 15 to 17] (Fig. 1A). To interrogate patterns of differential transcript expression (DTE) and differential transcript usage (DTU) during peak neurogenesis, samples were first microdissected into the neural progenitor–enriched germinal zone (GZ) and the neuron-enriched cortical plate (CP). After sequencing and comprehensive quality control (Fig. 1, B and C), we generated >33 million

[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, 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, TN 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 and 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.
*Corresponding author. Email: ldelatorreubieta@mednet.ucla.edu (L.d.l.T.-U.); michael.gandal@pennmedicine.upenn.edu (M.J.G.)
†These authors contributed equally to this work.

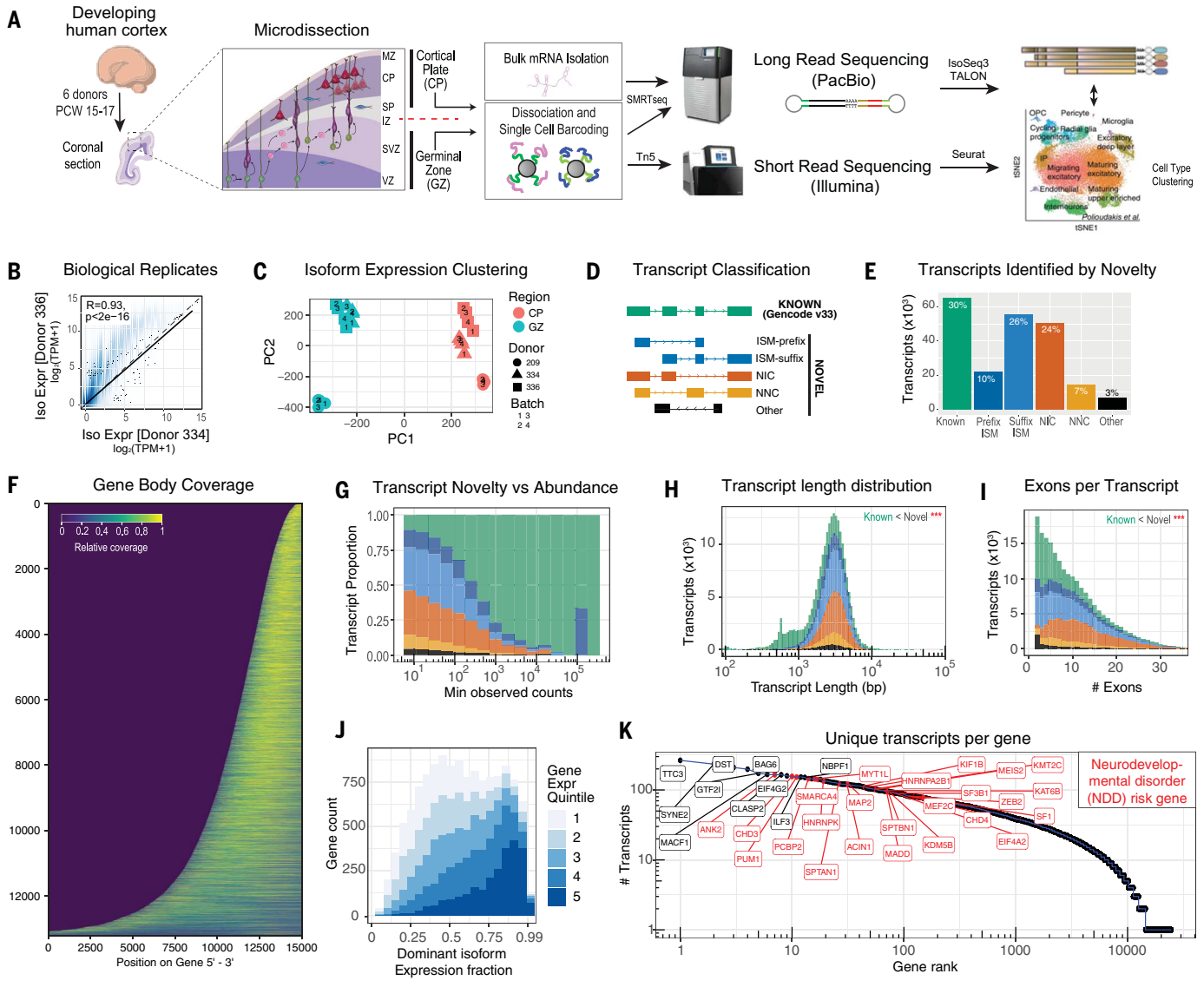


Fig. 1. The full-length cell type–specific transcriptome of the developing human neocortex at midgestation. (**A**) Experimental design for isoform-centric profiling of the developing human brain transcriptome at bulk and single-cell resolution. Briefly, microdissected samples from the progenitor-enriched GZ and neuronally enriched CP were profiled from six separate donors at midgestation. 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. MZ, marginal zone; SP, subplate; IZ, intermediate zone; SVZ, subventricular zone; VZ, ventricular zone. (**B** and **C**) Isoform expression quantifications demonstrated strong biological reproducibility across donors (B) and region-specific clustering through principal components analysis (C). (**D**) Transcript isoforms identified by Iso-Seq 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 is shown in dark blue and high coverage in yellow. (**G**) Abundance of the isoforms by each class as described in (D). The prefix ISM signal observed at ~10^5 minimum observed counts (*x* axis) largely corresponds to a highly expressed isoform of the *MAP1B* gene. (**H** and **I**) Compared with known isoforms, novel transcripts identified in this work were significantly longer ($P < 2 \times 10^{-16}$, Kruskal-Wallis) (H) and contained a significantly greater number of exons ($P < 2 \times 10^{-16}$, Kruskal-Wallis) (I). (**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) (*84*, *85*) had significantly more detected isoforms, controlling for total expression, gene length, and coding length (OR, 1.56; $P = 3.6 \times 10^{-3}$, logistic regression).

high-quality circular consensus sequence (CCS) reads across samples (figs. S1 and S2). Using minimap2 (*34*), >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. S2, A and B). Using the TALON pipeline (*35*), 214,516 distinct transcripts were identified in this bulk tissue transcriptome [data S1 in (*36*)], corresponding to 24,554 genes; of these, >175,000 isoforms from 17,299 genes were expressed at >0.1 transcripts per million (TPM) in at least half of the samples (Fig. 1, D and E). Isoform-level expression was highly reproducible 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 to 14,200 base pairs (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 [data S2 in (*36*)], merging data across samples. Compared with Gencode v33, only 65,006 (30.3%) of the observed isoforms matched existing transcripts (Fig. 1, D and E). We further classified isoforms based on their splice junction match to Gencode, strand specificity, and 5′ and/or 3′ overlap with known transcripts [Fig. 1D; see data S2 in (*36*)] (*37*). As a class, novel transcripts (those not previously annotated in Gencode v33) were more lowly expressed compared with known transcripts (Fig. 1G), although several were individually highly expressed, including *WASF1*, *CYFIP2*, *MAP1B*, *NEFL*, and *SMARCA4* (fig. S3, A to E).

Most isoforms not found in Gencode were classified as incomplete splice match (ISM) (Fig. 1, D and 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 and fig. S2A). Further, novel transcripts (including ISMs) were significantly longer ($P < 2 \times 10^{-16}$, Kruskal-Wallis; Fig. 1H) and contained more exons ($P < 2 \times 10^{-16}$, Kruskal-Wallis; Fig. 1I) compared with known transcripts. Consequently, we retained ISM transcripts with additional supporting evidence (*38*), and we note that many exhibited functional roles as hub isoforms in network analyses detailed below.

Given the 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 reads. Nevertheless, the number of detected isoforms per gene was strongly predictive of NDD risk gene status, accounting for gene length and total gene expression ($P = 5.2 \times 10^{-3}$, logistic regression; Fig. 1K). This association remained significant even when restricting only to novel transcripts [odds ratio (OR), 1.8; $P = 5.5 \times 10^{-4}$; 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 (*39*–*45*) and predicted potential downstream protein-coding consequences (Fig. 2). Altogether, ~80% of novel isoform TSSs were supported by proximal cap analysis of gene expression (CAGE) and/or assay for transposable-accessible chromatin with high-throughput sequencing (ATAC-seq) peaks. At the 3′ end, ~91% of novel transcripts were supported by nearby polyA sites or motifs—a rate higher than that for known transcripts (Fig. 2, A and B). Of 38,115 splice junctions not observed in Gencode, 74% were validated by Intropolis (Fig. 2C) (*45*). Finally, 53% of novel transcripts were validated when combining the latest version of Gencode (v43) with multiple independent long-read datasets, despite most representing nonneural tissues (fig. S3G). Together, these results provided broad orthogonal support for the novel isoforms discovered in this study.

In total, compared with Gencode, our data extended by ~27 Mb the transcribed portion of the human genome. Of this, 3.85 Mb comprised >7000 previously unannotated spliced exons, spanning >3500 distinct genes (Fig. 2D). Nearly 80% of these exons were supported by canonical splice junctions (table S1), and we validated several using reverse transcription polymerase chain reaction (RT-PCR) (Fig. 2E and table S1). Additionally, we identified 319 multiexonic genes not matching any existing gene model [data S2 in (*36*)], of which 256 were antisense to existing genes and 63 were located within intergenic regions. For example, we identified an 18-exon gene with four different 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 (*46*) provided peptide-level support for 35,467 novel transcripts with distinct proteoforms (table S1). 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 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–amino acid sequence (Fig. 2G). Notably, isoforms containing this microexon were only detected in neuronally enriched CP samples (fig. S4A). Extending these analyses beyond this single 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 and 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*, *39*), changes in 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 the GZ and CP [4475 of 24,554 genes, at false discovery rate (FDR) < 0.05; table S2]. Likewise, of genes with multiple expressed isoforms, a large proportion exhibited DTU across the GZ and CP (2679 of 10,809 genes at FDR < 0.05) (Fig. 3A and table S2), with the majority (57%) of significantly switching isoforms (5630 isoforms at FDR < 0.05) coming from novel transcripts (Fig. 3B). Although there was a significant overlap among DGE and DTU genes (1010 genes; OR, 2; $P < 10^{-46}$, Fisher's exact test; Fig. 3A), 1669 genes exhibited isoform switching without changes in overall expression. The majority of GZ and CP isoform-switching events had observed and/or predicted functional consequences (Fig. 3C), as summarized in table S2. For example, of 5630 isoform-switching events, 3204 (57%) were predicted to alter the ORF length, with longer ORFs predicted among CP–up-regulated isoforms (FDR-corrected $P < 0.0001$). However, CP–up-regulated isoforms were also more likely to exhibit predicted nonsense-mediated decay (NMD) sensitivity (243 versus 100 isoforms; FDR-corrected $P < 10^{-15}$). Together, these results implicate isoform switching as an important mechanism contributing to cell fate specification.

To further interrogate how the 3′ untranslated region (3′UTR) of transcripts differed between regions, we performed a complementary, read depth–based analysis of distal polyadenylation (polyA) usage (Fig. 3D). For the 9896 transcripts with multiple annotated polyA sites, we computed a distal polyA usage index (DPUI)—the fraction of total reads mapping to the longer (distal) 3′UTR (*47*). In total, 1013 transcripts exhibited significant differences in DPUI between the GZ and CP [repeated measures analysis of variance (ANOVA); FDR-corrected $P < 0.05$] (Fig. 3D and table S2). DPUIs were increased in the CP versus GZ for the majority (772) of transcripts, and DPUI was on average greater in the CP (two-sample paired t test, $P < 2.2 \times 10^{-16}$), indicating 3′UTR lengthening with neuronal maturation. Pathway enrichments were notable for RNA and/or mRNA binding as well as cytoplasmic stress granule gene ontologies. As a class, RNA binding proteins (RBPs) (*48*) were overrepresented among transcripts exhibiting significant DPUI changes (453/3527 RBP genes versus 560/6369 non-RBP genes; one-tailed Fisher's exact test, $P = 2.06 \times 10^{-10}$). These findings suggest a dynamic interplay between 3′UTR lengthening

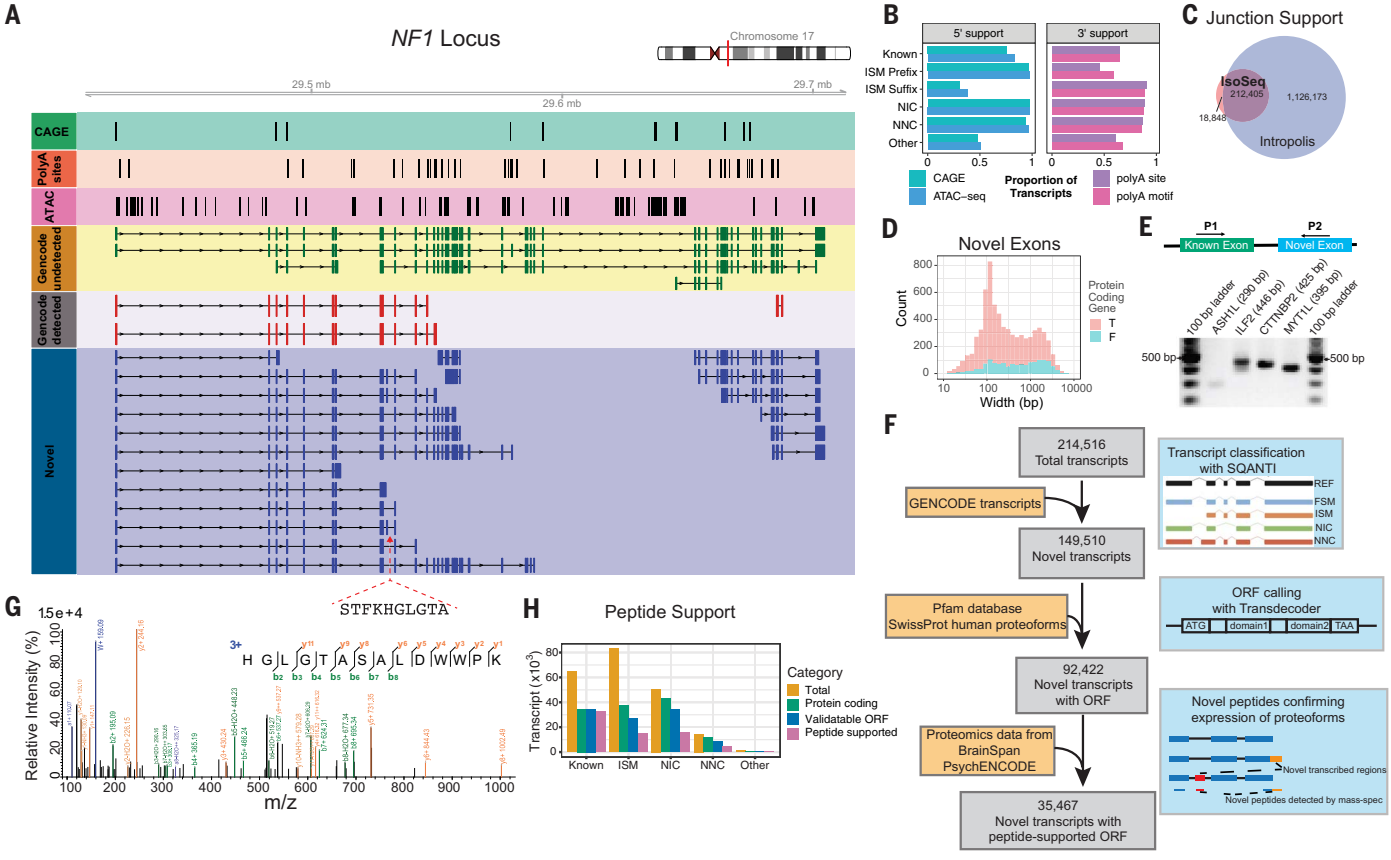


Fig. 2. Expanded transcriptomic and proteomic complexity. (**A**) The *NF1* gene locus with 20 previously unidentified brain-expressed isoforms. Tracks from top to bottom include CAGE clusters, 3′seq clusters, ATAC-seq peaks, Gencode isoforms that were not detected in our data, Gencode isoforms detected 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 midgestation cortex ATAC-seq (*39*–*41*); 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 majority of splice junctions identified by Iso-Seq 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 using 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. Of these novel transcripts, 92,422 were predicted to code for protein sequences, and 35,467 predicted ORFs were further confirmed by tandem mass spectrometry (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 are highlighted. m/z, mass/charge ratio. (**H**) Number of total transcripts, transcripts with ORFs, transcripts with novel ORFs compared with UniProt human protein sequences, and transcripts with ORFs validated by MS/MS proteomics, plotted per isoform structural category.

and RBP regulation in the transition from neural progenitor to neuron.

We next conducted pathway analyses for all genes exhibiting significant isoform switches across regions (DTU; one-sided Fisher's exact test, $P < 0.05$; table S2A). 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) (*49*, *50*). 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). Whereas overall expression of *SMARCC2* did not differ between the CP and GZ, two newly identified isoforms of *SMARCC2* showed preferential usage in the GZ and exhibited exon skipping compared with known transcripts. Other notable isoform-switching genes included known splicing regulators and RBPs (*SRRM1*, *SRRM4*, *CELF1*, *PTBP2*, *ELAVL2*, *ELAVL4*, and *RBFOX2*), chromatin modifiers (*KMT2E*, *KMT5B*, *SMARCA1*, *SMARCD3*, and *SMARCE1*), transcription factors (*FOXP2*), regulators of synaptic transmission (*GRIA3*, *VAMP1*, and *GAD1*), and synaptic cell-adhesion molecules (*NLGN4X* and *NRXN1*). Isoform-switching genes were broadly expressed across cell types, with particular enrichment for excitatory neuron lineages (Fig. 3G).

Putative RBP regulators of isoform-switching events

RBPs are a diverse class of proteins regulating the processing and fate of target mRNAs. Through regulation of splicing, RBPs alter isoform—and consequently protein—diversity and play important roles in mammalian neural development and function (*17*, *51*). To identify potential RBP regulators of isoform-switching events in the developing human brain, we compiled a set of experimentally defined RBP targets, including (i) those that regulate AS during neural development and/or maturation curated from previous work (i.e., brain-enriched) (Fig. 3H, black bar; fig. S4B, black bar; and fig. S4C) and (ii) targets defined by systematic RBP cross-linking immunoprecipitation (CLIP) in ENCODE (Fig. 3H, gray bar; fig. S4B, gray bar; and fig. S4C). Genes exhibiting regional isoform switching in our data showed significant overlap with targets of known RBP regulators of AS in the

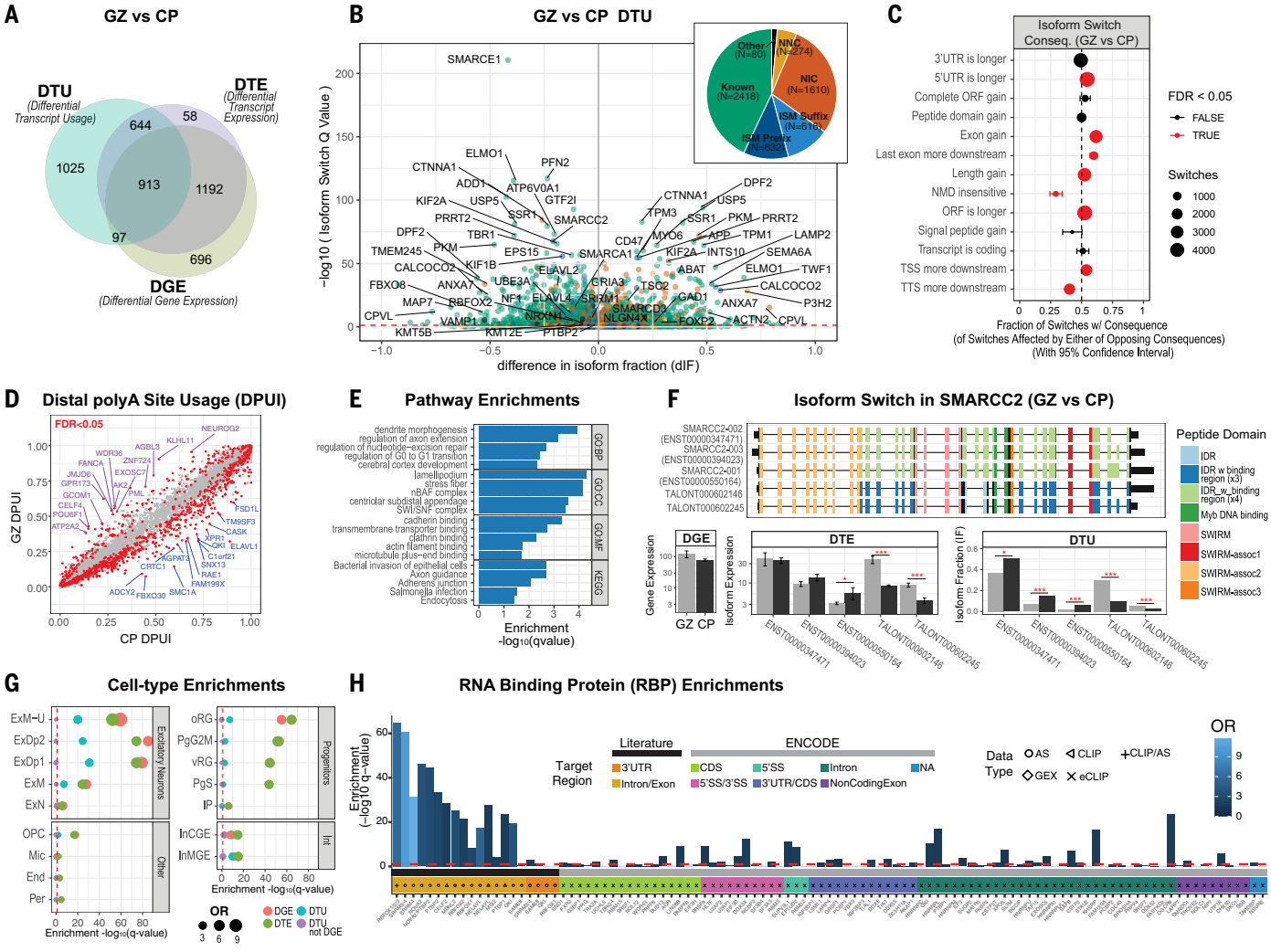


Fig. 3. The landscape of isoform switching during human corticogenesis. (**A**) Long-read RNA-seq data from GZ and CP samples were contrasted for patterns of DGE, DTE, and DTU. Venn diagram is shown depicting the overlap for genes exhibiting significant DGE, DTE, and DTU (FDR-corrected $P < 0.05$). (**B**) A volcano plot depicts isoform switching across the GZ and CP. The *x* axis depicts the difference in isoform fraction (dIF) for a given transcript in the CP versus GZ. (Inset) Most regionally variable DTU isoforms are not present in Gencode. (**C**) Functional consequences of isoform-switch events between the GZ and CP are shown. For example, CP–up-regulated isoforms were more significantly likely to have gained rather than lost an exon (2031 versus 1260 isoforms; FDR-corrected $P < 10^{-40}$). (**D**) Analysis of 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 the GZ and CP, significant switching was observed among DTU isoforms, with two isoforms exhibiting preferential usage in the 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 AS, gene expression (GEX), or direct binding (e/CLIP) are indicated. See also fig. S4C.

brain (one-sided Fisher's exact test, FDR-corrected $P < 0.05$; Fig. 3H) (*51*). These RBP splicing targets were more enriched among DTU compared with DTE or DGE genes. Conversely, gene expression targets—rather than AS targets—were more enriched in DGE and DTE compared with DTU genes (fig. S4B). Further, gene expression targets of RBPs that are more highly expressed in neurons than progenitors (e.g., *SRRM3/4*) (*51*) were enriched among genes that increase in expression over neural development, whereas those that are targets of progenitor-associated RBPs (*SAM68* and *PTBP1*) (*51*) showed the opposite trend (fig. S4C).

To identify RBPs important for—but not previously studied—during the developmental transition from progenitor to neuron, we repeated our enrichment analyses with a comprehensive set of RBP targets compiled from ENCODE (Fig. 3H) (*52*). As above, we found that ENCODE RBP targets were more enriched among DTU compared with DTE or DGE genes. Within DTU genes, ENCODE RBPs targeting introns, both 5′ and 3′ splice sites, and only 3′ splice sites were more enriched than those targeting the 3′UTR and coding sequence (CDS) or

only 3′UTRs (Fig. 3H; dark green, magenta, and aqua boxes versus green and dark blue boxes). Together, these findings highlight the importance of DTU analyses, with isoform switching 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 implicated in neural development and disease. These included RBPs known to regulate RNA metabolism and splicing, such as *LIN28B*, *EFTUD2*, *KHSRP*, and *DGCR8* (*53*–*55*). We observed a strong enrichment between DTU genes and targets of *DDX3X*, an X-linked RNA helicase where de novo mutations (DNMs) lead to sexually dimorphic ID and ASD (*56*, *57*). We also found strong enrichment of DTU genes among RBP targets with known roles in RNA metabolism, which have not yet been studied in the context of neural development. For example, DTU genes were enriched for targets bound by components of the exosome (the RNA degradation system) and those involved in ribosomal RNA biogenesis, such as *EXOSC5*, *UTP18*, and *SUPV3L1* (*58*–*60*), as well as the nuclear matrix protein *SAFB*, implicated in heterochromatin regulation (*61*). Altogether, these results indicate that although many GZ-CP isoform-switching events are likely regulated by brain-enriched RBPs through AS, many more 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 (*62*, *63*). We separately built unsupervised coexpression 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 on the basis of shared patterns of covariation across samples (fig. S5 and table S3), enabling in silico deconvolution of cell type isoform usage as well as functional inference through guilt by association (fig. S5). We first compared properties across the three networks to better understand which factors drove DTE, DGE, and DTU. Both geneExpr and isoExpr networks were strongly driven by cell type identity, with isoExpr modules in particular capturing specific progenitor and neuron subtypes (Fig. 4A and fig. S6). Multiple modules across isoExpr and isoUsage networks were enriched for RNA processing, cytoskeletal function, and chromatin regulation pathways (fig. S7, A to 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 the isoUsage network revealed expected enrichment patterns with modules exhibiting GZ or 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 $P < 0.05$) (*52*). 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 and M2 showed reciprocal GZ and CP specificity and concordant enrichments for progenitor cell function and neuronal morphogenesis pathways, respectively (Fig. 4, C to E). We also observed enrichment for neuronal markers in isoUsage.M1, reflecting the presence of neurons in the intermediate zone separating the GZ and CP. Across isoUsage.M1 and M2, we identified two novel (novel in catalog) 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 has demonstrated translational regulation of another ELAVL family member, *ELAVL4*, at its alternative 5′UTRs (*64*). Consistent with GZ or 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 (*65*) that mediates binding with higher-order complexes involved in chromatin remodeling and RNA splicing (*66*, *67*). 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 and M8, we observed complementary GZ and CP specificity and pathway enrichments (Fig. 4H) and identified multiple RBP hub transcripts (*ELAVL1*, *RBFOX2*, and *CELF2*) (Fig. 4, I and J) with concomitant enrichment of their targets (Fig. 4K). Whereas 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 (*68*, *69*) and that 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 (*70*) Both M3 and M8 *ELAVL1* isoforms contain a primate- and human-conserved, novel 3′UTR with a ~200-bp intron spliced out, potentially affecting autoregulation (Fig. 4L and fig. S7G). Altogether, these networks refine our understanding of the specific RNA regulons active in the developing brain (*48*).

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 single-cell isoform RNA-seq (ScISOr-Seq) (*20*) approach to profile >7000 single cells across an additional three separate donor samples derived from microdissected GZ and CP regions (Fig. 1A and fig. S8). Barcoded full-length single-cell cDNA libraries generated using Drop-seq, with incorporated unique molecular identifiers (UMIs) as published (*3*), were used as input to generate >26.4 million 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 7189 individual single-cell full-length transcriptomes, 4281 had matching barcodes from short-read sequencing (fig. S8, fig. S9A, and table S4A). All subsequent analyses were performed on this matched subset of the single-cell Iso-Seq (scIso-Seq) data.

After strict quality control and downstream processing (fig. S8 and fig. S9, A and B) (*38*), we detected on average 530 distinct transcripts per cell, mapping in aggregate to 18,541 genes and 138,497 isoforms (fig. S9C) [data S3 in (*36*)]. We observed high concordance between pseudo-bulk short-read– and long-read–based gene expression [correlation coefficient

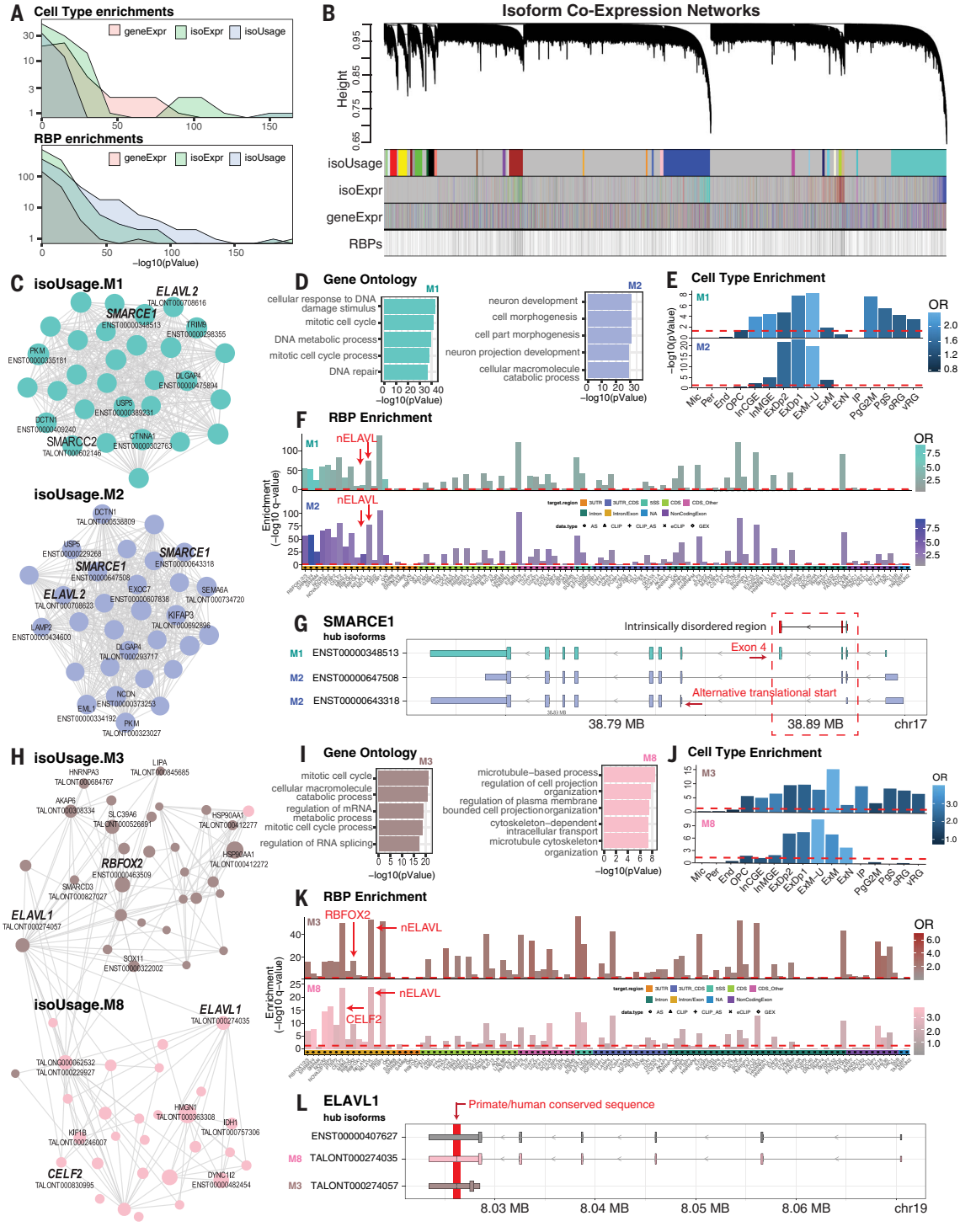


Fig. 4. Network-based contextualization of isoform usage. (**A**) The isoUsage network shows more and stronger enrichments for RBP targets compared with geneExpr and isoExpr. (Top) Density plot of cell type enrichments for the three networks. (Bottom) Density plot of RBP enrichments. (**B**) The isoUsage network is driven by RBP isoform usage. A dendrogram of the isoUsage network with isoforms (isoUsage and isoExpr) or genes (geneExpr) organized by their presence in isoUsage modules is plotted below. Isoforms of known RBPs 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**) GO 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 exon 3 (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, whereas 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**) GO for M3 and M8. (**J**) Cell type enrichments for M3 and M8. (**K**) RBP target enrichment for these modules includes targets of *RBFOX2*, *CELF2*, and *ELAVL2* (an nELAVL), which are hub isoforms in M3 and M8. (**L**) Transcript models for *ELAVL1* hub isoforms. The arrow highlights the primate- and human-conserved sequence missing from these 3′UTRs.

(r) = 0.9; $P < 2.2 \times 10^{-16}$] and detection (r = 0.92; $P < 2.2 \times 10^{-16}$) (fig. S9, D and E) and high interdonor reproducibility (r = 0.84 to 0.87; $P < 2.2 \times 10^{-16}$) (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 Iso-Seq 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 Iso-Seq (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 2452 out of 5165 previously unannotated spliced-in exons (table S4B), 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 (*71*) 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), which allowed us to construct cell type–specific isoform expression profiles (table S4 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*, *72*, *73*), with potential implications for neuropsychiatric disorders (*26*). 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 (table S4C) (*38*) to identify transcripts that define cellular identities and can serve as molecular markers (Fig. 5B). As expected, most 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 1040 transcripts enriched in specific cell types, 257 (24.7%) corresponded to isoforms newly identified in this study (table S4C). 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 and table S4C) (*74*–*79*). (Note, cell cluster abbreviations are defined in the Fig. 3 legend.)

We next conducted a pairwise DTE analysis to identify transcripts changing across specific cell type transitions, detecting 409 transcripts corresponding to 147 genes (table S4D; $P_{nominal} < 0.05$). 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 (*80*), 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 (Fig. 5D) (*81*, *82*). 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 the GZ and CP (Fig. 3), we sought to quantify similar events across individual cell types (*38*). We identified 1695 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 S4E). These instances represent switches in isoform utilization across cell types that may be missed by traditional DGE analyses. Of the 2284 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*, and *HNRNPA2B1*), cell division, regulation of synapse maturation (*NRXN1* and *YWHAZ*), and gene ontology (GO) categories related to cytoskeleton dynamics and vesicle transport (*MAP1B*, *ANXA6*, *TPM1*, and *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. Reclustering cells on the basis of isoform expression yielded 15 highly stable clusters largely mapping to many of the same cell classes as defined by gene-based clustering (Fig. 5, F to H, and fig. S10D) (*38*). 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 to ExN3), 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 and 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 S4C). Across ExN clusters, isoforms of *ENC1*, *ANKRD18CP*, and *RASD1* were enriched in ExN1, ExN2, and ExN3 cells, respectively (fig. S10, E and F). Moreover, vRG-ExN cells were defined by a large proportion of transcripts involved in mitochondrial metabolism, consistent with a recent study of the role of mitochondria in regulating neuronal maturation (*83*). 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 whole-genome sequencing studies of neurodevelopmental and psychiatric disorders, including ASD (*84*), NDD (*84*, *85*), SCZ (*86*), bipolar disorder (BIP) (*87*), and epilepsy (*88*). Risk genes for NDD, ASD, and DDD (developmental disabilities) had significantly more isoforms (log_2 scale; ORs, 1.14 to 1.4; q values < 3×10^{-4}) and exons (log_2 scale; ORs, 1.19 to 1.64; q values < 0.01) compared with nondisease genes (Fig. 6A, fig. S11A, and table S5A) (FDR-corrected P values from logistic regression, correcting for gene length and coding length). These associations were not observed for epilepsy (ORs, 1.14 and 1.22; not significant), BIP (ORs, 1.12 and 1.22; not significant), or SCZ (ORs, 1.4 and 1.56; not significant). 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 up-regulated in the CP (DGE.up, DTE.

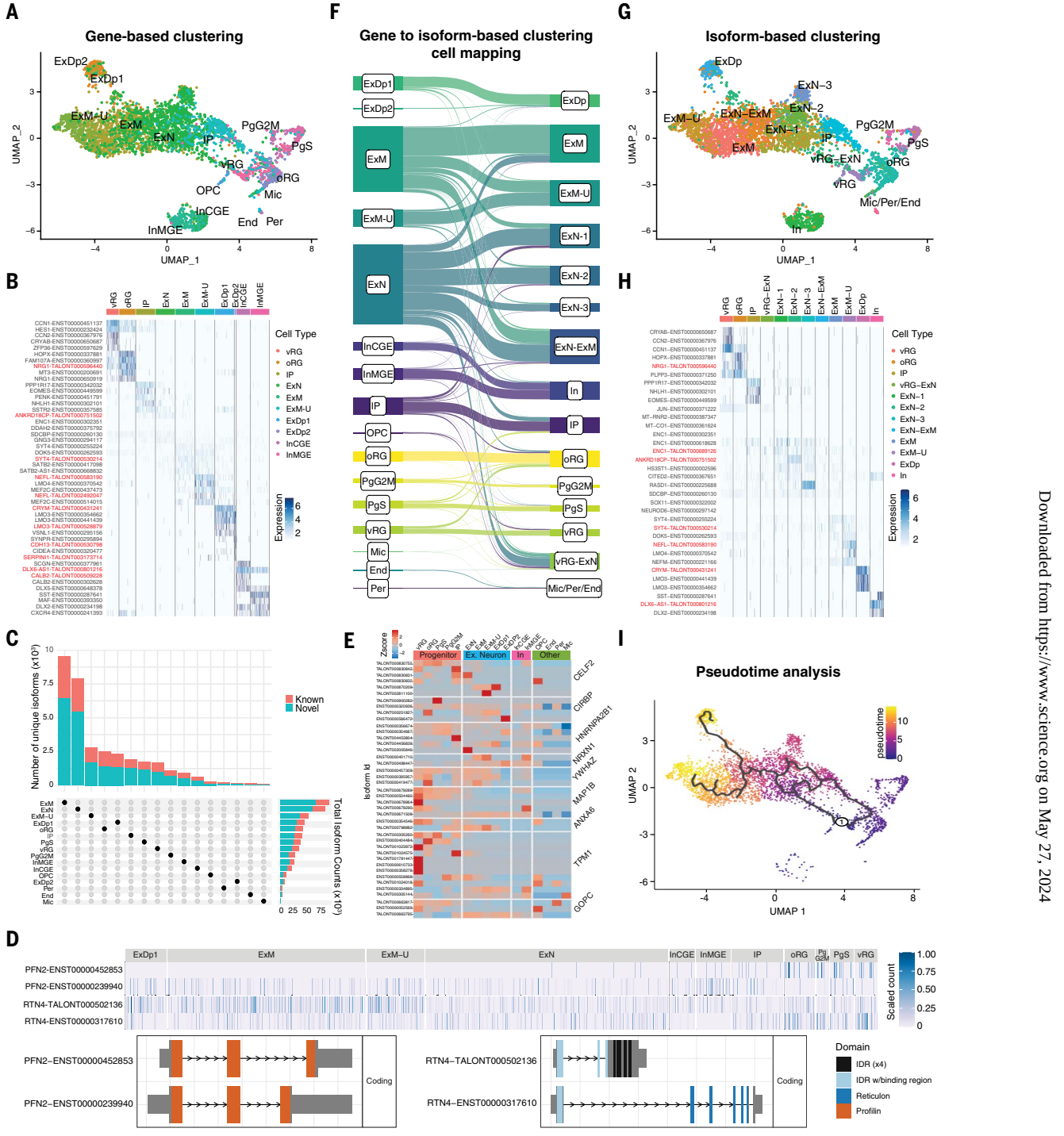


Fig. 5. Cell type–specific isoform diversity in the developing human cortex. (**A**) Uniform manifold approximation and projection (UMAP) plot of 4281 cells detected by both 3′ end short-read sequencing and by scIso-Seq. Each dot represents a single cell, colored by its corresponding cluster. UMAP position of the cells is calculated based on isoform expression, whereas cluster labels are as previously defined (*3*). (**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 with other cell types in midgestation 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 scIso-Seq. 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.

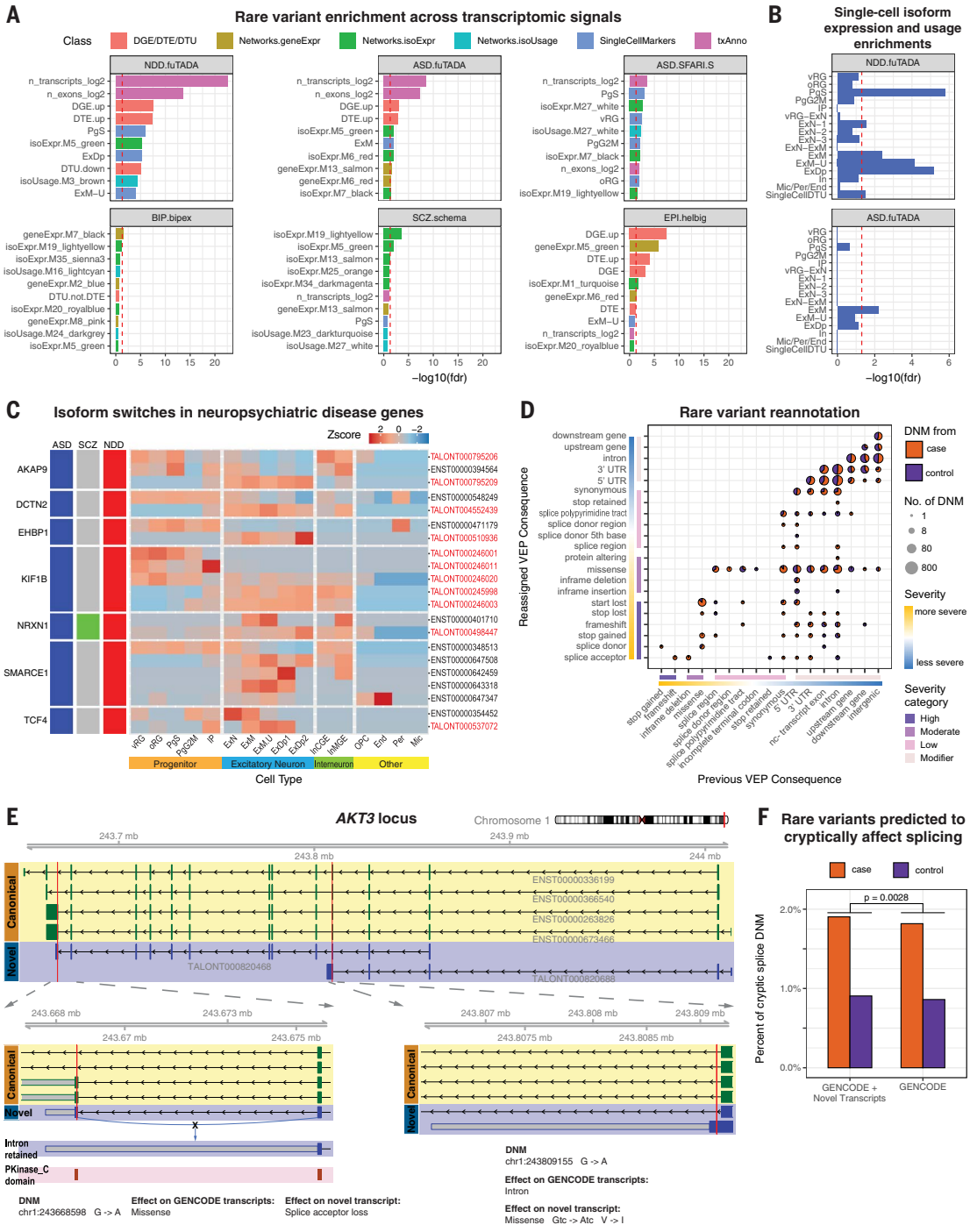


Fig. 6. Isoform-centric contextualization of neurogenetic risk mechanisms. (**A**) Enrichment of transcriptomic features, differential expression analyses across cortical regions and cell types, or isoform expression and usage networks with neuropsychiatric disorders. Red lines indicate the FDR-corrected significance thresholds. (**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) A DNM located in an intronic region of canonical protein isoforms leads to a 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 DNMs predicted to cryptically affect splicing, with or without the annotation of newly identified isoforms from this study.

up), indicative of neuronal expression, and were shared across NDD, nonsyndromic 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, "SingleCellDTU"), and many NDD and ASD genes showed DTU across cell types in the developing neocortex (Fig. 6C and fig. S12A). NDD, DDD, and ASD gene isoforms were primarily enriched in excitatory neurons (ExM-U, ExDp, or ExM) (Fig. 6B and 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 and 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 and isoform covariation 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 nonsyndromic ASD shared overlapping molecular signatures with modules mainly enriched for neuronal markers (fig. S13) and isoExpr modules regulating chromatin and histone modification (isoExpr.M7) (Fig. 6A and fig. S7B) and RNA metabolism and splicing (isoExpr.M11, M10, and M28) (table S5A and fig. S7, C to 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 and M24) (table S5A and fig. S7D) and ribosomal RNA processing and chromatin (isoUsage.M29) (table S5A and fig. S7C). These findings support a major role for isoform expression and diversification in neuropsychiatric disease mechanisms during development—consistent with recent findings (*26*)—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 (*89*–*91*) and intellectual disability/developmental disorders (ID/DD) (*91*). We reasoned that some variants previously disregarded as noncoding may actually fall within the >27 Mb 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 reannotated the set of compiled genetic variants (N_{total} = 272,187; N_{case} = 145,880; $N_{control}$ = 126,307) using Ensembl VEP (Variant Effect Predictor). Altogether, this annotation framework uncovered more severe consequences for 1.24% of all variants (Fig. 6D and table S5B). 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 DNM from the ASD cohort (*92*), previously classified as a benign intronic variant but now predicted to cause a missense mutation in the newly identified amino acid sequence. In another example, a reported DNM from the ID/DD cohort (*93*) 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 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 NDDs, 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 and table S5C) (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), whereas 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 (*94*), and dysregulation of *AKT3* is associated with NDDs (*95*). Our findings highlight that multiple DNMs may contribute to NDDs by affecting specific *AKT3* isoforms. More broadly, these results demonstrate that a more complete catalog of midgestation brain–expressed full-length isoforms provides more granular molecular insight into the genetic risk mechanisms underlying NDDs.

Discussion

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*, *26*), 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 >27 Mb 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 specific to the particular donors, developmental periods, and/or regions profiled in this work or, alternatively, may 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 provide considerable support for their functional importance during neurogenesis and in NDDs.

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 the 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 (*96*, *97*). Although there are many individual examples of APA in neuronal RBP genes (*96*, *98*, *99*), our analysis found that APA of RBP genes occurs on a large scale during human corticogenesis, which indicates that tight regulation of RBP activity is important for neurogenesis.

In network analyses, we found that although 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 a primate- and human-specific novel 3′UTR sequence of the RBP *ELAVL1*. Although many developmental processes are conserved across vertebrates in early neural development, primate- and human-specific differences are important to understand given the specific cell types found in the expanded cortices of these species (*100*, *101*). 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 (ExN1 to ExN3) 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 disease risk GWASs. 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 reannotate and reprioritize thousands of de novo ASD and IDD rare variants. The large number of previously unannotated transcripts identified in this work suggests that the functional consequences of many variants may have been missed using previous incomplete annotations. 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 institutional review board (IRB) and the UCLA Office of Human Research Protection regulations, with full informed consent from the parent donors. GZ and CP regions of PCW 15 to 17 cortices were microdissected and processed for bulk Iso-Seq or scIso-Seq. For bulk Iso-Seq, the 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 and isoforms. DGE, transcript usage, and APA site usage analyses were conducted using DESeq2, IsoformSwitchAnalyzeR, and DaPars2, respectively. Pathway enrichment was performed using gProfileR, and RBP target gene sets were analyzed for overlap with gene lists or module-associated genes or 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. scIso-Seq 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 DTU 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 or isoform modules, or differentially expressed genes or isoforms using logistic regression controlling for gene and coding length and transcript expression. ASD and IDD de novo variants (*89*–*91*) 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 *et al*., Radial glia require PDGFD-PDGFRβ signalling in human but not mouse neocortex. *Nature* **515**, 264–268 (2014). doi: 10.1038/nature13973; pmid: 25391964
2. R. L. Walker *et al*., Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750–771.e22 (2019). doi: 10.1016/j.cell.2019.09.021; pmid: 31626773
3. D. Polioudakis *et al*., A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron* **103**, 785–801.e8 (2019). doi: 10.1016/j.neuron.2019.06.011; pmid: 31303374
4. T. J. Nowakowski *et al*., Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017). doi: 10.1126/science.aap8809; pmid: 29217575
5. S. Zhong *et al*., A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018). doi: 10.1038/nature25980; pmid: 29539641
6. G. La Manno *et al*., RNA velocity of single cells. *Nature* **560**, 494–498 (2018). doi: 10.1038/s41586-018-0414-6; pmid: 30089906
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). doi: 10.1093/nar/gku555; pmid: 25053837
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). doi: 10.1186/s13059-020-01981-w; pmid: 32293520
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). doi: 10.1126/science.1228186; pmid: 23258891
10. C. Foord *et al*., The variables on RNA molecules: Concert or cacophony? Answers in long-read sequencing. *Nat. Methods* **20**, 20–24 (2023). doi: 10.1038/s41592-022-01715-9; pmid: 36635536
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). doi: 10.1073/pnas.1403244111; pmid: 24639501
12. M. Melé *et al*., The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015). doi: 10.1126/science.aaa0355; pmid: 25954002
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). doi: 10.1016/j.neuron.2015.05.004; pmid: 26139367
14. S. L. Zipursky, W. B. Grueber, The molecular basis of self-avoidance. *Annu. Rev. Neurosci.* **36**, 547–568 (2013). doi: 10.1146/annurev-neuro-062111-150414; pmid: 23841842
15. X. Zhang *et al*., Cell-Type-Specific Alternative Splicing Governs Cell Fate in the Developing Cerebral Cortex. *Cell* **166**, 1147–1162.e15 (2016). doi: 10.1016/j.cell.2016.07.025; pmid: 27565344
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). doi: 10.1038/s41593-019-0465-5; pmid: 31451803
17. S. M. Weyn-Vanhentenryck *et al*., Precise temporal regulation of alternative splicing during neural development. *Nat. Commun.* **9**, 2189 (2018). doi: 10.1038/s41467-018-04559-0; pmid: 29875359
18. Y. Song *et al*., Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. *Mol. Cell* **67**, 148–161.e5 (2017). doi: 10.1016/j.molcel.2017.06.003; pmid: 28673540
19. A. S. Booeshaghi *et al*., Isoform cell-type specificity in the mouse primary motor cortex. *Nature* **598**, 195–199 (2021). doi: 10.1038/s41586-021-03969-3; pmid: 34616073
20. I. Gupta *et al*., Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202 (2018). doi: 10.1038/nbt.4259; pmid: 30320766
21. Y. I. Li *et al*., RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016). doi: 10.1126/science.aad9417; pmid: 27126046
22. T. Raj *et al*., Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat. Genet.* **50**, 1584–1592 (2018). doi: 10.1038/s41588-018-0238-1; pmid: 30297968
23. K. Jaganathan *et al*., Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535–548.e24 (2019). doi: 10.1016/j.cell.2018.12.015; pmid: 30661751
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). doi: 10.1038/ncomms14519; pmid: 28240266
25. K. K. Chau *et al*., Full-length isoform transcriptome of the developing human brain provides further insights into autism. *Cell Rep.* **36**, 109631 (2021). doi: 10.1016/j.celrep.2021.109631; pmid: 34469739
26. M. J. Gandal *et al*., Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018). doi: 10.1126/science.aat8127; pmid: 30545856
27. N. Akula *et al*., Deep transcriptome sequencing of subgenual anterior cingulate cortex reveals cross-diagnostic and diagnosis-specific RNA expression changes in major psychiatric disorders. *Neuropsychopharmacology* **46**, 1364–1372 (2021). doi: 10.1038/s41386-020-00949-5; pmid: 33558674

28. M. J. Gandal *et al*., Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022). doi: 10.1038/s41586-022-05377-7; pmid: 36323788
29. A. J. Willsey *et al*., Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013). doi: 10.1016/j.cell.2013.10.020; pmid: 24267886
30. N. N. Parikshak *et al*., Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013). doi: 10.1016/j.cell.2013.10.031; pmid: 24267887
31. E. J. Lopez Soto *et al*., Mechanisms of Neuronal Alternative Splicing and Strategies for Therapeutic Interventions. *J. Neurosci.* **39**, 8193–8199 (2019). doi: 10.1523/JNEUROSCI.1149-19.2019; pmid: 31619487
32. R. Stark, M. Grzelak, J. Hadfield, RNA sequencing: The teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019). doi: 10.1038/s41576-019-0150-2; pmid: 31341269
33. A. Joglekar *et al*., A spatially resolved brain region- and cell type-specific isoform atlas of the postnatal mouse brain. *Nat. Commun.* **12**, 463 (2021). doi: 10.1038/s41467-020-20343-5; pmid: 33469025
34. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018). doi: 10.1093/bioinformatics/bty191; pmid: 29750242
35. D. Wyman *et al*., A technology-agnostic long-read analysis pipeline for transcriptome discovery and quantification. *bioRxiv*672931 [Preprint] (2020); .doi: 10.1101/672931
36. A. Patowary *et al*., Developmental isoform diversity in the human neocortex informs neuropsychiatric risk mechanisms, Dataset, *Dryad* (2024); .doi: 10.5061/dryad.0rxwdbs68
37. M. Tardaguila *et al*., SQANTI: Extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res.* **28**, 396–411 (2018). doi: 10.1101/gr.222976.117; pmid: 29440222
38. Materials and methods are available as supplementary materials.
39. L. de la Torre-Ubieta *et al*., The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289–304.e18 (2018). doi: 10.1016/j.cell.2017.12.014; pmid: 29307494
40. A. E. Trevino *et al*., Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. *Cell* **184**, 5053–5069.e23 (2021). doi: 10.1016/j.cell.2021.07.039; pmid: 34390642
41. R. S. Ziffra *et al*., Single-cell epigenomics reveals mechanisms of human cortical development. *Nature* **598**, 205–213 (2021). doi: 10.1038/s41586-021-03209-8; pmid: 34616060
42. S. Noguchi *et al*., FANTOM5 CAGE profiles of human and mouse samples. *Sci. Data* **4**, 170112 (2017). doi: 10.1038/sdata.2017.112; pmid: 28850106
43. C. J. Herrmann *et al*., PolyASite 2.0: A consolidated atlas of polyadenylation sites from 3′ end sequencing. *Nucleic Acids Res.* **48**, D174–D179 (2020). doi: 10.1093/nar/gkz918; pmid: 31617559
44. L. M. Pardo *et al*., Regional differences in gene expression and promoter usage in aged human brains. *Neurobiol. Aging* **34**, 1825–1836 (2013). doi: 10.1016/j.neurobiolaging.2013.01.005; pmid: 23428183
45. A. Nellore *et al*., 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). doi: 10.1186/s13059-016-1118-6; pmid: 28038678
46. B. C. Carlyle *et al*., A multiregional proteomic survey of the postnatal human brain. *Nat. Neurosci.* **20**, 1787–1795 (2017). doi: 10.1038/s41593-017-0011-2; pmid: 29184206
47. L. Li *et al*., An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.* **53**, 994–1005 (2021). doi: 10.1038/s41588-021-00864-5; pmid: 33986536
48. F. Gebauer, T. Schwarzl, J. Valcárcel, M. W. Hentze, RNA-binding proteins in human genetic disease. *Nat. Rev. Genet.* **22**, 185–198 (2021). doi: 10.1038/s41576-020-00302-y; pmid: 33235359
49. 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). doi: 10.3389/fnmol.2017.00243; pmid: 28824374
50. 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). doi: 10.3389/fnmol.2022.805158; pmid: 35185468
51. C. K. Vuong, D. L. Black, S. Zheng, The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* **17**, 265–281 (2016). doi: 10.1038/nrn.2016.27; pmid: 27094079
52. E. L. Van Nostrand *et al*., A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020). doi: 10.1038/s41586-020-2077-3; pmid: 32728246
53. M. Yang *et al*., Lin28 promotes the proliferative capacity of neural progenitor cells in brain development. *Development* **142**, 1616–1627 (2015). doi: 10.1242/dev.120543; pmid: 25922525
54. L. Lei *et al*., Spliceosomal protein eftud2 mutation leads to p53-dependent apoptosis in zebrafish neural progenitors. *Nucleic Acids Res.* **45**, 3422–3436 (2017). doi: 10.1093/nar/gkw1043; pmid: 27899647
55. C. W. Bird *et al*., KSRP modulation of GAP-43 mRNA stability restricts axonal outgrowth in embryonic hippocampal neurons. *PLOS ONE* **8**, e79255 (2013). doi: 10.1371/journal.pone.0079255; pmid: 24244461
56. A. L. Lennox *et al*., Pathogenic *DDX3X* Mutations Impair RNA Metabolism and Neurogenesis during Fetal Cortical Development. *Neuron* **106**, 404–420.e8 (2020). doi: 10.1016/j.neuron.2020.01.042; pmid: 32135084
57. M. L. Hoye *et al*., Aberrant cortical development is driven by impaired cell cycle and translational control in a *DDX3X* syndrome model. *eLife* **11**, e78203 (2022). doi: 10.7554/eLife.78203; pmid: 35762573
58. 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). doi: 10.1038/nrm3657; pmid: 23989960
59. P. Clemente *et al*., SUV3 helicase is required for correct processing of mitochondrial transcripts. *Nucleic Acids Res.* **43**, 7398–7413 (2015). doi: 10.1093/nar/gkv692; pmid: 26152302
60. M. Thoms *et al*., The Exosome Is Recruited to RNA Substrates through Specific Adaptor Proteins. *Cell* **162**, 1029–1038 (2015). doi: 10.1016/j.cell.2015.07.060; pmid: 26317469
61. X. Huo *et al*., 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). doi: 10.1016/j.molcel.2019.10.001; pmid: 31677973
62. 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). doi: 10.1038/nrg3934; pmid: 26149713
63. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). doi: 10.1186/1471-2105-9-559; pmid: 19114008
64. T. Popovitchenko *et al*., Translational derepression of *Elavl4* isoforms at their alternative 5′ UTRs determines neuronal development. *Nat. Commun.* **11**, 1674 (2020). doi: 10.1038/s41467-020-15412-8; pmid: 32245946
65. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015). doi: 10.1038/nrm3920; pmid: 25531225
66. C. A. Musselman, T. G. Kutateladze, Characterization of functional disordered regions within chromatin-associated proteins. *iScience* **24**, 102070 (2021). doi: 10.1016/j.isci.2021.102070; pmid: 33604523
67. S. Calabretta, S. Richard, Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends Biochem. Sci.* **40**, 662–672 (2015). doi: 10.1016/j.tibs.2015.08.012; pmid: 26481498
68. R. Ayala, T. Shu, L.-H. Tsai, Trekking across the brain: The journey of neuronal migration. *Cell* **128**, 29–43 (2007). doi: 10.1016/j.cell.2006.12.021; pmid: 17218253
69. M. Valiente, O. Marín, Neuronal migration mechanisms in development and disease. *Curr. Opin. Neurobiol.* **20**, 68–78 (2010). doi: 10.1016/j.conb.2009.12.003; pmid: 20053546
70. G. Ince-Dunn *et al*., Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability. *Neuron* **75**, 1067–1080 (2012). doi: 10.1016/j.neuron.2012.07.009; pmid: 22998874
71. T. Stuart *et al*., Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019). doi: 10.1016/j.cell.2019.05.031; pmid: 31178118
72. S. A. Hardwick *et al*., Single-nuclei isoform RNA sequencing unlocks barcoded exon connectivity in frozen brain tissue. *Nat. Biotechnol.* **40**, 1082–1092 (2022). doi: 10.1038/s41587-022-01231-3; pmid: 35256815
73. 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**, e2114326118 (2021). doi: 10.1073/pnas.2114326118; pmid: 34795060
74. R. S. Schmid *et al*., 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). doi: 10.1073/pnas.0630496100; pmid: 12649319
75. 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). doi: 10.1016/j.ydbio.2005.04.038; pmid: 15949792
76. A. R. Ypsilanti *et al*., Transcriptional network orchestrating regional patterning of cortical progenitors. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024795118 (2021). doi: 10.1073/pnas.2024795118; pmid: 34921112
77. J. A. Miller *et al*., Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014). doi: 10.1038/nature13185; pmid: 24695229
78. M. T. K. Kirkcaldie, S. T. Dwyer, The third wave: Intermediate filaments in the maturing nervous system. *Mol. Cell. Neurosci.* **84**, 68–76 (2017). doi: 10.1016/j.mcn.2017.05.010; pmid: 28554564
79. C. Dean *et al*., Synaptotagmin-IV modulates synaptic function and long-term potentiation by regulating BDNF release. *Nat. Neurosci.* **12**, 767–776 (2009). doi: 10.1038/nn.2315; pmid: 19448629
80. K. Murk, M. Ornaghi, J. Schiweck, Profilin Isoforms in Health and Disease – All the Same but Different. *Front. Cell Dev. Biol.* **9**, 681122 (2021). doi: 10.3389/fcell.2021.681122; pmid: 34458253
81. J. Yuasa-Kawada, M. Kinoshita-Kawada, Y. Tsuboi, J. Y. Wu, Neuronal guidance genes in health and diseases. *Protein Cell* **14**, 238–261 (2023). doi: 10.1093/procel/pwac030; pmid: 36942388
82. V. Chiurchiù, M. Maccarrone, A. Orlacchio, The role of reticulons in neurodegenerative diseases. *Neuromolecular Med.* **16**, 3–15 (2014). doi: 10.1007/s12017-013-8271-9; pmid: 24218324
83. R. Iwata *et al*., Mitochondria metabolism sets the species-specific tempo of neuronal development. *Science* **379**, eabn4705 (2023). doi: 10.1126/science.abn4705; pmid: 36705539
84. J. M. Fu *et al*., Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022). doi: 10.1038/s41588-022-01104-0; pmid: 35982160
85. J. Kaplanis *et al*., Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* **586**, 757–762 (2020). doi: 10.1038/s41586-020-2832-5; pmid: 33057194
86. T. Singh *et al*., Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509–516 (2022). doi: 10.1038/s41586-022-04556-w; pmid: 35396579
87. D. S. Palmer *et al*., Exome sequencing in bipolar disorder identifies *AKAP11* as a risk gene shared with schizophrenia. *Nat. Genet.* **54**, 541–547 (2022). doi: 10.1038/s41588-022-01034-x; pmid: 35410376
88. 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). doi: 10.1016/j.ajhg.2019.05.020; pmid: 31327507
89. D. M. Werling *et al*., An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat. Genet.* **50**, 727–736 (2018). doi: 10.1038/s41588-018-0107-y; pmid: 29700473
90. J.-Y. An *et al*., Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* **362**, eaat6576 (2018). doi: 10.1126/science.aat6576; pmid: 30545852
91. B. B. Cummings *et al*., Transcript expression-aware annotation improves rare variant interpretation. *Nature* **581**, 452–458 (2020). doi: 10.1038/s41586-020-2329-2; pmid: 32461655
92. F. K. Satterstrom *et al*., Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568–584.e23 (2020). doi: 10.1016/j.cell.2019.12.036; pmid: 31981491

93. S. H. Lelieveld *et al*., Meta-analysis of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* **19**, 1194–1196 (2016). doi: 10.1038/nn.4352; pmid: 27479843
94. R. M. Easton *et al*., Role for Akt3/protein kinase Bγ in attainment of normal brain size. *Mol. Cell. Biol.* **25**, 1869–1878 (2005). doi: 10.1128/MCB.25.5.1869-1878.2005; pmid: 15713641
95. L. Wang *et al*., Brain Development and Akt Signaling: The Crossroads of Signaling Pathway and Neurodevelopmental Diseases. *J. Mol. Neurosci.* **61**, 379–384 (2017). doi: 10.1007/s12031-016-0872-y; pmid: 28025777
96. 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). doi: 10.1038/s41467-021-25388-8; pmid: 34429411
97. P. J. Shepard *et al*., Complex and dynamic landscape of RNA polyadenylation revealed by PAS-Seq. *RNA* **17**, 761–772 (2011). doi: 10.1261/rna.2581711; pmid: 21343387
98. J. J. An *et al*., Distinct role of long 3′ UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons. *Cell* **134**, 175–187 (2008). doi: 10.1016/j.cell.2008.05.045; pmid: 18614020
99. B. Lackford *et al*., Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J.* **33**, 878–889 (2014). doi: 10.1002/embj.201386537; pmid: 24596251
100. J. H. Lui, D. V. Hansen, A. R. Kriegstein, Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011). doi: 10.1016/j.cell.2011.06.030; pmid: 21729779
101. 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). doi: 10.1146/annurev-cellbio-101011-155801; pmid: 25000993
102. PsychENCODE Knowledge Portal, UCLA-DevCtx, Synapse (Project SynID syn51090399, 2023); .doi: 10.7303/syn51090399
103. C. Jops, A. Patowary, P. Zhang, C. K. Vuong, X. Ge, K. Hou, M. Kim, N. Gong, M. Margolis, D. Vo, X. Wang, C. Liu, B. Pasaniuc, J. J. Li, M. J. Gandal, L. de la Torre-Ubieta, gandallab/Dev_Brain_IsoSeq: v1.0.0, version 1.0.0, Zenodo (2023); .doi: 10.5281/zenodo.8346777

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 D. 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 for Autism Research (SFARI Bridge to Independence Award and SFARI 957585 to M.J.G.), the National Institute of Mental Health (R01MH125252, R01MH121521, and R01MH123922 to M.J.G.; R01MH124018 to L.d.l.T.-U.; and T32MH073526 to M.K.), the National Science Foundation (NSF CBET-2225624 to M.J.G.), and the UCLA Medical Scientist Training Program (T32GM008042 to M.K.). 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 A. Abyzov, N. Ahituv, S. Akbarian, K. Brennand, A. Chess, G. Cooper, G. Crawford, S. Dracheva, P. Farnham, M.J.G., M. Gerstein, D. Geschwind, F. Goes, J. F. Hallmayer, V. Haroutunian, T. M. Hyde, A. Jaffe, P. Jin, M. Kellis, J. Kleinman, J. A. Knowles, A. Kriegstein, C.L., C. E. Mason, K. Martinowich, E. Mukamel, R. Myers, C. Nemeroff, M. Peters, D. Pinto, K. Pollard, K. Ressler, P. Roussos, S. Sanders, N. Sestan, P. Sklar, M. P. Snyder, M. State, J. Stein, P. Sullivan, A. E. Urban, F. Vaccarino, S. Warren, D. Weinberger, S. Weissman, Z. Weng, K. White, A. J. Willsey, H. Won, and P. Zandi. **Author contributions:** A.P., M.J.G., and L.d.l.T.-U. conceived and designed the study. C.K.V. and L.d.l.T.-U. collected and processed the tissue specimens and dissected the samples for processing. A.P. and L.d.l.T.-U. generated the library for sclso-Seq. A.P. processed the single-cell raw data and performed the validation experiments. C.J. and A.P. analyzed bulk tissue Iso-Seq 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 APA analysis; C.K.V. performed the RBP-associated analysis; C.K.V., L.d.l.T.-U., and M.J.G. conducted network analyses; A.P., C.J., and M.M. performed sclso-Seq analysis; and M.K., D.V., and A.P. performed single-cell isoform-switch analysis. X.G. and J.J.L. performed single-cell DTU analysis; A.P. and L.d.l.T.-U. analyzed the DTU results; and M.J.G., L.d.l.T.-U., P.Z., K.H., C.J., A.P., and B.P. performed the disease enrichment analysis. A.P., C.K.V., M.J.G., and L.d.l.T.-U. interpreted the data. A.P., C.K.V., P.Z., M.J.G., and L.d.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 sclso-Seq data are available at Synapse (*102*) 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 midgestation 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 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 (*103*).

SUPPLEMENTARY MATERIALS

Submitted 15 March 2023; accepted 13 March 2024
10.1126/science.adh7688