Long-read RNA sequencing identifies region- and sex-specific C57BL/6J mouse brain mRNA isoform expression and usage

Emma F. Jones[1], Timothy C. Howton[1], Victoria L. Flanary[1], Amanda D. Clark[1] and Brittany N. Lasseigne[1*]

Abstract

Alternative splicing (AS) contributes to the biological heterogeneity between species, sexes, tissues, and cell types. Many diseases are either caused by alterations in AS or by alterations to AS. Therefore, measuring AS accurately and efficiently is critical for assessing molecular phenotypes, including those associated with disease. Long-read sequencing enables more accurate quantification of differentially spliced isoform expression than short-read sequencing approaches, and third-generation platforms facilitate high-throughput experiments. To assess differences in AS across the cerebellum, cortex, hippocampus, and striatum by sex, we generated and analyzed Oxford Nanopore Technologies (ONT) long-read RNA sequencing (lrRNA-Seq) C57BL/6J mouse brain cDNA libraries. From > 85 million reads that passed quality control metrics, we calculated differential gene expression (DGE), differential transcript expression (DTE), and differential transcript usage (DTU) across brain regions and by sex. We found significant DGE, DTE, and DTU across brain regions and that the cerebellum had the most differences compared to the other three regions. Additionally, we found region-specific differential splicing between sexes, with the most sex differences in DTU in the cortex and no DTU in the hippocampus. We also report on two distinct patterns of sex DTU we observed, sex-divergent and sex-specific, that could potentially help explain sex differences in the prevalence and prognosis of various neurological and psychiatric disorders in future studies. Finally, we built a Shiny web application for researchers to explore the data further. Our study provides a resource for the community; it underscores the importance of AS in biological heterogeneity and the utility of long-read sequencing to better understand AS in the brain.

Keywords Long-read RNA sequencing, Brain, Sex, Transcript usage, Isoform usage

*Correspondence:
Brittany N. Lasseigne
bnp0001@uab.edu
[1]Department of Cell, Developmental and Integrative Biology, Heersink
School of Medicine, The University of Alabama at Birmingham,
Birmingham, AL, United States of America

Introduction

Alternative splicing (AS) of preRNAs to mRNAs can result in multiple transcript isoforms and proteins from a single gene. This process contributes to the biological heterogeneity between species [1], sexes [2], tissues [3, 4], and cell types [5]. Notably, AS is more abundant in the brain, and the brain has more tissue-specific transcript isoforms than other tissues [3]. AS is associated with many psychiatric and neurological disorders (e.g., Autism Spectrum Disorder (ASD), schizophrenia, and epilepsy [6, 7]). Furthermore, many psychiatric and neurological disorders differ in prevalence by sex [8, 9]. For example, ASD, more common in males, has been linked to multiple genetic changes, including disordered splicing [10, 11]. However, as biomedical research has historically failed to study sex as a biological variable [12], there is still a need to quantify AS in the brain by sex accurately.

Recent advances in third-generation long-read sequencing technologies (i.e., Pacific Biosciences and Oxford Nanopore Technologies - ONT) enable high-throughput sequencing of complete mRNA transcripts to more rigorously determine the expressed transcript isoforms in a given sample compared to short-read (i.e., next- or second-generation) sequencing approaches. The resulting "long reads" can measure novel transcripts missed with prior studies and reveal extensive isoform-level diversity. For example, Clark et al. applied long-read sequencing to the human psychiatric risk gene *CACNA1C* and discovered 38 novel exons and 241 novel transcripts [13]. While short-read gene expression AS data analysis can include calculating the percent spliced-in of exons or the splice junctions for a given gene, long reads enable researchers to quantify splicing across entire transcripts directly. Differential transcript usage (DTU), sometimes referred to as differential isoform usage, quantifies changes in expression of a specific transcript as a fraction of the overall expression of a particular gene (**Methods**), complementing differential gene expression (DGE) and differential transcript expression (DTE) analyses [14]. This fraction, referred to as the isoform fraction (IF), is essential for including information about a given isoform's expression in relation to other isoforms of the same gene. Please note that DTE and DTU are not mutually exclusive; a visual example of significant DTE with and without significant DTU is available in Fig. 1. Recently, researchers identified six candidate genes with novel DTU events in a schizophrenia cohort and developed a method to stratify patient populations using multi-gene DTU patterns [15], exemplifying that DTU can identify biologically relevant information in heterogeneous patient populations. These studies underscore how long-read sequencing approaches paired with novel analytical frameworks can identify and quantify AS patterns in the brain.

Due to known sex biases in healthy brain gene expression [2] and brain-related disease phenotypes [8, 9], we studied AS across brain regions and sexes. Thus, we sequenced the cDNA from C57BL/6J mouse cerebellum, cortex, hippocampus, and striatum RNA for each sex (n=5 each) using ONT and calculated DGE, DTE, and DTU between conditions. We generated over 85 million reads passing quality control metrics. We observed that the brain region with the highest DGE, DTE, and DTU is the cerebellum and that the most sex differences were in the cortex. We also built a web application hosting our data for use by the scientific community.

Results

Long-read RNA-Seq profiles across four mouse brain regions identified potentially novel genes and transcripts

We sequenced cDNA synthesized from total mRNA from the cerebellum, cortex, hippocampus, and striatum of 20-week-old male and female (n=5 each) C57BL/6J mice using an ONT GridION device (Fig. 2A). We obtained 85,909,493 reads passing quality control metrics (**Methods**), with each brain region receiving at least 16 million reads across the ten samples for that region (Fig. 2C). The hippocampus had the lowest number of total reads (n=16,739,487), potentially due to our reduced starting material as it is smaller than the other brain regions we assayed. We aligned and quantified our data using the nf-core [16] nanoseq pipeline and Bambu [17], a tool for performing machine-learning-based transcript discovery and quantification of long-read RNA-sequencing data with high precision and recall [18]. When visualizing our samples based on variance-stabilization transformed (VST) gene counts by principal component analysis (PCA), samples are separated by tissue (Fig. 2B). The difference in cerebellum samples to all other brain region samples drove the greatest gene expression variation in the data set (PC1, 33% of the total variance, Fig. 2B).

Next, we determined any potentially novel genes and transcripts we had captured with Bambu [17]. We identified 285 genes and 382 transcripts not previously annotated in mouse GENCODE release M31 (Fig. 2D and E) and considered them "novel." These 382 potentially novel transcripts correspond to 354 unique genes. Of the 382 novel transcripts, 309 (81%) transcripts belonged to novel genes, and 73 (19%) belonged to previously annotated genes (Additional File 1). Interestingly, when we examined the overall expression distributions of these novel transcripts compared to annotated transcripts, novel transcripts were expressed significantly more than annotated transcripts (one-sided Wilcoxon rank sum test, p=7.850072e-114, Additional File 8 - Supplementary Fig. 1A), potentially due to the stringent expression thresholds Bambu has to identify novel transcripts. Of these novel transcripts, 279 out of 382 (79%) had a mean
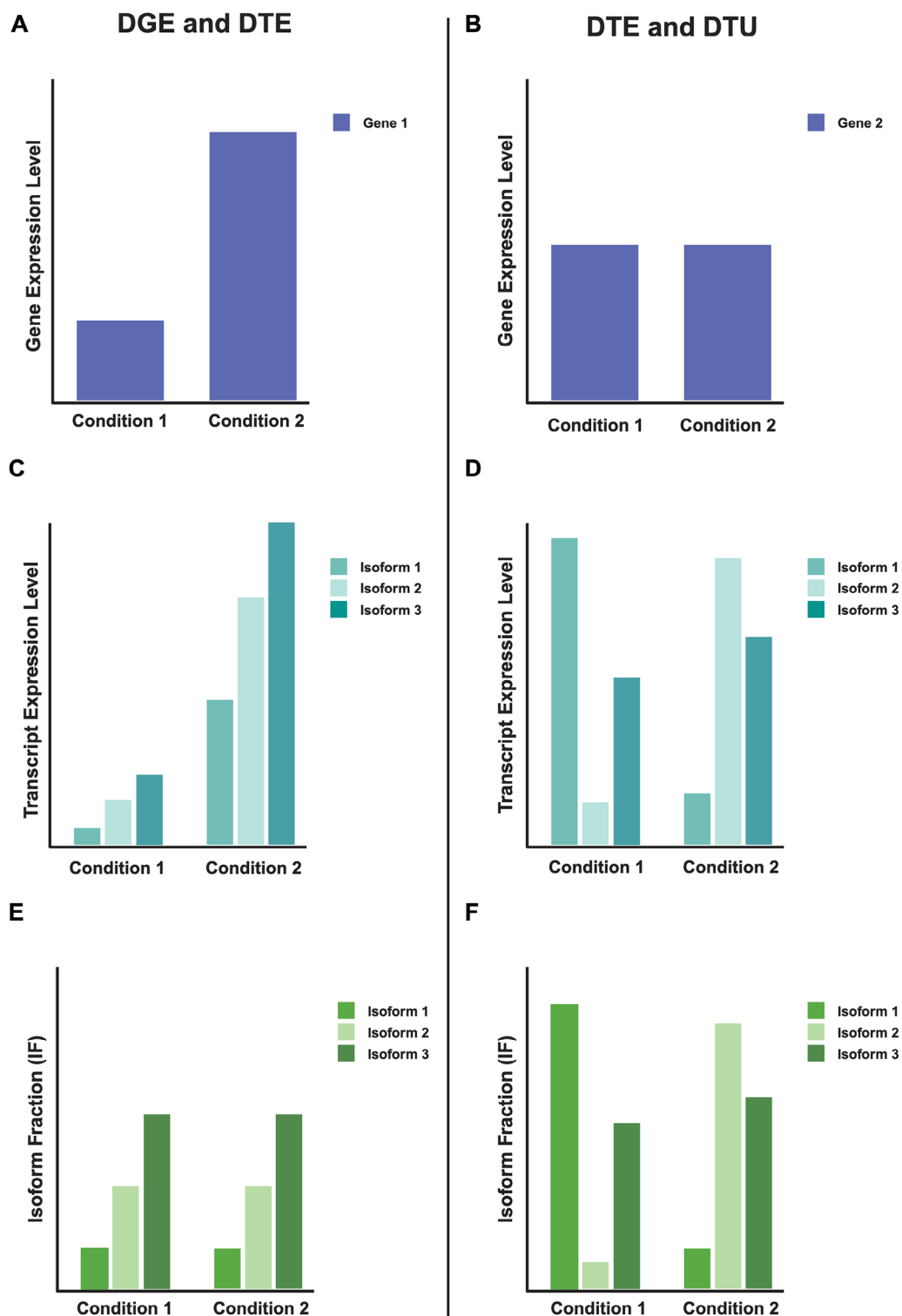
Fig. 1 Two examples of DGE, DTE, and DTU for two genes. These cartoon examples portray two genes with three transcript isoforms in two conditions. Both genes have significant DTE, but Gene 1 has no significant DTU (**A, C**, and **E**), while Gene 2 has significant DTU (**B, D**, and **F**). (**A** and **B**) Cartoon examples representing the overall gene expression: (**A**) shows down-regulation of Gene 1 in condition one compared to condition two, and (**B**) shows about equal expression of Gene 2 across both conditions. (**C** and **D**) Cartoon examples representing transcript isoform expression between the two conditions. (**E** and **F**) Cartoon examples showing isoform fraction (IF) in these two genes, where (**E**) shows no change in IF across the two conditions and, therefore, no significant DTU. (**F**) Cartoon showing significant changes in IF across conditions, revealing significant DTU. IF is calculated by the number of counts for a specific isoform divided by the total number of read counts for that gene (including all isoforms)
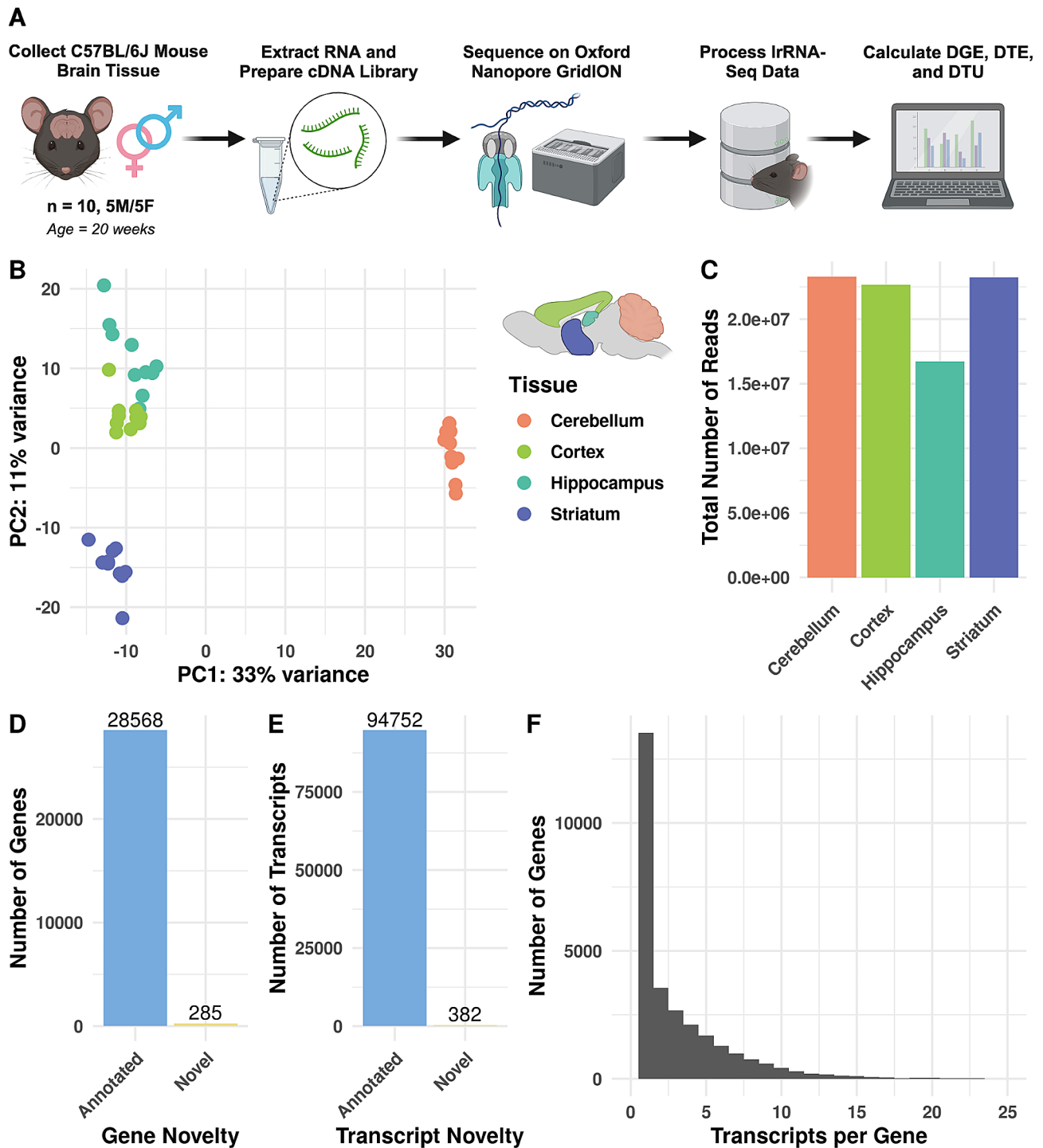
Fig. 2 Long-read Nanopore RNA sequencing across four mouse brain regions. (**A**) Overview of the study design. (**B**) PCA plot (PCs 1 and 2) of VST gene counts. Here, we colored samples by brain region. (**C**) Bar graph of the total number of long reads sequenced for each tissue. (**D** and **E**) Bar graphs of the number of novel and annotated genes and transcripts. (**F**) Histogram of the transcript counts per gene, truncated to 25 transcripts per gene. Supplementary File 2 includes the numbers of all transcripts measured for each gene

counts per million (CPM) of at least one across all samples. To test if transcript discovery thresholds contributed to the effect that novel transcripts were expressed higher, we performed Wilcoxon rank sum tests at four mean CPM thresholds (1 CPM, 2 CPM, 5 CPM, and 8 CPM - top 10%, Additional File 8 - Supplementary Fig. 1B-E), and the result was still significant each time (One-sided Wilcoxon rank sum tests, $p < 1e\text{-}11$). Therefore, on

average, novel transcripts were still expressed more than annotated transcripts. This differs from other studies that identified novel transcripts using long read data, though these studies used different transcript discovery tools [19–21], and bambu consistently discovers fewer false positives [18]. However, all genes had a mean of 3.2 transcripts, while novel genes had a mean of 1.1 transcripts, though a subset of all genes ($n=76$) had over 25 transcripts expressed (Fig. 2F, Additional File 2). Two long non-coding RNA (lncRNA) genes, *Gas5* and *Pvt1*, had the most transcripts (149 and 129, respectively). In short, we generated a lrRNA-seq dataset for four brain regions and both sexes of C57BL/6J mice, in which we identified potentially novel genes, transcripts, and patterns of gene expression variance across mouse brain regions.

Differential gene expression and differential transcript expression and usage identified across brain regions

We calculated DGE and DTE using the R package DESeq2 [22], a 'gold-standard' method that performs consistently well for DTE in long-read sequencing data [18]. We show a cartoon example representing significant DGE between two conditions in Fig. 3A. We found 8,055 (Wald test with Benjamini-Hochberg (BH) correction $p<0.05$) pairwise brain region DGE events involving 3,546 unique genes (Fig. 3B), where the cerebellum, compared to the striatum, had the most DGE ($n=2,229$, Wald test with BH correction $p<0.05$), and the cortex, compared to the hippocampus, had the least DGE ($n=349$, Wald test with BH correction $p<0.05$) (Fig. 3B). Consistent with our PCA (Fig. 2B), each brain region compared to the cerebellum had the most DGE, with 920 genes consistently differentially expressed in the cerebellum compared to the other regions (Fig. 3B). We calculated DTE for each expressed transcript, and we considered a gene to have DTE if it had at least one transcript with differential expression for that comparison (a cartoon example representing significant DTE between two conditions is in Fig. 3C). We identified 11,138 DTE events (Wald test with BH correction $p<0.05$) associated with 4,126 unique DTE genes (Fig. 3D). Unlike DGE, the greatest difference in DTE genes was between the cerebellum and cortex ($n=2,620$, Wald test with BH correction $p<0.05$), and
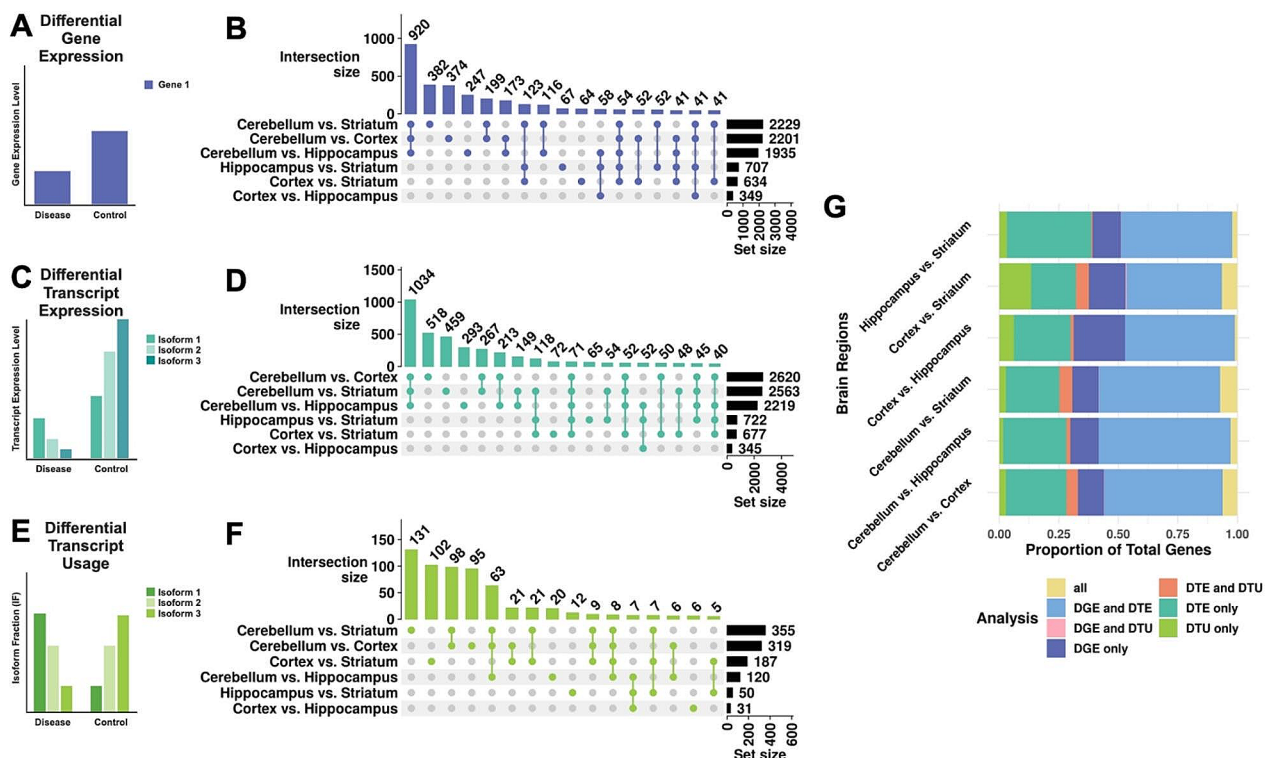

Fig. 3 DGE, DTE, and DTU across pairwise brain region comparisons. Cartoon representations of a gene with three isoforms (actual genes may have more or fewer isoforms) exemplifying (**A**) differential gene expression (DGE - violet), (**C**) differential transcript expression (DTE - turquoise), (**E**) and differential transcript usage (DTU - green). UpSet plots of the overlap of genes with (**B**) DGE (Wald test with BH correction $p<0.05$), (**D**) DTE (Wald test with BH correction $p<0.05$), and (**F**) DTU (t-test with BH correction $p<0.05$) between pairwise brain region comparisons. The bar plot above denotes intersection size, circles denote which comparisons have overlap, and the set size reflects the total number of genes with DTU for that comparison. We omitted intersections of fewer than 40 genes from the chart for legibility for panels B and D. We omitted intersections of fewer than five for legibility in panel F. (**G**) Stacked bar chart representing pairwise brain region comparison overlap across DGE, DTE, and DTU. Genes included in the chart must express at least two transcripts

the least was between the cortex and the hippocampus ($n=345$, Wald test with BH correction $p<0.05$) (Fig. 3D).

Next, we calculated DTU for each pair of brain regions using the DTU method SaTuRn [23] with the R package IsoformSwitchAnalyzeR [24]. We show a cartoon example representing significant DTU between two conditions in Fig. 3E. Here, we considered a gene to be a DTU gene if it had a t-test statistic (calculated from the log-odds ratio and variance of the quasi-binomial generalized linear model) BH-corrected p-value<0.05 for at least one of its transcripts where genes had at least two expressed transcripts (**Methods**). We analyzed DTU across brain regions and found 1,051 DTU events in 648 unique genes (Fig. 3F). The most DTU genes were in the cerebellum compared to the striatum ($n=355$, t-test with BH correction $p<0.05$), and the least were in the cortex compared to the hippocampus ($n=31$, t-test with BH correction $p<0.05$) (Fig. 3F). Consistent with our other analyses (65% for DGE and 71% for DTE), we identified the majority of DTU genes (66%) from comparisons including the cerebellum (Fig. 3B, D, F). Interestingly, the number of DTU genes ($n=63$, t-test with BH correction $p<0.05$) shared across all three comparisons including the cerebellum was a smaller percentage (10%) of the total unique DTU genes (Fig. 3F) than DGE (26%) or DTE (25%) (Fig. 3B, D), suggesting that DTU analysis is less driven by the cerebellum. We also directly compared which genes were identified for each analysis (DGE, DTE, and DTU) that expressed at least 2 transcripts and qualified for DTU analysis. We found that DGE and DTE genes had the most overlap across comparisons, with a small proportion of significant genes for each comparison identified by all three methods (Fig. 3G).

We also performed functional enrichment analysis using gprofiler2 [25] of DGE, DTE, and DTU genes for all comparisons (Additional Files 3–5). For the cortex compared to the cerebellum DGE, DTE, and DTU genes, we found enrichment (Fisher's exact test with g: SCS correction, $p<0.05$) for 1742, 2431, and 54 terms. Strikingly, we found a much larger percentage of terms associated with the neuronal synapse in DTU (24/54, 44%; e.g., synapse, glutamatergic synapse, post-synapse, synaptic signaling, neuron-to-neuron synapse, and postsynaptic membrane) compared to DGE (50/1742, 2.9%) and DTE (77/2431, 3.2%). Because a larger proportion of DTU genes were

enriched for pathways required for synaptic neurotransmission, this suggests that DTU potentially identifies biologically distinct molecular signatures from DGE and DTE. To see if any ontology terms were enriched in genes specific to DTU, we performed functional enrichment analysis on the DTU genes that did not have significant DGE or DTE. We identified multiple ontologies enriched for genes with DTU (Fisher's exact test with g: SCS correction, $p<0.05$), including histone deacetylation, TCF/WNT signaling, cytoplasmic ribosomal proteins, Kir4.1-dystrophin complex, and muscle-derived dystrobrevin-syntrophin complex (Additional File 6). The term with the most significant enrichment for DTU genes between cerebellum and cortex was histone deacetylation (adj. $p=0.014$), suggesting that these genes (*Mta1*, *Arid4b*, and *Suds3*) play an integral role in isoform-specific chromatin remodeling between the two regions (Additional File 6). Overall, a pairwise comparison of DGE, DTE, and DTU between brain regions revealed marked heterogeneity for each analysis per comparison, with a greater overlap in DGE and DTE than in either analysis with DTU. This underscores that isoform usage may be masked when only considering differential expression, hiding biologically distinct molecular signatures.

DTU sex differences are brain region-specific

Due to known sex biases in healthy brain gene expression [2] and in brain-related disease phenotypes [8, 9], we asked if there were sex-biased DGE, DTE, or DTU events by brain region. First, we measured DTU across sexes, combining brain regions, and identified four genes with DTU: *Zfp862-ps*, *Gm10605*, *Shisa5*, and *Zfp324* (t-test with BH correction $p<0.05$) (Additional File 8 - Supplementary Fig. 2). *Zfp862-ps* and *Zfp324* are a pseudogene and gene, respectively, for zinc finger proteins that contain a DNA-binding domain. While pseudogenes have traditionally been considered non-coding, they have been shown to regulate other genes and form viable proteins [26, 27]. Notably, the human ortholog of *Shisa5*, *SHISA5*, has been previously identified as having sex-biased splicing in human brain white matter [2], in line with our finding of sex-biased splicing in mouse brain regions. Finally, *Gm10605* is a predicted lncRNA gene. We did not identify any of these genes in our within-brain region analyses, suggesting that for these genes, we were underpowered to identify DTU in each region alone.

We next calculated DGE, DTE, and DTU across sexes within each brain region (Table 1; Fig. 4). We identified 23 region-specific genes with DTU by sex (analysis of deviance chi-squared test with BH correction $p<0.05$): 14 in the cortex, seven in the striatum, and two in the cerebellum (Table 1). Despite documentation of phenotypic sex differences in the hippocampus [28], we did not find sex DTU in the hippocampus (Fig. 4C). None of the

Table 1 Genes with brain-region-specific DTU across sexes

Brain Region	Genes with DTU across sexes
Cortex	*6430548M08Rik*, **Anxa7**, *Plppr2, Sel1l, Zmiz2, **Mtcl1**, Kifap3, Leprot, Bcar1, Arhgap12, Washc3, Fbxw2, Bmal1, Lmtk3*
Striatum	*Fbxo25, Dhrs4, **Rab28**, Cacnb2, Cstpp1, Rsrc1, Celf2*
Cerebellum	*Camk2d, Srgn*

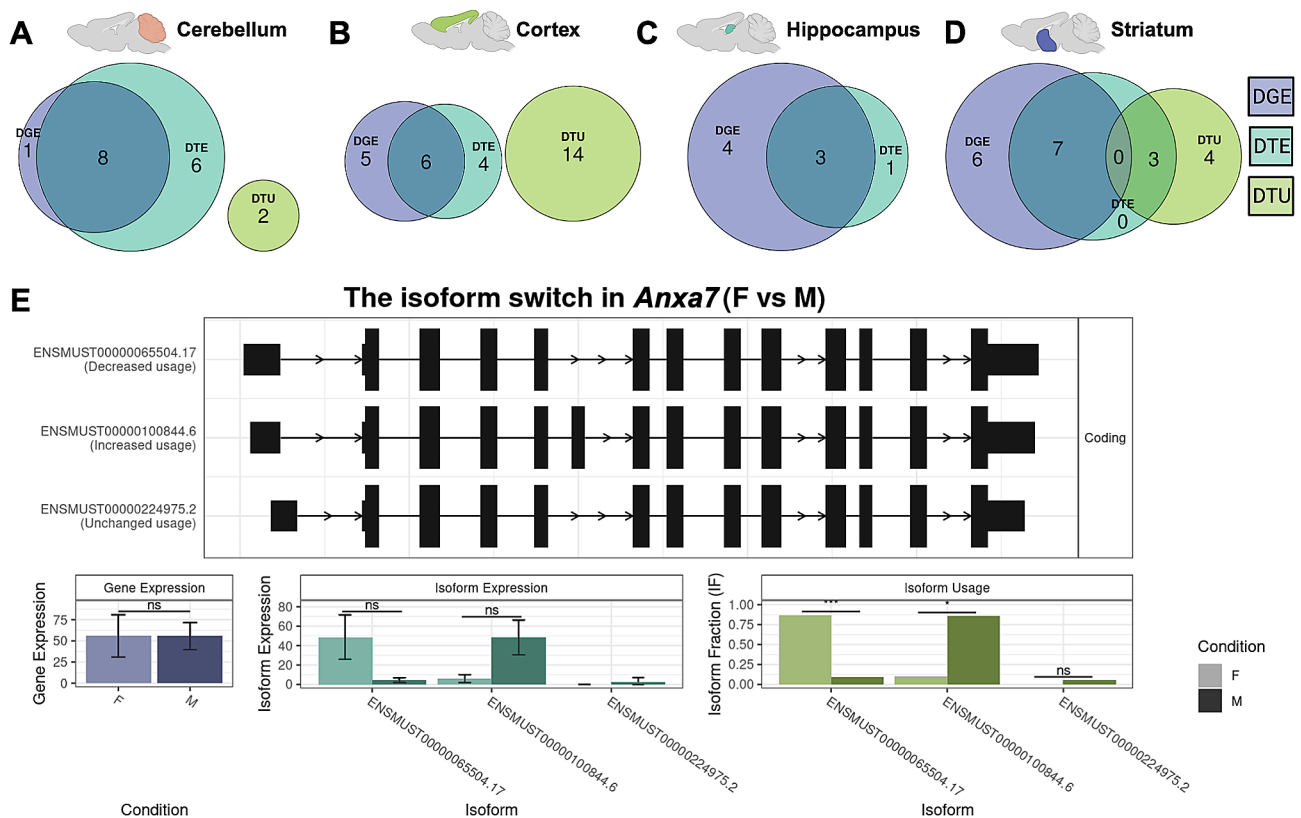Legend: Bolded genes are highlighted in the results and Figs. 4 and 5

Fig. 4 DGE, DTE, and DTU across sex within brain regions. (**A-D**) Euler diagrams represent the overlap of genes with significant DGE (Wald test with BH correction $p < 0.05$, purple), DTE (Wald test with BH correction $p < 0.05$, cyan), and DTU (analysis of deviance chi-squared test with BH correction $p < 0.05$, green). The brain regions represented are (**A**) cerebellum, (**B**) cortex, (**C**) hippocampus, and (**D**) striatum. (**E**) Switchplot displaying a transcript summary, gene expression, isoform expression, and isoform usage of the gene *Anxa7* across female (F; light color) and male (M; dark color) cerebral cortex. In the indicated comparison, ns denotes not significant, * denotes $P < 0.05$, ** denotes $P < 0.01$, and *** denotes $P < 0.001$

23 genes overlapped between brain regions, suggesting these sex differences may be brain region-specific. When we compared these DTU genes to DGE genes for each region, none overlapped (Fig. 4A-D), and only three of 23 overlapped with DTE. Therefore, by analyzing DTU, we identified 20 additional genes with differential sex effects.

We highlighted one of these sex-significant cortex DTU genes, *Anxa7*, for its many known connections to sex-associated phenotypes in humans (Fig. 4E). Human *ANXA7* is a member of the annexin family, and humans express this gene in all tissues [29]. *ANXA7* has multiple links to sex hormones; for example, *ANXA7* promoter activity is affected by estrogen and progesterone nuclear receptors [30]. In addition, patients with schizophrenia express this gene lower than healthy controls [31]. In our study, we measured three distinct *Anxa7* isoforms: ENMUST00000100844.6 (the Ensembl canonical transcript), ENMUST00000065504.7, and ENMUST00000224975.2 (Fig. 4E). When we aggregate transcript expression, *Anxa7* does not have differential gene expression between males and females in any region. However, there was DTU of ENMUST00000065504.7 and ENMUST00000100844.6 across sex (Fig. 4E). Males

expressed ENMUST00000100844.6 (the only transcript that included exon 5) higher than females. Humans have a documented clinical variant of uncertain significance (gnomAD variant 10-75143086-T-A) in the conserved male-biased exon [32]. Strikingly, 11/16 reported cases with this variant were in XY males and only 5/16 in XX females [32]. In the alternatively spliced exon 5, multiple transcription factor binding sites exist, including for FOXO1, which is strongly sex-associated and a key transcription factor associated with early pregnancy [33]. In summary, analysis of sex-significant DTU genes revealed differential isoform usage by sex within brain regions that would have otherwise been undetected by gene or transcript expression analyses, including genes with known sex-associated phenotypes.

There are two main patterns of sexually dimorphic transcript usage: sex-divergent and sex-specific

In addition, we noticed distinct patterns in sex DTU genes expressing two transcripts (Fig. 5A-B). First, we identified sex-divergent switches, i.e., sexually dimorphic transcript expression, where a single dominant transcript switch is in the opposite direction
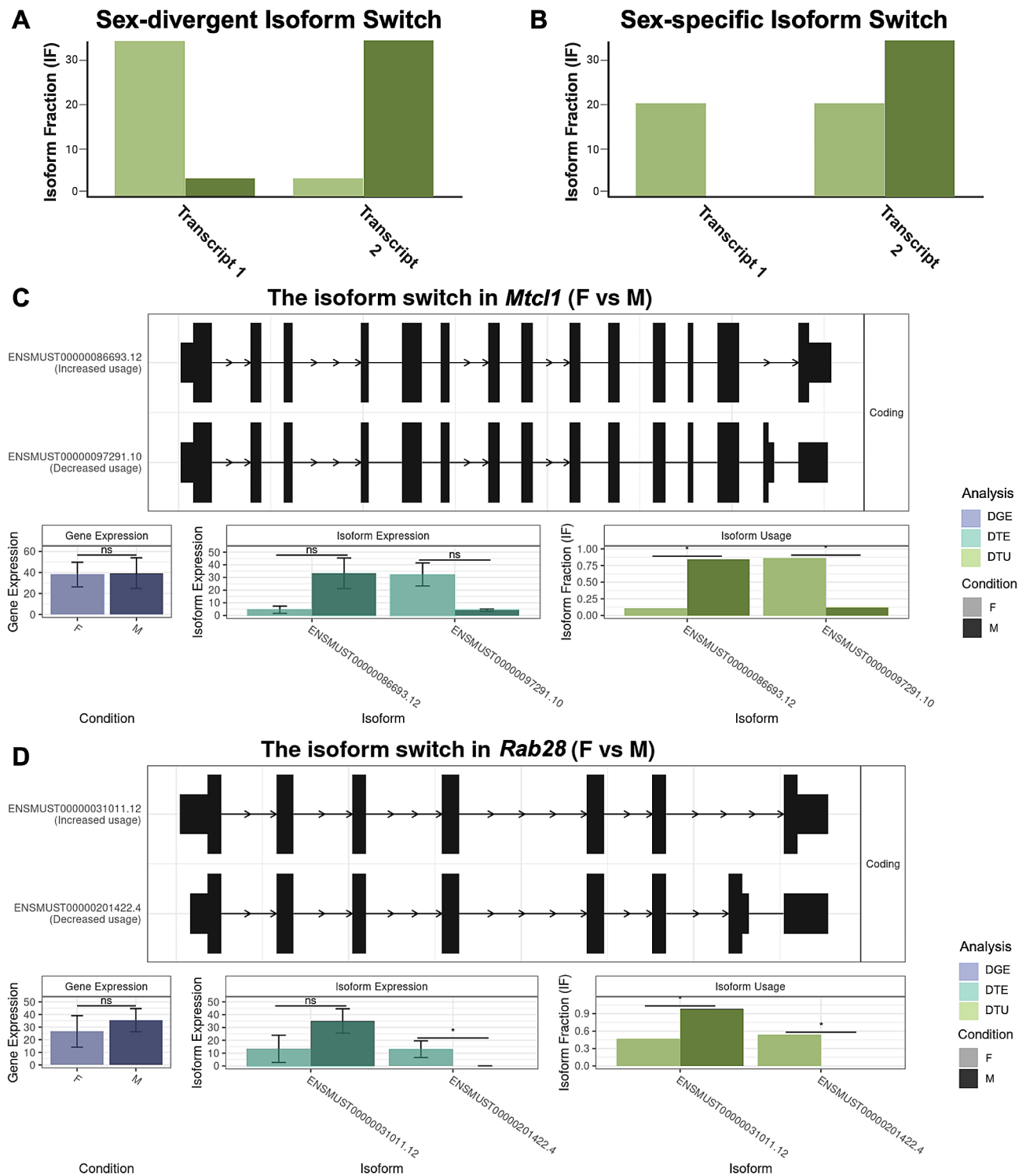
Fig. 5 Sex-divergent and sex-specific DTU. (**A-B**) Representative cartoons exemplify two transcript expression patterns of isoform switching: sex-divergent (**A**) and sex-specific (**B**). (**C**) Switchplot displaying a transcript summary, DGE (Wald test with BH correction $p < 0.05$, purple), DTE (Wald test with BH correction $p < 0.05$, cyan), and DTU (analysis of deviance chi-squared test with BH correction $p < 0.05$, green) of the sex-divergent gene *Mtcl1* in the cortex between females (F; light color) and males (M; dark color). (**D**) Switchplot displaying a transcript summary, DGE (Wald test with BH correction $p < 0.05$, purple), DTE (Wald test with BH correction $p < 0.05$, cyan), and DTU (analysis of deviance chi-squared test with BH correction $p < 0.05$, green) of the sex-specific gene *Rab28* in the striatum between females (F; light color) and males (M; dark color). Please note that these plots do not display all possible transcript structures of this gene, only the ones measured in our dataset. In the indicated comparison, ns denotes not significant and * denotes $P < 0.05$

for both sexes (Fig. 5A). We identified sex-divergent switches in *Mtcl1, Sel1l, 6430548M08Rik, Srgn*, and *Lmtk3*. For example, the sex-divergent gene *Mtcl1* has two transcripts, ENMUST00000086693.12 and ENMUST00000097291.10, where ENMUST00000086693.12 is dominant in males and ENMUST00000097291.10 in females (Fig. 5C). *Mtcl1* codes for Microtubule Crosslinking Factor 1 and is expressed highly in the cerebellum in the literature and our dataset [29]. Human *MTCL1* is known to be essential for the development of Purkinje neurons [34]. Despite its connections to the cerebellum, we only saw DTU in *Mtcl1* by sex in the cortex. We also identified sex-specific isoform switches, i.e., where one sex expresses one isoform, but the other sex had almost equal expression of both isoforms (Fig. 5B). We identified sex-specific isoform switches in *Rab28, Fbxo25, Leprot, Kifap3*, and *Plppr2. Rab28* (Fig. 5D) has a female-specific isoform, ENMUST000000201422.4, which had approximately equal expression as the other isoform, ENMUST00000031011.12, in females, while ENMUST00000031011.12 was the only isoform expressed in males. *RAB28* is an essential gene for vision, and loss of function mutations in *RAB28* cause cone-rod dystrophy in humans [35, 36]. Thus, in addition to identifying significant differences in isoform usage between sexes, we also found distinct patterns of sex DTU gene expression, with sex-significant DTU genes showing either sex-divergent or sex-specific transcript expression.

A web application for visualizing DGE, DTE, and DTU in mouse brain lrRNA-seq data

Finally, we built an R Shiny application for our data set. Users may create custom gene expression heatmaps (Fig. 6A) or examine switch plots for individual genes using the IsoformSwitchAnalyzeR package (Fig. 6B). We also provide the option for users to download the intermediate gene expression and isoform switch test result data and plots directly. Our Shiny application has been made publicly available at https://lasseignelab.shinyapps. io/mouse_brain_iso_div/.

Discussion

In summary, we produced a high-quality, publicly-available ONT lrRNA-Seq dataset across four brain regions from C57BL/6J mice, balanced for sex. We processed this data and identified 285 potentially novel genes and 382 novel transcripts, mostly (81%) associated with novel genes. We then calculated DGE, DTE, and DTU across brain regions and by sex. As expected, we identified DGE, DTE, and DTU between the four brain regions. The cerebellum had the most differences, potentially driven by cell type composition compared to the other three regions. Additionally, we found region-specific DTU

between sexes, with the most differences in DTU in the cortex. We also report two distinct patterns of sex DTU in our data: sex-divergent and sex-specific. Finally, we built a Shiny web application for researchers to explore our lrRNA-Seq results.

Our study aligns with multiple prior studies identifying changes in isoform regulation across brain regions in mice [37–39] and humans [13, 40–42]. Additionally, we found the most differences in bulk DGE in the cerebellum, which agrees with other studies examining AS across multiple brain areas [43]. For example, the gene with the highest DGE for all pairwise comparisons including the cerebellum is *Pcp2*, Purkinje cell protein 2. We suspect this reflects brain region-specific differences in cell type composition, as Purkinje neurons are unique to the cerebellum. However, confirmation of this hypothesis requires future studies at the single-cell level. Additionally, we found that some of these significant DTU genes across brain regions are known psychiatric risk genes (Additional File 7), potentially linking to region-specific differences in disease manifestation [44]. We were not surprised by the low amount of DTU we observed across sexes when we grouped all brain regions because of the variability between different brain regions' cell type compositions. Therefore, we also investigated AS across sexes within brain regions and found differences in the gene expression and transcript expression and usage of multiple brain-region-specific genes. Interestingly, we found the brain region with the most DTU by sex was the cortex, which is involved in high-level cognition. Many psychiatric phenotypes are associated with the cortex, and several of these are sex-biased in prevalence (e.g., ASD [8], schizophrenia [9], and major depressive disorder [45]). We also noticed that these DTU genes had two separate patterns of sex-significant transcript usage, either sex-divergent or sex-specific. These patterns demonstrate that while some transcripts are specific to one sex, others may shift in abundance between sexes, exemplifying nuanced sex differences. To see if these patterns could be related to epigenetic patterns of inheritance, such as genomic imprinting, we compared our list of sex DTU genes to a list of 261 known imprintied genes in mice [46], but found no overlaps. This suggests that other epigenetic mechanisms could be at play in regulating these sexually dimorphic patterns and could warrant further study.

We analyzed differences in gene expression on three fronts: DGE, DTE, and DTU, which together reveal more information on gene expression patterns by region and sex. While our work had many strengths, some limitations include using a bulk RNA-seq approach, read depth as a general limitation for transcript discovery, sample number constraints, and using mice instead of human tissues for translation to human disease. Future work
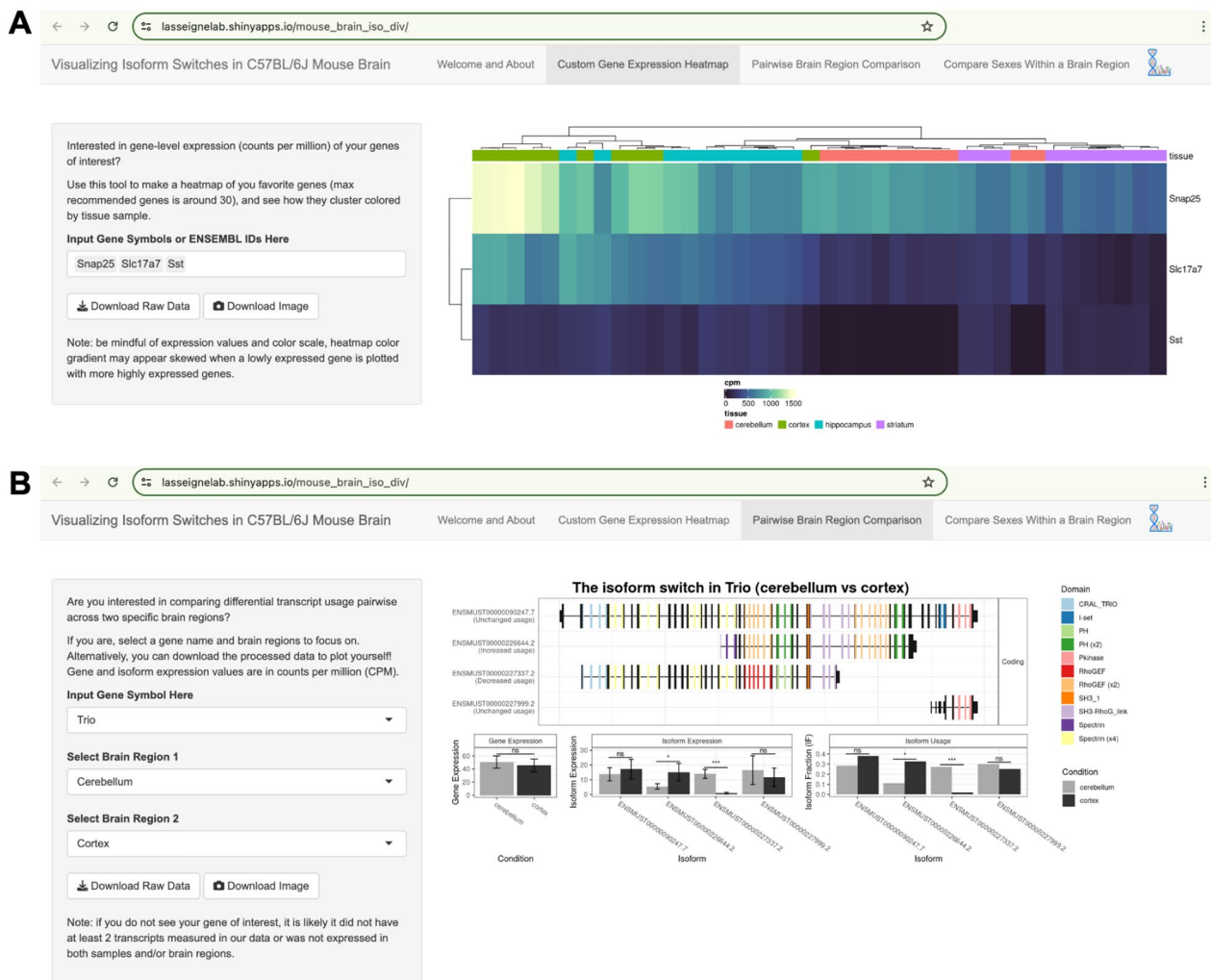
Fig. 6 Shiny app presents a user-friendly interface for exploring our mouse brain dataset. Screenshots of our web application (**A**) The "Custom Gene Expression Heatmap" lets users examine and download the gene-level expression of any gene(s) of interest in our dataset. Users can also download the expression and isoform switch test result data to analyze further or download the plots as-is. (**B**) In the "Pairwise Brain Region Comparison" tab, users can visualize their gene of interest in pairwise brain region comparisons in real-time and download expression and isoform switch test result data and plots

would benefit from single-cell resolution to determine the extent to which brain region differences stem from cell-type composition differences. Researchers could investigate this effect of cell-type composition through computational cell-type deconvolution, fluorescence-activated cell sorting (FACS), or new single-cell lrRNA-Seq methods, such as scISOr-Seq and scISO-Seq [47, 48]. While we sequenced an average of two million reads per sample and found 285 potentially novel genes and 382 novel transcripts, deeper sequencing depth may allow for greater novel isoform detection, as demonstrated by recently published Alzheimer's Disease (AD) data with extremely high-depth long-read sequencing (averaged 35.5 million reads per sample, discovered 3,394 new isoforms and 1,676 new gene bodies) [49]. We reported on novel genes and transcripts with any level of expression,

and additional work is needed for our study and others to confirm these ORFs are actually novel genes and not a result of sequencing bias or some other artifact. We attempted to reduce the number of false positives by using the stringent transcript quantification tool Bambu, which is specially designed for long-read sequencing data and found that novel transcripts were expressed more highly than annotated transcripts. We speculate that this finding suggests that long-read technologies enable the identification of additional biologically relevant transcripts at various expression levels.

Furthermore, more samples may allow greater statistical power to detect smaller expression differences approaching significance with our current sample size. Intuitively, we would expect most DTU genes to have DTE but not all DTE genes to have DTU. In our data,

this assumption was not correct. Although SatuRn and DEXSeq use transcript expression information as the basis for their DTU analyses, these inconsistencies in significance between DESeq2 and SatuRn/DEXSeq may stem from using different models to calculate statistical significance (**Methods**). Therefore, it is possible that larger sample sizes and thus increased statistical power to detect significant differences in transcript expression and usage may result in the two methods agreeing more often for genes with DTU. Additionally, while mice and humans share many genetic similarities, our findings may not be directly translatable to humans. Surprisingly, we could not detect sex differences in alternatively spliced transcripts in the hippocampus, despite known sex differences in humans with hippocampal diseases (e.g., AD [50]). This may have been due to the sample input amount, sample numbers, species, or sequencing depth. To investigate if this was due to statistical power, we examined the hippocampal genes approaching significance in our data, such as *Tsr2*. In the hippocampus, female samples expressed three transcripts of *Tsr2*, but males expressed only one transcript (Additional File 8 - Supplementary Fig. 3). We speculate the lower expression of these transcripts is why this is not significant (analysis of deviance chi-squared test with BH correction $p = 0.1652978$), but would require more follow-up data generation and analyses to confirm.

We aimed to examine and quantify differences across sexes and brain regions in C57BL/6J mouse brain tissue to better understand AS regulation. To our knowledge, this work is the first paper to use lrRNA-Seq to focus on brain-region-specific AS sex differences in a mammalian brain. We harnessed the power of lrRNA-Seq to investigate differences in AS with higher confidence than short-read and compared the results from three separate differential analyses. Here, we used novel sequencing technology to study sex as a biological variable, which is a necessary effort to resolve the long-standing practices of single-sex studies in preclinical biomedical research [12]. In addition to making all our data and code publicly available, we created an easily accessible web application for researchers to interact with the data. This research also serves as a launchpad for future directions involving additional time points, species, and disease contexts. Specifically, long-read spatial transcriptomics [51] and long-read ATAC [52] present opportunities for discerning patterns of AS and could be used to examine transcriptomic sex differences in isoform regulation at the spatial and epigenetic levels. Another future direction includes investigating classes of transcript diversity and structure (i.e., promoter usage and 3′ end choice) as done in ENCODE4 [53], but with an emphasis on studying differences across sexes in the brain. There is also a need to investigate sex differences in splicing across the lifespan,

including early development [54, 55] and aging [41]. Finally, future research could combine long-read transcriptomics with measures of neuronal activity to discern the effects of AS on signal transmission across sexes [56]. Our findings provide insight into sex differences in mammalian brains, and the data produced by this research can serve as a useful resource for the scientific community.

Materials and methods

Mouse sample collection and RNA isolation

We obtained flash-frozen hippocampus, striatum, cerebellum, and cortex C57BL/6J mouse brain tissues from The Jackson Laboratory (JAX #000664, age = 20 weeks) from five male and five female mice. The samples arrived on dry ice, and we stored them at -70 °C upon arrival. For each sample, we transferred ~ 30 mg of each brain region (or the entire brain region, in the case of hippocampus and striatum tissue) into an MP Biomedical Lysis D Matrix, 2 ml tube (#6913500) containing 500 µl of TRIzol reagent (Invitrogen #15596018) and lysed cells from each tissue on the FastPrep-24 5G bead beating grinder and lysis system (MP Biomedical #116005500). After lysis, we added 100 µl of chloroform to the tube, centrifuged at 12,000×g for 15 min, and then transferred the clear top layer of the supernatant into a fresh tube. We next added an equivolume amount of isopropanol and centrifuged at 12,000×g for 10 min. We decanted the supernatant, washed the pellet twice with 75% ethanol, and resuspended the air-dried pellet in RNAse-free water. We incubated the final RNA product with TURBO DNase (Invitrogen #AM1907) for 30 min and assessed for RNA quality using a Qubit fluorometer and Agilent Fragment Analyzer. All RNA samples had an RNA quality number (RQN) score > 7.

Oxford Nanopore Technologies lrRNA-Seq library preparation

We processed RNA samples for nanopore sequencing using the PCR-cDNA Barcoding Kit (SQK-PCB111.24) according to manufacturer instructions and prepared libraries in equimolar amounts based on fragment length and concentration to make 15 fmol of cDNA library per flow cell. Because the barcoding kit only included 24 barcodes and we had 40 samples, we prepped and pooled two batches with 20 samples each. We loaded 11 µl of each pooled library with 1 µl Rapid Adapter T (12 µl total) onto 12 R9.4 flow cells (FLO-MIN106D). Because the ONT GRIDion (GRD-MK1) sequencing device can sequence five flow cells simultaneously, we sequenced these libraries in three separate sequencing runs for 72 h each.

Nanopore settings and software versions

We ran our nanopore with active channel selection turned on, a 1.5-hour pore scan frequency, a -170 mV initial bias voltage, and a -185 mV final bias voltage. We selected to have reserved pores off with high-accuracy base calling turned on. We used the following GridION software versions: MinKNOW 22.05.7, Bream 7.1.3, Configuration 5.1.5, Guppy 6.1.5, and MinKNOW Core 5.1.0.

Raw sequencing data processing

We transferred demultiplexed FASTQ files to UAB's supercomputer cluster, Cheaha, merged FASTQs passing a minimum Phred quality score of 9 for each sample and processed using the nf-core [16] nanoseq pipeline (https://doi.org/10.5281/zenodo.1400710) with the following options: version 2.0.1, protocol cDNA, flow cell FLO-MIN106, kit SQK-PCB109, skip_basecalling, skip_demultiplexing, skip_differential_analysis, profile cheaha, and a custom configuration file specifying nanoplot version 1.32.1. The packages we used for alignment and transcript quantification in this pipeline framework were Minimap2 version 2.17 [57], samtools version 1.13 [58], and Bambu version 1.0.2 [17]. We mapped reads using the GENCODE mm39 release M31 (available at: https://www.gencodegenes.org/mouse/) primary assembly genome and annotation. We retrieved transcript counts from the Bambu outputs of the nextflow results for further analysis.

Data normalization

We processed and normalized data in R version 4.3.0 and RStudio version 2023.06.2+561. Because nanopore read lengths vary depending on the input cDNA length, we normalized by counts per million (CPM) instead of transcripts per million (TPM) since Bambu already accounts for length in its expression abundances. We calculated CPM by multiplying the number of read counts by 1 million and dividing by the sum of the total read counts for that sample. We found no outliers or batch effects by visual inspection when we performed principal component analysis (PCA).

Differential gene and transcript expression analysis

For DGE and DTE analysis, we used the R package DESeq2 version 1.40.0 [22] using the negative binomial Wald test function. We considered a differentially expressed gene or transcript significance with a BH-adjusted p-value of less than 0.05 and an absolute log2 fold change>1.5 value. Therefore, we used three models:

1. Region compared to another region (e.g., the cerebellum directly compared to the cortex).
2. Sex within a region (e.g., female compared to male in the cerebellum).

3. Sex across all regions (e.g., female compared to male).

We performed this analysis with gene-level counts for differential gene expression (DGE) and again with transcript-level counts for differential transcript expression (DTE). We then incorporated these results into the IsofrmSwitchAnalyseR switchList format for downstream plotting.

Differential transcript usage analysis

We performed Differential Transcript Usage (DTU) analysis with the R package IsoformSwitchAnalyzeR package version 1.99.17 [24], using the satuRn version 1.8.0 [23] algorithm, and within brain regions, the DEXSeq version 1.46.0 [59] algorithm. We chose to use DEXSeq for smaller sample sizes ($n=5$) because of its increased detection ability. Still, we did not use it for larger sample groups because it has a higher false discovery rate [18] and is computationally inefficient [23]. Therefore, we used three models:

1. Region compared to another region (e.g., cerebellum compared to cortex) (satuRn).
2. Sex within a region (e.g., female compared to male in cerebellum) (DEXSeq).
3. Sex across all regions (e.g., female compared to male) (satuRn).

First, we created a switchAnalyzeRlist object with the importRdata function. We used the raw counts from Bambu [60] for the count matrix. For normalized isoform abundance values, we calculated CPM as described above. We used the IsoformSwitchAnalyzeR [24] package to remove genes that do not have more than one transcript and no gene expression minimum and proceeded with the satuRn [23] or DEXSeq [59] isoform switch tests. The satuRn isoform switch test uses a quasi-binomial generalized linear model to model transcript usage and calculates the posterior variance using an empirical Bayes procedure [23]. Using this model, satuRn runs a t-test based on the model's log-odds ratio estimates with the posterior variance and uses BH correction to reduce FDR [23]. The DEXSeq isoform switch test uses a binomial generalized linear model and analyzes deviance for each "counting bin" based on a chi-squared likelihood ratio test (59). The IsoformSwitchAnalyzeR implementation of DEXSeq differs from other implementations of DEXSeq in that it uses full transcripts as the "counting bins" instead of exons so that it can detect DTU instead of only differential exon usage [24]. Our significance filtering thresholds were an isoform switch q value<0.05 and a differential isoform fraction (dIF) with an absolute value of at least 0.1, reflecting at least 10% change

in isoform fraction across conditions. We calculated IF values as the isoform expression divided by total gene expression.

Functional enrichment analysis

To infer pathways and diseases associated with the identified lists of significant genes with DGE/DTE/DTU, we performed a statistical enrichment analysis using gprofiler2 version 0.2.1 [25] with a custom set of background genes that passed filtering criteria (genes must have more than one transcript and be present in both conditions). We used the g: GOSt function, which uses a one-tailed Fisher's exact test to obtain statistical probabilities for each term, and the g: SCS method for multiple testing correction. The default data sources for the gprofiler2 g: GOSt function include Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, Transfac, mirTarBase, CORUM, Human Protein Atlas (HPA), and Human Phenotype Ontology (HPO). We then saved the results in Additional Files 3–5 and plotted these results, which passed our p-value threshold of <0.05 for each comparison. When we compared the proportions of synaptic enrichment terms across analyses, we returned the number of terms that included the character string "synap". We divided it by the total terms overall for that analysis.

Comparison of DGE, DTE, and DTU

After determining which genes had DGE, DTE, and DTU for each condition tested, we created Euler diagrams and UpSet plots using eulerr version 7.0.0 and ComplexHeatmap version 2.16.0 [61] packages, respectively, to visualize the overlap between these conditions. We identified genes with DTE by taking the unique list of gene IDs paired with transcripts identified as differentially expressed (adj $p < 0.05$) from DESeq2, where we only counted a gene with DTE in multiple transcripts once.

Neurological disease phenotype gene sets

We compared three main gene lists to our significant DTU gene lists to known neurological disease risk genes. First, we compared against a recent set of AD risk genes [62]. Next, we compared against multi-disorder psychiatric risk genes from the Cross-Disorder Group of the Psychiatric Genomics Consortium [44]. We listed psychiatric disorders if they have a posterior probability of association of above 0.9. Finally, we also compared active cases in UAB's Center for Precision Animal Modeling (C-PAM).

To facilitate conversion between mouse and human genes, we converted the human neurological gene lists into mouse genes using the biomaRt Bioconductor package [63] in R. We then identified genes that were present in both DTU lists and neurological gene lists and reported them in Additional File 7.

Protein domain family analysis

Following the package framework from the IsoformSwitchAnalyzeR package version 1.99.17, we extracted nucleotide and amino acid sequences from each gene's open reading frame (ORF). Using those amino acid sequences as input, we ran the pfamscan.pl perl script with Perl 5 version 34 obtained from ftp://ftp.ebi.ac.uk/pub/databases/Pfam/ to identify known protein domains from the Protein family database (Pfam) [64]. We incorporated these outputs into our R objects, and users can visualize select genes using our Shiny app.

Abbreviations

AS	Alternative splicing
AD	Alzheimer's Disease
ASD	Autism spectrum disorder
BH	Benjamini-Hochberg
CPM	Counts per million
dIF	Differential isoform fraction
DGE	Differential gene expression
DTE	Differential transcript expression
DTU	Differential transcript usage
GEO	Gene Expression Omnibus
lncRNA	Long non-coding RNA
lrRNA-Seq	Long-read RNA sequencing
ONT	Oxford Nanopore Technologies
ORF	Open reading frame
PCA	Principal component analysis
RQN	RNA quality number
SRA	Sequence Read Archive
TPM	Transcripts per million

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13041-024-01112-7.

Additional File 1: Novel transcripts information. This supplementary file is a CSV with general information about the novel transcripts. Columns as as follows: Location - which chromosome we found this gene on. Start - the genomic start position of the novel gene. Stop - the genomic end position of the novel gene. Strand - the direction of transcription for the novel gene. Gene_id - the gene identification number automatically assigned by Bambu. Transcript_id - the transcript identification number automatically assigned by Bambu.

Additional file 2. Transcripts per gene table. This supplementary file is a CSV with the number of all transcripts measured for each gene. Columns are as follows: GENEID - Either ENSEMBL gene identification number, if available, or Bambu assigned identification number. Transcript count - number of transcripts counted per gene with more than 0 counts.

Additional file 3. DGE Functional enrichment results. This supplementary file is a Microsoft Excel file with gprofiler results of DGE genes with a sheet for each comparison. Columns are as follows: Query - all results were processed as individual queries. Significant - all terms in this table were kept if they had a significance of below 0.05. P_value - p-value from Fisher's one-tailed test. Term_size - number of genes in this term size. Query_size - number of genes for this specific query. Intersection_size - number of genes in the intersection between term and query. Precision - statistical precision for this term. Recall - statistical recall for this term. Term_id - Identification number for this term. Source - Data source for this term. Term_name - Name for this term. Effective_domain_size - Domain size for this term. Source_order - Order for this term. Parents - Any parent terms for this term.

Additional file 4. DTE Functional enrichment results. This supplementary file is a Microsoft Excel file with gprofiler results of DTE genes with a sheet for each comparison. Columns are as follows: Query - all results were processed as individual queries. Significant - all terms in this table were kept if they had a significance of below 0.05. *P*_value - *p*-value from Fisher's one-tailed test. Term_size - number of genes in this term size. Query_size - number of genes for this specific query. Intersection_size - number of genes in the intersection between term and query. Precision - statistical precision for this term. Recall - statistical recall for this term. Term_id - Identification number for this term. Source - Data source for this term. Term_name - Name for this term. Effective_domain_size - Domain size for this term. Source_order - Order for this term. Parents - Any parent terms for this term.

Additional file 5. DTU Functional enrichment results. This supplementary file is a Microsoft Excel file with gprofiler results of all DTU genes with a sheet for each comparison. Columns are as follows: Query - all results were processed as individual queries. Significant - all terms in this table were kept if they had a significance of below 0.05. *P*_value - *p*-value from Fisher's one-tailed test. Term_size - number of genes in this term size. Query_size - number of genes for this specific query. Intersection_size - number of genes in the intersection between term and query. Precision - statistical precision for this term. Recall - statistical recall for this term. Term_id - Identification number for this term. Source - Data source for this term. Term_name - Name for this term. Effective_domain_size - Domain size for this term. Source_order - Order for this term. Parents - Any parent terms for this term.

Additional file 6. DTU-specific Functional enrichment results. This supplementary file is a Microsoft Excel file with gprofiler results of DTU-specifc genes with a sheet for each comparison. Columns are as follows: Query - all results were processed as individual queries. Significant - all terms in this table were kept if they had a significance of below 0.05. *P*_value - *p*-value from Fisher's one-tailed test. Term_size - number of genes in this term size. Query_size - number of genes for this specific query. Intersection_size - number of genes in the intersection between term and query. Precision - statistical precision for this term. Recall - statistical recall for this term. Term_id - Identification number for this term. Source - Data source for this term. Term_name - Name for this term. Effective_domain_size - Domain size for this term. Source_order - Order for this term. Parents - Any parent terms for this term.

Additional file 7. DTU genes that are known neurological disease risk genes. This supplementary file is an Excel file of neurological disease risk genes and which DTU design we used to identify them. We compared significant DTU genes to lists of AD, cross-psychiatric disorder, and UAB's Center for Precision Animal Modeling case genes (C-PAM) to identify disease genes.

Additional file 8. Supplementary Figures. This PDF contains supplementary figures and their figure legends.

Authors' contributions
EFJ and BNL conceptualized the project. EFJ, TCH, and VLF collected and generated the sequencing data set. All analyses were coded and performed by EFJ. EFJ developed and deployed the web application. TCH, VLF, and ADC reviewed and validated the code. BNL and TCH provided supervision and project administration. BNL acquired funding. EFJ wrote the first draft. EFJ, TCH, VLF, ADC, and BNL reviewed and edited the manuscript. All authors read and approved the final manuscript.

Data availability
The raw dataset supporting the conclusions of this article is available in the Gene Expression Omnibus (GEO) repository, with accession number GSE246705, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE246705. The data is also fully available on Sequence Read Archive (SRA) with accession number SRP469534 and BioProject with accession number PRJNA1034151.
The docker images, intermediate datasets, and code to reproduce all analyses and results in this article are available in the following Zenodo repositories: Docker images - https://zenodo.org/records/10480924, intermediate data -https://zenodo.org/records/10381745, GitHub code - https://zenodo.org/records/10481313.
The code supporting the conclusions and for reproducing analyses of this article is available in the GitHub repository, https://github.com/lasseignelab/230227_EJ_MouseBrainIsoDiv.
The interactive web browser application associated with this manuscript is available at https://lasseignelab.shinyapps.io/mouse_brain_iso_div/.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

References
1. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, et al. The evolutionary landscape of alternative splicing in vertebrate species. Science. 2012;338(6114):1587–93.
2. Trabzuni D, Ramasamy A, Imran S, Walker R, Smith C, Weale ME, et al. Widespread sex differences in gene expression and splicing in the adult human brain. Nat Commun. 2013;4:2771.
3. Xu Q, Modrek B, Lee C. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. Nucleic Acids Res. 2002;30(17):3754–66.
4. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456(7221):470–6.
5. Zhang X, Chen MH, Wu X, Kodani A, Fan J, Doan R, et al. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. Cell. 2016;166(5):1147–e6215.
6. Licatalosi DD, Darnell RB. Splicing regulation in neurologic disease. Neuron. 2006;52(1):93–101.
7. Gandal MJ, Zhang P, Hadjimichael E, Walker RL, Chen C, Liu S et al. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science [Internet]. 2018;362(6420). https://doi.org/10.1126/science.aat8127.
8. Werling DM, Geschwind DH. Sex differences in autism spectrum disorders. Curr Opin Neurol. 2013;26(2):146–53.
9. Ochoa S, Usall J, Cobo J, Labad X, Kulkarni J. Gender differences in schizophrenia and first-episode psychosis: a comprehensive literature review. Schizophr Res Treat. 2012;2012:916198.
10. Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell. 2014;159(7):1511–23.
11. Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature. 2016;540(7633):423–7.

12. Arnegard ME, Whitten LA, Hunter C, Clayton JA. Sex as a Biological Variable: a 5-Year Progress Report and call to action. J Womens Health. 2020;29(6):858–64.

13. Clark MB, Wrzesinski T, Garcia AB, Hall NAL, Kleinman JE, Hyde T, et al. Long-read sequencing reveals the complex splicing profile of the psychiatric risk gene CACNA1C in human brain. Mol Psychiatry. 2020;25(1):37–47.

14. Jones EF, Haldar A, Oza VH, Lasseigne BN. Quantifying transcriptome diversity: a review. Brief Funct Genomics [Internet]. 2023; https://doi.org/10.1093/bfgp/elad019.

15. Erdogdu B, Varabyou A, Hicks SC, Salzberg SL, Pertea M. Detecting differential transcript usage in complex diseases with SPIT [Internet]. bioRxiv. 2023 [cited 2023 Nov 9]. p. 2023.07.10.548289. https://www.biorxiv.org/content/https://doi.org/10.1101/2023.07.10.548289v1.full.

16. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020;38(3):276–8.

17. Chen Y, Sim A, Wan YK, Yeo K, Lee JJX, Ling MH et al. Context-aware transcript quantification from long-read RNA-seq data with Bambu. Nat Methods [Internet]. 2023; https://doi.org/10.1038/s41592-023-01908-w.

18. Dong X, Du MRM, Gouil Q, Tian L, Jabbari JS, Bowden R et al. Benchmarking long-read RNA-sequencing analysis tools using in silico mixtures. Nat Methods [Internet]. 2023; https://doi.org/10.1038/s41592-023-02026-3.

19. Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, Del Risco H, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. 2018;28(3):396–411.

20. Sun Q, Han Y, He J, Wang J, Ma X, Ning Q, et al. Long-read sequencing reveals the landscape of aberrant alternative splicing and novel therapeutic target in colorectal cancer. Genome Med. 2023;15(1):76.

21. Kiyose H, Nakagawa H, Ono A, Aikata H, Ueno M, Hayami S, et al. Comprehensive analysis of full-length transcripts reveals novel splicing abnormalities and oncogenic transcripts in liver cancer. PLoS Genet. 2022;18(8):e1010342.

22. Love MI, Huber W, Anders S. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

23. Gilis J, Vitting-Seerup K, Van den Berge K, Clement L, satuRn. Scalable analysis of differential transcript usage for bulk and single-cell RNA-sequencing applications. F1000Res. 2021;10(374):374.

24. Vitting-Seerup K, Sandelin A. IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. Bioinformatics. 2019;35(21):4469–71.

25. Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res [Internet]. 2020;9. https://doi.org/10.12688/f1000research.24956.2.

26. St-Germain J, Khan MR, Bavykina V, Desmarais R, Scott M, Boissonneault G et al. Functional Characterization of a Phf8 Processed Pseudogene in the Mouse Genome. Genes [Internet]. 2023;14(1). https://doi.org/10.3390/genes14010172.

27. Li W, Yang W, Wang XJ. Pseudogenes: pseudo or real functional elements? J Genet Genomics. 2013;40(4):171–7.

28. Yagi S, Galea LAM. Sex differences in hippocampal cognition and neurogenesis. Neuropsychopharmacology. 2019;44(1):200–13.

29. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. Science. 2015;348(6235):660–5.

30. Torosyan Y, Dobi A, Naga S, Mezhevaya K, Glasman M, Norris C, et al. Distinct effects of annexin A7 and p53 on arachidonate lipoxygenation in prostate cancer cells involve 5-lipoxygenase transcription. Cancer Res. 2006;66(19):9609–16.

31. Liu CM, Fann CSJ, Chen CY, Liu YL, Oyang YJ, Yang WC, et al. ANXA7, PPP3CB, DNAJC9, and ZMYND17 genes at chromosome 10q22 associated with the subgroup of schizophrenia with deficits in attention and executive function. Biol Psychiatry. 2011;70(1):51–8.

32. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q et al. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes [Internet]. bioRxiv. 2022 [cited 2023 Nov 8]. p. 2022.03.20.485034. https://www.biorxiv.org/content/https://doi.org/10.1101/2022.03.20.485034v2.

33. Adiguzel D, Celik-Ozenci C. FoxO1 is a cell-specific core transcription factor for endometrial remodeling and homeostasis during menstrual cycle and early pregnancy. Hum Reprod Update. 2021;27(3):570–83.

34. Satake T, Yamashita K, Hayashi K, Miyatake S, Tamura-Nakano M, Doi H, et al. MTCL1 plays an essential role in maintaining Purkinje neuron axon initial segment. EMBO J. 2017;36(9):1227–42.

35. Roosing S, Rohrschneider K, Beryozkin A, Sharon D, Weisschuh N, Staller J, et al. Mutations in RAB28, encoding a farnesylated small GTPase, are associated with autosomal-recessive cone-rod dystrophy. Am J Hum Genet. 2013;93(1):110–7.

36. Riveiro-Álvarez R, Xie YA, López-Martínez MÁ, Gambin T, Pérez-Carro R, Ávila-Fernández A, et al. New mutations in the RAB28 gene in 2 Spanish families with cone-rod dystrophy. JAMA Ophthalmol. 2015;133(2):133–9.

37. Joglekar A, Hu W, Zhang B, Narykov O, Diekhans M, Balacco J et al. Single-cell long-read mRNA isoform regulation is pervasive across mammalian brain regions, cell types, and development [Internet]. bioRxiv. 2023 [cited 2023 Apr 18]. p. 2023.04.02.535281. https://www.biorxiv.org/content/https://doi.org/10.1101/2023.04.02.535281v1.full.

38. Vaquero-Garcia J, Barrera A, Gazzara MR, González-Vallinas J, Lahens NF, Hogenesch JB, et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. Elife. 2016;5:e11752.

39. McMillan P, Korvatska E, Poorkaj P, Evstafjeva Z, Robinson L, Greenup L, et al. Tau isoform regulation is region- and cell-specific in mouse brain. J Comp Neurol. 2008;511(6):788–803.

40. Twine NA, Janitz K, Wilkins MR, Janitz M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. PLoS ONE. 2011;6(1):e16266.

41. Mazin P, Xiong J, Liu X, Yan Z, Zhang X, Li M, et al. Widespread splicing changes in human brain development and aging. Mol Syst Biol. 2013;9:633.

42. Zhang Y, Yang HT, Kadash-Edmondson K, Pan Y, Pan Z, Davidson BL, et al. Regional Variation of Splicing QTLs in human brain. Am J Hum Genet. 2020;107(2):196–210.

43. Chappell S, Patel T, Guetta-Baranes T, Sang F, Francis PT, Morgan K, et al. Observations of extensive gene expression differences in the cerebellum and potential relevance to Alzheimer's disease. BMC Res Notes. 2018;11(1):646.

44. Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mgh.harvard.edu, Cross-disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight Psychiatric disorders. Cell. 2019;179(7):1469–e8211.

45. Charlson FJ, Ferrari AJ, Santomauro DF, Diminic S, Stockings E, Scott JG, et al. Global Epidemiology and burden of Schizophrenia: findings from the global burden of Disease Study 2016. Schizophr Bull. 2018;44(6):1195–203.

46. Tucci V, Isles AR, Kelsey G, Ferguson-Smith AC. Erice Imprinting Group. Genomic imprinting and physiological processes in mammals. Cell. 2019;176(5):952–65.

47. Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nat Biotechnol [Internet]. 2018; https://doi.org/10.1038/nbt.4259.

48. Yang Y, Yang R, Kang B, Qian S, He X, Zhang X. Single-cell long-read sequencing in human cerebral organoids uncovers cell-type-specific and autism-associated exons. Cell Rep. 2023;42(11):113335.

49. Aguzzoli Heberle B, Brandon JA, Page ML, Nations KA, Dikobe KI, White BJ et al. Using deep long-read RNAseq in Alzheimer's disease brain to assess clinical relevance of RNA isoform diversity. bioRxiv [Internet]. 2023; https://doi.org/10.1101/2023.08.06.552162.

50. Barnes LL, Wilson RS, Bienias JL, Schneider JA, Evans DA, Bennett DA. Sex differences in the clinical manifestations of Alzheimer disease pathology. Arch Gen Psychiatry. 2005;62(6):685–91.

51. Boileau E, Li X, Naarmann-de Vries IS, Becker C, Casper R, Altmüller J, et al. Full-length spatial transcriptomics reveals the unexplored isoform diversity of the myocardium Post-MI. Front Genet. 2022;13:912572.

52. Hu Y, Jiang Z, Chen K, Zhou Z, Zhou X, Wang Y, et al. scNanoATAC-seq: a long-read single-cell ATAC sequencing method to detect chromatin accessibility and genetic variants simultaneously within an individual cell. Cell Res. 2023;33(1):83–6.

53. Reese F, Williams B, Balderrama-Gutierrez G, Wyman D, Çelik MH, Rebboah E et al. The ENCODE4 long-read RNA-seq collection reveals distinct classes of transcript structure diversity [Internet]. bioRxiv. 2023 [cited 2023 May 24]. p. 2023.05.15.540865. https://www.biorxiv.org/content/https://doi.org/10.1101/2023.05.15.540865v1.

54. Patoway A, Zhang P, Jops C, Vuong CK, Ge X, Hou K et al. Cell-type-specificity of isoform diversity in the developing human neocortex informs mechanisms of neurodevelopmental disorders [Internet]. bioRxiv. 2023 [cited 2023 Apr 18]. p. 2023.03.25.534016. https://www.biorxiv.org/content/https://doi.org/10.1101/2023.03.25.534016v2.

55. Torre D, Francoeur NJ, Kalma Y, Gross Carmel I, Melo BS, Deikus G, et al. Isoform-resolved transcriptome of the human preimplantation embryo. Nat Commun. 2023;14(1):6902.

56. Denkena J, Zaisser A, Merz B, Klinger B, Kuhl D, Blüthgen N, et al. Neuronal activity regulates alternative exon usage. Mol Brain. 2020;13(1):148.

57. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.

58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.

59. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res. 2012;22(10):2008–17.

60. Chen Y, Sim A, Wan Y, Goeke J, bambu. Reference-guided isoform reconstruction and quantification for long read RNA-Seq data [Internet]. 2022. https://github.com/GoekeLab/bambu.

61. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847–9.

62. Bellenguez C, Küçükali F, Jansen IE, Kleineidam L, Moreno-Grau S, Amin N, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet. 2022;54(4):412–36.

63. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 2009;4(8):1184–91.

64. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, et al. The pfam protein families database. Nucleic Acids Res. 2012;40(Database issue):D290–301.

Publisher's Note