Cell-type-specificity of isoform diversity in the developing human neocortex informs mechanisms of neurodevelopmental disorders

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

Affiliations

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

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

Abstract:

Human brain development is under tight molecular genetic control and the recent advent of single-cell genomics has revolutionized our ability to elucidate the diverse underlying cell-types and states. Although RNA splicing is highly prevalent in the brain and has strong links to neuropsychiatric disorders, previous work has not systematically investigated the role of cell-type-specific splicing or transcript-isoform diversity during human brain development. Here, we leverage single molecule long-read sequencing to deeply profile the full-length transcriptome of the germinal zone (GZ) and cortical plate (CP) regions of the developing human neocortex at tissue and single-cell resolution. We identify 214,516 unique isoforms, corresponding to 22,391 genes. Remarkably, we find that 72.6% of these are novel and together with >7,000 novel-spliced exons expands the proteome by 92,422 proteoforms. We uncover myriad novel isoform switches during cortical neurogenesis, implicating previously-uncharacterized RNA-binding protein-mediated and other regulatory mechanisms in cellular identity and disease. Early-stage excitatory neurons exhibit the greatest isoform diversity and isoform-based single-cell clustering identifies previously uncharacterized cell states. Leveraging this resource, we re-prioritize thousands of rare *de novo* risk variants associated with neurodevelopmental disorders (NDDs) and reveal that risk genes are strongly associated with the number of unique isoforms observed per gene. Altogether, this work uncovers a substantial contribution of transcript-isoform diversity in cellular identity in the developing neocortex, elucidates novel genetic risk mechanisms for neurodevelopmental and neuropsychiatric disorders, and provides a comprehensive isoform-centric gene annotation for the developing human brain.

One-Sentence Summary:

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

Main Text

The development of the human brain is a tightly coordinated process under precise molecular and genetic control. Transcriptomics has provided substantial insights into the cellular and molecular processes regulating human brain development (*1*, *2*), including systematic characterization of the underlying heterogeneous cell-types, states, and lineages through single-cell and single-nucleus RNA-sequencing (scRNA-Seq) (*3–5*) as well as their developmental trajectories (*6*) and cell-to-cell variability (*7*). However, due to technological limitations of short-read scRNA-Seq, previous genomic characterization has been largely limited to changes at the gene-level, unable to capture the full complexity of alternative splicing and resulting isoform diversity present in the human brain (*8*).

Alternative splicing is a fundamental form of tissue-specific gene regulation which is present in >90% of multi-exon genes in humans (*9*). RNA transcript diversity is a result of the combinatorial effects of alternative transcription start sites (TSS), regulated exon inclusion/exclusion or intron retention, and distinct transcript termination sites due to alternative polyadenylation (APA) (*10*). In individual cases, synaptic genes such as *Nrnx1* have been shown to have thousands of unique isoforms (*11*). Human brain-expressed genes, which are longer and contain the most exons, undergo the greatest degree of splicing compared with other tissues and species -- a mechanism contributing to the vast proteomic, phenotypic, and evolutionary complexity of the human brain (*12*, *13*). Indeed, alternative splicing plays an important role in synaptogenesis and synapse specification (*14*) and brain development more broadly (*15*). Substantial cell-type specificity of splicing and resulting isoform diversity has been observed in the mouse brain, even among closely related neurons, and often with precise temporal regulation (*16–20*).

Emerging evidence further implicates alternative splicing as a critical mechanism linking genetic variation and neuropsychiatric disease (*2*, *21–24*). For example, recent work has identified significant splicing and isoform expression dysregulation in the brains of individuals with ASD and schizophrenia (SCZ) -- a signal that was substantially larger and more widespread than gene expression changes and showed greater enrichment for genetic risk (*25–27*). Common variants associated with ASD and SCZ identified in genome-wide association studies (GWAS) show substantial enrichment among splicing-QTLs observed in the developing and adult human brain (*2*, *24*). Likewise, rare *de novo* variants identified in individuals with ASD and intellectual disability are significantly enriched for predicted splice-altering consequences (*23*). These genetic signals are also known to exhibit convergent effects during mid-fetal brain development, highlighting the importance of this critical developmental period for understanding genetic risk mechanisms (*28*, *29*). Finally, a comprehensive understanding of the complex splicing patterns and resulting isoform diversity can have direct translational relevance, as demonstrated recently for the rare neurogenetic disorder spinal muscular atrophy (SMA) (*30*).

The recent development of third-generation long-read sequencing technologies now enables highly accurate, in-depth characterization of the full-length alternatively spliced transcriptome at scale (*10*, *31*). Coupled with single-cell barcoding technologies and unique molecular indexing, this can be used to catalog the isoform-centric transcriptome with single cell resolution (*20*, *32*). Here, we leverage this approach to deeply profile the major cell-types of the developing human neocortex at mid-gestation. We uncover ~150k previously unannotated transcript-isoforms and thousands of

novel spliced exons, many within high-confidence neurodevelopmental disorder risk genes. To functionally annotate these novel 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, these data provide an unprecedented view of the complexity of transcript-isoform diversity during human brain development with critical implications for understanding mechanisms of cell-fate specification and genetic risk for neurodevelopmental disorders.

Full-length transcriptome of the developing human brain

We performed high-depth, high-fidelity long-read sequencing (PacBio HiFi IsoSeq) to comprehensively profile the full-length polyadenylated transcriptome in developing human neocortex across six unique donors at mid-gestation, spanning post-conception weeks (PCW) 15-17 (**Fig. 1A**). To capture how transcript-isoform expression and usage change during neurogenesis, samples were microdissected into the neural progenitor-enriched germinal zone (GZ) and the neuron-enriched cortical plate (CP) and subjected to high-depth long-read sequencing in replicate **(Fig. 1B-C**; **fig. S1; Methods**). Following sequencing and comprehensive quality control, we generated over 33.6 million high-quality (>Q20) circular consensus sequence (CCS) reads from bulk tissue samples (**fig. S1-2; Methods).** Using minimap2, over 99% of the full-length reads were confidently aligned to the human reference genome, a marked improvement over the 85% mapping rate of short read RNA-Seq, enhancing discovery of novel splice isoforms (**fig. S2A-B)**. Using the TALON pipeline (*33*), we identify a total of 214,516 unique transcripts in the bulk tissue transcriptome, corresponding to 24,554 genes; of these, >175k isoforms from 17,299 genes were expressed at >0.1 TPM in at least half of the samples (**Fig. 1D-E)**. Isoform-level expression showed strong reproducibility across technical and biological replicates (**Fig. 1B** and **fig. S2C**), with clear biological separation along the first principal component between CP/GZ (**Fig. 1C**). The median read length from bulk tissue samples was 2.99 kb, with a range of 80bp to 14.2kb, consistent with the high quality of RNA samples and expected distribution of mammalian mRNA transcript lengths, and that observed in human cell lines (**Fig. 1F,H).**

We next sought to generate a high quality custom reference transcriptome (**table S1**) for the developing human brain, merging data across the bulk tissue samples. Remarkably, compared with the Gencode v33 reference, only 65,006 (30.3%) of observed isoforms corresponded with existing annotations (**Fig. 1D-E**). Novel isoforms were further classified based on their splice junction match to the reference transcriptome, strand specificity, and overlap with known transcripts at 5' and/or 3'ends. A total of 65,184 novel isoforms exhibited completely new internal splicing patterns (novel-in-catalog, NIC, and novel-not-in-catalog, NNC), corresponding to 10,175 unique genes (**table S1**). As expected, novel transcripts were expressed at lower levels than known transcripts, with 72% of total reads mapping to known isoforms **(Fig. 1G**). Nevertheless, we observed several novel transcripts expressed at high levels, including *WASF1*, *CYFIP2*, *MAP1B*, *NEFL*, and *SMARCA4* (**fig. S3A-E**).

The majority of novel isoforms were classified as incomplete splice match (ISM), with reads overlapping either the 5' (prefix) or the 3' (suffix) ends of known transcripts. Although ISM transcripts are often disregarded as artifacts of RNA degradation or internal priming, analysis of gene body coverage did not show 3' bias consistent with RNA degradation, and was highly

concordant with short-read total RNA sequencing from ribozero libraries (**Fig. 1F; fig. S2A**). Indeed, novel transcripts (including ISMs) were actually found to be significantly longer than known transcripts (Known: 2276+/-2224 bp, Novel: 3072+/-1169 bp; mean +/- sd; P<2e-16, Kruskal-Wallis; **Fig. 1H**). Further, novel transcripts contained significantly more exons than known transcripts (11.0 +/- 7.25 vs 7.16 +/- 6.88; mean +/- sd; P<2e-16, Kruskal-Wallis; **Fig. 1I**). As such, we retained ISM transcripts with additional supporting evidence (**Methods**), and further note that many exhibited functional roles as "hub" isoforms in network analyses detailed below.

Finally, in light of this large number of isoforms detected in the developing human brain, we sought to characterize their patterns of usage and potential functional relevance to neurodevelopmental disorders (NDD). To assess patterns of isoform usage, for multi-isoform genes, we characterized the proportion of gene expression attributable to the top (dominant) isoform, stratified by total gene abundance (**Fig. 1J**). There was a relatively even distribution of isoforms among more lowly expressed genes, in which isoforms tended to be expressed at similar proportions. As genes became more highly expressed, however, the dominant isoform tended to capture the majority of expressed reads. Nevertheless, the number of detected isoforms for a given gene was strongly predictive of NDD risk genes, even when gene length and total gene expression are accounted for (OR 3.03, P=3.6e-08, logistic regression; **Fig. 1K**). This association remained significant even when restricting only to novel transcripts (**fig. S3F**).

Expanded transcriptomic and proteomic complexity in the developing human brain

To contextualize the substantially expanded transcriptomic complexity observed in the developing human brain, we next sought to integrate several orthogonal genomic annotations (*34–40*) and predict potential downstream protein coding consequences (**Fig. 2; Methods)**. In aggregate, we find that ~80% of novel isoform TSS's are supported by the existence of proximal, annotated CAGE or ATAC-Seq peaks. At the 3' end, ~91% of novel transcripts are supported by nearby polyA sites or motifs -- a rate higher than for known transcripts (**Fig. 2A-B**). Of the 38,115 novel splice junctions identified here not present in Gencode, 74% were validated by Intropolis (**Fig. 2C**). Together, these results demonstrate broad orthogonal support for novel isoforms identified here.

In total, we identify 27 Mb of the human genome that is transcriptionally active in the developing human brain but not currently annotated to any Gencode gene model. Of this, 3.85Mb is comprised of >7000 novel spliced exons, spanning >3500 unique genes, with a median width 207bp and exhibiting no overlap with Gencode (**Fig. 2D**). We validated several of these exons using RT-PCR (**Fig. 2E**; **table S2)**. Additionally, we identify 319 novel multi-exonic genes, not consistent with any existing gene model (**table S1**). Most of those are antisense to existing genes (N=256), although 63 are found in intergenic regions. For example, we identify a novel 18-exon gene on chr 8 with 4 unique splice isoforms antisense to the high-confidence ASD risk gene *DPYSL2* (**fig. S3G**).

We next sought to determine the functional significance of this expanded compendium of novel exons and transcripts at the protein-coding level (**Fig. 2F**; **Methods)**. Of the novel transcripts observed, 92,422 exhibited protein coding potential, with at least one complete putative open reading frames. Integration of proteomics data from the adult human brain (*41*) provided peptide-

level support for 35,467 novel transcripts with unique proteoforms (**table S2**). For example, a novel, alternatively spliced 30bp microexon found in the neurodevelopmental disorder risk gene *NF1* was predicted to add an additional 10 amino acids (AA) to the known protein (**Fig. 2A**). Searching against mass-spec based proteomics data identified spectra (n=15) that confidently support the existence of this novel 10 AA sequence (**Fig. 2G**). Notably, isoforms containing this microexon were only detected in neuronally-enriched CP samples (**fig. S4A**). Extending these analyses beyond a single gene locus, we observed broad peptide-level support for novel transcripts across all classification categories (**Fig. 2H**).

Isoform Switching during Neurogenesis

We next sought to contrast gene and isoform usage across GZ and CP samples to identify genes exhibiting significant isoform-switching events associated with neurogenesis or neuronal maturation. Whereas differential gene expression signatures of neuronal maturation have been extensively characterized in the developing human brain within bulk tissue and at single-cell resolution (*3*, *4*, *34*), transcriptome-wide alternative splicing and isoform usage have not been systematically investigated during this critical developmental period.

Consistent with previous work, we observed a large number of genes exhibiting significant differential gene expression (DGE) patterns between GZ and CP (N=4,475 of 24,554, at FDR < 0.05; **table S3**). Likewise, of genes with multiple expressed isoforms, a large proportion exhibited differential transcript usage (DTU) across GZ and CP (2,679 of 10,809 genes at FDR <0.05; **Fig 3A**; **Methods**; **table S3**), with the majority (57%) of significantly switching isoforms (n=5,630 isoforms at FDR<0.05) coming from novel transcripts (**Fig 3B**). Although there was a significant overlap among genes exhibiting both DGE and DTU (N=1010, OR=2, $P<10^{-46}$, Fisher's exact test; **Fig. 3A**), there were also 1,411 genes displaying isoform switching without changing in overall gene expression. The majority of isoform switching events between GZ and CP had observed and/or predicted functional consequences (**Fig. 3C)**, summarized in **table S3**. For example, CP-upregulated isoforms were significantly more likely to have gained than lost an exon, although both outcomes were observed (N=2,031 vs 1,260; FDR-corrected $P<10^{-40}$). Of 5,630 isoform switching events, 3,204 (57%) were predicted to alter the ORF length, which was slightly more likely to result in longer ORFs among CP-upregulated isoforms (FDR-corrected P<0.0001). However, CP-upregulated isoforms were also more likely to exhibit predicted NMD sensitivity (243 vs 100; FDR-corrected $P<10^{-15}$).

To further interrogate how the 3' UTR of transcripts differ between regions, we performed a complementary, read-depth based analysis of distal polyadenylation usage (**Methods**; **Fig. 3D**). For the 9,896 transcripts with multiple annotated polyA sites, we used DaPars2 to compute a distal polyA usage index (DPUI) -- the fraction of total reads mapping to the longer (e.g., distal) 3'UTR (*42*). In total, 1,013 transcripts exhibited significant differences in DPUI between GZ and CP (repeated measures ANOVA; FDR-corrected P < 0.05; **Fig. 3D; table S3**). DPUIs were increased in CP versus GZ for the majority (772) of transcripts, and DPUI was on average increased in the CP (two-sample paired T-test, p < 2.2e-16), indicating overall increased 3'UTR lengths in transcripts in the CP compared to the GZ. Pathway enrichment among genes with regional DPUI differences was notable for RNA/mRNA binding as well as cytoplasmic stress granule ontologies. Finally, we found genes encoding RNA-binding proteins (RBPs) (*43*) were overrepresented for

transcripts exhibiting significant DPUI changes (453/1,013 genes vs 3,074/8,883 genes, respectively; one-tailed Fisher's Exact Test, p = 2.06e-10).

We next conducted pathway analyses for all genes exhibiting significant isoform switches across regions. Enriched biological pathways included dendrite morphogenesis, cadherin binding, and the nBAF and SWI/SNF complexes, which are known to harbor convergent genetic risk for neuropsychiatric disorders (**Fig. 3E**). For example, among the top genes was *SMARCC2*, a high-confidence ASD risk gene encoding a subunit of the SWI/SNF chromatin remodeling complex (**Fig. 3F**). Although overall expression of *SMARCC2* did not differ between CP and GZ, two novel isoforms of *SMARCC2* showed preferential usage in GZ and exhibited exon skipping compared with known transcripts. Other notable isoform-switching genes included several known splicing regulators and RNA-binding proteins (e.g., *SRRM1*, *SRRM4*, *CELF1*, *PTBP2*, *ELAVL2*, *ELAVL4*, *RBFOX2*), chromatin modifiers (e.g., *KMT2E*, *KMT5B*, *SMARCA1*, *SMARCD3*, *SMARCE1*), patterning transcription factors (*FOXP2*), regulators of synaptic transmission (*GRIA3*, *VAMP1*, *GAD1*), and synaptic adhesion molecules (*NLGN4X*, *NRXN1*). Isoform-switching genes were broadly expressed across cell-types, with particular enrichment for excitatory neuron lineages (**Fig. 3G**).

Putative RNA-binding protein regulators of isoform switching events

RBPs are a diverse class of proteins regulating the processing and fate of their target mRNAs. Through a variety of mechanisms including alternative splicing (AS), RBPs increase isoform and consequently protein diversity, and play important roles in mammalian neural development and function (*17*, *44*). To identify potential RBP regulators of isoform switching events in the developing human brain, we assessed DGE, DTE and DTU genes for enrichment for RBP targets identified through AS changes or direct binding by cross-linking immunoprecipitation (CLIP) using two datasets – 1) targets of RBPs known to regulate AS during neural development and/or maturation curated from previous work ("brain-enriched"; black bar, **Fig. 3H and fig. S4B-C;**), and 2) targets defined by systematic profiling of RBPs in the ENCODE repository (grey bar; **Fig. 3H and fig. S4B-C;** see also **Methods**).

Examining isoform diversity, we find that DTU genes significantly overlap with targets of known regulators of alternative splicing in the brain (*44*) (**Fig. 3H**). We performed enrichment analysis of genes containing alternative exons and/or CLIP-binding sites targeted by well-known brain-enriched RBPs and assessed their overlap with DTU, DTE (differential transcript expression) and DGE genes. Brain-enriched RBP targets were much more enriched in DTU genes compared to DTE or DGE genes – this result supports previous work which has characterized many of these RBPs as regulators of alternative splicing in the brain. Complementary to this, gene expression targets, rather than alternative splicing targets, were more enriched in DGE and DTE compared to DTU genes (**fig. S4B**). We also observed that gene expression targets of RBPs that are more highly expressed in neurons as compared to progenitors (*SRRM3/4*; (*44*)) are enriched in genes that increase in expression over neural development, while those that are targets of progenitor-associated RBPs (*SAM68*, *PTBP1*; (*44*)) show the opposite trend (**fig. S4C**).

To identify RBPs important for but not previously studied during the developmental transition from progenitor to neuron, we performed enrichment analysis of DTU, DTE and DGE genes compared to the most currently comprehensive set of RBP targets available from the ENCODE repository (**Fig 3H**; (*45*)). Similar to our results comparing target genes of brain-enriched RBPs, we find that targets of RBPs from the ENCODE dataset are more enriched in DTU genes compared to DTE and DGE. Interestingly, within DTU genes, ENCODE RBPs targeting introns, 5' splice sites and 3' splice sites are more enriched as compared to those targeting the CDS or 3'UTRs (**Fig. 3H**; teal, magenta and aqua boxes vs. green and dark blue boxes). This result matches the overall trend for RBP target enrichment in DTU over DGE, the latter of which is often mechanistically regulated through 3' UTR binding.

Of the ENCODE RBPs showing significant target overlap with DTU genes, we noted several that have been increasingly recognized to play critical roles in neural development and disease. Among these are RBPs known to regulate RNA metabolism and splicing in the brain but less studied in the context of neural development, such as *LIN28B*, *EFTUD2*, *KHSRP* and *DGCR8* (*46–48*). Notably, we observed a strong enrichment between DTU genes and targets of *DDX3X*, an X-linked RNA helicase where *de novo* mutations lead to sexually dimorphic intellectual disability and ASD (*49, 50*). Most interestingly, we also find strong enrichment of DTU genes with targets of RBPs for which novel roles in RNA metabolism have been recently identified, but which have not yet been studied in the context of neural development. For example, targets bound by components of the exosome (the RNA degradation system) and those involved in rRNA biogenesis, such as *EXOSC5*, *UTP18* and *SUPV3L1* (*51–53*), and the nuclear matrix protein *SAFB*, implicated in heterochromatin regulation (*54*), are enriched in DTU genes. Altogether, these results indicate that while many of the isoform switching events observed between GZ and CP are likely regulated by brain-enriched RBPs through alternative splicing, many more switching events are expected to be produced through diverse mechanisms regulated by novel RBPs.

Network Context of Developmental Isoform Regulation

Given the large number of isoform switch events observed, we next leveraged the weighted gene correlation network analysis (WGCNA) framework to place these results within a systems-level context during human brain development (*55, 56*). We separately built unsupervised co-expression networks for gene and transcript expression levels ($\log_2$TPM normalized, herein referred to as geneExpr and isoExpr networks). We also built networks using transcript usage quantifications (isoUsage networks), the proportion of each gene's total abundance attributable to a given isoform. For each network, genes or isoforms were then assigned to modules based on shared patterns of covariation across samples (**Fig. 4A**; **Methods; table S4**). As expected, network modules were highly enriched for cell-type-specific marker genes, enabling an *in silico* deconvolution of cell-type isoform usage and expression (**fig. S5**).

We next compared properties across the three networks, contrasting cell-type enrichments within modules. We observed that while both geneExpr and isoExpr networks are strongly driven by cell-type identity, geneExpr modules are more broadly defined and shared across multiple cell types. For example, of the six progenitor modules all are shared across multiple cell types, and of the eight excitatory modules only one (geneExpr.M7) distinguishes a particular cell type (ExDp1) but is also shared with inhibitory neurons (**fig. S6A, geneExpr**). In contrast, while the isoExpr network

also contains modules shared across broad cell types (e.g. isoExpr.M2 – progenitors, isoExpr.M1 – excitatory neurons), we observe a number of modules within progenitors, excitatory neurons, and inhibitory neurons that uniquely distinguish specific cell types from their broader cell categories (**fig. S6A, isoExpr**). Within the progenitor category, PgS, PgG2M and oRG are defined by unique modules (isoExpr.M19, M29, and M23 and isoExpr.M26, respectively). Similarly, within excitatory neurons, ExM-U can be distinguished from ExM (isoExpr.M9 and M39 vs isoExpr.M48) and newborn neurons, ExN, can be distinguished from all other excitatory cells and progenitors (isoExpr.M16). Finally, though inhibitory neurons also share multiple modules, InMGE can be distinguished by isoExpr.M13. Overall, geneExpr and isoExpr networks show more cell type enrichments compared to isoUsage, with even more enrichment in isoExpr compared to geneExpr (**Fig. 4B, Cell Type enrichments**).

Interestingly, while geneExpr and isoExpr networks are highly driven by cell-type identity, the isoUsage network is more directly defined by RBP isoform usage patterns. Plotting the expression of known RBPs (*57*) against the isoUsage network, we observe that RBP target enrichments, rather than cell type markers, better define these modules (**Fig. 4A** and **fig. S6B**). To assess how isoUsage modules are driven by RBP isoform usage and understand their functional consequences, we performed enrichment analyses of the overlap between brain-enriched RBP and ENCODE RBP target genes (as described earlier) with those in each module. We also performed this analysis for the geneExpr and isoExpr networks, finding that isoExpr and isoUsage modules show more global enrichment for RBP targets compared to those from geneExpr, with isoUsage showing even more enrichment compared to isoExpr (**Fig. 4B, RBP enrichments** and **fig. S6B**). A detailed examination of isoUsage modules revealed both expected and novel module-RBP enrichment patterns. For example, modules in the isoUsage network represented by eigengenes with GZ-to-CP or CP-to-GZ expression patterns are enriched for targets of brain-enriched RBPs such as *SRRM4*, *PTBP2* and *RBFOX1/2/3*, as well as progenitor-enriched *PTBP1* (**Fig. 4C-D**). Many modules were also significantly enriched for targets of RBPs less studied in brain development, from the ENCODE dataset. Of note is isoExpr.M11, a progenitor and early neuron module whose hub transcripts include an isoform of *DDX3X* and is involved in RNA processing and cytoskeletal dynamics (**fig. S7A**). To contextualize these modules in relation to human brain development, we assess their enrichment for genes containing rare variants causing neuropsychiatric disorders and focus on modules significantly enriched for those associated with ASD. Below, we highlight several modules containing known and novel hub transcripts for which isoform-level information furthers our understanding of the cellular processes involved in neural development.

In the two largest isoUsage modules, isoUsage.M1 and M2, we observed three different hub isoforms of the BAF (SWI/SNF) chromatin remodeling complex gene, *SMARCE1* (**Fig. 4C**). These isoforms are all found in Gencode and differ most in the first four exons which encode an intrinsically disordered region (IDR) upstream of the DNA-binding domain. IDRs are protein regions of low complexity often containing polar/charged repeats resulting in unstable secondary structure (*58*) and are involved in cellular processes involving higher-order complexes such as chromatin remodeling and RNA splicing (*59, 60*) and function in both protein-protein and protein-nucleic acid interactions. The *SMARCE1* isoform found in isoUsage.M1 encodes the fullest version of the protein. In contrast, the two isoforms in isoUsage.M2 lack either all or a portion of the first IDR – in ENST00000647508, exclusion of exon 4 truncates the IDR, while in ENST00000643318 a downstream translational start in combination with exon 4 exclusion entirely

removes the IDR. In addition, we identified a novel (NIC) hub isoform of the RBP *ELAVL2* in isoUsage.M1 containing an alternative 5' start site. *ELAVL2* isoforms form roughly three groups with different 5' start sites which do not alter the reading frame or the predicted protein domains, suggesting that these alternative starts may serve a regulatory function. Indeed, recent work demonstrated translational regulation of another ELAVL family member, *ELAVL4*, at its alternative 5' UTRs (*61*). Notably, the eigengenes of these two modules show complementary patterns - GZ-enriched in isoUsage.M1 and CP-enriched in isoUsage.M2. Gene ontology (GO) analysis for these two modules revealed enrichment for mitotic cell cycle and other processes matching progenitor cell function in isoUsage.M1, and for neuronal morphogenesis in isoUsage.M2 (**Fig. 4C, Gene Ontology**).

In two other modules, isoUsage. M3 and M8, different isoforms of another member of the ELAVL family of RBPs, *ELAVL1*, were identified as hub transcripts. *ELAVL1* is broadly expressed and binds the 3'UTRs of target transcripts to increase their stability (*62*). Both *ELAVL1* hub transcripts contain a novel 3'UTR, with a short ~200bp intron in the center of the UTR spliced out, which interestingly is conserved only between primates and human. In addition, the *ELAVL1* hub isoform found in isoUsage.M3 encodes a version of the last exon containing only one of the three RNA-recognition motifs found in the full-length transcript. These modules contain other RBP isoforms as hub transcripts including those of *RBFOX2* and *HNRNPA3* (isoUsage.M3) and *CELF2* (isoUsage.M8). Indeed, isoUsage.M3 is noticeably driven by RBP isoform usage (**Fig. 4A, brown module**) and is enriched for GO terms involving mRNA metabolism and RNA splicing (**Fig. 4D, Gene Ontology**). In contrast, isoUsage.M8 is more enriched for biological processes involving cytoskeletal function and cell projection organization, processes important in neuronal migration and maturation (*63*, *64*). Altogether, these networks appear to capture and refine our understanding of the specific "RNA regulons" active in the developing brain (*43*).

In the four isoUsage modules we highlight here, isoform-level information is critical in refining interpretations of the cellular mechanisms contributing to disease during brain development. For example, while the BAF complex is already known to be involved in neural development and has been implicated in ASD (*65*, *66*), our analysis revealed that specific exon combinations and consequently protein domains are altered between modules driving different cellular processes. In another example, isoform-level analysis identified a novel 3'UTR sequence in *ELAVL1* which is conserved specifically in primates and human. While many developmental processes are conserved across vertebrates in early neural development, primate- and human-specific differences are important to understand given the unique cell types (e.g. oRGs in human) found in the expanded cortices of these species (*67*, *68*)

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, gene isoform expression has yet to be characterized. To gain single cell resolution, we leveraged the recently developed ScISOrSeq (*20*) approach to profile >7,000 single cells across an additional 3 unique donor samples derived from microdissected GZ and CP regions (**Fig. 1A**; **fig. S8; Methods**). Barcoded full-length single-cell cDNA libraries generated using SMRT-seq v2, with incorporated UMIs as published (*3*) were used as input to generate >26.4M high-quality

PacBio CCS reads. To obtain cell-type specificity, cell-barcodes were matched to the high-depth, short-read sequencing dataset previously published on the same libraries (*3*). Of 7,189 individual single cell full-length transcriptomes, 4,281 had matching barcodes from short-read sequencing (**figs. S8 and S9A; Methods**). All subsequent analyses were performed on this matched subset of the scIso-Seq data.

Following strict quality control and downstream processing (**Methods, figs. S8,** and **S9A-B**), we detected on average 530 unique transcripts per cell, mapping in aggregate to 18,541 genes and 138,497 unique isoforms (**fig. S9C**). We observed high concordance between pseudo-bulk short-read and long-read based gene expression ($R = 0.9$, $p<2.2e^{-16}$) and detection ($R = 0.92$, $p<2.2e^{-16-}$) (**fig. S9, D** and **E**), and high inter-donor reproducibility ($R = 0.84\text{-}0.87$, $p<2.2e^{-16-}$) (**fig. S9F)**, demonstrating the robustness of the data. Similar to the bulk tissue transcriptome, the majority of detected isoforms (71.7%) were novel (**fig. S9G**). We found broad support for these isoforms in bulk tissue Iso-Seq with ~80% matching both 5' and 3' end termini (**fig. S9H)**, and ~87% of splice junctions detected (**fig. S9I**) in bulk. Altogether, 67,183 isoforms detected by scISo-Seq (49% of total) fully match isoforms detected bulk tissue, including those containing 1,808 out of 5,165 novel spliced-in exons (**table S5B**), which were also validated at a high rate by RT-PCR (**fig. S9J**).

We next connected single cells with their specific cellular identities. Previous unsupervised graph-based clustering in Seurat (*69*) using high-depth short-read RNA-Seq identified 16 transcriptionally distinct cell-type clusters in the developing human neocortex (**Fig. 1A**) (*3*). Through barcode matching, we detected cells from all 16 clusters **(Fig. 5A**), allowing us to construct cell-type-specific isoform expression profiles (**table S5** and **Fig. 5B)**. Comparing isoform expression diversity across cell types, we observed that excitatory neuron clusters, in particular those corresponding to newly-born migrating (ExN) and maturing neurons (ExM), harbored the largest number of isoforms (**Fig. 5C**)**.** This was not due to differences in sequencing depth across clusters (**fig. S9K**) or gene detection (**fig. S10A**). Remarkably, these same cell types exhibited the greatest diversity of novel expressed isoforms (**Fig. 5C**), highlighting a role for these novel transcripts in early neuronal maturation processes.

Selective isoform expression across different cell types has been reported in the adult brain for a few genes (*19*, *70*, *71*) with potential implications for neuropsychiatric disorders (*25*). Thus, we next sought to systematically characterize gene isoform expression and utilization across cells in the developing neocortex. We first conducted differential transcript expression between each cell type cluster and all other clusters (**Methods, table S5C**) to identify transcripts defining the identity of each cell type and that can serve as molecular markers (**Fig. 5B**). As expected, the majority of these transcripts belong 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. Interestingly, 257/1,040 transcripts enriched in specific cell types (24.7%) correspond to novel isoforms identified in this study, indicating a major role for these transcripts as a group in defining cellular identity. Among the top enriched transcripts for each cluster, we identify novel isoforms of *NRG1*, *LMO3*, *NEFL* and *SYT4* enriched in oRG, ExDp, ExM-U, ExM cell classes, respectively, all of which have established roles in brain development and function (**Fig. 5B, table S5C**).

We also conducted a pairwise DTE analysis to identify transcripts changing across specific cell type transitions, detecting 409 transcripts corresponding to 147 genes (**table S5D,** $P_{nominal}<0.05$). Focusing on genes with multiple isoforms showing dynamic expression between progenitors and neurons, we observed isoforms of *PFN2*, which functions in actin polymerization dynamics and morphogenesis (*72*), 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 (*73*, *74*) (**Fig. 5D**). Together, these examples highlight changes in isoform expression across cell types and developmentally-relevant transitions with putative consequences to the structure or stability of their encoded protein products.

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

Novel Cell-Types Uncovered from Isoform-level Clustering

Given the broad changes in isoform diversity and expression observed across cells, we leveraged these data to expand current cell type classification catalogs. Re-clustering cells based on isoform expression yielded 15 highly-stable clusters largely mapping to many of the same cell classes as defined by gene-based clustering (**Methods**, **Fig. 5 F-H** and **fig. S10D**). However, progenitors transitioning into neurons and early-born excitatory neurons were split into additional clusters providing higher-resolution cell maturation stages than those observed by traditional gene-based clustering. In particular, newborn migrating neurons (ExN) split into three clusters (ExN1-3), encompassing cells previously annotated to IP, ExN and ExM clusters, and two additional new clusters, vRG-ExN and ExN-ExM, represent cells in states on either side of a maturity spectrum centered around ExN cells (**Fig. 5 F-G)**. This is supported by pseudotime lineage inference analysis whereby vRG-ExN cells represent a path of direct neurogenesis from vRG cells distinct from another path through oRG cells, and where ExN-ExM cells precede the most mature neuronal clusters (**Fig. 5I**). To better understand the molecular programs across these new cell states, we conducted differential isoform expression analysis (**table S5C**). Across ExN clusters, isoforms of *ENC1*, *ANKRD18CP*, and *RASD1* are enriched in ExN1, EnX2 and ExN3 cells, respectively (**fig S10E-F**). Moreover, vRG-ExN cells are defined by a majority of transcripts involved in mitochondrial metabolism, consistent with recent reports of the role of mitochondria in regulating neuronal maturation (*75*). 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 Unveiling of Neuropsychiatric Disease Mechanisms

Lastly, we sought to leverage our cell-type specific isoform-centric atlas of human cortical development to better contextualize the genetic risk mechanisms contributing to neuropsychiatric disorders and gain new insights into underlying cellular and molecular pathophysiology. We performed enrichment analyses to localize rare-variant association signals from large-scale whole exome and genome sequencing studies of neurodevelopmental and psychiatric disorders, including ASD (*76*), NDD (*76*, *77*), SCZ (*78*), bipolar disorder (BIP) (*79*), and epilepsy (*80*). Risk genes for NDD, ASD, DDD, and to some extent, epilepsy, had a significantly greater number of isoforms (log2 scale; OR's 1.23-1.65; q's<3e-5) and exons (log2 scale; ORs 1.40-2.45; q's<2e-5) compared to non-disease genes (**Fig. 6A**, **fig. S11A, table S6B;** FDR-corrected p-values from logistic regression, correcting for gene length). These associations were not observed for bipolar disorder (ORs 1.13, 1.21; NS) and were less significant in SCZ (OR 1.50, 1.79; P<0.03), albeit with similar effect sizes suggesting this attenuation was driven by statistical power. Notably, disease-associated genes showed significant overlap with those exhibiting DGE, DTE, DTU during cortical neurogenesis (**Fig. 6A** and **fig. S11B**). Overall, these associations were observed only for genes and isoforms upregulated in the CP (DGE.up, DTE.up), indicative of neuronal expression, and were shared across NDD, non-syndromic ASD, DDD and epilepsy (Helbig), but not other diseases. In particular, NDD genes were enriched for those changing isoform usage (DTU), showing enrichment in genes that are DTE or DTU, but not changing in overall gene expression (DTU.not.DGE, DTE.not.DGE). Consistent with these findings, we observed many NDD and ASD genes showed differential isoform utilization across cell types in the developing neocortex (**Fig. 6B** and **fig. S12)** with NDD genes significantly enriched in this form of regulation (**Fig. 6C,** "SingleCell-DTU"). NDD, DDD, and ASD gene isoforms were primarily enriched in excitatory neurons (ExM-U, ExDp, or ExM) (**Fig. 6C** and **fig. S11C)**. However, isoforms of NDD genes and those causing syndromic forms of ASD were also enriched in mitotic progenitors and radial glia (NDD/ASD.SFARI.S: PgS; ASD-S: PgG2M, vRG and oRG), as expected from the broader phenotypic spectrum in these disorders. Finally, we observed NDD genes were enriched in ExN1, a newly-defined cell state based on isoform-level quantifications (**Fig. 6C** and **Fig. 5**), highlighting the usefulness of this data.

Next, we leveraged our gene and isoform co-variation networks to dissect disease pathophysiology at the molecular and cellular level (**fig. S11D)**. First, we observed that NDD, DDD and non-syndromic ASD broadly shared overlapping molecular signatures as evidenced by shared enrichment across the same modules and clustering. These modules are enriched for cellular processes and markers found in both progenitors and neurons (**fig. S11D**). Remarkably, the large majority of disease-associated modules were defined by isoform expression (isoExpr, 77.3% $p_{nominal}$<0.05), followed by isoform usage (isoUsage, 62.1%), as compared to gene expression (geneExpr, 40%). This finding supports a major role for isoform expression and diversification in neuropsychiatric disease mechanisms during development consistent with recent findings (*25*), and presents an opportunity to dissect disease mechanisms at a more granular level.

In that regard, we observed that a module regulating chromatin and histone modification (isoExpr.M7; **Fig. 6A and fig. S7B**) is enriched across NDD, DDD and both syndromic and non-syndromic ASD. Interestingly, this module is also shared among all neuronal cell types tested, suggesting that chromatin regulation is a key cellular process dysregulated in NDD and ASD.

Among NDD, DDD and non-syndromic ASD, we observe enrichment of disease-associated genes in a number of modules driven by transcripts/genes involved in RNA metabolism and splicing (isoExpr.M11, M10 and M28; **table S6B and fig. S7C**). Indeed, isoExpr.M11 contains an isoform of the ASD-associated RNA helicase *DDX3X* as a hub transcript (**fig. S7A**). Overall, modules enriched for RBP targets include those known to be regulated by alternative splicing in neurons, as well as by RBPs with emerging roles in neural development.

We also observed enrichment of NDD, DDD and ASD genes in modules involved in a variety of cellular processes, sometimes in surprisingly specific cell types. For example, disease-associated transcripts are enriched in isoExpr.M29, a cell adhesion module active across cycling progenitors, excitatory (ExM) and inhibitory neurons. NDD and ASD genes are also enriched across multiple neuronal modules regulating cytoskeleton, synaptic vesicles and neurite morphogenesis (isoExpr.M30 andM24, **table S6B and fig. S7D**), as well as in a module involved in ribosomal RNA processing and chromatin specifically in excitatory neurons (isoUsage.M29, **table S6B and fig. S7E**). Overall, we find that NDD, DDD and ASD share overlapping modules involved in many cellular processes but with an emphasis on chromatin remodeling, cytoskeletal dynamics as it relates to neurite and synapse development, and RNA processing.

To move from population-level to individual genetic risk mechanisms, we next used our atlas to reinterpret *de novo*, noncoding genetic variations reported to be associated with ASD (*81–83*) and intellectual disability/developmental disorders (ID/DD) (*83*). We reasoned that some previously identified noncoding variants may carry newfound transcriptional activity based on the >27MB of new annotations. To test this hypothesis, we complemented the Gencode v33 annotation with our newly identified protein-coding isoforms, and re-annotated the collected genetic variants (N_{total}=272,187; N_{case}=145,880; $N_{control}$=126,307) using VEP. The most severe consequences for a variant changed for 1.24% of all variants (**Fig. 6D**, **table S6A**). Slightly higher proportion of variants from cases (1.27%) were re-annotated compared to those from controls (1.21%) (Table S6), although the difference did not reach statistical significance (P = 0.14, test of proportions).

Among these variants, a previously reported de novo mutation (DNM) from ID/DD cohort (*84*) was predicted to cause a missense mutation in the *KLC1* protein based on known isoforms from Gencode. The *KLC1* gene encodes a member of the kinesin light chain family, involved in microtubule cargo transport. We observed a KLC1 isoform with a novel TSS, which is supported by overlapping CAGE peaks (**Fig. 6E**). This isoform is predicted to code for a protein with a novel start codon (TALONT000423578.p1). Given the structure of this novel isoform, the DNM would lead to the loss of the start codon, which is a much more severe consequence compared to a missense mutation. More interestingly, TALONT000423578.p1 protein has an alternative carboxy termini which is not present in the Gencode database (v33). Our search against proteomics data found strong evidence for the existence of this alternative carboxy termini peptide sequence (**fig. S13A**). Taken together, we identified a highly confident novel protein-coding *KLC1* isoform which could better explain the deleterious effect of a DD/ID DNM.

We also observed a novel *AKT3* isoform with an alternative last exon extending into previously-annotated intronic region and encoding a novel AA sequence (**fig. S13B**). A reported DNM from the ASD cohort (*85*) is predicted to cause a missense mutation in the novel AA sequence, while this DNM was previously considered an intronic variant. *AKT3* is a key regulator of the PI3K-

AKT-mTOR pathway, is most active in the nervous system (*86*), and dysregulation of AKT3 is associated with neurodevelopmental disorders (*87*). Our finding suggests that *de novo* mutations may contribute to ASD by affecting specific AKT3 isoforms. Altogether, these results indicate that a more complete catalog of mid-gestation-brain-expressed full-length isoforms provides more granular mechanistic insight into how common and rare genetic variants convey risk for neurodevelopmental diseases.

Discussion

Here, we provide an unprecedented view of the full-length, alternatively spliced transcriptome in the developing human neocortex at midgestation, with regional and single-cell specificity. Although splicing and resulting isoform changes have been strongly implicated in neurodevelopmental disorder risk (*21*, *23*, *25*), as well as a critical component for proper neural development, 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 find that only ~30% of the transcriptome was previously known and identify a remarkable number of novel transcripts (149,510) significantly expanding the previously known proteomic diversity of the human brain. Examining the functional consequences of these isoforms is an important endeavor of future studies, however our analyses here provide considerable support for their functional importance.

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. Notably, enrichment analyses of genes changing isoform usage during corticogenesis implicate known neuronal splicing regulators as well as RBPs previously not studied in context of brain development. In network analyses, the addition of isoform-level information improved our ability to define cell types compared to relying solely on overall gene expression. Correspondingly, isoform-level single-cell transcriptomics enabled the identification of novel cell states in newborn excitatory neurons (ExN 1-3) as well as states encompassing the transition from progenitor to neuron and neuronal maturation (vRG-ExN and ExN-ExM). Together, these results significantly increase the catalog of isoforms expressed during corticogenesis and strongly implicate splicing and RBP regulation of isoform expression and usage in cell identity.

The new data generated in this study can now be used to inform current and future genome-wide disease risk association studies. For example, using changes in isoform expression, we identified a progenitor-early neuron module driven by an isoform of the ASD-associated gene, *DDX3X*. While recent studies have identified a role for *DDX3X* in translational regulation in progenitors (*50*), our observations suggest that this RNA helicase plays an additional role in splicing regulation during the transition from progenitor to neuron. We also show here 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 new isoform-centric atlas to re-annotate and re-prioritize thousands of *de novo* rare variants of NDD and ASD. The large number of novel transcripts we identified here suggests that the

functional consequences of many variants may have been missed using previous incomplete annotations.

In sum, our results have broad implications for understanding cell fate specification in the developing human brain and for comprehensive interpretation of the genetic risk mechanisms underlying developmental brain disorders.

References

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

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

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

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

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

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

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

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

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

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

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

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

Science. **348**, 660–665 (2015).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

27. J. R. Haney, B. Wamsley, G. T. Chen, S. Parhami, P. S. Emani, N. Chang, G. D. Hoftman, D. de Alba, G. Kale, G. Ramaswami, C. L. Hartl, T. Jin, D. Wang, J. Ou, Y. E. Wu, N. N. Parikshak, V. Swarup, T. Grant Belgard, M. Gerstein, B. Pasaniuc, M. J. Gandal, D. H. Geschwind, Broad transcriptomic dysregulation across the cerebral cortex in ASD. *bioRxiv* (2020), p. 2020.12.17.423129.

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

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

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

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

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

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

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

35. A. E. Trevino, F. Müller, J. Andersen, L. Sundaram, A. Kathiria, A. Shcherbina, K. Farh, H. Y. Chang, A. M. Pașca, A. Kundaje, S. P. Pașca, W. J. Greenleaf, Chromatin and gene-regulatory

dynamics of the developing human cerebral cortex at single-cell resolution. *Cell*. **184**, 5053–5069.e23 (2021).

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

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

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

39. L. M. Pardo, P. Rizzu, M. Francescatto, M. Vitezic, G. G. R. Leday, J. S. Sanchez, A. Khamis, H. Takahashi, W. D. J. van de Berg, Y. A. Medvedeva, M. A. van de Wiel, C. O. Daub, P. Carninci, P. Heutink, Regional differences in gene expression and promoter usage in aged human brains. *Neurobiol. Aging*. **34**, 1825–1836 (2013).

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

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

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

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

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

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

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

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

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

49. A. L. Lennox, M. L. Hoye, R. Jiang, B. L. Johnson-Kerner, L. A. Suit, S. Venkataramanan, C. J. Sheehan, F. C. Alsina, B. Fregeau, K. A. Aldinger, C. Moey, I. Lobach, A. Afenjar, D. Babovic-Vuksanovic, S. Bézieau, P. R. Blackburn, J. Bunt, L. Burglen, P. M. Campeau, P. Charles, B. H. Y. Chung, B. Cogné, C. Curry, M. D. D'Agostino, N. Di Donato, L. Faivre, D. Héron, A. M. Innes, B. Isidor, B. Keren, A. Kimball, E. W. Klee, P. Kuentz, S. Küry, D. Martin-Coignard, G. Mirzaa, C. Mignot, N. Miyake, N. Matsumoto, A. Fujita, C. Nava, M. Nizon, D. Rodriguez, L. S. Blok, C. Thauvin-Robinet, J. Thevenon, M. Vincent, A. Ziegler, W. Dobyns, L. J. Richards, A. J. Barkovich, S. N. Floor, D. L. Silver, E. H. Sherr, Pathogenic DDX3X Mutations Impair RNA Metabolism and Neurogenesis during Fetal Cortical Development. *Neuron.* **106**, 404–420.e8 (2020).

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

51. 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).

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

53. M. Thoms, E. Thomson, J. Baßler, M. Gnädig, S. Griesel, E. Hurt, The Exosome Is Recruited to

RNA Substrates through Specific Adaptor Proteins. *Cell*. **162**, 1029–1038 (2015).

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

55. 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).

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

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

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

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

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

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

62. G. Ince-Dunn, H. J. Okano, K. B. Jensen, W.-Y. Park, R. Zhong, J. Ule, A. Mele, J. J. Fak, C. Yang, C. Zhang, J. Yoo, M. Herre, H. Okano, J. L. Noebels, R. B. Darnell, Neuronal Elav-like (Hu) proteins regulate RNA splicing and abundance to control glutamate levels and neuronal excitability. *Neuron*. **75**, 1067–1080 (2012).

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

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

65. L. Ho, G. R. Crabtree, Chromatin remodelling during development. *Nature*. **463**, 474–484 (2010).

66. C. Aguilera, E. Gabau, A. Ramirez-Mallafré, C. Brun-Gasca, J. Dominguez-Carral, V. Delgadillo, S. Laurie, S. Derdak, N. Padilla, X. de la Cruz, N. Capdevila, N. Spataro, N. Baena, M. Guitart, A. Ruiz, New genes involved in Angelman syndrome-like: Expanding the genetic spectrum. *PLoS One*. **16**, e0258766 (2021).

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

68. 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).

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

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

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

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

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

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

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

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

77. J. Kaplanis, K. E. Samocha, L. Wiel, Z. Zhang, K. J. Arvai, R. Y. Eberhardt, G. Gallone, S. H. Lelieveld, H. C. Martin, J. F. McRae, P. J. Short, R. I. Torene, E. de Boer, P. Danecek, E. J. Gardner, N. Huang, J. Lord, I. Martincorena, R. Pfundt, M. R. F. Reijnders, A. Yeung, H. G. Yntema, S. Borras, C. Clark, J. Dean, Z. Miedzybrodzka, A. Ross, S. Tennant, T. Dabir, D. Donnelly, M.

Humphreys, A. Magee, V. McConnell, S. McKee, S. McNerlan, P. J. Morrison, G. Rea, F. Stewart, T. Cole, N. Cooper, L. Cooper-Charles, H. Cox, L. Islam, J. Jarvis, R. Keelagher, D. Lim, D. McMullan, J. Morton, S. Naik, M. O'Driscoll, K.-R. Ong, D. Osio, N. Ragge, S. Turton, J. Vogt, D. Williams, S. Bodek, A. Donaldson, A. Hills, K. Low, R. Newbury-Ecob, A. M. Norman, E. Roberts, I. Scurr, S. Smithson, M. Tooley, S. Abbs, R. Armstrong, C. Dunn, S. Holden, S.-M. Park, J. Paterson, L. Raymond, E. Reid, R. Sandford, I. Simonic, M. Tischkowitz, G. Woods, L. Bradley, J. Comerford, A. Green, S. Lynch, S. McQuaid, B. Mullaney, J. Berg, D. Goudie, E. Mavrak, J. McLean, C. McWilliam, E. Reavey, T. Azam, E. Cleary, A. Jackson, W. Lam, A. Lampe, D. Moore, M. Porteous, E. Baple, J. Baptista, C. Brewer, B. Castle, E. Kivuva, M. Owens, J. Rankin, C. Shaw-Smith, C. Turner, P. Turnpenny, C. Tysoe, T. Bradley, R. Davidson, C. Gardiner, S. Joss, E. Kinning, C. Longman, R. McGowan, V. Murday, D. Pilz, E. Tobias, M. Whiteford, N. Williams, A. Barnicoat, E. Clement, F. Faravelli, J. Hurst, L. Jenkins, W. Jones, V. K. A. Kumar, M. Lees, S. Loughlin, A. Male, D. Morrogh, E. Rosser, R. Scott, L. Wilson, A. Beleza, C. Deshpande, F. Flinter, M. Holder, M. Irving, L. Izatt, D. Josifova, S. Mohammed, A. Molenda, L. Robert, W. Roworth, D. Ruddy, M. Ryten, S. Yau, C. Bennett, M. Blyth, J. Campbell, A. Coates, A. Dobbie, S. Hewitt, E. Hobson, E. Jackson, R. Jewell, A. Kraus, K. Prescott, E. Sheridan, J. Thomson, K. Bradshaw, A. Dixit, J. Eason, R. Haines, R. Harrison, S. Mutch, A. Sarkar, C. Searle, N. Shannon, A. Sharif, M. Suri, P. Vasudevan, N. Canham, I. Ellis, L. Greenhalgh, E. Howard, V. Stinton, A. Swale, A. Weber, S. Banka, C. Breen, T. Briggs, E. Burkitt-Wright, K. Chandler, J. Clayton-Smith, D. Donnai, S. Douzgou, L. Gaunt, E. Jones, B. Kerr, C. Langley, K. Metcalfe, A. Smith, R. Wright, D. Bourn, J. Burn, R. Fisher, S. Hellens, A. Henderson, T. Montgomery, M. Splitt, V. Straub, M. Wright, S. Zwolinski, Z. Allen, B. Bernhard, A. Brady, C. Brooks, L. Busby, V. Clowes, N. Ghali, S. Holder, R. Ibitoye, E. Wakeling, E. Blair, J. Carmichael, D. Cilliers, S. Clasper, R. Gibbons, U. Kini, T. Lester, A. Nemeth, J. Poulton, S. Price, D. Shears, H. Stewart, A. Wilkie, S. Albaba, D. Baker, M. Balasubramanian, D. Johnson, M. Parker, O. Quarrell, A. Stewart, J. Willoughby, C. Crosby, F. Elmslie, T. Homfray, H. Jin, N. Lahiri, S. Mansour, K. Marks, M. McEntagart, A. Saggar, K. Tatton-Brown, R. Butler, A. Clarke, S. Corrin, A. Fry, A. Kamath, E. McCann, H. Mugalaasi, C. Pottinger, A. Procter, J. Sampson, F. Sansbury, V. Varghese, D. Baralle, A. Callaway, E. J. Cassidy, S. Daniels, A. Douglas, N. Foulds, D. Hunt, M. Kharbanda, K. Lachlan, C. Mercer, L. Side, I. K. Temple, D. Wellesley, L. E. L. M. Vissers, J. Juusola, C. F. Wright, H. G. Brunner, H. V. Firth, D. R. FitzPatrick, J. C. Barrett, M. E. Hurles, C. Gilissen, K. Retterer, Deciphering Developmental Disorders Study, Evidence for 28 genetic disorders discovered by combining healthcare and research data. *Nature* (2020), doi:10.1038/s41586-020-2832-5.

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

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

80. Epi25 Collaborative. Electronic address: s.berkovic@unimelb.edu.au, 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).

81. D. M. Werling, H. Brand, J.-Y. An, M. R. Stone, L. Zhu, J. T. Glessner, R. L. Collins, S. Dong, R. M. Layer, E. Markenscoff-Papadimitriou, A. Farrell, G. B. Schwartz, H. Z. Wang, B. B. Currall, X. Zhao, J. Dea, C. Duhn, C. A. Erdman, M. C. Gilson, R. Yadav, R. E. Handsaker, S. Kashin, L. Klei, J. D. Mandell, T. J. Nowakowski, Y. Liu, S. Pochareddy, L. Smith, M. F. Walker, M. J. Waterman, X. He, A. R. Kriegstein, J. L. Rubenstein, N. Sestan, S. A. McCarroll, B. M. Neale, H. Coon, A. J. Willsey, J. D. Buxbaum, M. J. Daly, M. W. State, A. R. Quinlan, G. T. Marth, K. Roeder, B. Devlin, M. E. Talkowski, S. J. Sanders, An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat. Genet.* **50**, 727–736 (2018).

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

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

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

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

86. R. M. Easton, H. Cho, K. Roovers, D. W. Shineman, M. Mizrahi, M. S. Forman, V. M.-Y. Lee, M.

Szabolcs, R. de Jong, T. Oltersdorf, T. Ludwig, A. Efstratiadis, M. J. Birnbaum, Role for Akt3/protein kinase Bgamma in attainment of normal brain size. *Mol. Cell. Biol.* **25**, 1869–1878 (2005).

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

88. M. Tardaguila, L. de la Fuente, C. Marti, C. Pereira, F. J. Pardo-Palacios, H. Del Risco, M. Ferrell, M. Mellado, M. Macchietto, K. Verheggen, M. Edelmann, I. Ezkurdia, J. Vazquez, M. Tress, A. Mortazavi, L. Martens, S. Rodriguez-Navarro, V. Moreno-Manzano, A. Conesa, SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Res.* **28**, 396–411 (2018).

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

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

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

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

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

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

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

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

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

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

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

100. A. T. L. Lun, D. J. McCarthy, J. C. Marioni, A step-by-step workflow for low-level analysis of

single-cell RNA-seq data with Bioconductor. *F1000Res*. **5**, 2122 (2016).

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

Acknowledgments

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

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

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

Author Contributions: A.P., M.J.G and L.T.U. conceived and designed the study. C.V. and L.T.U. collected and processed the tissue specimens and dissected the samples for processing. A.P. and L.T.U. generated the library for single cell iso-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 alternative polyadenylation analysis, C.V. performed the RBP associated analysis, C.V., L.T.U. and M.J.G. conducted network analyses, A.P., C.J., M.M. performed single cell iso-Seq analysis, M.K., D.V., A.P. performed single cell isoform switch analysis. X.G., J.J.L. perform single cell DTU analysis, A.P and L.T.U analyzed the DTU results,

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

Competing Interests: M.J.G. receives grant funding from Mitsubishi Tanabe Pharma America that is unrelated to this current project.

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

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

Code availability: Code used to process the data and generate all figures in this manuscript is available at https://github.com/gandallab/Dev_Brain_IsoSeq

Supplementary Materials

Materials and Methods

Figs. S1 to S13

Tables S1 to S8

References

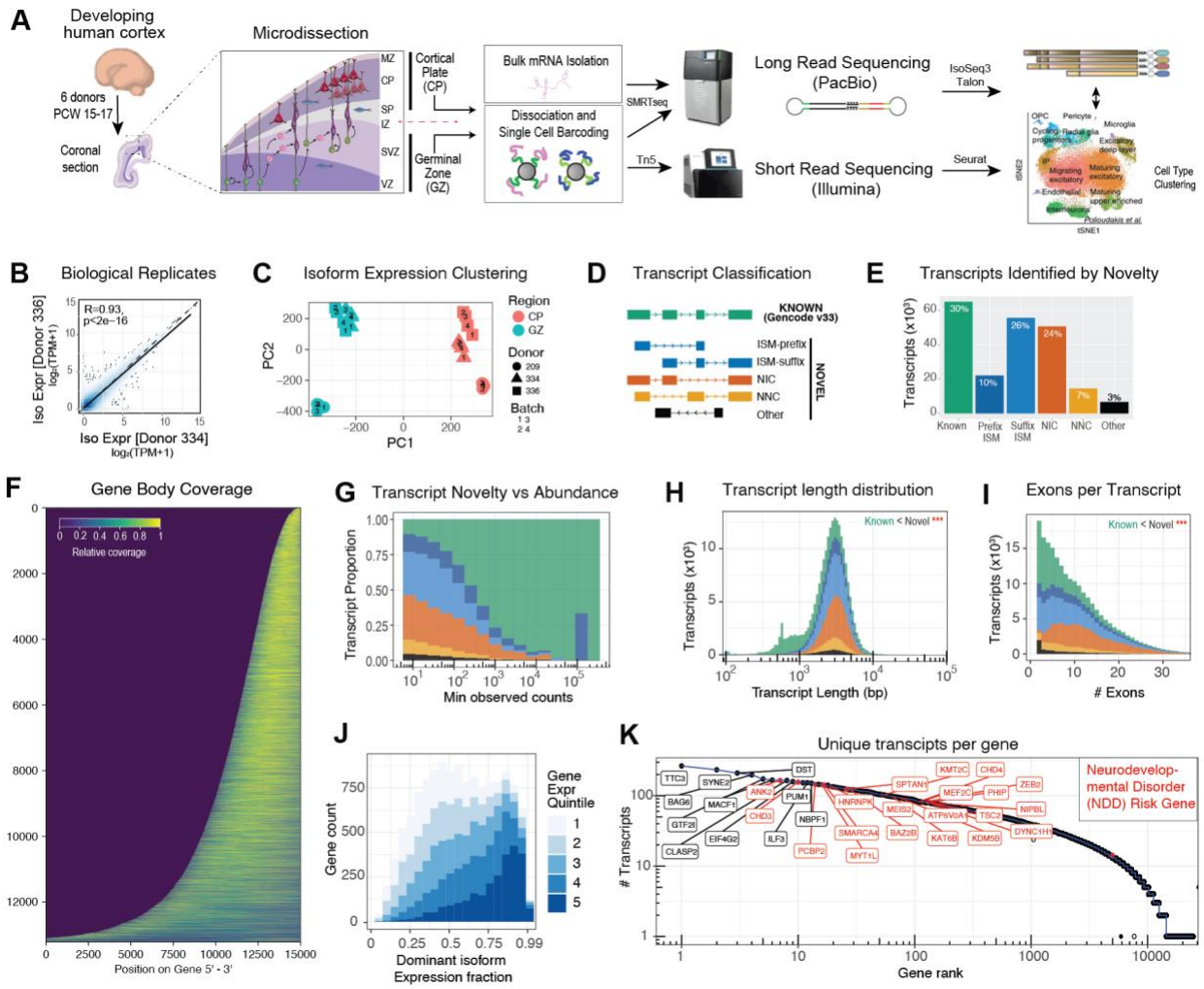

Fig. 1. The full-length cell-type specific transcriptome of the developing human neocortex at mid-gestation.

(**A**) Experimental design for isoform-centric profiling of the developing human brain transcriptome at bulk and single-cell resolution. Briefly, microdissected samples from the progenitor-enriched germinal zone (GZ) and neuronally-enriched cortical plate (CP) were profiled from 6 unique donors at mid-gestation. Full-length cDNA libraries were generated from homogenate tissue as well as from dissociated, barcoded single-cells with incorporation of UMIs. Single molecule long-read sequencing (PacBio) was used to quantify transcript-isoforms and integrated with matched short-read scRNA-Seq. Isoform expression quantifications demonstrated (**B**) strong biological reproducibility across unique donors and (**C**) region-specific clustering via principal component analysis. (**D**) Transcript isoforms identified by IsoSeq were compared against the Gencode v33 reference. Novel transcripts were further classified by their splice junction matching to annotated Gencode isoforms as described by TALON: ISM (incomplete splice match); NIC (novel in catalog); NNC (novel not in catalog). The 'other' category denotes isoforms belonging to antisense, genomic and intergenic classes. (**E**) Number of isoforms identified based on classes described in Fig 1D. (**F**) Heatmap shows uniform patterns of relative read-depth coverage across genes, arranged by length. Low coverage in dark blue, high coverage in yellow. (**G**) Abundance of the isoforms by each class as described in Fig 1D. Compared with known isoforms, novel transcripts identified here were (**H**) significantly longer and (**I**) contained a significantly greater number of exons. (**J**) Proportion of the dominant isoforms for each gene by gene expression percentile. For highly expressed genes, the dominant isoform contributed maximum to the gene expression. (**K**) Genes ranked by the number of unique transcript isoforms detected. Neurodevelopmental disorder (NDD) risk genes (red) had significantly more detected isoforms, controlling with total expression and length.

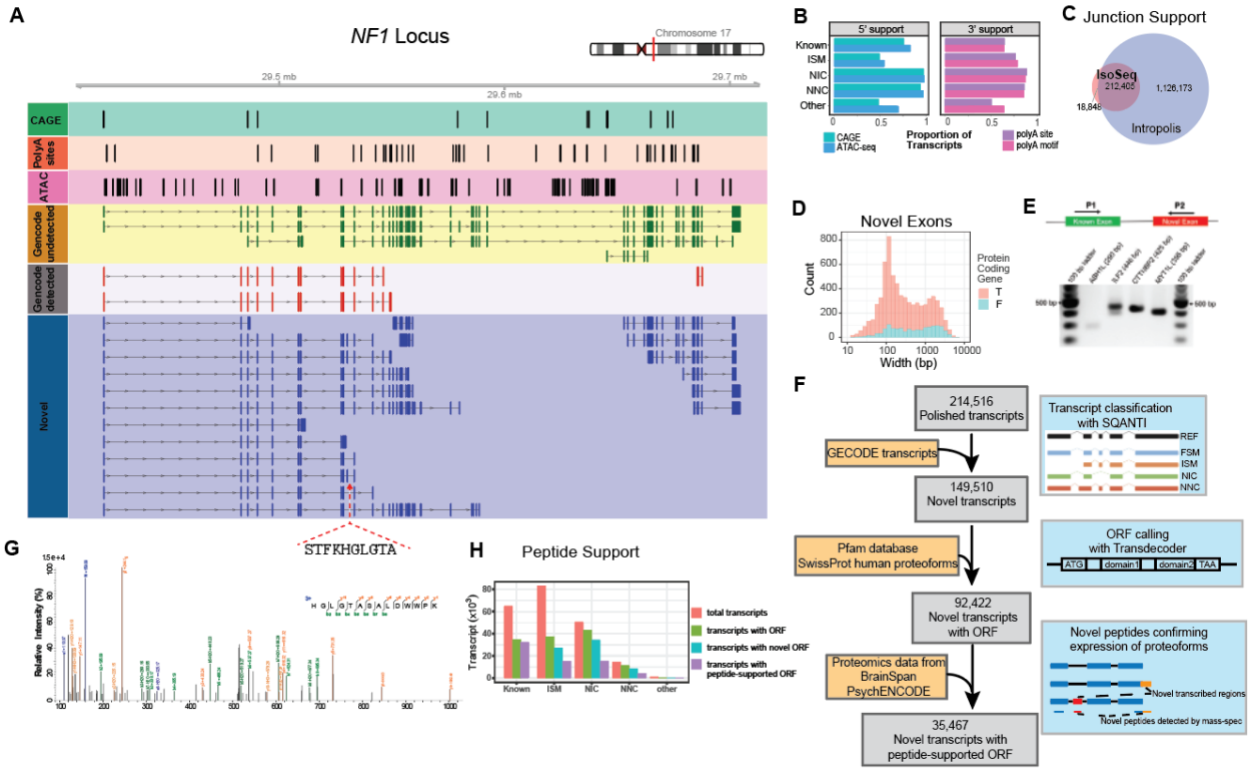

Fig. 2. Expanded Transcriptomic and Proteomic Complexity

(A) The *NF1* gene locus with 20 previously unidentified brain-expressed isoforms. Tracks from top to bottom include: CAGE clusters, 3′seq clusters, ATAC-Seq peaks, Gencode isoforms that were not detected in our data, Gencode isoforms detected in our data, and novel isoforms identified from our data. A novel microexon is highlighted. **(B)** External validation of isoforms by independent datasets: 5' end validation was performed by presence of peak from CAGE (FANTOM5 and Fetal brain cortex) and mid-gestation cortex ATAC-Seq (*34–36*); 3' end validation was performed by presence of polyA motif or peak from polyAsite database. Percentage of transcript with support at either end is highlighted. **(C)** The vast majority of splice junctions identified by IsoSeq are supported by the external Intropolis splice junction database. **(D)** Length distribution of the >7000 novel spliced exons uncovered here. **(E)** Validation of novel exons via RT-PCR. Expected product size is shown in parenthesis along the name of the gene with the exon. Each novel exon was amplified with the primer sets shown in the schematic (see Methods). **(F)** Characterization of novel protein-coding transcripts. Long-read sequencing identified a total of 214,516 transcripts, 149,510 of which were novel compared to Gencode v33. 92,422 novel transcripts were predicted to code for protein sequences, and 35,467 predicted ORFs were further confirmed by MS/MS proteomics data. **(G)** Representative mass spectrum of peptide HGLGTASALDWWPK, which confirms the translation of the identified NF1 microexon. Matched b, y, a and immonium ions were highlighted. **(H)** Number of total transcripts, transcripts with ORF, transcripts with novel ORF compared to UniProt human protein sequences, and transcripts with ORFs validated by MS/MS proteomics, plotted per isoform structural category.

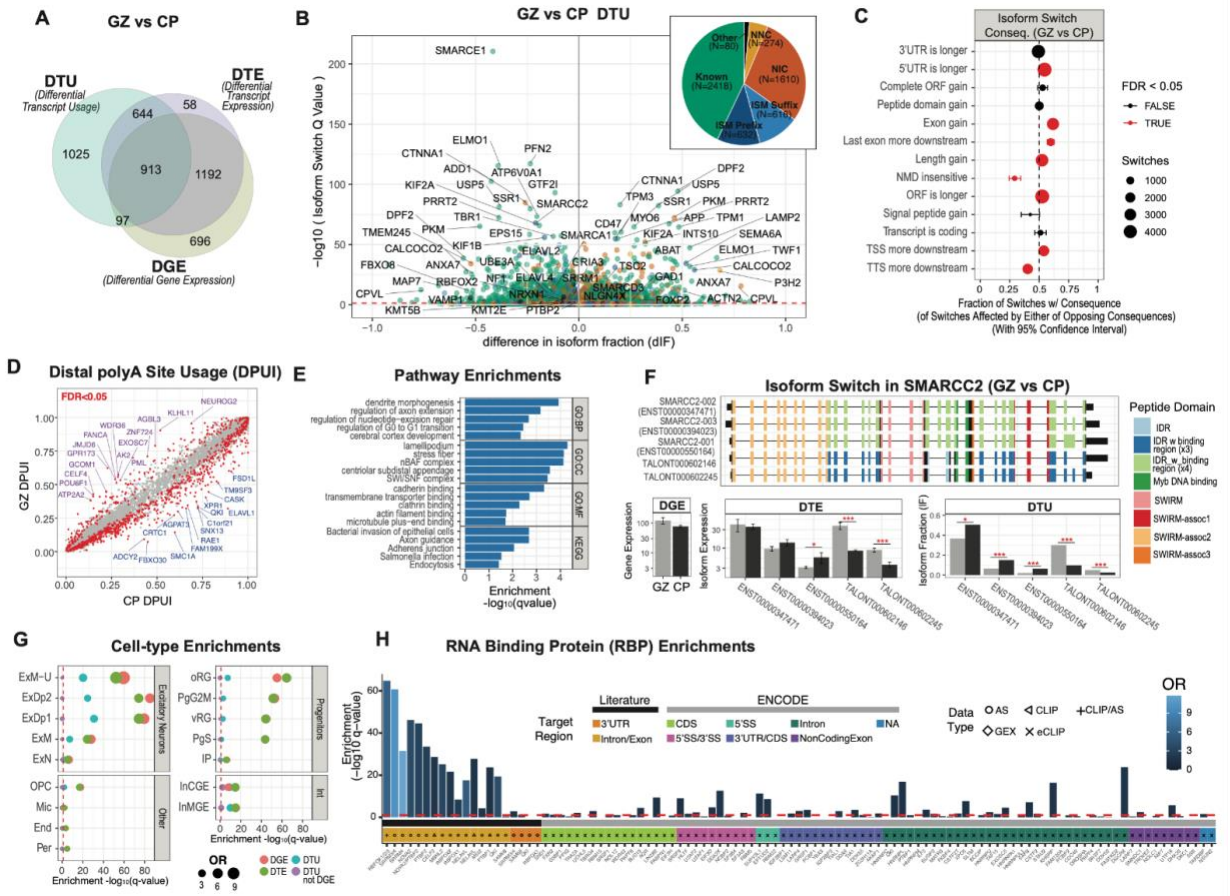

Fig. 3. The landscape of isoform switching during human corticogenesis.

(**A**) Long-read RNA-Seq data from GZ and CP samples was contrasted for patterns of differential gene expression (DGE), differential transcript expression (DTE), and differential transcript usage (DTU). Venn diagram is shown depicting the overlap for genes exhibiting significant DGE, DTE, and DTU (FDR-corrected P<0.05). (**B**) A volcano plot depicts isoform switching across GZ and CP. The x-axis depicts the difference in isoform fraction (dIF) for a given transcript in CP versus GZ. Inset, most regionally variable DTU isoforms are novel compared to Gencode. (**C**) Functional consequences of isoform switch events between GZ and CP are shown. For example, CP-upregulated isoforms were more significantly likely to have gained (than lost) an exon. (**D**) Analysis of distal polyA site usage index (DPUI) for known transcripts between GZ and CP samples. On average, CP transcripts have higher DPUIs indicative of longer 3' UTRs. **(E)** Pathways enrichments for genes exhibiting cross-region DTU are notable for dendrite morphogenesis and SWI/SNF complex genes, among others. (**F**) An example of isoform-switching observed within the ASD risk gene *SMARCC2*. Although total gene expression was not different between GZ and CP, significant switching was observed with two novel isoforms exhibiting preferential usage in GZ. (**G**) Regionally variable genes were significantly enriched for cell-type specific marker genes from scRNA-Seq. (**H**) Genes containing DTU isoforms were also highly enriched for targets of RNA binding proteins known to regulate alternative splicing in the brain (black bar), and for targets of RBPs profiled in the ENCODE database (grey bar). See also **fig. S4C**.

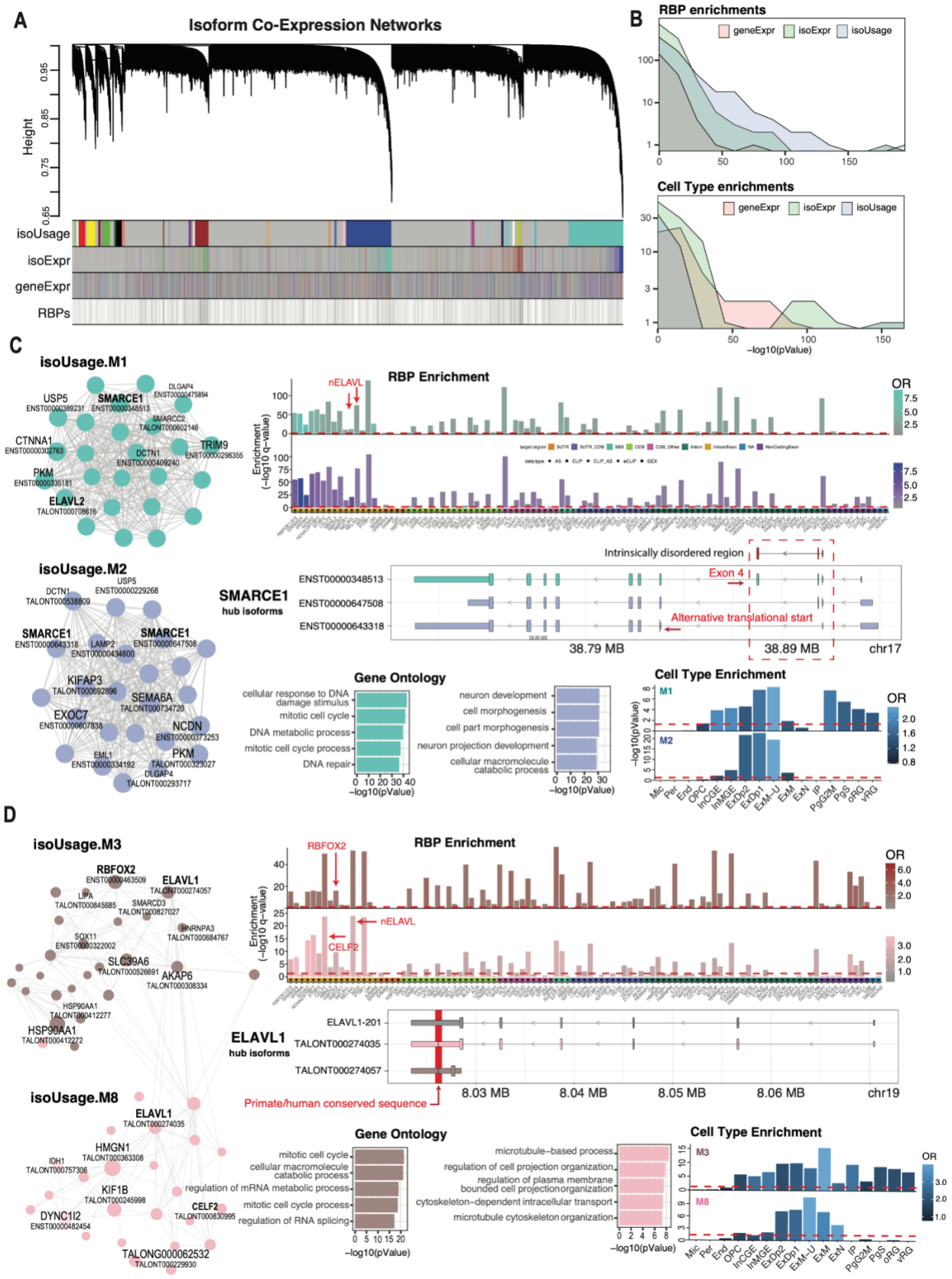
A
Isoform Co-Expression Networks
Height
isoUsage
isoExpr
geneExpr
RBPs
B
RBP enrichments
geneExpr
isoExpr
isoUsage
Cell Type enrichments
−log10(pValue)
C
isoUsage.M1
RBP Enrichment
nELAVL
OR
Enrichment (−log10 q-value)
isoUsage.M2
SMARCE1
hub isoforms
Intrinsically disordered region
Exon 4
Alternative translational start
ENST00000348513
ENST00000647508
ENST00000643318
38.79 MB
38.89 MB
chr17
Gene Ontology
cellular response to DNA damage stimulus
mitotic cell cycle
DNA metabolic process
mitotic cell cycle process
DNA repair
neuron development
cell morphogenesis
cell part morphogenesis
neuron projection development
cellular macromolecule catabolic process
−log10(pValue)
Cell Type Enrichment
M1
M2
Mic
Per
End
OPC
InCGE
InMGE
ExDp2
ExDp1
ExM-U
ExM
ExN
IP
PgG2M
PgS
oRG
vRG
D
isoUsage.M3
RBP Enrichment
RBFOX2
nELAVL
CELF2
ELAVL1
hub isoforms
ELAVL1-201
TALONT000274035
TALONT000274057
Primate/human conserved sequence
8.03 MB
8.04 MB
8.05 MB
8.06 MB
chr19
isoUsage.M8
Gene Ontology
mitotic cell cycle
cellular macromolecule catabolic process
regulation of mRNA metabolic process
mitotic cell cycle process
regulation of RNA splicing
microtubule-based process
regulation of cell projection organization
regulation of plasma membrane bounded cell projection organization
cytoskeleton-dependent intracellular transport
microtubule cytoskeleton organization
Cell Type Enrichment
M3
M8

Fig. 4. Network-based contextualization of isoform usage

(A) The isoUsage network is driven by RNA-binding protein isoform usage. *Top*, a dendrogram of the isoUsage network with isoforms (isoUsage and isoExpr) or genes (geneExpr) organized by their presence in isoUsage modules plotted below. Isoforms of known RNA-binding proteins (RBPs) are plotted at *bottom*. (**B**) The isoUsage network shows more and stronger enrichments for RBP targets compared to geneExpr and isoExpr. *Top*, density plot of RBP enrichments for the three networks; *bottom*, density plot of cell type enrichments. (**C**) *SMARCE1* hub isoforms inform different cellular processes associated with progenitors and neurons. *Left*, module plots highlighting hub isoforms in isoUsage.M1 and M2. *Top*, RBP target enrichments for these modules. Note enrichment for neuronal ELAVL targets in isoUsage.M1. *Middle*, transcript models for the *SMARCE1* hub isoforms. Box highlights exon3 (turquoise) and the shifted reading frame driven by an alternative translational start (blue). (**D**) *ELAVL1* hub isoforms containing a primate/human-specific 3'UTR. *Left*, hairballs for hub isoforms in the M3 and M8 isoUsage modules. *Top*, RBP target enrichments showing strong enrichments for targets of ELAVL and CELF2. *Middle*, isoform models for *ELAVL1* – arrow highlights the primate/human-conserved sequence missing from these 3'UTRs. C and D) *Bottom*, gene ontology (biological process) and cell type enrichments for the four highlighted modules.

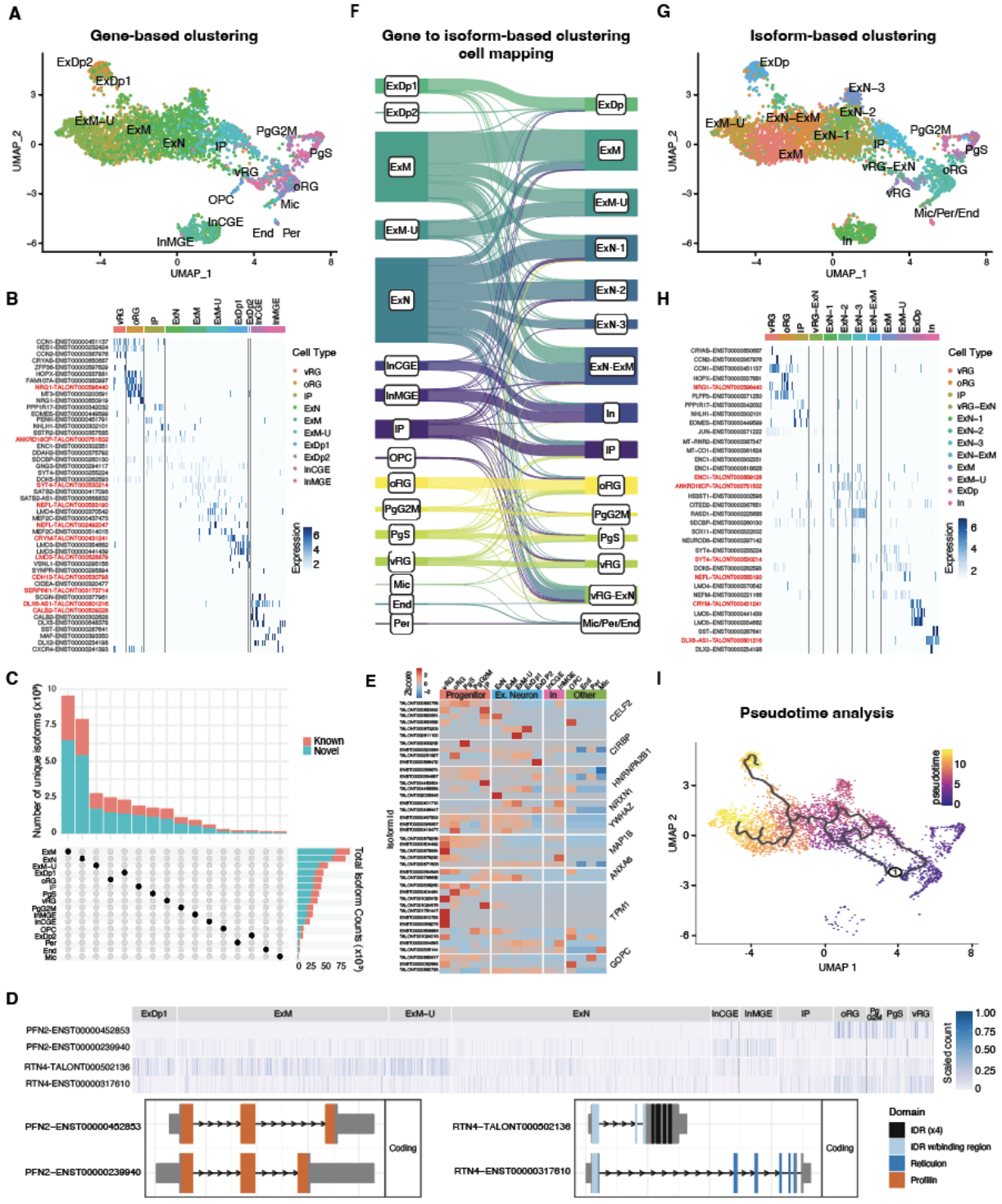
A
Gene-based clustering
ExDp2
ExDp1
ExM-U
ExM
ExN
IP
PgG2M
PgS
vRG
oRG
OPC
Mic
InCGE
InMGE
End
Per
UMAP_1
UMAP_2
F
Gene to isoform-based clustering cell mapping
ExDp1
ExDp2
ExM
ExM-U
ExN
InCGE
InMGE
IP
OPC
oRG
PgG2M
PgS
vRG
Mic
End
Per
ExDp
ExM
ExM-U
ExN-1
ExN-2
ExN-3
ExN-ExM
In
IP
oRG
PgG2M
PgS
vRG
vRG-ExN
Mic/Per/End
G
Isoform-based clustering
ExDp
ExN-3
ExN-2
ExN-ExM
ExM-U
ExN-1
ExM
IP
PgG2M
PgS
vRG-ExN
oRG
vRG
Mic/Per/End
In
B
Cell Type
vRG
oRG
IP
ExN
ExM
ExM-U
ExDp1
ExDp2
InCGE
InMGE
Expression
H
Cell Type
vRG
oRG
IP
vRG-ExN
ExN-1
ExN-2
ExN-3
ExN-ExM
ExM
ExM-U
ExDp
In
Expression
C
Number of unique isoforms (x10⁴)
Known
Novel
Total Isoform Counts (x10⁵)
E
Zscore
Progenitor
Ex. Neuron
In
Other
Isoform Id
CELF2
CIRBP
HNRNPA2B1
NRXN1
YWHAZ
MAPT
ANXA6
TPM1
GOPC
I
Pseudotime analysis
pseudotime
UMAP 1
UMAP 2
D
ExDp1
ExM
ExM-U
ExN
InCGE
InMGE
IP
oRG
PgG2M
PgS
vRG
PFN2-ENST00000452853
PFN2-ENST00000239940
RTN4-TALONT000502136
RTN4-ENST00000317610
Scaled count
PFN2–ENST00000452853
PFN2–ENST00000239940
Coding
RTN4–TALONT000502136
RTN4–ENST00000317610
Coding
Domain
IDR (x4)
IDR w/binding region
Reticulon
Profilin

Fig. 5. Cell type-specific isoform diversity in the developing human cortex

(**A**) Uniform manifold approximation and projection (UMAP) plot of 4,281 cells detected by both 3' end short-read sequencing and by scIso-Seq. Each dot represents a single cell, colored by its corresponding cluster. UMAP position of the cells is calculated based on isoform expression while cluster labels are as previously defined ((*3*)). vRG = ventricular radial glia, oRG = Outer Radial Glia, PgG2M = Cycling progenitors (G2/M phase), PgS = Cycling progenitors (S phase), IP = Intermediate Progenitors, ExN = Migrating excitatory, ExM = Maturing excitatory, ExM-U = Maturing excitatory upper enriched, ExDp1 = Excitatory deep layer 1, ExDp2 = Excitatory deep layer 2, InMGE = Interneuron MGE, InCGE = Interneuron CGE, OPC = Oligodendrocytes Precursor Cells, End = Endothelial, Per = Pericyte, Mic = Microglia (**B**) Heatmap showing differentially expressed isoforms across cell types defined by gene-based clustering. Novel isoforms are shown in red. (**C**) Distribution of isoforms across cell types shows greater diversity of isoforms in newborn migrating (ExN) and maturing excitatory neurons (ExM). (**D**) Isoforms of PFN2 and RTN4 differentially expressed across cell types along with their predicted functional consequences. Isoform ENST00000239940, predominantly expressed in neurons, is predicted to encode IDR protein domains not found in the progenitor-enriched ENST00000452853 isoform. The novel isoform TALONT000502136 is enriched in neurons, whereas the progenitor enriched ENST00000317610 isoform is longer and contains multiple reticulon protein domains. (**E**) Heatmap showing a subset of isoforms with differential usage across cell types (DTU). (**F**) River plot showing mapping of cells from gene-based (left) to isoform-based (right) clustering. Each line represents a single cell. (**G**) UMAP of cells clustered based on isoform expression as measured by scIsoSeq. Additional stages of excitatory neuron maturation can be defined using isoform-level data. (**H**) Heatmap showing differentially expressed isoforms across cell types defined by isoform-based clustering. Novel isoforms are shown in red. (**I**) Cell lineage trajectory analysis (Monocle3) shows direct neurogenesis through vRG-ExN cells and indirect path through IP cells.

A
Rare variant enrichments across transcriptomic signals
NDD.fuTADA
n_transcripts_log2
n_exons_log2
DTU.down
isoUsage.M1_turquoise
isoUsage.M3_brown
isoExpr.M5_green
n_transcripts_log2ltpm
DTE.up
DGE.up
isoExpr.M19_lightyellow
ASD.fuTADA
n_transcripts_log2
n_exons_log2
isoExpr.M5_green
n_transcripts_log2ltpm
DTE.up
DGE.up
isoExpr.M7_black
geneExpr.M13_salmon
DTU.up
isoUsage.M2_blue
ASD.SFARI.S
n_transcripts_log2
n_exons_log2
isoExpr.M19_lightyellow
isoExpr.M7_black
isoExpr.M27_white
PgS
vRG
PgG2M
isoUsage.M27_white
oRG
DDD.kaplanis
n_transcripts_log2
n_exons_log2
isoUsage.M3_brown
isoExpr.M5_green
n_transcripts_log2ltpm
DTE.up
DGE.up
isoExpr.M7_black
geneExpr.M13_salmon
isoExpr.M20_royalblue
SCZ.schema
n_transcripts_log2
n_exons_log2
isoExpr.M5_green
isoExpr.M19_lightyellow
geneExpr.M13_salmon
isoUsage.M27_white
isoExpr.M13_salmon
isoExpr.M25_orange
isoUsage.M23_darkturquoise
isoExpr.M34_darkmagenta
EPI.helbig
n_transcripts_log2
n_exons_log2
DTE.up
DGE.up
geneExpr.M5_green
DGE
isoExpr.M1_turquoise
isoExpr.M20_royalblue
geneExpr.M6_red
isoUsage.M8_pink
0 10 20 30 40
−log10(fdr)
class
DGE/DTE/DTU
Networks.geneExpr
Networks.isoExpr
Networks.isoUsage
SingleCellMarkers
txAnno
B
Isoform switches in neuropsychiatric disease genes
Cell Type
ASD
SCZ
NDD
Progenitor
Excitatory Neuron
Interneuron
Other
Zscore
2
0
−2
AKAP9
DCTN2
EHBP1
KIF1B
NRXN1
SMARCE1
TCF4
C
Single-cell isoform expresssion and usage enrichments
NDD.fuTADA
ASD.fuTADA
PgS
ExM−U
ExDp
ExM
SingleCellDTU
vRG
PgG2M
ExN−1
oRG
ExN−3
ExN−2
In
vRG−ExN
Mic/Per/End
IP
ExN−ExM
0 1 2 3 4 5
−log10(fdr)
D
Rare variant reannotation
Reassigned VEP Consequence
downstream gene
upstream gene
intron
3' UTR
5' UTR
synonymous
stop retained
splice polypyrimidine tract
splice donor region
splice donor 5th base
splice region
protein altering
missense
inframe deletion
inframe insertion
start lost
stop lost
frameshift
stop gained
splice donor
splice acceptor
stop gained
frameshift
inframe deletion
missense
splice region
splice donor region
splice polypyrimidine tract
incomplete terminal codon
stop retained
synonymous
5' UTR
3' UTR
nc- transcript exon
intron
upstream gene
downstream gene
intergenic
Previous VEP Consequence
DNM from
case
control
Number of DNM
1
8
80
800
E
KLC1 locus
Chromosome 14
CAGE
PolyA sites
ATAC
Canonical
Novel
TALONT000423578.p1
DNM
chr14:104129236 A -> G
Effect on GENCODE transcripts:
Missense Atg -> Gtg M -> V
Effect on novel transcript:
Start lost Atg -> Gtg M -> V

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

(A) Enrichment of transcriptomic features, differential expression analyses across cortical regions and cell types or isoform expression and usage networks with neuropsychiatric disorders. Red line indicates the FDR-corrected significance threshold. (**B**) 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. (**C**) Cell type enrichments indicate differential isoform expression and utilization in NDD and ASD. (**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. (**E**) The *KLC1* gene locus with representative canonical isoforms and a novel isoform identified in this study. Red vertical line indicates the position of a case DNM that affects this locus. The affected region is highlighted in the lower panel. This DNM leads to a missense mutation in the canonical proteoforms, which causes a start lost in the novel proteoform.