Published in final edited form as:

Nat Neurosci. 2013 April ; 16(4): 499–506. doi:10.1038/nn.3332.

Messenger RNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex

Allissa A. Dillman[1,2], **David N. Hauser**[1,3], **J. Raphael Gibbs**[1], **Michael A. Nalls**[1], **Melissa K. McCoy**[1], **Iakov N. Rudenko**[1], **Dagmar Galter**[2], and **Mark R. Cookson**[1,CA]

[1]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

[2]Department of Neuroscience, Karolinska Institutet, 171 77 Stockholm, Sweden

[3]Brown University/National Institutes of Health Graduate Partnership Program, Department of Neuroscience, Brown University, Providence, Rhode Island, USA

Abstract

The complexity of the adult brain is a result of both developmental processes and experience-dependent circuit formation. One way to look at the differences between embryonic and adult brain is to examine gene expression. Previous studies have used microarrays to address this in a global manner. However, the transcriptome is more complex than gene expression levels alone, as alternative splicing and RNA editing generate a diverse set of mature transcripts. Here, we developed a high-resolution transcriptome dataset of mouse cerebral cortex at embryonic and adult stages using RNA sequencing (RNA-Seq). We found many differences in gene expression, splicing and RNA editing between embryonic and adult cerebral cortex. Each dataset was validated technically and biologically, and in each case we found our RNA-Seq observations to have predictive validity. We propose this dataset and analysis to be a helpful resource for understanding gene expression in the embryonic and adult cerebral cortex.

Neurons are produced during embryogenesis then migrate and form circuits during development and into adulthood. This circuit formation and optimization is critical for the fine-tuning of behaviors and is necessary for information processing in the adult brain [1,2]. Of the many molecular events during brain development, gene expression levels are known to undergo substantial changes. Using microarrays, it is estimated that approximately 17% of

[CA]Correspondence to: Mark R. Cookson, Cell Biology and Gene Expression Section, Laboratory of Neurogenetics, National Institute on Aging, 35 Convent Drive, Bethesda MD 20892-3707, USA. Phone: +1 301 451 3870. Fax: +1 301 451 7295. cookson@mail.nih.gov.

Author Contributions

A.D. performed the RNA-Seq experiments and analyzed the data. J.R.G. and M.A.N. provided additional analytical approaches. D.H., M.K.M. and I.N.R. performed mouse dissections and contributed additional validation results. M.R.C. and D.G. supervised the project. M.R.C. and A.D. wrote the paper with contributions from all authors.

Accession codes

Raw data is available at Gene Expression Omnibus, series GSE39866

assayed genes are differentially expressed between embryonic day 16 (E16) and postnatal day 30 (P30)[3].

However, gene abundance represents only part of the complexity of the transcriptome. Specific exons are included or excluded from mature mRNAs by alternative splicing, which adds proteomic diversity by generating unique isoforms from the same gene. Alternative splicing also influences the regulation of transcript expression by changing the stability of the mRNA, affecting the efficiency of translation, altering the number of miRNA sites or switching localization signals [4,5]. Splicing plays an important role in neurodevelopment and synaptic plasticity and strength[6]. For instance, isoforms of *Dscam* mediate axon guidance[7], while neurexin splice variants are involved in synaptogenesis[8].

Additional complexity is generated at the single base level via RNA editing by adenosine deaminase (ADAR), which convert adenosine to inosine (A to I editing)[9], or APOBEC1, which replaces cytosine with uracil[10]. These base modifications can result in a change in the amino acid sequence or differences in splicing or nuclear retention of the transcript[11-13]. Of the three known genes encoding ADARs, *Adar* and *Adar1b* are ubiquitously expressed, but their expression levels are highest in the brain. *Adarb2* is exclusively expressed in the brain[9]. Editing creates multiple isoforms of neurotransmitter receptors, including AMPA-subtype glutamate channels such as *Gria2*. Mice unable to edit *Gria2* die of seizures shortly after birth[14] and many other examples of A to I editing change the properties of proteins important in neuronal function[15]. There are substantial differences in the proportion of edited RNA between adult and embryonic brain tissue for specific genes[16].

Although expression, splicing, and editing all occur in the brain and are regulated during development and aging, few studies have examined these at a genome wide scale and in concert. This means that while the adult brain may be more complicated at a circuit level, we do not have an unbiased view of whether the adult brain transcriptome is inherently more complex, or simply different from, the embryonic brain.

Here, we used RNA sequencing (RNA-Seq) to develop a high-resolution transcriptome data set of the embryonic and adult brain mouse cerebral cortex. Compared to microarray technologies, RNA-Seq has an improved dynamic range in estimates of gene expression levels and better precision[17] and allows for estimation of exon-specific and single base pair events. We used the rich literature of developmental expression as an estimation of the accuracy of results and performed technical and biological validation of various levels of gene expression changes. We show that there are large numbers of differences in gene expression levels and exon utilization as the brain develops as well as a clear tendency for more complete editing in the adult brain.

RESULTS

Overall quality parameters of the RNA-Seq dataset

We used RNA-Seq to measure the transcriptome of three adult (3-4 months old) female mice and four embryonic day 17 (E17) female mice. We chose the inbred C57Bl/6J strain used for the mouse genome (http://www.ncbi.nlm.nih.gov/assembly/165668/) to minimize

genetic heterogeneity as genetic background effects can influence gene expression[18] and might complicate interpretation of single base pair sequences.

We generated ~30million reads with high quality scores (Supplementary Fig. 1) and mapped 75-80% of those reads to the mouse genome (Supplementary Table 1) equivalently for all samples. Values for base mean, i.e. sequencing depth for each transcript normalized to library size [19], used an estimate of the overall expression levels, varied over a wide range, reflecting the dynamic nature of RNA expression (Supplementary Figs. 2 and 3). Because the libraries used were based on poly-adenylated RNA, we were concerned that coverage might vary across features of the RNA. To address this, we separated features into the most 5′ exons, (generally 5′ untranslated region [UTR]), 3′ exons (i.e., 3′ UTR) and intervening coding sequences (Supplementary Figs. 2). Both 3′ UTR and coding sequences (cds) were represented with a similar (~10^5-fold) range of base mean values compared to 5′ UTR sequences, suggesting that the library construction method was able to capture transcriptional complexity across different elements of genes.

Differential Gene Expression

We first considered overall gene expression for transcripts irrespective of alternate exon usage. Using this analysis, we could completely separate the two groups of embryonic and adult mice (Fig. 1a). Similar to the analysis of 5′, 3′ and coding features, base mean values for genes varied over a 10^6-fold range (Supplementary Fig. 3). Although the overall number of expressed genes was similar in the two groups, there was a modest excess of moderate to low abundance mRNAs in the adult animals, assuming that genes with higher base mean values are more expressed than those with lower base mean.

We next compared differential expression of genes between the two groups (Fig. 1b) using a negative binomial test[19]. There were 4125 genes with differential expression of 4-fold or greater and $p<0.05$ after false discovery rate (FDR) correction. Of these, 1152 had higher expression levels in embryonic brain, while 2973 were more highly expressed in adult brain (Supplementary Table 2). The overall distribution of differences between embryonic and adult brains was slightly asymmetric with a tendency for higher fold differences in the adult (Supplementary Fig. 4). We did not find a distinct group of genes with on/off expression, but rather a smooth distribution from equivalence to high divergence between groups.

The single gene that showed the largest difference between embryo and adult was *Mobp* (myelin-associated oligodendrocytic basic protein), which was ~9000 fold higher in adult compared to embryonic brain (FDR adjusted $p=6\times10^{-137}$). Myelination is a major event that occurs postnatally in the mouse brain, reflected by the higher expression in the adult brain of other myelin genes including myelin associated glycoprotein (*Mog*, 440-fold, $p_{adj}=4.2\times10^{-32}$) or myelin basic protein (*Mbp*, 147-fold, $p_{adj}=7.3\times10^{-36}$;Supplementary Table 2).

We performed technical validation using quantitative reverse transcription-PCR (qRT-PCR). We selected a series of eight genes that showed a range of differential expression as well as different estimated expression levels in embryonic or adult tissue. These were *Vax1, Igf2bp1* and *Wipf1* as low expression genes, *Draxin, Nrp1* and *Caly* as moderately expressed genes

and *Ttr* and *Mobp* as highly expressed genes. We also chose two reference genes that showed low variance and were not differentially expressed between groups; *Ppid*, encoding cyclophilin 40 or cyclophilin D, or *Ubc*, encoding the ubiquitin C gene. We found high agreement between measurements of relative gene expression with RNA-Seq and qRT-PCR normalized to *Ppid* for all of the eight tested genes (Fig. 1c). We obtained similar results using *Ubc* as a reference (Supplementary Fig. 4). In either case, our estimates of fold difference, on a log2 scale, were sufficiently close with two techniques to have an overall r^2 value for regression of 0.979 (n=8, p=2.8×10^{-6}).

We also examined a group of eight genes whose expression differences were close to the 4-fold cutoff used to generate the differential expression list. Four genes were higher in the adult (*ATP10a, Grm4, Sparc, Baiap3*) and four were higher in the embryonic brain (*Ncapg2, Tet1, Ccnd2, Ooep*). We confirmed differential expression in all cases by qRT-PCR and the estimates of fold expression were consistent between both techniques (Supplementary Fig. 5). These results show that the false positive rate is low even around the cutoff values we chose for differential gene expression.

For biological validation, we extracted RNA from multiple developmental stages (E15, E17, P0, P14, P28 and adult; E are embryonic samples, P are postnatal samples; adult animals were 3-4 months old), from an independent set of animals. We were able to confirm that each of the eight genes tested showed changes consistent with the original RNA-Seq dataset using qRT-PCR (Fig. 1d). In most cases there was a substantial change between P0 and P14. Comparing the E17 and adult data points only, we saw concordance between quantitative estimates of difference between RNA-Seq and qRT-PCR (Supplementary Fig. 6) with an overall r^2 value for regression of 0.992 (n=8,p=9.4×10^{-7}). Together, these results show that the RNA-Seq dataset accurately captures estimates of expression differences and these can be replicated in independent biological cohorts.

To provide some information on regional brain expression, we examined expression of our eight validated genes in the Allen brain atlas (http://mouse.brain-map.org/). Although most of the genes were not present in this database, we were able to recover probes for *Mobp* and *Draxin* (Supplementary Fig. 7). There was a clear difference in expression of *Mobp*, with low expression at E18.5 and robust expression at P28. This was seen in multiple brain regions suggesting that *Mobp* expression increases in the cortex can be generalized. The differences in *Draxin* also matched the RNA-Seq pattern of decreased expression at P28 compared to embryonic brain, but here expression in the cortex was relatively low and a more dramatic difference was seen in the developing hindbrain. This suggests that as well as developmental differences in gene expression, there are regional differences, but that RNA-Seq is sufficiently sensitive to detect genes expressed at low levels in the cerebral cortex such as *Draxin*.

Gene ontology analysis using DAVID (Database for Annotation, Visualization and Integrated Discovery)[19] performed separately on genes more highly expressed in the embryonic brain or in the adult brain showed that the former were enriched in genes involved in cell division whereas the latter included genes involved in neurotransmission or ion homeostasis (Table 1). These differences presumably reflect the expected change from

the developing embryonic brain towards the mature postmitotic nervous system with stable synaptic transmission. Although we did not select genes for validation because of gene ontology category membership, *Caly* is involved in endocytosis, both *Nrp1* and *Vax1* are found in the gene ontology category of neuron differentiation.

Alternative Exon Utilization

Alternative splicing and UTR usage was estimated by quantifying counts for each exon relative to total gene expression count to correct for changes in overall expression. There were 387 (Supplementary Table 3) significant differential alternative events (FDR adjusted $p<0.05$), approximately equally divided between those that are higher in adult or in embryo (Fig. 2a). These changes in the ratio of specific exons to overall expression of the gene were separated into 3′UTR, 5′UTR, single exon inclusion, and multiple exon inclusions with alternate 5′ or 3′UTR. The number of single exon inclusion events and multiple exon inclusions with alternative 5′UTR usage were similar in embryonic and adult cortex (86 to 98 and 21 to 23 respectively). However, there was more alternative 3′UTR usage and multiple exon inclusions including the 3′UTR in embryonic brain versus adult (17 to 13 and 11 to 3 events respectively). We found that 123 events (31%) were within genes that did not show any statistically significant differential gene expression.

Gene ontology analysis was performed for the list of genes containing differentially utilized exons, separately considering those that had higher expression of given exons relative to overall gene expression in the adult and those that had higher exon inclusion in the embryo. We found that categories of genes with higher exon ratios in the adult, i.e. where more exons tended to be expressed, included several aspects of the actin cytoskeletal system (Table 2). Genes with higher exon inclusion ratios in the embryonic brain included regulation of small GTPases and several aspects of signal transduction in the nervous system (Table 2). The categories of genes with differential exon usage (Table 2) showed very little overlap with the categories of genes with differential overall expression (Table 1), suggesting independence of splicing and expression (see Discussion).

We next validated these results by comparison to the literature. One well-studied gene with differential splicing is *Mapt*, coding for microtubule associated protein tau. This gene was included in the gene ontology category of actin filament-based processes in the adult enriched exons (Table 2). Three exons in *Mapt* showed statistically significant differences in inclusion between embryo and adult, after correcting for overall *Mapt* expression. Exons 2 and 3, which code for N-terminal repeats, are not utilized in embryonic tau, which we confirmed in our data (Supplementary Fig. 9). Previous data suggest that, in mouse brain, incorporation of exon 10 rises from less than 10% in fetal tau to nearly 100% by postnatal day 24 [20] and our data shows 11-fold higher usage of this exon in adults than embryonic animals.

Having established that this dataset includes known events, we validated novel examples of differential alternative splicing. An alternative exon of *Abr*, a GTPase activating family member that is in the gene ontology category of regulation of small GTPase mediated signal transduction, was expressed at approximately 50-fold (adjusted $p=6.9\times10^{-5}$) higher levels in the embryonic brain (Fig. 2b). We interpret this exon as an alternate 5′UTR because it

contains a predicted ATG start site and because individual RNA-Seq reads did not contain junctions to more 5′ regions (Supplementary Fig. 9). Although not present in the RefSeq gene model, this sequence was present in the UCSC expressed sequence tag (EST) database (Supplementary Fig. 10). In RT-PCR validation using technical replicates, this candidate alternative 5′UTR was only amplified in embryonic samples (Fig. 2c) and in E15, E17 and P0 brains but not in the P14, P28, or adult brain samples from the biological replicate group (Fig. 2d). We confirmed that the identity of the RT-PCR products matched the RNA-Seq data by sequencing (data not shown).

In contrast, a 42bp insert after exon 12 of the kinesin family member *Kif1b* was included in adult brains but excluded in the embryonic tissue (Fig. 2e). As for *Abr*, there was EST support for this exon (Supplementary Fig. 10). We confirmed this difference using RT-PCR in technical (Fig. 2f) and biological replication (Fig. 2g). Sequencing these products identified spliced mRNA identical to exons 12-13 of NM_008441.2, but also products containing inserts, consistent with the RNA-Seq data (Supplementary Fig. 11). These data show that we could validate both of the alternate events that we examined.

RNA Editing

Because RNA-Seq provides sequence information, we next considered RNA editing. We identified and quantified candidate RNA edit sites, limiting our search to A to I changes, as these account for over 90% of editing events in recent studies of the human transcriptome[21], and further restricted candidate edit sites by limiting to those shared by all samples within at least one group. In total, 176 A to I candidate edit sites were found (Supplementary Table 4) of which only 19 were in the Darned database (http://darned.ucc.ie/). Most sites were, however, supported by the EST database on UCSC suggesting that they are robust even if not widely recognized as editing.

Candidate edits showed a tendency to be higher in the adult compared to the embryonic tissues (Fig. 3a). Of all candidate edit sites, 77% were found in UTRs, while 23% were within the coding region (Fig. 3a). All but three of the UTR candidate edit sites were in the 3′UTR, an enrichment previously described in adult mouse[22]. All but 5 of the candidate coding edit sites were non-synonymous, also consistent with previous observations[23]. The majority of the 3′UTR candidate edit sites were in repeat regions, particularly rodent specific Alu sequences. The propensity for A-to-I edits to be found in highly repetitive regions such as Alu elements has been noted previously [24-26].

We chose ten candidate differential edit sites for validation using cDNA cloning and sequencing (Fig. 3b-c), including amino acid changing and synonymous edit sites plus three sites in the heavily edited UTR of *Rpa1* that encodes for Replication protein A1. In all cases, the cloned cDNA sequences were a mixture of A/G but sequencing genomic DNA revealed Adenosine, demonstrating that these are not polymorphisms (Fig. 3d-f and Supplementary Fig. 11). We also confirmed editing at the Q/R site of *Gria2* mRNA by sequencing (data not shown). We found quantitative agreement with the estimates of editing in the RNA-Seq and cloning approaches (Fig. 3b-c; r^2=0.984, p=8.83×10^{-09} for embryonic samples; r^2=0.987, p<2.49×10^{-09} for adult; n=12 sites). We further confirmed these effects in the biological replicate series for three edit sites; a highly edited site that results in a Lysine substitution

(K/E) at position 320 of the *Cyfip2* protein (Cytoplasmic FMR1 interacting protein 2; sequences are shown in Fig. 3d and quantitation from the biological replicates in Fig. 3g), a non-synonymous amino acid (R/G) edit site in the DNA binding protein *Son* (encoding the protein son/Negative regulatory element-binding protein (*Nrebp*); Fig. 3e, h) and a synonymous site in the same gene (Fig. 3f, i).

Candidate amino acid changing edits were more fully edited in adult versus embryonic brains but candidate edits in untranslated regions were not. We tested whether the current resource could be used to examine expression and editing in the same samples. Some edited transcripts also showed differential expression but there was no relationship between editing and expression levels or differential expression (Supplementary Fig. 13). Therefore, editing and expression could not explain the tendency towards higher levels of editing in the adult animals. We were able to measure expression of ADAR genes in the RNA-Seq data. In the initial dataset, *Adarb1* but not *Adar* had statistically significant differential expression between embryonic and adult brain after correction for multiple testing. However, in our biological replicate set we noted that both enzymes showed increased expression with maturation of the brain, although the increase in expression was much greater for *Adarb1* than *Adar* (Fig. 3j,k). This analysis shows that multiple aspects of editing and expression can be queried in this dataset.

DISCUSSION

We have used RNA-Seq to measure differential gene expression, splicing and editing between embryonic and adult mouse cerebral cortex. We present this as a resource for examining these events in brain development.

We saw a very large number of expression changes between E17 and adult and many genes and pathways overlap with those identified using microarrays. For example, we found that many of the genes more highly expressed in embryonic tissue are involved in cell division [3,27-29]. We additionally found pathways that have not previously been reported, such as DNA damage response and repair in embryonic brain, and immune response in adult brain. This RNA-Seq dataset can therefore be used to generate novel insights into biological processes and testable hypotheses. For example, the association with DNA damage might be a consequence of apoptosis in development.

Part of utility of this resource is the high rate of both technical and biological validation of proposed gene expression changes. We found quantitative agreement between RNA-Seq and qRT-PCR for eight genes across a range of expression levels and an additional eight genes with smaller differences in expression, suggesting that the RNA-Seq dataset and analysis are both accurate and predictive.

We also examined differential alternative exon utilization events and replicated known events such as the alternative splicing of *Mapt*[20] but, additionally, a number of novel examples of splicing and UTR utilization. Although we used libraries generated using poly-adenylated RNA with oligo-dT priming, which might be biased towards the 3′ end of transcripts, we also recovered and validated 5′ UTR usage differences. Approximately one

third of the splicing events were in genes that did not have statistically significant differential gene expression in agreement with previous estimates[30-32]. This suggests that differential splicing and exon usage are independent ways to influence gene expression.

We identified some novel patterns of the types of genes where usage of specific exons differed between adult and embryonic brain. Exons more utilized in the adult brain were categorized using gene ontology analysis as related to the cytoskeleton, including *Mapt*, which has more exons in the adult compared to the embryonic brain. Exons more expressed in the embryonic brain included several aspects of signal transduction related to small GTPases (including the valudated gene *Abr*) and synaptic transmission. This was surprising given that neurotransmission is a prominent characteristic of the adult brain. We speculate that multiple variants of genes may be tolerated in development but become more restricted in the adult as function becomes specialized.

Finally, RNA-Seq allowed us to extract sequence data from expressed genes to examine RNA editing in the same samples. In looking for potential edits, we compared the mRNA sequence against the mm9 reference sequence, which was created with the C57Bl/6J strain used in this study. Because the mice are highly inbred, we can be confident that these are not polymorphic DNA sites. In general, we found that editing tended to increase with the shift from embryonic to adult brain, although editing at some sites were already at 100% in the embryonic brains. This increase in editing over development further supports these are edit sites rather than polymorphisms. We confirmed nine out of eleven previously reported[16] editing events, although two genes, (*Grik1* and *Adarb1*) were not in our final list of edits due to low expression levels. Taken together, our results show that we can discover and validate known and novel editing sites in genes expressed in the murine cortex.

We noted that edited sites in UTRs tended to have similar editing levels in both embryonic and adult cortex, whereas those in the coding region were more likely to be differentially edited. This may suggest a different mechanism in editing noncoding versus coding edit sites. Because there is a larger increase in *Adarb1* than *Adar* expression in the adult compared to embryonic brain, we suggest that *Adarb1* could be responsible for those edits that are differential between embryo and adult. We did not note any specific sequence similarity between groups of differential and non-differential edits. ADARs bind double stranded RNA, suggesting that RNA structure rather than sequence affects the efficiency of editing[33].

One limitation of this study is that due to the current costs of RNA-Seq we chose only two stages to generate the dataset and cannot estimate if there are additional changes at specific developmental stages. However, in our biological validation we included additional stages and we noted that for many of the observed E17-adult differences in expression, there was a sharp transition between P0 and P14. This shift around P0 was also noted in validation of alternate splicing and UTR usage. Therefore, we predict that our RNA-Seq dataset comparing E17 with adult captures a substantial proportion of the changes in the transcriptome as the brain develops. Future studies to query a more finely graded series of developmental stages might allow detection of more transient changes in expression.

There are additional limitations of the splicing data. For example, the list of differentially expressed alternative events may be a conservative estimate, as the large number of exons requires substantial correction for multiple testing. Additionally, we have only accounted for relatively simple events and did not consider more complex rearrangements as the short reads available in the RNA-Seq platform lead to sampling around exon/intron junctions rather than full length RNA species. It is expected that future developments of RNA-Seq technology including longer reads and greater sequencing depth will minimize these problems.

The proportion of bases in the genome that are edited has been controversial, with early estimates suggesting a very high proportion [34] but more recent studies showing that mapping and sequencing errors in RNA-Seq lead to overestimation [35]. We therefore only considered the best characterized example of adenosine to inosine editing[21] and only accepting changes where there were at least five reads for a given base. This limits our ability to detect rare edits, but overall our estimates of editing are similar to previous studies[16]. In the case of inbred mice, such sequence differences are unlikely to be due to DNA polymorphisms and for eight validated sites we confirmed the genomic sequence was Adenosine. The tendency of editing to increase in the adult also argues that these changes are not genomic variants.

We limited the current analysis to only female animals, a single brain region and and a single species. Data from the Allen brain atlas to confirm two gene changes and provide some evidence that there are also regional differences in expression. These might be tested in the future using additional RNA-Seq datasets from different brain regions. It would also be of interest to compare gene expression in rodents and humans, which might allow us to test whether the human brain is inherently more complex than other species at the RNA level. Previous experiments using human brain samples [36,37] have shown a very large number of gene expression changes in development. Integrating such findings would be challenging due to species differences in developmental staging and differences in the genes themselves, but may be approached with RNA-Seq in the future.

In many of the measures that we have used, the adult brain does not appear to be inherently more complex than the embryonic brain. For the total number of genes expressed and splicing ratios, differences between the two groups of samples were symmetrical, implying that similar numbers of genes are used. There was a small excess of moderate to low expression genes in the adult brain compared to the embryonic brain, suggesting some possible additional complexity in the adult, but this was not a dramatic difference. In the case of editing, the tendency is for the adult brain to be more fully edited, i.e. the adult brain tolerates less variation than the embryonic brain. Of interest is that one of the most well characterized examples of editing, *Gria2*, is required for neuronal survival as electric activity is established. Our observation that there is a strong tendency of edited sites to be more highly edited in the adult brain suggests that editing is critical in maintaining the correct function of the adult brain.

A final limitation is that as the brain is heterogeneous, we cannot distinguish changes in gene expression in specific cellular populations (e.g. neurons) from changes in cellularity. In

some cases, such as *Mapt*, the transcript is restricted to neurons and therefore changes in exon inclusion are not influenced by cellularity, especially when exon inclusion levels are normalized to the overall transcript expression. In contrast, accumulation of *Mobp* represents the maturation and division of mature oligodendrocytes from precursor cells in the embryonic cerebrum [38]. The current dataset will underestimate differences in those cases where a gene is widely expressed but shows expression or splicing differences in a specific cell type. How frequently this occurs is not known, but the adult brain might be more complex than we estimate here if there are expression/splicing events that occur in specific neuronal populations of the adult brain. Future studies using single cell profiling may improve resolution.

Overall, this resource provides a gene expression set with validated estimates of abundance, splicing and editing of RNA during the development of the mouse brain. We envisage that these data are useful for providing an initial map of expression of genes of interest and their regulation.

Online Methods

RNA Extraction and sequencing

All mouse work followed the guidelines approved by the Institutional Animal Care and Use Committees of the National Institute of Child Health and Human Development. Adult female mice were housed in pairs in a vivarium on a 12h light:dark cycle. We extracted total RNA from 3 adult cerebral cortices and 4 embryonic day 17 cerebral cortices of female C57BL/6J mice using a tissue homogenizer. We measured RNA quality using the Agilent 2100 Bioanalyzer RNA Nano Chip and found that samples had a mean RIN of 9. We purified Poly(A)+ RNA from 10ug total RNA and cDNA libraries were synthesized using the mRNA-Seq prep kit with oligo(dT) priming (Illumina cat# RS-100-0801) as per the manufacturer's protocol (http://mmjggl.caltech.edu/sequencing/mRNA-Seq_SamplePrep_1004898_D.pdf). We hybridized 5 pM of each library to a flow cell, with a single lane for each sample and used an Illumina cluster station for cluster generation. Finally, we generated 80bp single end sequences using an Illumina GA-IIx sequencer.

Statistical analysis of Gene Expression

We used the standard Illumina pipeline with default options to analyze the images and extract base calls in order to generate fastq files. Overall quality and total read counts can be found in Supplementary table 1. We aligned the fastq files to the mm9 mouse reference genome using Tophat (v1.0.13) and Bowtie (0.12.7.0). We annotated and quantified reads to specific genes was using the Python module HT-SEQ with the NCBIM37.61 gtf to provide reference boundaries. We used the R/Bioconductor package DESeq for comparison of aligned reads across samples19. The aligned and aggregated counts underwent a variance stabilizing transformation then the Poisson distributions of normalized counts for each transcript were compared across adult and embryonic groups using a negative binomial test. We corrected for multiple testing using the Benjamini-Hochberg procedure.

We uploaded differentially expressed genes into DAVID[40,41] as two gene lists, one set with higher expression in the adult and one set with higher expression in embryos, with offical_gene_symbol selected for *Mus musculus* genes. We used the functional annotation tool with default parameters, and the biological processes gene ontology terms selected. We calculated P-values for geneenrichment using a modified Fisher's Exact Test, and a Benjamini-Hochberg multiple test correction was performed using tools within DAVID[40,41].

Alternate Exon Utilization

We obtained a track of exons plus bed from the UCSC table browser using the NCBIM37 assembly found in the genes and gene prediction option track. We removed exons with the same start or end position so only one set of coordinates for each exon was obtaine and used this file as the reference gtf file for HT-SEQ as described above. We analyzed junction reads separately, with the start position and end position of each junction being treated as independent measures using a custom java script to map junction start and stop to an exon. We then added both sets of quantification for the total read count per exon. We merged the exon data with total gene count data and then filtered with each gene having a 10% or greater read count of the total bp count to filter out genes with low and/or uneven coverage. We then used the ratio of gene count to exon count as input for DE-SEQ, specifically using gene/exon rather than exon/gene as the latter would be fractions and not suitable for DE-SEQ which requires integer numbers. However, for display purposes we plotted the inverse ratio, ie exon/gene. As for exon data, p-values were based on a negative binomial test and corrected using the Benjamini-Hochberg procedure. Finally, we classified each alternative splicing as exon inclusion/exclusion, alternative UTR, or exon exclusion due to an alternate 3′ or 5′ UTR (shorter/longer isoform). We performed gene ontology analysis as for differential gene expression, with two gene sets of differential exons more expressed in adult or more expressed in embryonic tissue.

Editing

To find edit sites in a genome-wide manner, we called variants for each sample with SAMtools pileup (samtools-0.1.7a) using the mm9 as a reference. Next, we filtered for depth of coverage of at least 5 reads per base at each edit site21. We created a merged table to include only variants shared by all samples within the adult or embryonic group. We extracted only variants within a gene boundary with A to I changes, noting that Inosine is sequenced as Guanosine in the RNASeq platform. Finally, we submitted sequences containing the candidate edit sites to UCSC BLAT to distinguish reads with single mismatches to the genomic sequence due to RNA editing rather than those due to inappropriate mapping to another part of the genome. We accepted candidate edits if the best scored alignment of the read included the site of the edited base from the original alignment; if there were multiple genomic alignments of equal or higher score, then the candidate edit site was discarded.

qRT-PCR Expression Validation

We synthesized cDNA from trizol extracted RNA using the SuperScript III First-Strand Synthesis System from Invitrogen. We measured cDNA abundance using Sybr Green on an

Applied BioSystems HT-7900 qRT-PCR system. We performed serial dilutions of cDNA to find primer pairs with 100% efficiency and a single product on the disassociation curve. We used *Ppid* or *Ubc* used as the normalization gene (primer pairs are listed in Supplementary Table 5). For biological replication, we used samples from developmental stages of E15 (n=4), E17 (n=5), P0 (n=6), P14 (n=4), P28 (n=4), and adult (n=4). Regression was used to compare estimates of fold difference between RNA-Seq and validation.

RNA editing and splicing validation by RT-PCR and sequencing

We used cDNA made as described above from two biological replicates for PCR with primer pairs (primer sequences are listed in Supplementary Table 6) designed to span an exon junction on one side to and approximately 150 base pairs on the other side of the edit site using Fast Start Master Mix (Roche). We amplified genomic DNA in the same manner. We separated products on a 1.5% agarose gel then cut out and extracted DNA using the QIAquick gel extraction kit (Qiagen). We cloned DNA into the PCR8-TA cloning vector (Invitrogen) in topo 10 competent cells and sequenced twenty colonies from each biological replicate for each edit site and 12 colonies from genomic DNA using a 3730 capillary sequencer. We repeated the same process for splicing validation where 3 colonies from each PCR product were sequenced (primer sequences are listed in Supplementary Table S7). Regression was used to compare estimates of proportion of bases edited between RNA-Seq and validation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (Project number AG000947) and by the Swedish Research Council and Swedish brain power. The authors would like to thank Eva Lindquist for excellent technical assistance. We would also like to thank Michael Do for assistance with cloning experiments.

References

1. Innocenti GM, Price DJ. Exuberance in the development of cortical networks. Nat. Rev. Neurosci. 2005; 6:955–965. [PubMed: 16288299]
2. Price DJ, et al. The development of cortical connections. Eur. J. Neurosci. 2006; 23:910–920. [PubMed: 16519656]
3. Mody M, et al. Genome-wide gene expression profiles of the developing mouse hippocampus. Proc. Natl. Acad. Sci. U.S.A. 2001; 98:8862–8867. [PubMed: 11438693]
4. Kalsotra A, Cooper TA. Functional consequences of developmentally regulated alternative splicing. Nat. Rev. Genet. 2011; 12:715–729. [PubMed: 21921927]
5. Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. Nat. Rev. Genet. 2010; 11:75–87. [PubMed: 20019688]
6. Li Q, Lee J-A, Black DL. Neuronal regulation of alternative pre-mRNA splicing.]Nat. Rev. Neurosci. 2007; 8:819–831. [PubMed: 17895907]
7. Wojtowicz WM, Flanagan JJ, Millard SS, Zipursky SL, Clemens JC. Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell. 2004; 118:619–633. [PubMed: 15339666]

8. Ullrich B, Ushkaryov YA, Südhof TC. Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. Neuron. 1995; 14:497–507. [PubMed: 7695896]
9. Hogg M, Paro S, Keegan LP, O'Connell MA. RNA editing by mammalian ADARs. Adv. Genet. 2011; 73:87–120. [PubMed: 21310295]
10. Blanc V, Davidson NO. APOBEC-1-mediated RNA editing. Wiley Interdiscip Rev Syst Biol Med. 2010; 2:594–602. [PubMed: 20836050]
11. Rueter SM, Dawson TR, Emeson RB. Regulation of alternative splicing by RNA editing. Nature. 1999; 399:75–80. [PubMed: 10331393]
12. Serra MJ, Smolter PE, Westhof E. Pronounced instability of tandem IU base pairs in RNA. Nucleic Acids Res. 2004; 32:1824–1828. [PubMed: 15037659]
13. Zhang Z, Carmichael GG. The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. Cell. 2001; 106:465–475. [PubMed: 11525732]
14. Higuchi M, et al. Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. Nature. 2000; 406:78–81. [PubMed: 10894545]
15. Rosenthal JJC, Seeburg PH. A-to-I RNA editing: effects on proteins key to neural excitability. Neuron. 2012; 74:432–439. [PubMed: 22578495]
16. Wahlstedt H, Daniel C, Enstërö M, Ohman M. Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. Genome Res. 2009; 19:978–986. [PubMed: 19420382]
17. 't Hoen PAC, et al. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res. 2008; 36:e141. [PubMed: 18927111]
18. Keane TM, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011; 477:289–294. [PubMed: 21921910]
19. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. [PubMed: 20979621]
20. McMillan P, et al. Tau isoform regulation is region- and cell-specific in mouse brain. J. Comp. Neurol. 2008; 511:788–803. [PubMed: 18925637]
21. Peng Z, et al. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat. Biotechnol. 2012; 30:253–260. [PubMed: 22327324]
22. Gu T, et al. Canonical A-to-I and C-to-U RNA editing is enriched at 3′UTRs and microRNA target sites in multiple mouse tissues. PLoS ONE. 2012; 7:e33720. [PubMed: 22448268]
23. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem. 2010; 79:321–349. [PubMed: 20192758]
24. Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. PLoS Biol. 2004; 2:e391. [PubMed: 15534692]
25. Neeman Y, Levanon EY, Jantsch MF, Eisenberg E. RNA editing level in the mouse is determined by the genomic repeat repertoire. RNA. 2006; 12:1802–1809. [PubMed: 16940548]
26. DeCerbo J, Carmichael GG. SINEs point to abundant editing in the human genome. Genome Biol. 2005; 6:216. [PubMed: 15833131]
27. Semeralul MO, et al. Microarray analysis of the developing cortex. J. Neurobiol. 2006; 66:1646–1658. [PubMed: 17013924]
28. Kagami Y, Furuichi T. Investigation of differentially expressed genes during the development of mouse cerebellum. Brain Res. Gene Expr. Patterns. 2001; 1:39–59. [PubMed: 15018818]
29. Matsuki T, Hori G, Furuichi T. Gene expression profiling during the embryonic development of mouse brain using an oligonucleotide-based microarray system. Brain Res. Mol. Brain Res. 2005; 136:231–254. [PubMed: 15893606]
30. Bland CS, et al. Global regulation of alternative splicing during myogenic differentiation. Nucleic Acids Res. 2010; 38:7651–7664. [PubMed: 20634200]

31. Ramsköld D, Wang ET, Burge CB, Sandberg R. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput. Biol. 2009; 5:e1000598. [PubMed: 20011106]
32. McKee AE, et al. Exon expression profiling reveals stimulus-mediated exon use in neural cells. Genome Biol. 2007; 8:R159. [PubMed: 17683528]
33. Stefl R, Allain FH-T. A novel RNA pentaloop fold involved in targeting ADAR2. RNA. 2005; 11:592–597. [PubMed: 15840813]
34. Li M, et al. Widespread RNA and DNA sequence differences in the human transcriptome. Science. 2011; 333:53–58. [PubMed: 21596952]
35. Pickrell JK, Gilad Y, Pritchard JK. Comment on 'Widespread RNA and DNA sequence differences in the human transcriptome'. Science. 335:1302. author reply 1302 (2012). [PubMed: 22422963]
36. Kang HJ, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011; 478:483–489. [PubMed: 22031440]
37. Colantuoni C, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature. 2011; 478:519–523. [PubMed: 22031444]
38. Itoh K. Culture of oligodendrocyte precursor cells (NG2(+)/O1(−)) and oligodendrocytes (NG2(−)/O1(+)) from embryonic rat cerebrum. Brain Res. Brain Res. Protoc. 2002; 10:23–30. [PubMed: 12379434]
39. Tian N, Wu X, Zhang Y, Jin Y. A-to-I editing sites are a genomically encoded G: implications for the evolutionary significance and identification of novel editing sites. RNA. 2008; 14:211–216. [PubMed: 18094120]
40. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. [PubMed: 19131956]
41. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37:1–13. [PubMed: 19033363]

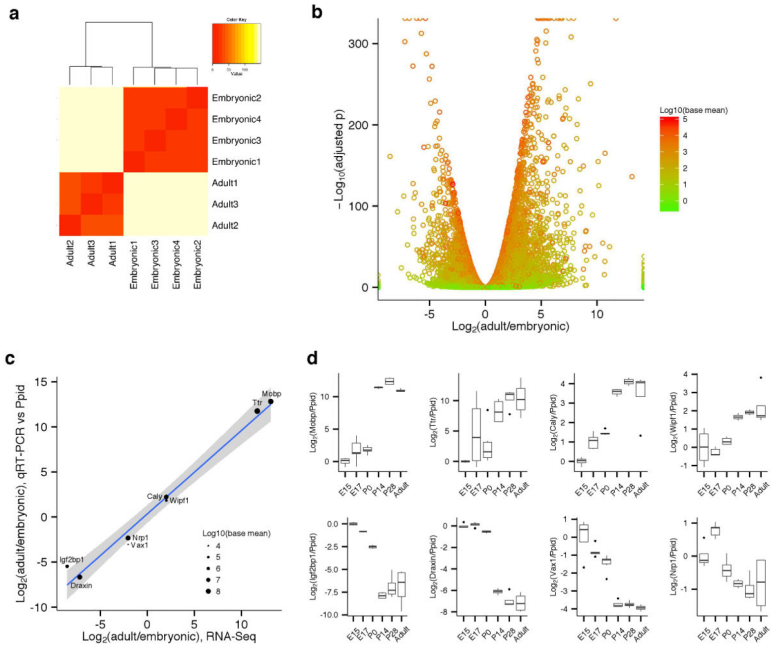

Figure 1.

Gene expression

(a) Expression Heatmap. A heatmap of sample-to-sample distances on the matrix of variance stabilized data for overall gene expression is shown where darker red colors indicate more similar expression, as shown on the color key to the upper right of the panel. Clustering, drawn above the heatmap, demonstrates that the adult samples are very similar to each other, but show complete separation from the embryonic samples. (b) Differential expression. Plot of differential gene expression with fold difference of $\log_2$ normalized expression in adult cerebral cortex (n=3) versus embryonic cerebral cortex (n=4) on the x-axis and $-\log_{10}$ adjusted p-value on the y-axis. Each gene is colored based on the $\log_{10}$ base mean expression, i.e. more highly expressed genes are in darker colors. (c) Technical validation of differential expression. The x-axis shows the $\log_2$ fold expression (Adult/Embryonic) using the RNA-Seq data compared to qRT-PCR data for eight genes (y axis). *Ppid* was used as the normalization gene. The size of each point represents the base mean expression level from the RNA-Seq data as in (b). The grey shaded area indicates the 95% confidence interval for the regression. (d) Boxplots showing biological validation of differential expression. The same eight genes were used for validation with a new set of animals and an expanded set of developmental stages. On the y-axes are the $\log_2$ expression values normalized to E15 measured with qRT-PCR with *Ppid* used as the normalization gene. The x-axes show each developmental stage with E15 (n=4), E17 (n=5), P0 (n=6), P14 (n=4), P28 (n=4), and adult (n=4). The boxes represent the range between first and third quartiles and whiskers indicate highest value and lowest values within 1.5 multiples of the inter-quartile range; outliers from this range are plotted as individual dots.

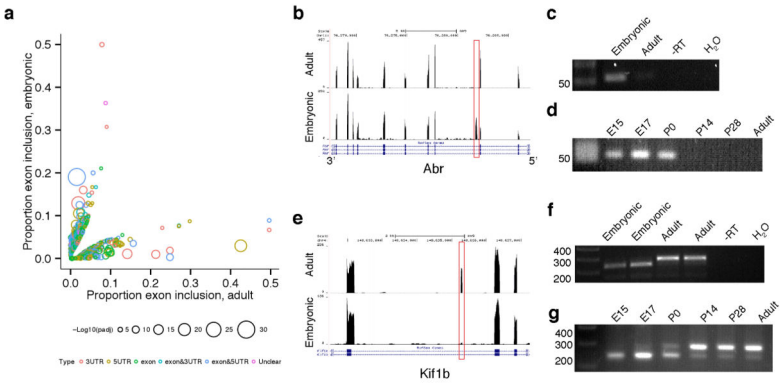

Figure 2.

Alternative Exon Utilization

(a) Differential alternative event ratios. A graph of the proportion of exon inclusion (counts per exon/counts per gene) with adult ratios (n=3) on the x-axis and embryonic ratios (n=4) on the y-axis. The values are colored based on the type of event, and sized according to the $-\log_{10}$ of the adjusted p-value. For clarity, only events that were significantly different between groups (FDR adjusted $p<0.05$) are shown. (b) *Abr* alternative 5′UTR. Plot of RNA-Seq reads generated using the UCSC genome browser covering an alternative 5′UTR of the *Abr* gene, boxed in red; note that *Abr* is on the negative DNA strand and the 5′UTR is on the right of the plot. (c) Validation of *Abr* alternative isoform using RT-PCR of technical replicates. Using PCR, primers for the alternative UTR only amplified the product in the embryonic tissue. Reactions are representative of triplicate biological samples. (d) *Abr* alternative 5′UTR biological replication. Using RT-PCR, primers for the alternative UTR amplified the product in the developmental stages E15, E17, and P0. In contrast, there is no detectable product at P14, P28, and adult. Gel is representative of duplicate biological replicates. (e) Exon inclusion in *Kif1b*. The adult brain had a relatively higher proportion of an inclusion sequence expressed after exon 12 (boxed in red) compared to embryonic brain. (f) *Kif1b* exon inclusion technical replication. RT-PCR with gene specific primers confirmed the inclusion of a novel sequence shown by a shift in expected size from embryo to adult brain using duplicate biological replicates. (g) *Kif1b* exon inclusion biological replication. A similar RT-PCR approach in independent biological samples shows the same pattern of inclusion of the 42bp sequence towards adult animals. Gels are representative of three reactions with independent biological replicates.

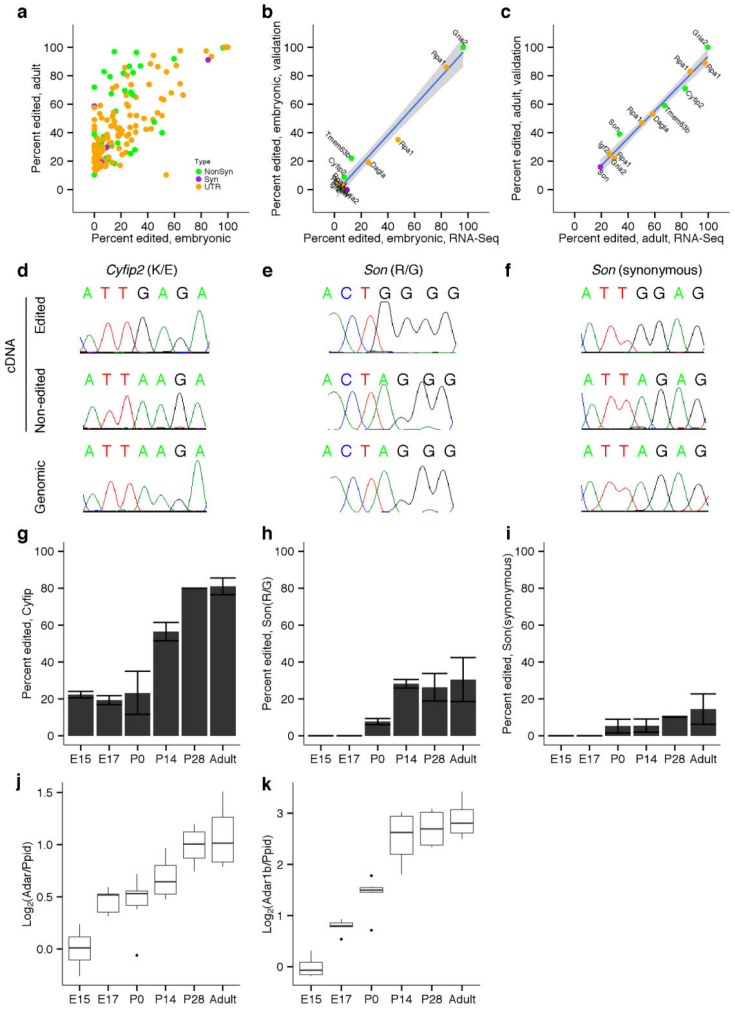

Figure 3.

Adenosine to Inosine RNA editing

(a) Adenosine to Inosine editing. Average percent edited in adult mouse cerebral cortex (*n*=3) on the x-axis and average percent editing in the embryonic mouse cerebral cortex (*n*=4) on the y-axis. Yellow indicates the edit is found in the UTR, green indicates the edit results in a non-synonymous amino acid changes, and purple indicates the edit results in no change in the amino acid sequence. (b, c) Technical validation of RNA editing. Eleven sites, both coding and noncoding, were chosen for validation. The x-axis shows the total average percent of the edited site using the RNA-Seq data compared to validation using cloning and sequencing on the y-axis for embryonic samples (b) and adult data (c). The grey shaded areas indicate the 95% confidence interval for the regression. (d-i) Biological validation of RNA editing. We took sites in *Cyfip2* (d), a non-synonymous amino acid change (e) and a synonymous edit (f) in *Son* and sequenced both cDNA and genomic DNA clones for each site. For each panel in d-f an example of edited and non-edited cDNA clones are shown in the top two chromatograms and the genomic sequence is shown in the lowest chromatogram. We then quantified the percentage of cDNA clones that showed editing for *Cyfip2* (g), the non-synonymous amino acid change (h) and synonymous edit (i) in *Son*

across different developmental stages as indicated on the x-axes for the proportion of clones with the edited sequence (y axis, error bars indicate range across replicates). (j,k) Expression of genes coding for ADAR enzymes. Expression of *Adar* (j) and *Adar1b* (k) was estimated by qRT-PCR with *Ppid* as a reference gene using the biological replicates at the indicated developmental stages on the x-axis. Note that both y-axes are on $\log_2$ scales but that the proportional increase in expression is higher for *Adar1b* than for *Adar*. The boxes represent the range between first and third quartiles and whiskers indicate highest value and lowest values within 1.5 multiples of the inter-quartile range; outliers from this range are plotted as individual dots.

Table 1

Pathway analysis for differentially expressed genes

Adult>Embryo			Embryo>Adult		
Gene OntologyTerm	**Gene Count**	***P***	**Gene Ontology Term**	**Gene Count**	***P***
ion transport	182	5.1×10^{-19}	cell cycle	121	5.9×10^{-49}
immune response	135	3.2×10^{-18}	cell cycle process	89	1.1×10^{-39}
metal ion transport	129	3.4×10^{-18}	cell cycle phase	80	5.0×10^{-38}
cell-cell signaling	95	1.2×10^{-16}	M phase	75	5.6×10^{-38}
cation transport	139	1.4×10^{-16}	cell division	69	1.4×10^{-32}
transmission of nerve impulse	80	5.5×10^{-16}	mitotic cell cycle	64	7.0×10^{-32}
regulation of system process	72	1.4×10^{-14}	M phase of mitotic cell cycle	57	5.5×10^{-31}
synaptic transmission	66	5.3×10^{-14}	nuclear division	55	1.7×10^{-29}
chemical homeostasis	99	1.7×10^{-11}	mitosis	55	1.7×10^{-29}
ion homeostasis	85	2.1×10^{-11}	organelle fission	55	1.2×10^{-28}
monovalent inorganic cation transport	85	1.5×10^{-10}	DNA metabolic process	66	1.2×10^{-19}
potassium ion transport	56	1.5×10^{-10}	microtubule-based process	44	3.5×10^{-17}
cellular ion homeostasis	76	3.3×10^{-10}	DNA replication	35	2.0×10^{-14}
cellular chemical homeostasis	77	4.5×10^{-10}	chromosome segregation	22	2.8×10^{-12}
antigen processing and presentation of peptide antigen	23	8.3×10^{-10}	microtubule-based movement	25	1.0×10^{-10}
behavior	100	3.3×10^{-9}	response to DNA damage stimulus	39	5.9×10^{-9}
regulation of neurological system process	43	3.4×10^{-9}	DNA repair	33	2.2×10^{-8}
regulation of transmission of nerve impulse	41	8.0×10^{-9}	microtubule cytoskeleton organization	23	3.4×10^{-8}
response to wounding	88	1.2×10^{-8}	cellular response to stress	43	9.0×10^{-7}

Adult>Embryo			Embryo>Adult		
Gene OntologyTerm	**Gene Count**	***P***	**Gene Ontology Term**	**Gene Count**	***P***
cellular metal ion homeostasis	38	5.5×10^{-8}	chromosome organization	42	2.6×10^{-6}

We performed Gene ontologyanalyses using DAVID[40,41] on two sets of significantly regulated genes, those more highly expressed in adult and those more highly expressed in embryonic brain. Differentially expressed genes had an adjusted $p<0.05$ and a 4-fold or greater expression difference between the embryonic and adult. Out of 1152 genes for the embryonic group, 534 were accepted for gene ontology analysis. Out of the 4125 genes for the adult group, 1694 genes were accepted by DAVID for gene ontology analysis.

Table 2

Pathway analysis in alternate exon utilization data

Adult>Embryo			Embryo>Adult		
Gene Ontology Term	**Gene Count**	***P***	**Gene Ontology Term**	**Gene Count**	***P***
Actin Filament-based Process	10	1.7×10^{-2}	Regulation of Small GTPase Mediated Signal Transduction	12	3.7×10^{-3}
Actin Cytoskeleton Organization	10	2.1×10^{-2}	Transmission of Nerve Impulse	12	6.8×10^{-3}
Actin Filament-based Process	10	1.7×10^{-2}	Regulation of Ras Protein Signal Transduction	10	1.3×10^{-2}
			Vesicle-mediated Transport	15	2.2×10^{-2}
			Synaptic Transmission	9	4.1×10^{-2}
			Cell-Cell Signaling	11	4.3×10^{-2}
			Endocytosis	9	5.0×10^{-2}
			Membrane Invagination	9	5.0×10^{-2}

We performed gene ontology analyses using DAVID[40,41] on two sets of genes, those where specific exons were more highly expressed in adult and those with specific exons more highly expressed in embryonic brain, each relative to overall gene expression. Differential alternative events were defined as an adjusted $p<0.05$ and with a 4-fold or greater expression difference between the embryonic and adult. Out of 172 genes for the embryonic group, 127 were accepted for gene ontology analysis. Out of the 184 genes for the adult group 129 genes were accepted by DAVID for gene ontology analysis.