

# An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex

Ye Zhang,<sup>1\*</sup> Kenian Chen,<sup>2,3\*</sup> Steven A. Sloan,<sup>1\*</sup> Mariko L. Bennett,<sup>1</sup> Anja R. Scholze,<sup>1</sup> Sean O'Keeffe,<sup>4</sup> Hemali P. Phatnani,<sup>4</sup> Paolo Guarneri,<sup>7,9</sup> Christine Caneda,<sup>1</sup> Nadine Ruderisch,<sup>5</sup> Shuyun Deng,<sup>2,3</sup> Shane A. Liddelow,<sup>1,6</sup> Chaolin Zhang,<sup>4,7,8</sup> Richard Daneman,<sup>5</sup> Tom Maniatis,<sup>4</sup> Ben A. Barres,<sup>1,†</sup> and Jia Qian Wu<sup>2,3,†</sup>

<sup>1</sup>Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305-5125, <sup>2</sup>The Vivian L. Smith Department of Neurosurgery, University of Texas Medical School at Houston, Houston, Texas 77057, <sup>3</sup>Center for Stem Cell and Regenerative Medicine, University of Texas Brown Institute of Molecular Medicine, Houston, Texas 77057, <sup>4</sup>Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, New York 10032, <sup>5</sup>Department of Anatomy, University of California, San Francisco, San Francisco, California 94143-0452, <sup>6</sup>Department of Pharmacology and Therapeutics, University of Melbourne, Parkville, Victoria, Australia 3010, and <sup>7</sup>Department of Systems Biology, <sup>8</sup>Center for Motor Neuron Biology and Disease, and <sup>9</sup>Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York 10032

The major cell classes of the brain differ in their developmental processes, metabolism, signaling, and function. To better understand the functions and interactions of the cell types that comprise these classes, we acutely purified representative populations of neurons, astrocytes, oligodendrocyte precursor cells, newly formed oligodendrocytes, myelinating oligodendrocytes, microglia, endothelial cells, and pericytes from mouse cerebral cortex. We generated a transcriptome database for these eight cell types by RNA sequencing and used a sensitive algorithm to detect alternative splicing events in each cell type. Bioinformatic analyses identified thousands of new cell type-enriched genes and splicing isoforms that will provide novel markers for cell identification, tools for genetic manipulation, and insights into the biology of the brain. For example, our data provide clues as to how neurons and astrocytes differ in their ability to dynamically regulate glycolytic flux and lactate generation attributable to unique splicing of *PKM2*, the gene encoding the glycolytic enzyme pyruvate kinase. This dataset will provide a powerful new resource for understanding the development and function of the brain. To ensure the widespread distribution of these datasets, we have created a user-friendly website ([http://web.stanford.edu/group/barres\\_lab/brain\\_rnaseq.html](http://web.stanford.edu/group/barres_lab/brain_rnaseq.html)) that provides a platform for analyzing and comparing transcription and alternative splicing profiles for various cell classes in the brain.

**Key words:** alternative splicing; astrocytes; microglia; oligodendrocytes; transcriptome; vascular cells

## Introduction

Together with neurons, glia (astrocytes, oligodendrocytes, and microglia) and vascular cells (endothelial cells and pericytes) are essential for the development and function of the nervous system (Allen and Barres, 2009; Molofsky et al., 2012). Each of these major cell types express a distinct repertoire of genes, and to-

gether these transcriptomes provide fundamental insights into the development and function of the brain. A transcriptome database of purified brain cell types has a variety of potential uses, including identifying cell type-specific markers, helping to develop tools that allow cell type-specific genetic manipulations (e.g., cell ablation, gene silencing, and optogenetic manipulations), revealing the metabolic division of labor between different cell types and suggesting novel cell-specific functions.

Previous groups have performed transcriptome studies of one or more purified cell types in the brain (Dugas et al., 2006; Rossner et al., 2006; Lovatt et al., 2007; Cahoy et al., 2008; Doyle et al., 2008; Lau et al., 2008; Daneman et al., 2010; Wylie et al., 2010; Bracko et al., 2012; Friedrich et al., 2012; Lerch et al., 2012; Bandopadhyay et al., 2013; Beutner et al., 2013; Chiu et al., 2013; Orre et al., 2014), but because of the technical complexity involved in purifying numerous cell types and varying foci of previous studies, a complete transcriptome dataset that allows direct comparison of gene expression across all of the major glial and vascular brain cell types is lacking. Moreover, the majority of previous transcriptome studies were performed using microarrays, in which transcript abundance is indirectly deduced from fluorescence signal intensities.

Received May 7, 2014; revised July 11, 2014; accepted July 22, 2014.

Author contributions: Y.Z., S.A.S., M.L.B., R.D., T.M., B.A.B., and J.Q.W. designed research; Y.Z., S.A.S., M.L.B., A.R.S., C.C., N.R., and S.D. performed research; S.A.L. and C.Z. contributed unpublished reagents/analytic tools; Y.Z., K.C., S.A.S., M.L.B., S.O., H.P.P., and P.G. analyzed data; Y.Z., K.C., S.A.S., B.A.B., and J.Q.W. wrote the paper.

This work is supported by National Institutes of Health Grants R01 MH0955501 and R01 NS08170301 (B.A.B.), T32GM007365 with additional support from Stanford School of Medicine and its Medical Scientist Training Program (S.S.), and R00 GM05713 (C.Z.), a Berry Postdoctoral Fellowship (Y.Z.), and a National Health and Medical Research Council of Australia C.J. Martin Fellowship (S.L.). J.Q.W., K.C., and S.D. are supported by National Institutes of Health and the Staman Ogilvie Fund–Memorial Hermann Foundation. We thank Dr. Dritan Agalliu for *Cldn5* expression plasmid and Vincent and Stella Coates for their generous support.

\*Y.Z., K.C., and S.A.S. contributed equally to this work. †B.A.B. and J.Q.W. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to either of the following: Ye Zhang, Stanford University School of Medicine, Department of Neurobiology, 299 Campus Drive, Fairchild Building, Stanford, California 94305-5125, E-mail: [zhangye@stanford.edu](mailto:zhangye@stanford.edu); or Jia Qian Wu, University of Texas Medical School, 1825 Pressler Street, SRB-630.01, Houston, TX 77030. E-mail: [jiaqian.wu@uth.tmc.edu](mailto:jiaqian.wu@uth.tmc.edu).

DOI:10.1523/JNEUROSCI.1860-14.2014

Copyright © 2014 the authors 0270-6474/14/3411929-19\$15.00/0

RNA sequencing (RNA-Seq) is a method that profiles the transcriptome by deep sequencing of isolated RNAs. Transcript abundance is directly proportional to the number of sequencing reads that map to a specific transcript, resulting in lower false-negative and false-positive discovery rates in addition to a greater linear range (Bainbridge et al., 2006; Cloonan et al., 2008; Mariotti et al., 2008; Mortazavi et al., 2008; Nagalakshmi et al., 2008; Sultan et al., 2008; Wilhelm et al., 2008; Trapnell et al., 2010; Wu et al., 2010). Direct sequencing of RNA libraries also provides the opportunity to explore alternative splicing, a key mechanism that contributes to transcriptome diversity (Pan et al., 2008; Wang et al., 2008). Aberrant alternative splicing has been implicated in many neurological diseases (Buckanovich et al., 1996; Yang et al., 1998), yet the complex landscape of differential splicing among glia, neurons, and vascular cells has not been investigated.

In this study, we purified neurons, astrocytes, microglia, endothelial cells, pericytes, and various maturation states of oligodendrocytes from mouse cortex and used RNA-Seq to generate a high-resolution transcriptome database of >22,000 genes. We identified thousands of novel cell type-enriched genes and developed a library of thousands of cell type-specific splicing events. To ensure the widespread distribution of these datasets, we have created a user-friendly website ([http://web.stanford.edu/group/bares\\_lab/brain\\_rnaseq.html](http://web.stanford.edu/group/bares_lab/brain_rnaseq.html)) that provides readily accessible platforms for analyzing and comparing transcription and alternative splicing profiles for various cell classes in the brain.

## Materials and Methods

### Purification of glia, neurons, and vascular cells

All procedures involving animals were conducted in conformity with Stanford University guidelines that are in compliance with national and state laws and policies. The purification procedures are modified from previously described dissociation, immunopanning, and fluorescence-activated cell sorting (FACS) purification protocols (Dugas et al., 2006; Cahoy et al., 2008; Daneman et al., 2010; Foo et al., 2011).

**Purification of astrocytes.** To purify astrocytes, we used a bacterial artificial chromosome (BAC) transgenic mouse line expressing EGFP under the control of regulatory sequences in Aldh1l1–BAC (Heintz, 2004). This line has been characterized previously to have complete astrocyte-specific labeling throughout the brain (Cahoy et al., 2008). For all cell types, one biological replicate consists of pooled cells from a litter of 3–12 mice. The cortices were dissected out, and the meninges were removed. The tissue was enzymatically dissociated to make a suspension of single cells as described previously (Dugas et al., 2006; Cahoy et al., 2008). Briefly, the tissue was incubated at 33°C for 45 min in 20 ml of a papain solution containing Earle's balanced salt solution (EBSS; catalog #E7510; Sigma), D(+)-glucose (22.5 mM), NaHCO<sub>3</sub> (26 mM), DNase (125 U/ml; catalog #LS002007; Worthington), papain (9 U/ml; catalog #LS03126; Worthington), and L-cysteine (1 mM; catalog #C7880; Sigma). The papain solution was equilibrated with 5% CO<sub>2</sub> and 95% O<sub>2</sub> gas before and during papain treatment. After papain treatment, the tissue was washed three times with 4.5 ml of inhibitor buffer containing BSA (1.0 mg/ml; catalog #A-8806; Sigma) and ovomucoid (also known as trypsin inhibitor; 1.0 mg/ml; catalog #109878; Roche Diagnostics) and then mechanically dissociated by gentle sequential trituration using a 5 ml pipette. Dissociated cells were layered on top of 10 ml of a high-concentration inhibitor solution with 5 mg/ml BSA and 5 mg/ml ovomucoid and centrifuged at 130 × g for 5 min. The cell pellet was then resuspended in 12 ml of Dulbecco's PBS (DPBS; catalog #14287; Invitrogen) containing 0.02% BSA and 12.5 U/ml DNase and filtered through a 20 μm Nitex mesh (Lab Pak 03-20/14; Sefar America) to remove undissociated cell clumps. Cell health was assessed by trypan blue exclusion. Only single-cell suspensions with >85% viability were used for purification experiments.

Propidium iodide (PI; 1 μg/ml; catalog #P4864; Sigma) was added to the single-cell solution to label dead cells. Cells were sorted on a BD Aria

II cell sorter (BD Bioscience) with a 70 μm nozzle. Dead cells and debris were gated first by their low forward light scatter and high side light scatter and second by high PI staining. Doublets were removed by high side light scatter. Cell concentration and flow rate were carefully adjusted to maximize purity. Astrocytes were identified based on high EGFP fluorescence. FACS routinely yielded >99% purity based on reanalysis of sorted cells. Purified cells were harvested by centrifugation at 2000 × g for 5 min. The cell pellet was then used for RNA extraction.

**Purification of endothelial cells.** To purify endothelial cells, we used Tie2–EGFP transgenic mice available from The Jackson Laboratory. These mice express EGFP under the pan-endothelial Tie2 promotor (Motoike et al., 2000; Daneman et al., 2010). Single-cell suspensions and FACS were performed as described above.

**Preparation of panning plates.** To prepare panning plates for immuno-panning, Petri dishes (150 × 15 mm; catalog #351058; BD Biosciences) were incubated with 22 ml of Tris-HCl buffer solution (50 mM, pH 9.5) and 150 μg secondary antibody overnight at 4°C. Each dish was then washed three times with 10–20 ml of DPBS and incubated with the corresponding primary antibodies diluted in 12 ml of DPBS/0.2% BSA solution per dish for at least 2 h at room temperature. Lectin-coated panning plates were prepared by adding 22 ml of DPBS and 50 μg of *Banderiaea simplicifolia* lectin 1 (BSL-1; catalog #L-1100; Vector Laboratories) and incubating at 4°C overnight. All panning dishes were washed three times with 10–20 ml of DPBS immediately before use. The secondary antibodies (Jackson ImmunoResearch) we used included affinity-purified goat anti-mouse IgG + IgM [heavy and light (H + L) chain; catalog #115-005-044], goat anti-mouse IgM μ-chain (catalog #115-005-02), and goat anti-rat IgG (H + L chain; catalog #112-005-167). All immunopanning was performed at room temperature.

**Purification of neurons.** To purify neurons, a single-cell suspension was prepared as described above and incubated at 34°C for 1 h to allow expression of cell-surface protein antigens digested by papain and then incubated on two sequential panning plates coated with BSL-1 to deplete endothelial cells (10 min each), followed by a 30 min incubation on a plate coated with mouse IgM anti-O4 hybridoma (Bansal et al., 1989; 4 ml of hybridoma supernatant diluted with 8 ml of DPBS/0.2% BSA) to deplete oligodendrocyte precursor cells (OPCs), and then incubated for 20 min on a plate coated with rat anti-mouse cluster of differentiation 45 (CD45) (catalog #550539; BD Pharmingen; 1.25 μg in 12 ml of DPBS/0.2% BSA) to deplete microglia and macrophages. Finally, cells were added to a plate coated with rat anti-mouse L1 neural cell adhesion molecule (L1CAM; 30 μg in 12 ml of DPBS/0.2% BSA; catalog #MAB5272; Millipore) to bind neurons. The adherent cells on the L1CAM plate were washed eight times with 10–20 ml of DPBS to remove all antigen-negative nonadherent cells and then removed from the plate by treating with trypsin (1000 U/ml; catalog #T-4665; Sigma) in 8 ml of Ca<sup>2+</sup> and Mg<sup>2+</sup> free EBSS (catalog #9208; Irvine Scientific) for 3–10 min at 37°C in a 10% CO<sub>2</sub> incubator. The trypsin was then neutralized with 20 ml of fetal calf serum (FCS) solution containing 30% FCS (catalog #10437-028; Gibco), 35% DMEM (catalog #11960-044; Invitrogen), and 35% Neurobasal (catalog #21103-049; Gibco). The cells were dislodged by gentle squirting of FCS solution over the plate and harvested by centrifugation at 200 × g for 10 min.

**Purification of microglia and oligodendrocyte-lineage cells.** Both microglia and peripheral macrophages express the surface protein CD45. Therefore, macrophages are potential contaminants in the microglia preparation isolated with anti-CD45 antibody. To minimize macrophage contamination, we first perfused the mice with 10 ml of PBS to wash away blood from the brain. A single-cell suspension was then prepared as described above and incubated 20 min on a plate coated with rat anti-mouse CD45 (1.25 μg in 12 ml of DPBS/0.2% BSA; catalog #550539; BD Pharmingen) to harvest microglia. To purify oligodendrocyte-lineage cells, we then incubated cells not bound to the anti-CD45 antibody-coated plate sequentially on four BSL-1-coated plates (8 min each) to deplete endothelial cells and remaining microglia. The remaining cells were next incubated for 30 min on a rat anti-PDGFR receptor α (PDGFRα; 10 μg in 12 ml of DPBS/0.2% BSA; catalog #10R-CD140aMS; Fitzgerald) coated plate to harvest OPCs and then incubated on an additional PDGFRα plate and mouse A2B5 monoclonal antibody ascites

(American Type Culture Collection) coated plate for 30 min each to deplete remaining OPCs. The cell suspension was next incubated on an anti-myelin oligodendrocyte glycoprotein (MOG) hybridoma-coated plate for 30 min to harvest myelinating oligodendrocytes (MOs), followed by an additional anti-MOG hybridoma-coated plate for 30 min to deplete any remaining MOs. Finally, the cell suspension was incubated on an anti-GalC hybridoma-coated plate for 30 min to harvest newly formed oligodendrocytes (NFOs). For purification of RNA, the cells were lysed while still attached to the panning plate with Qiazol reagent (catalog #217004; Qiagen), and total RNA was purified as described below.

**Purification of pericytes.** To purify pericytes, a single-cell suspension was prepared as described above and incubated at 34°C for 1 h to allow expression of cell-surface protein antigens digested by papain. The cells were incubated on three sequential panning plates coated with anti-CD45 (1.25 µg in 12 ml of DPBS/0.2% BSA) antibody to deplete microglia and macrophages (30 min each) and then incubated 30 min on a panning plate coated with anti-CD31 antibody (6 µg in 12 ml of DPBS/0.2% BSA; catalog #10RCD31gRT; Fitzgerald) to deplete endothelial cells. The cells were finally incubated 45 min on a panning plate coated with anti-PDGFR $\beta$  antibody (20 µg in 12 ml of DPBS/0.2% BSA; catalog #ab91066; Abcam) to harvest pericytes. Cells were removed from the plate by trypsin digestion as described above.

#### RNA library construction and sequencing

The polyadenylated fraction of RNA isolated from brain cell types was used for 100 bp paired-end RNA-Seq. Total RNA was extracted using the miRNeasy kit (Qiagen) under the protocols of the manufacturer. The quality was assessed by Bioanalyzer. Samples with high RNA integrity number (>8) were used for library construction. One hundred nanograms of total RNAs were used for each sequencing library. We used the TruSeq RNA Sample Prep Kit (Illumina) to construct poly(A) selected paired-end sequencing libraries following the instructions in the TruSeq RNA Sample Preparation V2 Guide (Illumina). All libraries were then sequenced using the Illumina HiSeq 2000 Sequencer. Two replicates of pooled animals for each cell type were sequenced. To minimize batch effects in library preparation and sequencing, samples were collected and sequenced in the largest feasible group size and performed by the same individuals.

#### Read mapping, transcript assembly, and expression level estimation

Mapping of Illumina 100 bp paired-end reads to the mouse reference genome [University of California, Santa Cruz (UCSC) Genome Browser version mm9] was performed using TopHat software (version 1.3.3; Trapnell et al., 2010), which invokes Bowtie (version 0.12.7) as an internal read mapper (Langmead et al., 2009). TopHat was designed to align reads from an experiment to a reference genome without relying on known splice sites and can be used to identify novel splice variants of genes. Read mapping (TopHat) was run with default settings and -G option, which supplies TopHat with gene model annotation of known transcripts (Illumina iGenome UCSC mm9.gtf annotation file; downloaded from <http://cufflinks.ccb.umd.edu/igenomes.html>) to facilitate read mapping. After read mapping, transcripts were then assembled using Cufflinks software (version 1.3.0; Trapnell et al., 2010). Expression level estimation was reported as fragments per kilobase of transcript sequence per million mapped fragments (FPKM) value together with confidence intervals for each sample. An in-house pipeline was developed to automate the abovementioned mapping and assembly process.

#### Determination of the threshold for reliable FPKM estimation

To facilitate downstream analyses (such as gene fold change analysis, etc.) and to assess the accuracy and reliability of our RNA-Seq experiments, we applied a widely used procedure to determine the threshold for reliable FPKM estimation based on optimizing the intersection of false-positive and false-negative rates (Ramskold et al., 2009; Lerch et al., 2012; Chen et al., 2013). We conducted an analysis based on the 95% confidence intervals of FPKM values as calculated by Cufflinks (Trapnell et al., 2010). Genes whose estimated FPKM values had lower confidence bounds of 0 were labeled as “unreliable.” We calculated the numbers of reliable and unreliable FPKM values at various FPKM thresholds and

produced false-positive and false-negative rate curves to identify an FPKM value that is optimized for both false positives and negatives. We determined an FPKM value of ~0.04 as a threshold for minimum gene expression according to the above procedure and chose a more conservative FPKM threshold of 0.1 for the following analyses. Therefore, genes expressed at FPKM values >0.1 are expressed at statistically significant levels (>99% confidence), although this represents the bottom of an exceedingly large dynamic range (>5 orders of magnitude) of FPKM values. Any FPKM that is <0.1 were set to 0.1 for fold enrichment calculations to avoid ratio inflation (Quackenbush, 2002). The RNA-Seq data have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) and is accessible through GEO Series accession number GSE52564.

#### Analysis of long noncoding RNAs

To identify cell type-enriched long noncoding RNAs (lncRNAs), we obtained the most current lncRNA annotations from the GENCODE project and created an annotation file that included both Ref-Seq and lncRNA genes. We then recalculated expression levels with Cufflinks using the amended annotation file and extracted corresponding FPKM values for lncRNAs.

#### Analysis of cell type-enriched genes, transcription factors, signaling pathways, and metabolic pathways

Differential expression was calculated as the FPKM of a given cell type divided by the average FPKM of all other cell types. Genes were ranked by their fold enrichment in each cell type, and top enriched genes for each cell type were identified (see Fig. 3B). To identify cell type-enriched transcription factors, we filtered our data with a previously annotated transcription factor database (Zhang et al., 2012), and top enriched transcription factors of each cell type were identified as described above.

Compared with the other seven cell types, the pericyte samples are confounded by a small number of contaminating astrocytes and endothelial cells. Accordingly, we did not include these data in any differential expression calculations of the remaining cell types. Nevertheless, when calculating pericyte-enriched genes, small astrocyte and endothelial cell contaminants do not contribute to the list of pericyte-specific transcripts because astrocyte and endothelial genes are included in the denominator of the enrichment calculation. Therefore, we have included pericyte data exclusively in lists of top cell type-specific genes (see Fig. 4; Table 1). Considering the current dearth of RNA-Seq pericyte transcriptome data, the pericyte gene expression presented here provides novel opportunities to identify molecular handles to study this important cell population.

We used Ingenuity Pathway Analysis (Ingenuity) to identify signaling pathways and metabolic pathways enriched in each cell type. Genes with FPKM >5 and fold enrichment >5 in each cell type were fed into Ingenuity Pathway Analysis to highlight those pathways with substantial expression values. Information of gene functions in specific signaling pathways and metabolic pathways with experimental evidence from the Ingenuity knowledge base were used for analysis.

#### Microarray analysis

Identical cell samples were used for RNA-Seq and microarray analyses. Each sample was purified with miRNeasy kit (Qiagen) and equally split into two portions, one for RNA-Seq and the other for microarray analysis. Mouse genome 430 2.0 Array (Affymetrix) was used for analysis. The .cel files from the arrays were analyzed with Arraystar 4.0 software (DNAstar) using robust multichip average processing and quantile normalization.

#### qRT-PCR

Forty genes found by RNA-Seq to be enriched in specific cell types were selected for qRT-PCR validation. We designed primers using NCBI primer blast software (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and selected primer pairs with the least probabilities of amplifying nonspecific products as predicted by NCBI primer blast. We designed primer pairs to amplify products that span exon-exon junctions to avoid amplification of genomic DNA. We tested the specificity of the primer pairs by PCR with mouse whole-brain cDNA (catalog #MD201; Zyagen), and we examined the PCR product by agarose gel electrophoresis.

We performed qRT-PCR with the Fluidigm BioMark system using methods modified from a previously described protocol (Citrì et al., 2012). Briefly, purified cell samples for qRT-PCR were prepared identically to the methods described for the RNA-Seq samples. Cells were harvested and lysed with SuperscriptIII CellsDirect Onestep qPCR kit (catalog #11732-020; Invitrogen) according to the instructions of the manufacturer. A 500 nm primer mixture of all primer pairs was mixed with cell lysates and components of the Superscript III CellsDirect Onestep qPCR kit components according to the instructions of the manufacturer. Reverse transcription and 8–13 cycles of specific target amplification were performed in a single program. Specific target amplification products were then treated with 1 U/ $\mu$ l Exonuclease I (catalog #M0293L; New England Biolabs) to degrade unbound primers. The products and exonuclease were incubated at 37°C for 30 min, followed by inactivation at 80°C for 15 min. Exonuclease I-treated samples were diluted 1:1 with DNA suspension buffer (catalog #T0223; Teknova) and mixed with 2 $\times$  SsoFast EvaGreen Supermix (catalog #172-5211; Bio-Rad) and 20 $\times$  DNA Binding Dye Sample Loading Reagent (catalog #100-0388; Fluidigm) before loading to preprimed 96.96 Dynamic Array IFC (Fluidigm). Primer pairs were mixed with 2 $\times$  Assay loading reagent (catalog #85000736; Fluidigm) and DNA suspension buffer (Teknova) and loaded to the same 96.96 Dynamic Array IFC (Fluidigm). In addition to purified cell samples, we included unpurified mouse whole-brain samples as positive controls and reactions without primer pairs or cDNA as negative controls. Thirty cycles of qPCR amplification were performed on the Fluidigm BioMark system, followed by a melting-curve analysis. A total of 9216 (96  $\times$  96) qPCR reactions were performed on a single chip. The Fluidigm BioMark software was used for data analysis and visualization. As a quality-control step, we manually examined the melting curves for each reaction and included only those reactions with a single peak at the expected melting temperature for final data analysis. Threshold was determined automatically, and threshold cycle (Ct) values were calculated by the Fluidigm BioMark software. Gene enrichment was calculated using the  $\Delta\Delta Ct$  method in relation to the housekeeping gene *Gapdh*.

#### Analysis of alternative splicing

We developed a pipeline, Quantas, for analysis of alternative splicing information from RNA-Seq data (<http://zhanglab.c2b2.columbia.edu/index.php?Quanta>). Briefly, we first performed *de novo* mapping of raw reads to the reference genome (mm9), as well as known and novel exon junctions using small seeds with the OLego algorithm (Wu et al., 2013). We then used gapless to infer transcript structure of alternative isoforms between ambiguously located paired-end reads using a simple Bayes model that considers size constraints of each cDNA fragment and previous isoform abundance estimations from mapped junction reads. We then used countit, a library of codes that quantifies the gapless output against a comprehensive database of alternative splicing events (six in our case; Zhang et al., 2010). We derived the information relative to alternatively spliced events and corresponding genes from the output files. We computed an inclusion score estimating the probability of an exon to be included or excluded for each event type in each sample. The score is calculated as the ratios of the OLego computed counts for the form including the exon over the total. Most splicing events exhibit a bimodal distribution in which the majority of the minor isoforms constitute <20% of the reads. Based on these data, we considered an event to be alternatively spliced only if it had a score between 0.2 and 0.8 to isolate those splicing events with the most biological significance. Finally, we also performed statistical tests (Fisher's test) to determine significance of the splicing events across different cell types. We then formatted output results into html and visualized both aligned reads and mapped junction reads on a web browser.

#### Double-fluorescent *in situ* hybridization

Full-length mouse cDNA expression plasmids were obtained for *Eno2*, *Glast*, *Plp1*, *Cx3cr1*, *Gdpd2*, *Pgapdc1a*, *Eltd1*, *Olfml3*, and *Lppr3* (Thermo Fisher Scientific). A full-length *Cldn5* expression plasmid was a gift from D. Agalliu (University of California, Irvine, Irvine, CA). Digoxigenin (DIG)- or fluorescein-labeled single-stranded antisense and sense riboprobes were prepared by linearizing plasmids and transcription with T7, Sp6, or T3 RNA polymerases and an RNA labeling kit (Roche) according

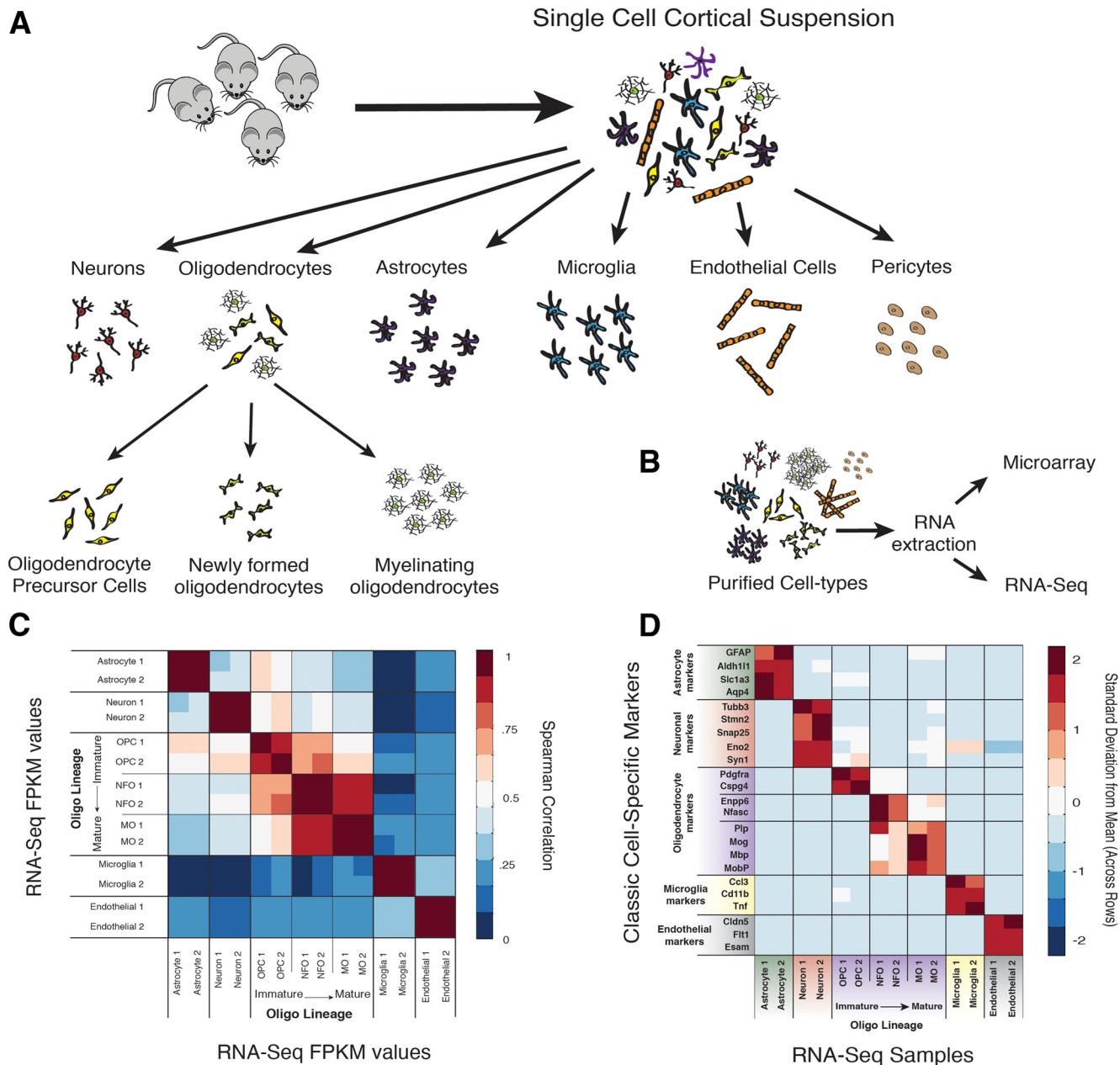
to the instructions of the manufacturer. Fresh frozen 16- $\mu$ m-thick brain sections were processed as described previously (Schaeren-Wiemers and Gerfin-Moser, 1993), with the following modifications for fluorescent *in situ* hybridization: after hybridization, sections were treated three times (20 min each) in 3% hydrogen peroxide (Sigma-Aldrich) in PBS to block endogenous peroxidase activity, washed three times in buffer B1 (0.1 M Tris-HCl, pH 7.5, and 150 mM NaCl), and blocked in blocking reagent (PerkinElmer Life and Analytical Sciences) in buffer B1 plus 0.1% Triton X-100 for 1 h. Anti-DIG POD or anti-FITC POD (Roche) antibodies were cross-absorbed against lightly fixed brain tissue to decrease nonspecific binding, diluted in blocking buffer (1:400), and applied to tissue overnight at 4°C. Slides were developed using a fluorescein or Cy3 Tyramide Signal Amplification Plus kit (PerkinElmer Life and Analytical Sciences). For double-fluorescent *in situ* hybridization, the protocol was repeated starting with the peroxidase blocking step, using the alternate antibody, and a different fluorophore. Slides were extensively washed in PBS and coverslipped with Vectashield with DAPI (Vector Laboratories) for imaging.

## Results

#### Creation of an RNA-Seq transcriptome library

To maximize biological diversity, we pooled dissected cerebral cortices from two litters of mice for each cell type that we profiled. Purified cell populations were obtained by immunopanning with cell type-specific cell-surface antibodies and FACS of transgenically labeled cell populations (astrocytes with Aldh11-GFP mice and endothelial cells with Tie2-GFP mice). Despite the fact that our protocols for immunopanning and FACS differ only in the final step of cell isolation, we wanted to ensure that the two methods did not affect gene expression differently. Therefore, we used both immunopanning and FACS to isolate astrocytes from mouse brain cortices and directly compared expression profiles of the two populations. As expected, both methodologies produced highly pure cell populations with minimal variation in gene expression (data not shown). Selecting an ideal time point for transcriptome profiling is a balance between maximizing cell maturation with increasing difficulty of cell isolation over time. We chose P7 because it is an age when differentiated populations of neurons, astrocytes, and endothelial cells are present and can be acutely purified with minimal activation. At P7, astrocytes are robustly growing and promoting synapse formation. Although these astrocytes have not reached their final mature size and morphology, their gene expression profiles closely resemble that of mature astrocytes isolated from P30 brains (Cahoy et al., 2008). Oligodendrocyte-lineage cell isolation occurred at P17 because it represents the earliest time point when the full collection of oligodendrocyte-lineage cells are present. We extracted RNA from purified cell populations and performed RNA-Seq with the Illumina HiSeq 2000 platform. We obtained  $65.6 \pm 5.4$  million (data represent mean  $\pm$  SD unless otherwise specified) 100 bp reads per sample. We successfully mapped  $86.9 \pm 6.4\%$  of the fragments to the genome. Few fragments were mapped to multiple locations of the genome ( $1.5 \pm 0.7\%$ ).

To assess the reproducibility of our data and conservation across biological replicates, we calculated correlations across all RNA-Seq samples and found high correlations among cell type replicates (Spearman's rank correlation, mean  $r = 0.979 \pm 0.004$ ) and lower correlations across differing cell types (Spearman's rank correlation, mean  $r = 0.714 \pm 0.008$ ). A notable and expected exception is that NFOs and MOs exhibit a tighter correlation (mean  $r = 0.873 \pm 0.010$ ) than typically observed across differing cell types (Fig. 1C).



**Figure 1.** Purification of neurons, glia, and vascular cells. **A**, We purified eight cell types—neurons, astrocytes, OPCs, NFOs, MOs, microglia, endothelial cells, and pericytes—from mouse cerebral cortex with a combination of immunopanning and FACS procedures (for details, see Materials and Methods). **B**, RNA was extracted from purified cells and analyzed by microarray and RNA-Seq. **C**, Spearman's rank correlation of RNA-Seq biologically independent replicates. Each replicate consists of pooled cortices from 3–12 animals. **D**, Expression of classic cell-specific markers in purified glia, neurons, and vascular cells samples determined by RNA-Seq. Two biological replicates of each cell type are shown. Specific expression of known cell-specific markers demonstrates the purity of the glial, neuronal, and vascular samples.

### Purity of neuronal, glial, and vascular samples

We next wanted to validate the purity of the isolated brain cell types. We probed the transcriptome data for expression of well known cell type-specific genes for astrocytes (e.g., *Gfap*, *Aldh1l1*, *Slc1a3*, *Aqp4*), neurons (e.g., *Tubb3*, *Stmn2*, *Snap25*, *Eno2*, *Syn1*), OPCs (e.g., *Pdgfra*, *Cspg4*), NFOs (e.g., *Enpp6*, *Nfasc*), MOs (e.g., *Plp*, *Mog*, *Mbp*, *Mobp*), microglia (e.g., *Ccl3*, *Cd11b*, *Tnf*), and endothelial cells (*Cldn5*, *Flt1*, *Esam*) (Fig. 1D; Dugas et al., 2006; Cahoy et al., 2008; Daneman et al., 2010; Beutner et al., 2013; Chiu et al., 2013). These classic cell type-specific markers each exhibited high expression levels in their corresponding cell types and undetectable or extremely low expression levels by the remaining cell populations (Fig. 1D). As expected, some of the

oligodendrocyte-lineage markers are expressed in a graded manner at high levels during one maturation stage and at lower levels at a different maturation stage (see OPC, NFO, and MO data in Fig. 1D; Dugas et al., 2006). These data confirmed the purity of the various isolated cell types and established the feasibility of constructing a high-quality transcriptome database.

### Improved sensitivity of RNA-Seq versus microarray

We chose to use RNA-Seq to construct our transcriptome database because it offers several substantial advantages over microarray-based platforms. Comparing these two technologies, RNA-Seq is believed to have increased sensitivity, improved linearity, and a vastly larger dynamic range (Marioni et al., 2008;

Mortazavi et al., 2008; Wang et al., 2009). Furthermore, microarray experiments also rely on species-specific predesigned probes of varying quality that preclude analysis of unannotated genes, which substantially limits data comparison across different studies or species. In contrast, RNA-Seq generates results at single-nucleotide resolution without a priori knowledge and includes results of both known and novel genes. The high reproducibility and platform agnostic nature of RNA-Seq allows direct comparison of data from different studies across diverse species.

Expression estimation by RNA-Seq can be biased by factors including non-uniform distribution of cDNA fragments and differences in fragment GC content. To correct for these biases, we adopted the widely accepted Cufflinks algorithm to estimate FPKM values. Cufflinks implements a likelihood-based bias correction step that considers fragment size and GC content bias, among others, and has been shown to significantly improve expression estimation (Roberts et al., 2011).

To directly compare the quality of our RNA-Seq transcriptome with microarray-generated data, we compared the two technologies on identical RNA samples. We evenly divided the RNA isolated from several of the brain cell types into two groups. Half of the RNA from each cell type was analyzed by RNA-Seq, whereas the other identical half was analyzed using microarray. Comparing identical samples by the two platforms back-to-back made it possible to validate the RNA-Seq data with an orthogonal methodology and to confirm previously reported advantages of RNA-Seq over the microarray platform.

To directly compare the RNA-Seq and microarray data, we calculated the correlation between expression of those transcripts that were present in both platforms (Fig. 2A). As expected, samples from identical cell types demonstrated high correlations (Spearman's rank correlation, mean  $r = 0.945 \pm 0.014$ ) in contrast to those samples from different cell types (Spearman's rank correlation, mean  $r = 0.721 \pm 0.009$ ), indicating that both RNA-Seq and microarray are consistent in identifying characteristic gene expression signatures of distinct cell classes. Although most correlation values between mismatched cell types were low, we again found slightly higher correlations between oligodendrocytes of different maturation stages (OPCs, NFOs, and MOs). These trends correctly represented the developmental and functional similarities and differences of glia, neurons, and vascular cells in the brain. Biological replicates of identical cell types showed high correlation as determined from either method, indicating that our cell purification methods are highly reproducible.

To compare the sensitivities of the two technologies, we compared the distributions of cell type-specific gene enrichments. For a given cell type, we calculated the fold enrichment of each gene—defined as expression of the gene in the given cell type divided by the average expression level of the gene in all other cell types—for both RNA-Seq and microarray datasets. We plotted these fold enrichment values for each gene to compare the two platforms directly (Fig. 2B–E). The resulting scatter plots did not reveal the 1:1 trend line expected if both technologies were equally sensitive but rather produced plots that are stretched across the RNA-Seq axis ( $x$ -axis). These data indicate that RNA-Seq identified a greater number of differentially expressed genes than did the microarray platform, likely because of the higher sensitivity of RNA-Seq in detecting low-abundance transcripts and the increased linear range of RNA-Seq compared with the microarrays.

We next examined the overlap between genes that were identified as differentially expressed by RNA-Seq and microarrays.

We identified differentially expressed genes (using a fourfold enrichment or depletion cutoff) in both modalities and found that the majority of genes classified as differentially expressed by microarray were similarly identified by RNA-Seq (mean,  $92.5 \pm 2.6\%$ ; for cell type-specific counts, see Fig. 2B–E). Conversely, RNA-Seq consistently identified a significant number of differentially expressed genes that were not classified as differentially expressed in the microarray dataset [ranging from 1471 (oligodendrocytes) to 3529 (microglia) genes; Fig. 2B–E]. This trend was consistent across all fold enrichment cutoffs (from 1.5 to 50; data not shown). Together, these data indicate that the RNA-Seq platform accurately identified many of the expression differences observed with the microarray analysis but also detected a substantially larger number of new genes with previously unknown cell type-specific distributions.

To investigate the relationship between enrichment and transcript abundance, we plotted the fold enrichment of all genes against their transcript abundance in the whole cortex. Consistent among all cell types, we observed a random distribution of enrichment scores across all transcript abundance levels. In other words, differential gene expression (as determined by RNA-Seq) is not significantly biased by transcript abundance in the cortex. We have included one such enrichment/expression plot for the astrocyte data in Figure 2F as a representative example.

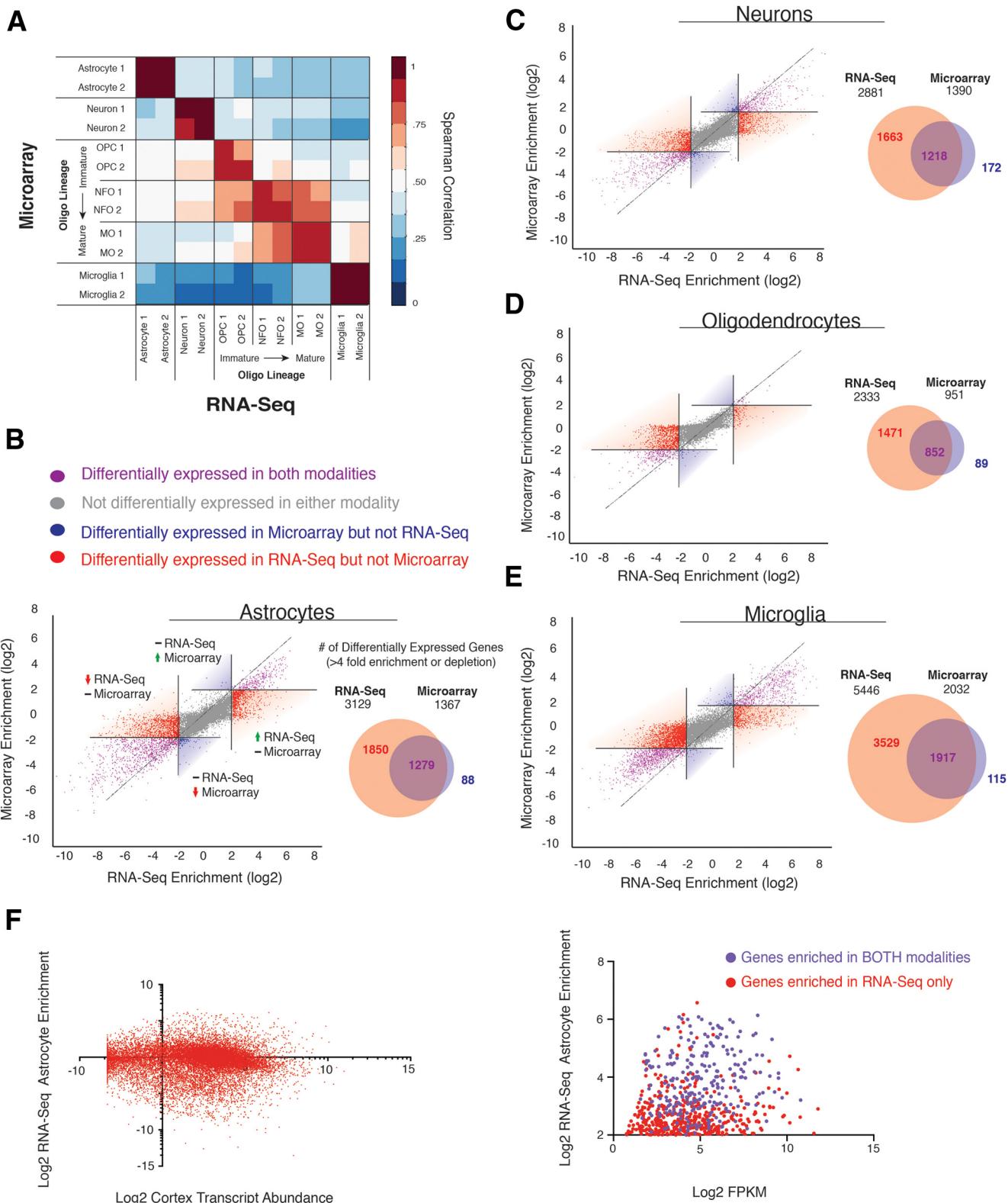
To further examine the genes detected as specifically enriched in the RNA-Seq data, we compared the expression levels of cell type-enriched genes identified by both RNA-Seq and microarray versus those identified as cell type-enriched by RNA-Seq alone (Fig. 2G). Here we observed a slight trend where genes enriched in the RNA-Seq data alone were clustered toward lower FPKM expression levels compared with those genes that were classified as enriched in both modalities. This likely reflects a consequence of the improved sensitivity and linearity of RNA-Seq compared with microarray, and it is from this pool of lesser expressed genes that we chose to validate new candidates with qPCR (see Fig. 6).

### RNA-Seq analysis reveals distinct gene expression signatures of each cell type

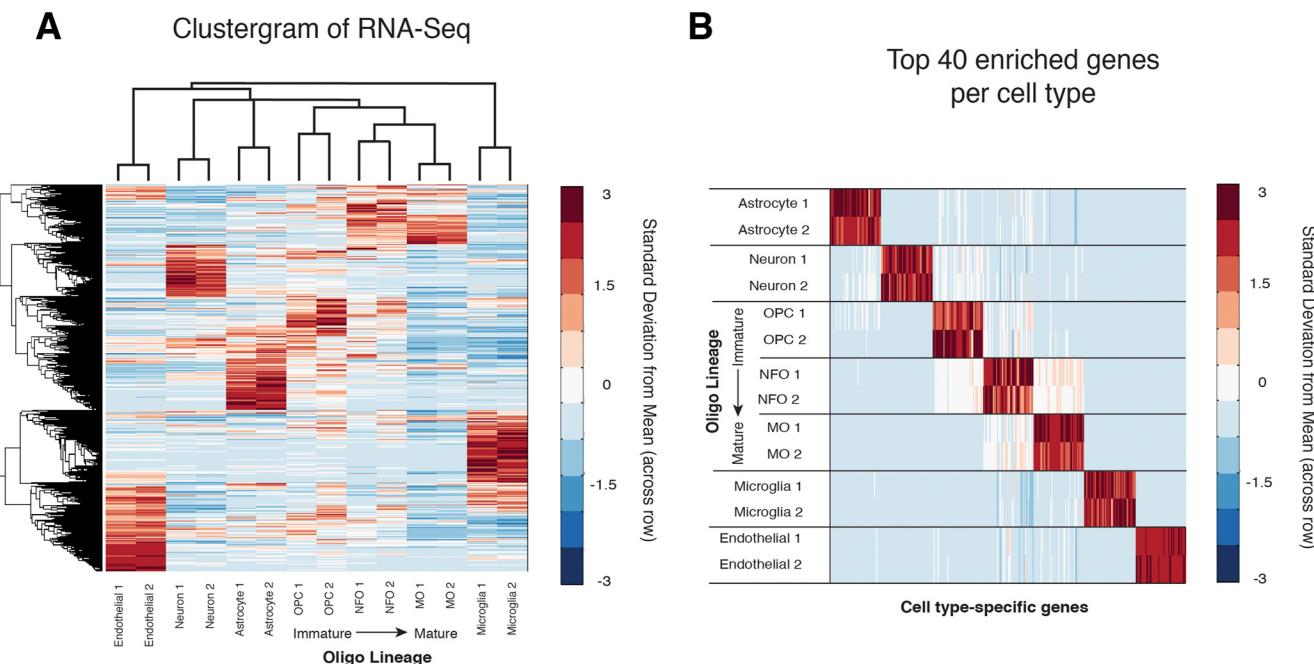
To assess the relationship between gene expression profiles of different cell types, we performed unsupervised hierarchical clustering of our complete RNA-Seq transcriptome data (Fig. 3A). As expected, different maturation stages within the oligodendrocyte lineage clustered together. Although astrocytes, oligodendrocytes, and microglia are all considered glia, they do not cluster distinctly from neurons as is typically assumed based on their misleading nomenclature. These glial cell types are actually as dissimilar from each other as they are from neurons (Cahoy et al., 2008). These differences are consistent with their distinct functions and unique developmental origins. Indeed, ectoderm-derived neurons, astrocytes, and oligodendrocyte-lineage cells cluster more closely together than do the mesoderm-derived microglia and endothelial cells. The clustering heat map reveals distinct groups of genes specifically expressed by each cell type that appear as red blocks in Figure 3A. Principal component analysis revealed similar distinctions between cell classes (data not shown).

### RNA-Seq analysis identifies novel cell type-enriched genes

We analyzed the most highly enriched genes for each cell type (Figs. 3B, 4). These genes are highly expressed (FPKM >20) by one cell type and undetectable or expressed at low levels by other cell types in the brain. As expected, several genes on this list are well known cell type-specific markers, e.g., *Aqp4*, *Aldh1l1*, and



**Figure 2.** Comparison of RNA-Seq and microarray analyses showed that RNA-Seq identified more differentially expressed genes across cell types than microarray. **A**, Spearman's rank correlation between gene expression data obtained by the RNA-Seq and microarray methods across cell types. Gene expression profiles of the same cell type obtained by the RNA-Seq and microarray methods showed a high degree of correlation. **B**, Numbers of differentially expressed genes identified by the RNA-Seq and microarray methods. A snapshot of the data is summarized in the Venn diagram. Using a fourfold difference as the cutoff, RNA-Seq analysis identified 3129 genes as differentially expressed by astrocytes, whereas microarray identified only 1367 genes. The majority of genes identified by microarray as differentially expressed are similarly classified by RNA-Seq (1279). RNA-Seq identified an additional 1850 genes as differentially expressed that were not identified by microarray as differentially expressed. **C-E**, RNA-Seq versus microarray comparisons for neurons, oligodendrocytes, and microglia. **F**, The relationship between fold enrichment and expression level in astrocytes. There is a sharp cutoff line on the left ( $\text{FPKM} = 0.1$  or  $-3.3$  on the log<sub>2</sub> scale) because we set any FPKM value  $<0.1$  to 0.1 to avoid ratio inflation in fold enrichment calculations. **G**, The relationship between fold enrichment and expression level in astrocytes.



**Figure 3.** RNA-Seq analysis revealed cell type-specific gene expression profiles. **A**, Dendrogram and unsupervised hierarchical clustering heat map (using Euclidean distance) of purified cortical glia, neurons, and vascular cells. The vertical distances on each branch of the dendrogram represent the degree of similarity between gene expression profiles of various samples. Biological replicates showed the highest degree of correlation represented by short vertical distances. Cells in the oligodendrocyte lineage cluster closely together, and the order of the three oligodendrocyte-lineage cell types corresponds to their maturation stages (OPC–NFO–MO). Although astrocytes and oligodendrocytes are both glial cells, their gene expression profiles are as different between each other as they are different from neurons. Consistent with their embryonic origin, mesodermal-derived endothelial cells and microglia cluster farther away from ectoderm-derived neurons, astrocytes, and oligodendrocytes. **B**, The top 40 enriched genes per cell type are shown in a heat map. Only highly expressed genes with FPKM > 20 are included in this analysis. Fold enrichment is calculated as FPKM of one cell type divided by the average FPKM of all other cell types. The majority of these genes showed specific expression by only one cell type, with the exception that some are expressed during more than one maturation stage in the oligodendrocyte lineage.

*Fgf3* for astrocytes, *Dlx1* and *Stmn2* for neurons, *Pdgfra* and *Cspg4* for OPCs, *Mbp* and *Mog* for MOs, *Tnf* and *C1qα* for microglia, and *Cldn5* for endothelial cells (Fig. 4). Additionally, however, we discovered a large number of cell type-enriched genes that have not yet been described previously as cell type-specific in the literature to the best of our knowledge. The restricted expression pattern of these genes reveals potential cell type-specific roles in brain development, signaling, metabolism, and disease. For example, we found astrocyclic enrichment of the autism and schizophrenia-associated gene *Tspan7*, neuronal enrichment of the gene encoding a novel transmembrane protein *Tmem59l* that is homologous to proteins implicated in the processing of amyloid precursor protein, and OPC enrichment of the gene encoding *Neurexophilin1*, a protein involved in synapse formation. Additional novel cell type-specific genes are included in Figure 4.

Transcription factors are key regulators of gene expression. Cell type-specific transcription factor expression is important for cell fate determination and cell differentiation (e.g., *NeuroD* for specification of neuron cell fate, *Olig2* for specification of oligodendrocyte fate, and *GM98/Myrf* for oligodendrocyte differentiation; Lee et al., 1995; Zhou and Anderson, 2002; Emery et al., 2009). Our RNA-Seq analysis identified numerous cell type-specific or enriched transcription factors (Fig. 4). These include well known transcription factors (e.g., *Sox9* specific to astrocytes, *Dlx1*, *Dlx5*, and *Tbr1* specific to neurons, *Olig1* and *Olig2* specific to OPCs, *Nkx2-2*, *Nkx6-2*, and *Myrf* specific to oligodendrocytes) and a large number of transcription factors that were not recognized previously to be cell type-specific. Additional investigation of these genes is likely to be a fruitful approach to understanding the development of the CNS.

What are the global differences in the signaling and metabolic pathways among cell types of the brain? Historically, this question has been difficult to address without a highly linear transcriptome database. Here, we performed Ingenuity Pathway Analysis enriched by each cell type among all expressed genes from our samples and identified distinct signaling and metabolic signatures for neurons, glia and vascular cells of the brain. As expected, our analysis identified the enrichment of a collection of neurotransmitter receptor signaling pathways in neurons, immune cell signaling pathways in microglia, and epithelial adherence junction signaling pathways in endothelial cells. Moreover, we found the enrichment of NF- $\kappa$ B, Wnt/ $\beta$ -catenin, and sonic hedgehog signaling pathways in astrocytes, suggesting the importance of these major signaling pathways in the development of astrocytes and/or the induction of reactive astrogliosis. We detected the enrichment of the inhibition of matrix metalloproteases and intrinsic prothrombin activation pathways in OPCs, suggesting that OPCs have unique properties in their interactions with the extracellular matrix compared with other cell types of the brain. Metabolic differences among unique cell types of the brain are poorly understood. We systematically analyzed enrichment of metabolic pathways in each cell type of the brain. Highlights include the enrichment of the  $\gamma$ -linolenate biosynthesis pathway in astrocytes, 3-phosphoinositide (IP3) biosynthesis and degradation pathways in oligodendrocytes, the pyridoxal 5'-phosphate salvage pathway in microglia, the citrulline–nitric oxide cycle in endothelial cells, and the retinoate biosynthesis pathway in pericytes. In addition, we conducted weighted gene coexpression network analysis (WGCNA) to cluster genes into coexpression modules using our RNA-Seq data (data available on our RNA-Seq data website; see Notes).

	Astrocytes	Neurons	OPCs	NFO	MO	Microglia	Endothelial	Pericytes
Overall Enriched Genes (ranked by fold-change)	Hgf	Reln	Pdgfr1	Gp1bb	Gjb1	Sifn2	Cldn5	Fmod
	Aqp4	Nhlh2	Lnx1	Tmem108	Ndrg1	Gpr84	Ttr	Rps2
	Itih3	Slc17a6	Dcn	Fyn	Ppp1r14a	Ccr7	Ly6a	Igf2
	Bmpr1b	Trp73	Mmp15	Ust	Adss1	Bcl2a1d	Macam1	Gpc3
	Itga7	Lhx5	Cdo1	Mical3	Aspa	Tnf	8430408G22Rik	Ogn
	Plcd4	Lhx1as	Sapcd2	Kif19a	Acy3	Ncf1	Akr1c14	Lrrc32
	Grm3	Dlx6os1	Kcnk1	1810041L15Rik	Trp53inp2	Gdf15	Ly6c2	Fln
	Slc14a1	Sst	Rasgrf1	9630013A20Rik	Pla2g16	Osm	Meox1	Gjb2
	Phkg1	5330417C22Rik	Pcdh15	Nfasc	Efhd1	Lrrc25	Ly6c2	Itih2
	Pla2g3	Mab2111	Chrna4	Ssh3	Itgb4	H2-Oa	Car4	Rdh10
	Cbs	Snhg11	Dli3	Pik3r3	Haplh2	Cd83	Bsg	Bmp6
	Paqr6	Mrap2	Col1a2	Enpp6	Mbp	Ccl3	Aplnr	Aldh1a2
	Aldh1l1	Dlx1	Fam70b	Tns3	Hcn2	Slamf8	Sigirr	Postn
	Cth	Tmem90a	Sstr1	Bmp4	Nmra1	Ccl4	Slc01a4	Sidt1
	Ccdc80	Islr2	Pnlip	Mcl1	Cdc42ep2	Gna15	Slc16a1	Lamc3
	Fmo1	Igfbp1l	Cspg4	Cdv3	Mal	Il1b	Icam2	Slc22a6
	Slc30a10	Gdf5	Lppr1	Tmem163	Mog	Plau	Kank3	Clec3b
	Slc6a11	Stmn2	Ppapdc1a	Rap2a	Slco3a1	Ccl9	Slc19a3	Slc6a13
	Fgfr3	Ecef1	Nxph1	Tmem2	Apod	Tmem119	Fam101b	Bicc1
	Slc4a4	Robo2	Pid1	Cnksr3	Gsn	C1qa	Slc16a4	S100a10
	Gdpd2	Dlx1as	Ugddh	Cyfip2	Pdlim2	Irf8	Nostrin	Rps18
	Ppp1r3c	Celf4	Slitrk1	Fmd4a	Prr18	1810011H11Rik	Sdpr	Serp1
	Grhl1	Celf6	Shc4	Slc12a2	Inf2	Pla2g15	Ptgis	Col1a1
	Entpd2	Nxph4	Smoc1	Itpr2	Tppp3	Cxcl16	Myct1	Dcn
	Egfr	Grm2	Emid1	Rnf122	Tbc1d9b	Ch25h	Vwa1	Col1a2
	AI464131	Npy	Rlbp1	Lims2	Nol3	Hck	Ankrd37	Pcolce
	Otx1	Tbr1	Dcaf12l1	Samd4b	Cenpb	Ccl12	Sox18	Cyp1b1
	Nwd1	Slc32a1	Lypd6	Chn2	Slc45a3	Ptafr	Prnd	Cited1
	Atp13a4	Dlx2	Lhfpl3	Ppp2r3a	Carns1	Cd300a	Dok4	Emp1
	Kcnn3	Npas4	Myt1	Strn	Opalin	Irf5	Serpinb6b	C4b
	Ptx3	Ebf3	Gfia3	GlrB	Arg	Sfpi1	Efna1	Ahnak
	Sorc2	Bcl11a	C1ql1	Rras2	Rlttn1	Selplg	Cd34	S1pr3
	Tnc	Cacna2d2	Tmem179	Fmnl2	Adap1	Sash3	Egfl7	Col3a1
	Sox9	C1stn2	Megf11	Sema5a	Plekhb1	Pltp	Pglyrp1	Fstl1
	Abcd2	Dpysl5	Ncald	Fam3c	Trf	Trem2	Slc35f2	Col4a5
	Fzd10	Vstmn2l	Sdc3	Cdc371	Insc	Tlr2	Cdkn2b	Vtn
	Lrig1	Tmem130	Rprm	Fam73a	Cryab	P2ry6	Fam129a	Lama2
	Mic1	Nppc	Cacng4	Elovl6	Kif5a	Cdl4	Sgms1	Mfap4
	Chrdl1	Vgf	Grin3a	Atrn	Trak2	Bcl2a1a	Flt1	Kcne4
	Aifm3	Bhlhe22	Fam5c	Lrrc42	Cldn11	Bcl2a1c	Tie1	Errfi1
	Astrocytes	Neurons	OPCs	NFO	MO	Microglia	Endothelial	Pericytes
Transcription Factors	Gli1	Nhlh2	Sox10	Myrf	Nkx6.2	Sfpi1	Erg	Tbx15
	Gli2	Trp73	Gsx1	Nkx6.2	Myrf	Irf8	Sox17	Foxc2
	Gli3	Lhx5	Olig1	Sox10	Sox10	Irf5	Foxq1	Twist1
	Otx1	Dlx1	Myt1	Barx2	Sp7	Irf4	Mecom	Tbx18
	Grhl1	Ebf3	Pou3f1	Olig1	Barx2	Batf	Foxf2	Foxd1
	Sox9	Dlx5	Sox8	Nkx2.2	Olig1	Runx1	Sox18	Fosl1
	Hes5	Tbr1	Olig2	Sox8	Pou3f1	Ikzf1	Bcl6b	Heyl
	Rfx4	Dlx2	Sox3	Olig2	Sox8	Cebpa	Sox7	Hic1
	Pax6	Lhx6	Nkx2.2	Sox3	Carhsp1	Mlxip	Meox1	Foxc1
	Dbx2	Bcl11a	Sox6	Mycl1	Nfe2l3	Batf3	Zic3	Prrx2

**Figure 4.** Cell-specific markers and transcription factors. The top 40 genes and top 10 transcription factors ranked by fold enrichment of each cell type are listed. The most highly expressed genes are highlighted. Green, FPKM > 150; blue, FPKM > 50 for transcription factors. Fold enrichment of astrocytes, neurons, microglia, and endothelial cells are calculated as FPKM of one cell type divided by the average FPKM of all other cell types. Fold enrichment of OPC, NFO, and MO are calculated as FPKM of one cell type divided by the average FPKM of all non-oligodendrocyte-lineage cells, to highlight top genes specifically expressed by a particular maturation stage during oligodendrocyte development. Only highly expressed genes with FPKM > 20 are included in the ranking to highlight genes that are most likely to have significant cell type-specific functions.

**Table 1.** Top cell type-enriched lncRNAs

Gene name	Expression (FPKM)								Enrichment
	Astrocyte	Neuron	OPC	NFO	MO	Microglia	Endothelial	Pericyte	
<b>Astrocyte</b>									
Gm3764	773.4	78.3	91.5	21.5	2.1	6.5	1.1	103.4	23.1
Malat1	419.9	161.2	44.8	63.0	12.2	25.0	32.4	120.8	7.4
C130071C03Rik	156.8	48.4	42.3	9.5	2.0	0.3	0.5	25.1	9.1
AW047730	100.3	9.7	42.8	5.5	2.3	1.9	0.5	40.1	9.6
Gm26924	54.9	0.1	0.1	0.1	0.1	1.7	4.9	7.5	46.8
<b>Neuron</b>									
Meg3	6.2	729.7	48.6	11.2	3.3	24.9	1.5	105.5	45.7
Rian	3.8	446.8	35.4	16.6	26.7	5.4	6.1	72.7	28.5
6330403K07Rik	23.4	141.1	69.3	14.0	6.5	2.5	1.2	27.8	7.2
Dlx1os	2.0	29.0	0.9	0.4	0.2	0.1	0.4	2.3	44.6
Dlx6os1	0.1	28.0	0.1	0.1	0.1	0.1	0.2	2.4	63.5
<b>OPC</b>									
2810468N07Rik	5.9	3.3	97.4	199.9	249.5	2.4	0.4	4.7	32.7
Gm11650	1.0	0.8	29.1	6.1	0.1	0.1	0.1	0.6	59.1
Gm16277	1.9	11.4	28.1	14.9	6.5	0.2	0.1	6.7	8.3
Gm4876	1.5	0.8	25.8	19.3	1.4	0.1	1.2	2.7	28.7
3110043A19Rik	0.6	0.9	18.2	3.0	0.1	0.2	0.7	0.4	30.5
<b>NFO</b>									
2810468N07Rik	5.9	3.3	97.4	199.9	249.5	2.4	0.4	4.7	67.1
9630013A20Rik	0.1	0.2	2.4	110.7	26.4	0.2	0.1	0.1	739.3
2410006H16Rik	20.2	8.9	2.7	67.0	30.1	8.1	3.1	19.0	6.7
Sox2ot	8.8	21.7	20.9	49.4	33.2	0.2	0.1	5.0	6.4
1700047M11Rik	0.1	0.1	3.2	30.9	24.4	0.2	0.1	0.2	240.9
<b>MO</b>									
2810468N07Rik	5.9	3.3	97.4	199.9	249.5	2.4	0.4	4.7	83.8
Gm10687	0.5	0.2	0.8	16.0	26.5	0.1	1.5	0.8	45.2
9630013A20Rik	0.1	0.2	2.4	110.7	26.4	0.2	0.1	0.1	176.2
1700047M11Rik	0.1	0.1	3.2	30.9	24.4	0.2	0.1	0.2	190.2
Gm21984	0.2	2.4	0.3	3.1	18.4	0.2	0.1	0.4	25.3
<b>Microglia</b>									
Gm13889	21.7	14.1	43.0	17.2	29.7	575.6	24.3	66.0	23.0
Gm26532	8.0	3.1	5.2	0.5	0.2	101.6	6.6	9.8	25.7
A430104N18Rik	0.1	0.2	1.4	0.1	0.1	60.3	0.1	0.3	181.2
Gm11974	21.6	12.6	9.2	5.0	3.0	49.3	6.6	11.0	5.1
Gm13476	1.4	1.0	0.1	4.4	3.4	38.7	0.1	0.1	22.2
<b>Endothelial cell</b>									
Gm20460	0.9	0.3	0.3	0.2	0.2	0.4	14.9	2.3	40.4
Gm14207	0.2	0.1	0.1	0.1	0.1	0.1	12.8	1.2	106.7
Gm20748	0.2	0.1	0.1	0.1	0.2	0.3	10.9	1.2	70.3
Gm16104	0.3	0.2	0.2	0.1	0.1	0.2	9.4	1.4	51.8
Gm9581	0.2	0.1	0.4	0.7	0.9	0.5	9.1	3.0	19.6
<b>Pericyte</b>									
Mir22hg	12.6	5.1	3.6	2.2	1.3	11.2	7.9	48.1	7.7
Gm17750	17.9	29.0	9.9	1.5	0.5	0.2	0.5	47.5	5.6
Gm17120	0.1	0.2	1.6	0.1	0.1	0.1	3.7	21.0	24.5
Gm14964	9.2	3.1	0.8	0.2	0.3	0.2	1.3	17.3	8.0
H19	2.0	0.7	0.7	0.2	0.1	0.2	1.7	16.3	20.1

The top expressed lncRNAs are listed for each cell type with an enrichment threshold >5.

The eukaryotic genome is transcribed in a developmentally regulated manner to produce large numbers of lncRNA or large intervening noncoding RNAs (Guttman et al., 2009; Mercer et al., 2009), the function of which in gene regulation and cancer pathogenesis are increasingly being recognized. However, the expression landscape of lncRNAs across different cell types has not been characterized in a complex organ such as the brain. We detected the expression of 811 lncRNAs with an inclusive criterion (FPKM >1) and 227 lncRNAs with a stringent criterion (FPKM >5) in at least one cell type. Some lncRNAs are among the highest expressed transcripts in the brain, with 12 having FPKM values >100, placing them among the 7% most highly expressed genes in our dataset. To ask whether we could observe cell type-specific expression of lncRNAs, we compared the numbers of lncRNAs expressed by each cell type. We found that astrocytes and neurons

express large numbers of lncRNAs ( $109 \pm 2$  and  $92 \pm 3$ , respectively), whereas MOs and endothelial cells express fewer lncRNAs ( $44 \pm 3$  and  $48 \pm 3$ , respectively). Some lncRNAs are expressed in cell type-specific or enriched manners (Table 1). Their function in cell type-specific gene regulation warrants future investigation.

How do cells of the brain communicate with each other to coordinate their unique roles in brain development and function? Much of our knowledge of cell–cell interactions is derived from studies of secreted ligands and their transmembrane receptors. For example, astrocytes secrete thrombospondins, which bind their receptors  $\alpha 2\delta 1$  on neurons to stimulate synapse formation (Christopherson et al., 2005; Eroglu et al., 2009). Endothelial cells secrete heparin-binding EGF-like growth factor, which binds its receptor epidermal growth factor receptor on astrocytes to promote astrocyte survival (Foo et al., 2011). Bio-

**Table 2.** Cell type-enriched ligands and receptors

	Astrocyte	Neuron	OPC	NFO	MO	Microglia	Endothelial	Pericyte
Enriched transmembrane receptors	Ptpz1	Gpc1	Pdgfra	Gpr17	Efnb3	Csf1r	Tfrc	Pdgfrb
	Ednrb	Ptpn	Gpr17	Sema4d	Gpr37	Cd83	Pglyrp1	Colec12
	S1pr1	Caly	Itgav	Plxnb3	Sema4d	Tyrobpb	H2-D1	Sfrp1
	Fgfr3	Grin1	Omg	Ddr1	Lpar1	Ccr12	Elt1	S1pr3
	Gabbr1	Ptpn2	Gfra1	Efnb3	Ddr1	B2m	Kdr	Abcc9
	Tnfrsf19	Gria1	Gria3	Gpr37	Omg	Trem2	Eng	Rarres2
	Vcam1	Cnr1	Sstr1	Lpar1	Ephb1	Sirpa	Tie1	Pdgfrl
	Adcyap1r1	Opcml	Il1rap	Pdgfra	Gpr5b	Fcer1g	Cav1	Mrc2
	Gria2	Stx1b	Adora1	Omg	S1pr5	Cd14	Flt1	Fas
	F3	Cxadr	Lypd1	Ephb1	Gpr17	Icam1	Fcgtr	Ednra
	Grm3	Nptxr	Gria4	Sema5a	Plxnb3	Gx3cr1	Sema7a	Lepr
	Dag1	Grm2	Chrna4	Il1rap	Gpr62	Lag3	Lsr	Osmr
	Plxnb1	Robo2	Sema5a	S1pr5	Il1rap	Gpr56	Acvrl1	Gprc5a
	Ntsr2	Kit	Calcr1	Prkcz	Prkcz	Itgam	Tek	Ifitm1
	Fgfr1	Gabrb3	Gabra3	Erbb3	Itgb4	Fcgr3	Gpr116	Ddr2
	Ptch1	Gabrg2	Grin3a	Gpr62	Erbb3	P2ry12	Fzd6	Scarf2
	Fgfr2	Sarm1	Oprl1	Ptpre	Sema5a	P2ry13	Ptprb	Tgfb3
	Gpr19	Gabra2	Grik4	Grik3		Tnfrsf12a	Aplnr	Fzd5
	Itga6	Darc	Plxnb3	Grik2		H2-K1	Ptprg	Npy1r
	Gabbr2	Celsr3	Grik3	Casr		Gpr183	Plxnd1	Celsr1
Enriched ligands	Sparc1	Reln	Matn4	Gsn	Trf	Cst3	Sparsc	Igf2
	Cpe	Sst	Scrg1	Lgi3	Gsn	C1q1a	Sepp1	Vtn
	Cyr61	Npy	Olfm2	Scrg1	Apod	Cd4	Pltp	Cxcl12
	Mfge8	Olfm1	Vcan	Enpp6	Lgi3	Cd3	Igfbp7	Col4a1
	Clu	Dkk3	Emid1	Matn4	Metrn	C1qb	Spock2	Col1a2
	Htra1	Cd27a	Tnr	Tnr	Endod1	C1qc	Ctla2a	Bgn
	Igfbp2	Cx3cl1	Nxph1	Ddr1	Adams4	Selpig	Pglyrp1	Dcn
	Vegfa	Cck	Tim4	Adamts4	Cntn2	Ctsb	Col4a1	Ptgds
	Scg3	Vgf	Spon1	Metrn	Enpp6	B2m	Egfl7	Cxcl1
	Ncan	Vstm2l	Igsf21	Fam3c	Haplh2	Gdf15	AU021092	Col1a1
	Pla2g7	Chgb	Gsn	C1q1	Il1rap	Olfm3	Srgn	Fst1
	Fjx1	Scg2	Fam5c	Vcan	Erbb3	Tnf	Fn1	Col3a1
	Tim3	C1qtnf4	Qpct	Tim4	Slpi	Pla2g15	Kdr	Mdk
	Il18	Cxadr	C1ql3	Il1rap	Klk6	Tcn2	Apln	Igfbp5
	Btbd17	Col6a2	Smoc1	Col11a2	Col11a2	Ly86	Wfdc1	Serpinf1
	Itih3	Resp18	Gpc5	Bmp4	Matn4	Plod1	Angptl4	Nbl1
	Hapl1	Vstm2a	Il1rap	Elfn2	Dlk2	Il1a	Htra3	Nid2
	Lcat	Car11	Dscam	Dlx2	Il23a	Tgfb1	Smpd3a	Islr
	Chrd1	Igfbpl1	Chga	Spon1	Wnt3	Lgals9	Lama4	Ptx3
	Pla2g3	Nppc	Nptx2	Dscam	Npb	Ccl2	Emcn	Vasn

The top 20 transmembrane receptors and top 20 ligands ranked by fold enrichment of each cell type are listed here. Fold enrichment of astrocytes, neurons, microglia, and endothelial cells are calculated as FPKM of one cell type divided by the average FPKM of all other cell types. Fold enrichment of OPCs, NFOs, and MOs are calculated as FPKM of one cell type divided by the average FPKM of all non-oligodendrocyte-lineage cells to highlight top genes specifically expressed by a particular maturation stage during oligodendrocyte development.

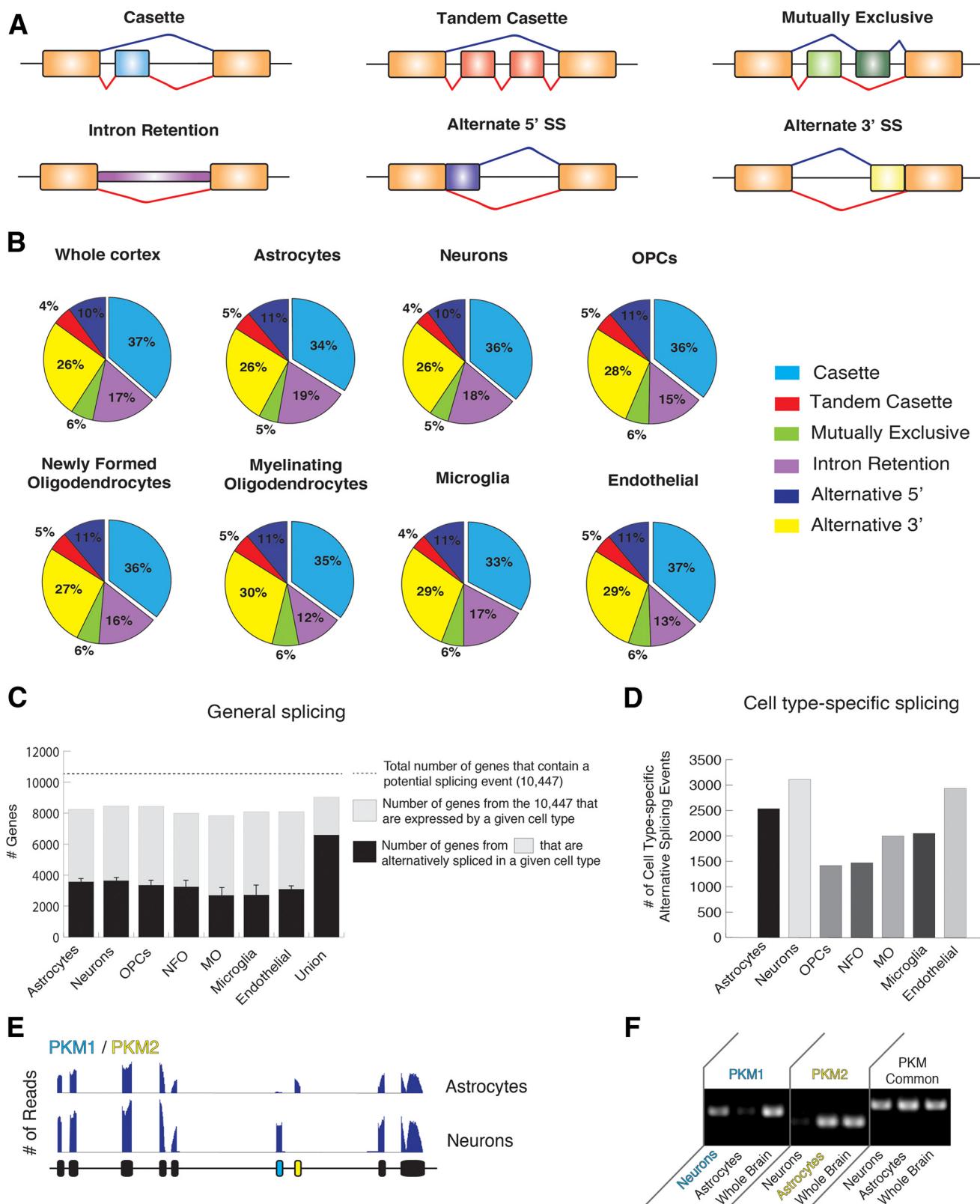
chemical and genetic identification of receptor–ligand pairs in cell types of the brain has been cumbersome and time consuming. Our transcriptome database provides an unprecedented opportunity to systematically identify cell type-specific expression of ligands and their corresponding receptors (Table 2). For example, we identified specific expression of *Lgr6* by astrocytes and its ligands, the *R-spondins*, by neurons. Stimulation of *Lgr6* by R-spondins potentiates Wnt/β-catenin signaling (Gong et al., 2012). Our finding of the cell type-specific expression of this ligand-receptor pair suggests a role of neurons in regulating Wnt/β-catenin signaling in astrocytes. Investigation of the function of these receptors and ligands will tremendously expand our knowledge of neuroglial, neurovascular, and gliovascular interactions in the brain.

### An alternative splicing database of glia, neurons, and vascular cells of the brain

Alternative splicing generates enormous transcriptome complexity by producing multiple mRNA isoforms from a single gene. To capture this transcriptome diversity, we constructed the first

RNA-Seq alternative splicing database of collective populations of glia, neurons, and vascular cells. A daunting task in analyzing alternative splicing from RNA-Seq data is to map hundreds of millions of short reads (usually 50–150 nt in length) back to the reference genome and to detect known or novel splice junctions. For this purpose, we recently developed an algorithm, OLEgo, that has an improved sensitivity and accuracy compared with previously published programs (Wu et al., 2013). The false-negative rate (8.2%) of OLEgo almost halved the false-negative rate of the widely used programs TopHat (15.4%) and PASSion (14.8%). OLEgo is capable of detecting six common alternative splicing event types (detailed in Fig. 5A), including alternative 3' and 5' splice sites, retained introns, and various types of exon inclusion/exclusion events.

To address the global question of how many genes in the brain are alternatively spliced, we mapped RNA-Seq reads from neurons, astrocytes, OPCs, NFOs, MOs, microglia, and endothelial cells with the OLEgo algorithm. This analysis identified 6588 total genes that are alternatively spliced in at least one cortical cell type. We next considered how alternative splicing varies across each of these differing cell types. We found that the overall frequency of



**Figure 5.** Alternative splicing analysis of neurons, glia and vascular cells. **A**, Six types of alternative splicing events are detected by RNA-Seq. Boxes and black lines represent exons and introns, respectively. Blue lines and red lines represent alternative splicing events detected in the dataset. The 5' end is to the left, and the 3' end is to the right. Cassette, The inclusion or exclusion of an exon; Tandem Cassette, the inclusion or exclusion of two or more tandem exons; Mutually Exclusive, the inclusion of one exon in one transcript and inclusion of a different exon in another transcript; Intron Retention, the inclusion or exclusion of a segment previously annotated to be an intron; Alternative 5' SS, the alternative usage of a splicing site on the 5' end of an exon; Alternative 3' SS, the alternative usage of a splicing site on the 3' end of an exon. **B**, Frequencies of the six types of alternative splicing events detected in the entire dataset and in individual cell types. In all cell types, cassette exon events, i.e., the inclusion or exclusion of an exon, are the most frequently detected alternative splicing events. **C**, The numbers of genes that are alternatively spliced in each cell type and the union of these samples. The dotted line represents the total number of genes that are known to contain a potential splicing event in the mouse genome. The number of these genes that are expressed in a given cell type are represented by gray bars. The black bars indicate the numbers of genes that are alternatively spliced in a given cell type based on criteria (Figure legend continues.).

alternative splicing ( $5549 \pm 674$  events corresponding to  $3187 \pm 347$  unique genes per cell type) and the distribution of various splicing event types were similar across all of our samples (Fig. 5*B,C*). Cassette exon events consistently comprised the most frequent alternative splicing event ( $35.3 \pm 1.3\%$ ), followed by a nearly identical distribution of the remaining five splicing event types (Fig. 5*B*). Despite the overall similarity in the frequency of alternative splicing, we detected thousands of cell type-specific alternatively spliced genes. Neurons contain the greatest number of specific alternative splicing events (3110) in contrast to oligodendrocyte-lineage cells, which contained less than half as many (1469 unique splicing events in OPCs; Fig. 5*D*). This observation likely reflects the significant heterogeneity that is present among various neuronal subtypes. The most significant splicing events enriched in each cell type can be found in Table 3. How many alternative splicing events of a gene, on average, are present in a given cell type? Per cell type, we detected  $5549 \pm 275$  alternative splicing events in  $3171 \pm 142$  alternatively spliced genes, resulting in  $\sim 1.7$  alternative splicing events per alternatively spliced gene. To provide easy access to the alternative splicing data that we generated, we deposited the complete splicing dataset in an interactive web browser and database (see Notes).

To test the accuracy of our splicing database, we looked for genes reported previously to be differentially spliced in neurons and glia. Two particularly interesting examples include the differential expression of neuronal and glial isoforms of *Nfasc* and *Sgce* genes (Zonta et al., 2008; Ritz et al., 2011), both of which are confirmed in our splicing database, among others. Next, to validate the accuracy of the novel cell type-specific splicing events in our dataset, we searched for a biologically significant gene candidate to validate via PCR. One of our top cell type-specific splicing events involved the gene pyruvate kinase (*Pkm*), which catalyzes the final step of glycolysis (Tani et al., 1988). Our RNA-Seq data indicated that neurons and glial cells express distinct splicing isoforms of the *Pkm* gene, *Pkm1* and *Pkm2*, respectively, and we confirmed this result using RT-PCR with isoform-specific primers (Fig. 5*E,F*). Unlike *Pkm1*, the *Pkm2* isoenzyme is allosterically regulated to dynamically regulate the relative rate of energy production (see below, Regulation of energy metabolism in neurons and astrocytes).

#### qRT-PCR and *in situ* hybridization validate RNA-Seq results

We next used two different methods, qRT-PCR and *in situ* hybridization, to determine the accuracy of the cell type-enriched genes identified by RNA-Seq. We selected 40 genes that had not been identified previously in the literature as cell type-specific for our validation. Some of these genes are expressed at moderate levels and others at low levels. We chose genes with low expres-

&lt;

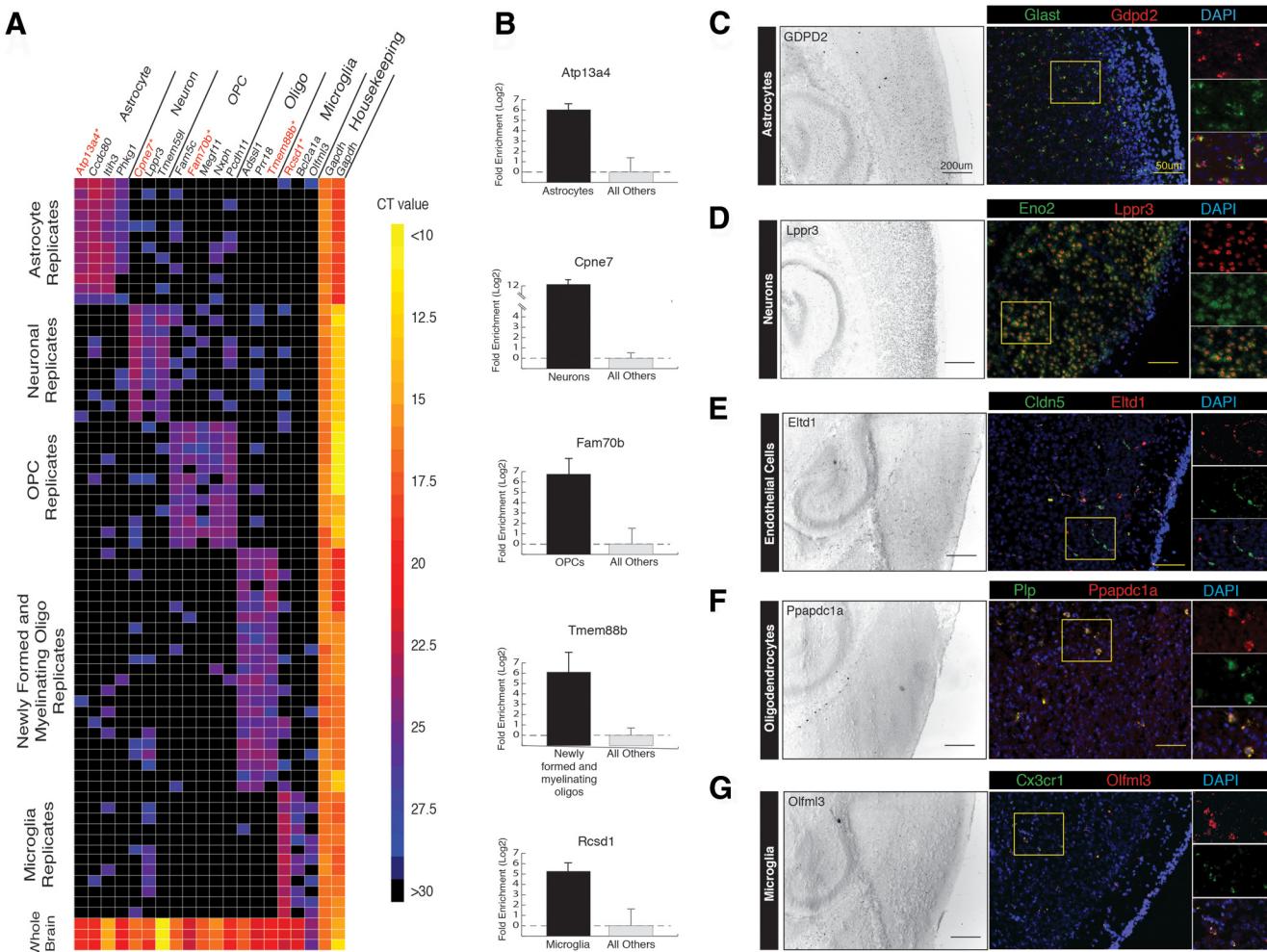
(Figure legend continued.) outlined in Materials and Methods. *D*, The numbers of statistically significant cell type-specific alternative splicing events in each cell type. Neurons have the highest number of specific splicing events, whereas oligodendrocyte-lineage cells have the least amount of specific splicing events. *E*, *Pkm2* is an example of a gene spliced uniquely in astrocytes and neurons. The traces represent raw data of the number of reads mapped to the *Pkm2* gene from astrocytes and neurons. The height of the blue bars represents number of reads. The bottom schematic is the transcript model of *Pkm2* gene from the UCSC Genome Browser. Boxes represent exons, and black lines represent introns. The exon shown in blue is predominantly included in neurons, whereas the exon shown in yellow is included only in astrocytes. This is an example of a mutually exclusive event. *F*, Validation of *PKM1/2* splicing differences by PCR. We designed primers targeting exons unique to *PKM1*, unique to *PKM2*, and exons common to *PKM1* and *PKM2*. PCR products were detected from neuron, astrocyte, and whole-brain samples in patterns predicted by the RNA-Seq data.

**Table 3. Cell type-specific splicing events**

	Gene	Coverage	dl	p value	FDR
Astrocyte	<i>Fyn</i>	771	-0.27	1.15 E-176	1.95 E-174
	<i>Prom1</i>	251	0.86	1.12 E-133	4.59 E-131
	<i>Ncam1</i>	3576	0.24	1.63 E-132	5.93 E-130
	<i>Ptpnfr</i>	480	-0.58	1.03 E-120	2.98 E-118
	<i>Srgap3</i>	308	0.69	1.99 E-84	1.44 E-82
	<i>Kif1a</i>	180	0.75	3.12 E-66	1.76 E-64
	<i>Ptk2</i>	134	-0.47	1.33 E-63	6.73 E-62
	<i>Camk2g</i>	903	-0.44	5.75 E-56	2.43 E-54
	<i>Mapk8</i>	270	-0.55	6.98 E-37	2.36 E-35
	<i>Pkm2</i>	1112	-0.23	2.85 E-25	6.58 E-24
Neurons	<i>Agrn</i>	907	-0.5	<1 E-300	<1 E-300
	<i>App</i>	5181	0.2	<1 E-300	<1 E-300
	<i>Atp6v0a1</i>	1815	-0.66	<1 E-300	<1 E-300
	<i>Cltc</i>	3032	-0.81	<1 E-300	<1 E-300
	<i>Dync1i2</i>	1618	-0.82	<1 E-300	<1 E-300
	<i>Nfasc</i>	821	-0.94	<1 E-300	<1 E-300
	<i>Rab6</i>	3058	0.34	<1 E-300	<1 E-300
	<i>Mtss1</i>	1244	-0.6	5.19 E-216	7.02 E-214
	<i>Srgap3</i>	834	0.62	1.69 E-132	1.52 E-130
	<i>Lrp8</i>	328	-0.75	3.23 E-119	2.51 E-117
Oligodendrocytes	<i>Phd1b1</i>	2947	-0.58	<1 E-300	<1 E-300
	<i>Aplp2</i>	1358	0.48	<1 E-300	<1 E-300
	<i>Capzb</i>	1350	-0.58	5.97 E-263	3.54 E-260
	<i>Add1</i>	1515	-0.51	1.41 E-231	5.95 E-229
	<i>Mpzl1</i>	1165	0.45	4.37 E-227	3.15 E-223
	<i>Cldnd1</i>	1187	-0.6	2.49 E-209	8.2 E-207
	<i>Enpp2</i>	1550	0.23	1.63 E-209	2.35 E-191
	<i>H2afy1</i>	320	0.43	1.31 E-35	5.11 E-34
	<i>Mtss1</i>	181	0.44	1.58 E-24	4.54 E-23
	<i>Snap25</i>	100	0.39	8.8 E-22	2.1 E-20
Microglia	<i>Clns1</i>	588	0.91	<1 E-300	<1 E-300
	<i>H13</i>	1263	0.2	6.2 E-296	3.94 E-283
	<i>Sema4d</i>	768	0.6	1.19 E-282	6.79 E-280
	<i>App</i>	705	-0.6	5.33 E-240	2.04 E-237
	<i>Add1</i>	509	0.57	9.97 E-188	3.17 E-185
	<i>Lass5</i>	408	0.65	2.49 E-174	7.12 E-172
	<i>Rapgef1</i>	355	0.45	2.22 E-173	5.79 E-171
	<i>Fmn1l</i>	493	0.44	1.32 E-158	3.15 E-156
	<i>Fez2</i>	853	0.36	1.61 E-139	3.08 E-137
	<i>Fyn</i>	131	0.68	7.07 E-89	9.31 E-87
Endothelial	<i>Adam15</i>	893	0.7	<1 E-300	<1 E-300
	<i>Mcf2l</i>	629	0.74	<1 E-300	<1 E-300
	<i>Palm</i>	942	-0.65	<1 E-300	<1 E-300
	<i>Ablim1</i>	1025	0.47	<1 E-300	<1 E-300
	<i>Mprip</i>	3292	-0.51	<1 E-300	<1 E-300
	<i>Actn4</i>	1805	-0.31	<1 E-300	<1 E-300
	<i>Ktn1</i>	809	-0.7	3.84 E-226	1.02 E-223
	<i>Arhgef1</i>	865	0.36	2.23 E-219	5.66 E-217
	<i>Eif4h</i>	1199	0.46	1.14 E-197	2.29 E-195
	<i>Pkp4</i>	577	0.63	1.68 E-195	3.26 E-193

The coverage, delta inclusion (dl), p value, and false discovery rate (FDR) are listed. To obtain dl values, we first calculated the ratio of inclusion junction tags to inclusion plus skipping junction tags in each cell type and then determined the differences of the ratios between cell types. dl essentially quantifies the magnitude of the difference between the splicing of the two groups being compared (1 or -1 represents maximum difference, whereas 0 represents no difference).

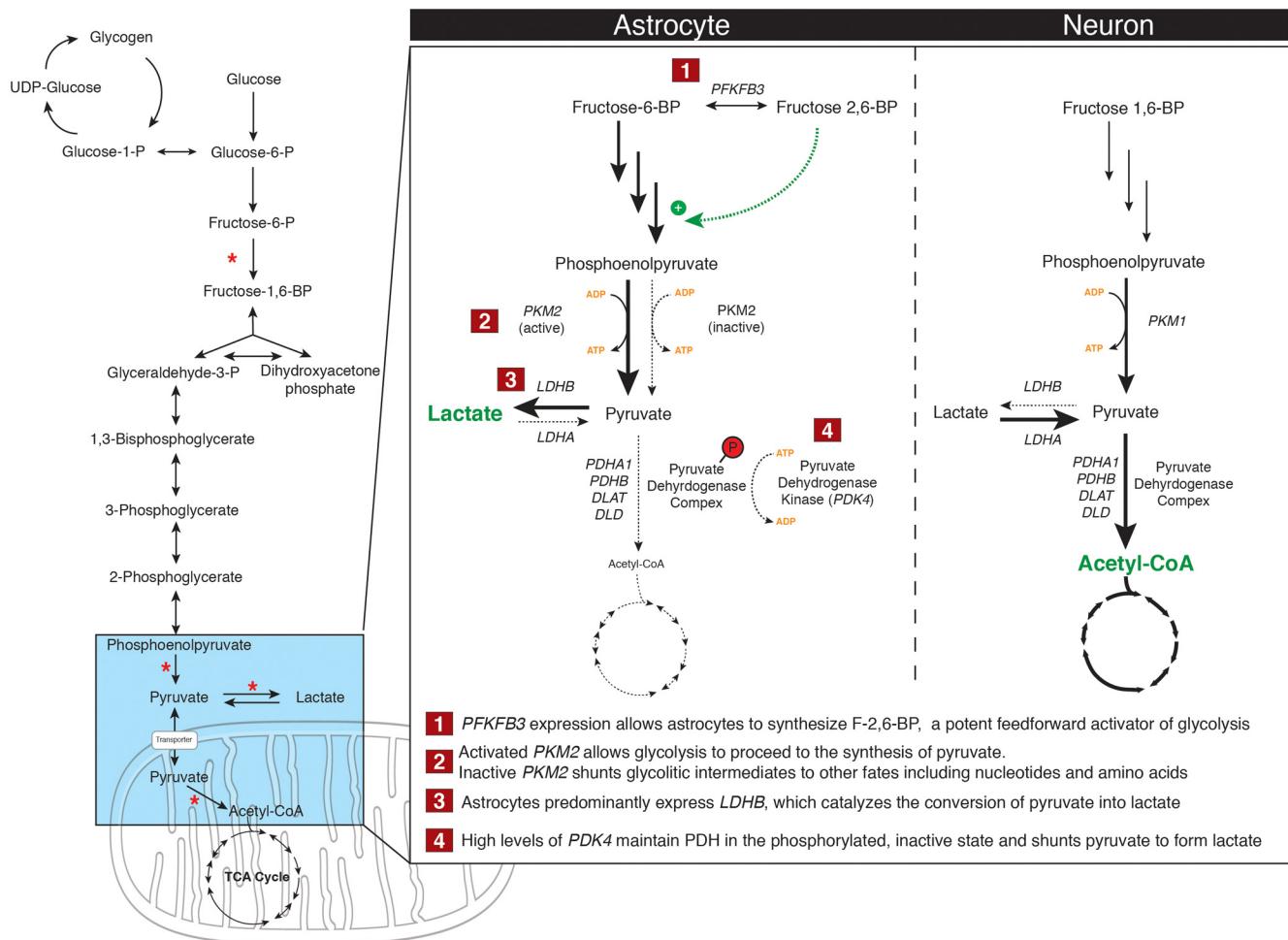
sion values because their validation provides confidence in the overall accuracy of the RNA-Seq dataset, whereas preferentially validating highly expressed genes is more straightforward but less informative. Additionally, these cell type-enriched genes have not been well characterized and are therefore novel markers worth validating. We used the Fluidigm Biomark microfluidics technology for qRT-PCR analysis. Using the Biomark microfluidics system, we loaded small volumes of 96 primer pairs and 96 samples to a microfluidics chip, in which each sample and each



**Figure 6.** Validation of RNA-Seq results by qRT-PCR and *in situ* hybridization. **A**, qRT-PCR validation of cell type-enriched genes identified by RNA-Seq. We performed qRT-PCR with Fluidigm BioMark microfluidic technology. Expression of several genes identified by RNA-Seq as enriched in each cell type was examined by qRT-PCR. The housekeeping gene *Gapdh* was included for comparison. Twelve replicates of each purified cell type and three replicates of whole-brain samples were analyzed. Warmer colors represent lower Ct values (higher abundance of transcripts), and cooler colors represent higher Ct values (lower abundance of transcripts). Black indicates no amplification. Data of genes labeled in red were quantified in **B**. **B**, Ct differences of *Atp13a4*, *Cpne7*, *Fam70b*, *Tmem88b*, and *Rcsd1* compared with *Gapdh* were plotted on a log<sub>2</sub> scale. Error bar represents SD. RNA-Seq analysis showed that *Atp13a4*, *Cpne7*, *Fam70b*, *Tmem88b*, and *Rcsd1* are enriched in astrocytes, neurons, OPCs, oligodendrocytes, and microglia, respectively. qRT-PCR validated these results. **C–G**, *In situ* hybridization validated novel cell type-specific genes. Left, Low-magnification image of the cortex and hippocampus. Scale bar, 200  $\mu$ m. Right, High-magnification image of the cortex. Scale bar, 50  $\mu$ m. Fluorescence *in situ* hybridization signals with probes against novel cell type-specific genes (red) and known cell type-specific markers (green) are shown. The regions in the yellow boxes are enlarged and shown as single channel and merged images on the right.

primer pair were mixed together to produce 9216 individual nanoliter-scaled qRT-PCR reactions in a single experiment. This eliminates the plate-to-plate variation seen commonly in traditional qRT-PCR experiments and allowed us to easily assay the expression of a large number of genes with many replicates. Thirty-four of the 40 genes we assayed were detected by qRT-PCR in the identical cell-enriched pattern as determined by RNA-Seq (Fig. 6). Six genes were not consistently amplified across replicates in any cell type. This is likely attributable to the lower sensitivity of running nano-scaled qRT-PCR reactions on microfluidics chips compared with RNA-Seq. None of the 40 genes we tested were reproducibly amplified in a cell type that contradicted the RNA-Seq data. Of the 34 genes we validated by PCR, our RNA-Seq data indicated that each was  $>10$ -fold enriched in a given cell type. Using the same threshold, only 20 of these genes were similarly identified as cell type-enriched by our microarray data. At a less-stringent twofold enrichment cutoff, only 29 were identified as cell type-enriched by microarray, thus demonstrating the increased sensitivity of the RNA-Seq data.

Finally, we sought to examine the accuracy of our transcriptome database by performing *in situ* hybridization on several relatively unknown cell-specific genes from our dataset. We generated specific probes to these candidate genes and cohybridized them with classic cell type-specific markers to investigate their cell specificity (Fig. 6C–G). We identified novel markers for astrocytes (*Gdpd2*), neurons (*Lppr3*), oligodendrocytes (*Ppapdc1a*), and microglia (*Olfml3*). Consistent with this finding, *Olfml3* was reported recently to be a microglia-specific marker in the spinal cord (Chiu et al., 2013). These results suggest that our database includes a substantial number of novel cell type-specific genes that could have wide-ranging significance for understanding brain function. In Figure 6C, the *in situ* hybridization signals of *Gdpd2* and *Glast* are located in close proximity but do not overlap. When we examined *Gdpd2* and *Glast* signals together with DAPI stains, it was clear that clusters of *Gdpd2*- and *Glast*-positive puncta localized to the same cell. In our experience, many astrocyte-specific genes produce *in situ* hybridization signals that are localized to processes instead off/in addition to the cell body. *In situ* hybrid-



ization signals in astrocytes often appear more diffuse than those in other cell types. The function of mRNA in astrocyte processes is unknown. One possibility is the presence of distinct RNA granules containing Gdpd2 and Glast mRNAs within astrocytes.

#### Regulation of energy metabolism in neurons and astrocytes

To demonstrate the application of this database toward understanding fundamental cellular interactions in the brain, we used our expression and splicing data to investigate the cellular division of labor in brain energy metabolism. We began by looking at glycolytic enzymes, because the primary energy consumers in the brain, neurons and astrocytes, each exhibit diametric patterns of glucose consumption (for review, see Bélanger et al., 2011; Bouzier-Sore and Pellerin, 2013); neurons heavily rely on oxidative metabolism, whereas astrocytes exhibit high glycolytic rates *in vivo* and *in vitro* (Lebon et al., 2002; Itoh et al., 2003; Bouzier-Sore and Pellerin, 2013). Under physiologic conditions, the majority of glucose entering the glycolytic pathway in astrocytes is metabolized to lactate rather than oxidized in the mitochondria. This phenomenon forms the foundation of the astrocyte–neuron lactate hypothesis (Walz and Mukerji, 1988), which proposes that astrocytes synthesize and subsequently expel lactate so that neurons can use it as an energy source. Subsequently, nu-

merous studies have demonstrated that neurons efficiently use lactate as an energy source (Schurr et al., 1997; Bouzier et al., 2000; Qu et al., 2001) and may even prefer it over glucose (Bouzier et al., 2000; Itoh et al., 2003).

We identified expression and splicing differences in various glycolytic enzymes that help distinguish astrocytes as primary glycolytic cells (summarized in Fig. 7). The first of these enzymes is 6-phosphofructose-2-kinase/fructose-2,6-bisphosphatase-3 (*Pfkfb3*), which synthesizes fructose 2,6-bisphosphate, a potent glycolytic activator. Protein levels of *Pfkfb3* are low in neurons because of its constant proteasomal degradation (Herrero-Mendez et al., 2009; Almeida et al., 2010), and, consistent with these observations, we found that *Pfkfb3* mRNA is enriched by an order of magnitude in astrocytes compared with neurons. Another key regulated step of glycolysis is the conversion of phosphoenol pyruvate into pyruvate through the enzyme pyruvate kinase (PK). PK is found in two primary isoforms, *Pkm1* and *Pkm2*. We identified and validated the exclusive expression of *Pkm1* in neurons and *Pkm2* in astrocytes (and other glial cells; Fig. 5E,F). *Pkm2* contains an inducible nuclear translocation signal that allows a cell to regulate the amount of glycolytic flux in response to the local energy state. However, *Pkm1* is not allosterically regulated and locks neurons in a steady state of glycolysis. After PK catalyzes the formation of pyruvate,

its flux into oxidative metabolism is carefully controlled by the activity of the enzyme pyruvate dehydrogenase (*Pdh*). Phosphorylation-mediated inactivation of *Pdh* in astrocytes (Itoh et al., 2003; Halim et al., 2010) prevents glycolytic pyruvate from entering oxidative phosphorylation. Analysis of our transcriptome data demonstrates that pyruvate dehydrogenase kinase 4 (*Pdk4*) transcripts are >30 times enriched in astrocytes compared with neurons, suggesting that high *Pdk4* expression in astrocytes might be responsible for subsequent *Pdh* inactivation. Diminished *Pdh* activity in astrocytes requires pyruvate to be shunted to the formation of lactate by the enzyme lactate dehydrogenase (*Ldh*). Our data indicate that astrocytes and neurons preferentially express different versions of the lactate dehydrogenase enzymes that interconvert pyruvate and lactate, *Ldha* and *Ldhb*. *Ldha* and *Ldhb* are biased toward the production of pyruvate or lactate, respectively, and we find that *Ldhb* is enriched in astrocytes (twofold), whereas *Ldha* is primarily expressed in neurons (1.7-fold).

The expression and splicing patterns that we identified in *Pfkfb3*, *Pkm2*, *Pdk4*, and *Ldha/b* are each consistent with metabolic observations in astrocytes and neurons. Although the majority of the primary glycolytic and oxidative phosphorylation enzymes are expressed at similar levels in neurons and astrocytes, the differential expression of these key regulatory enzymes helps to explain how astrocytes perform such high rates of aerobic glycolysis. Our data suggest that, in contrast to neurons, astrocytes express a collection of regulatory enzymes that are sensitive to the energy state of the cell.

## Discussion

### An RNA-Seq transcriptome database of purified cell classes of the brain

In this study, we refined purification protocols for neurons, astrocytes, various maturation stages of oligodendrocytes, microglia, endothelial cells, and pericytes from mouse cortex and obtained an RNA-Seq transcriptome dataset of these cell types. Although a large number of neuronal subtypes have already been recognized (and glial subpopulations likely exist), grouping these populations into major cell type classifications allows for the investigation of interactions between primary cell classes. We obtained data on the expression level of >22,000 genes, and we determined the abundance of each alternative splicing variant of each gene for each of our samples.

The systematic transcriptional dissection of an organ to its primary cell type components for the purpose of establishing a transcriptome database provides an essential step toward understanding a complex tissue such as the brain. Understanding brain function (as well as dysfunction) requires an understanding of how distinct cell classes in the brain interact in a dynamic environment. For example, recent advances in understanding disease mechanisms in the neurodegenerative disease amyotrophic lateral sclerosis have revealed that the disease is not simply a motor neurons disease. Rather, it is a disease of the spinal cord in which astrocytes, microglia, and oligodendrocytes all contribute to disease progression (Ling et al., 2013). Thus, the database we have established should be invaluable in future studies of complex cell–cell interactions in other neurological diseases.

Our dataset contains a tremendous quantity of novel gene expression data, and exploration of this dataset by the neuroscience research community will generate hundreds of new testable hypotheses with biological significance. We noted that the vast majority of well annotated cell type-specific genes exhibit the predicted cell-enrichment patterns in our dataset (Fig. 1). Fur-

thermore, RNA-Seq allows for the detection of novel genes that are unannotated in the current genome; we detected 26 novel transcripts with expression levels of FPKM >5 that were previously unannotated in UCSC Genome Browser (Meyer et al., 2013), Ensembl (Flicek et al., 2012), Vega (Ashurst et al., 2005), and AceView (Thierry-Mieg and Thierry-Mieg, 2006) databases (data not shown).

When directly compared with the microarray platform, our RNA-Seq library identified a dramatic increase in the number of cell type-specific genes (Fig. 2). This is in part attributable to the higher sensitivity of RNA-Seq in detecting low-abundance RNAs. One caveat regarding the interpretation of large-enrichment scores is to consider the effects of ratio inflation at lower FPKM levels. The biological relevance of a gene that is 20-fold enriched at an FPKM value of 2 should be weighed with more caution than a similarly enriched gene with an FPKM of 200.

### An alternative splicing database of purified cell types of the brain

Alternative pre-mRNA splicing is an essential mechanism for generating proteome complexity. Previous expressed sequence tagged mapping studies have led to the conclusion that the brain has the highest number of alternative splicing events when compared with other organs (Yeo et al., 2004), and different brain regions are associated with complex patterns of alternative splicing (Johnson et al., 2009). However, the cellular source of this complexity is not known because analyses were performed on brain tissues comprising mixed cell types. We sought to answer the question of how many genes are alternatively spliced in the mouse cortex with our highly sensitive RNA-Seq data. Of the ~22,000 genes that comprise the mouse genome, we examined a large collection of RNA-Seq data across several tissue types and identified 10,447 cases in which there is direct evidence of alternative splicing or predicted alternatively spliced isoforms. Of these genes that are potentially alternatively spliced in the mouse genome, 9036 are expressed in the mouse cortex above a minimum expression threshold. Interestingly, we found that 6588 of these genes (73%) are alternatively spliced in at least one major brain cell type. This demonstrates that the majority of genes that are known to undergo alternative splicing in the mouse genome are alternatively spliced in the brain. Thus, an understanding of cellular transcriptomes would not be complete unless alternative splicing information accompanies gene-level expression data.

Tissue-specific alternative splicing is well established (Xu et al., 2002). However, the regulation of alternative splicing in a cell type-specific manner in a complex tissue like the brain has not been addressed. We report that alternative splicing occurs as frequently in glia and vascular cells as it does in neurons. In contrast to the overall similarity in the frequency of alternative splicing, each cell type contains its own specific repertoire of thousands of alternatively spliced RNAs. These observations suggest that the distinct functions of each cell type may result as much from the presence of specific protein isoforms generated by alternative splicing as from global differences in the levels of gene expression. Moreover, previous studies have shown that disruption of alternative splicing can lead to neurological diseases in human and animal models (Buckanovich et al., 1996; Yang et al., 1998; Gehman et al., 2011), and our splicing dataset provides a resource for understanding the cell type-specific contributions to these disorders.

Alternative splicing of specific genes has also been shown to play important roles in the development and function of the nervous system. For example, the mechanisms by which differ-

entially spliced isoforms of neuroligin and neurexin pre-mRNA mediate synaptic adhesion offer a well characterized example of how cell-specific splicing can have considerable functional consequences (Chih et al., 2006). Before systematic methods for the analysis of alternative splicing became available, characterizing alternative splicing of specific molecules (e.g., neuroligin and neurexin) and their role in the nervous system was cumbersome and time consuming. In the present study, we generated a high-quality splicing dataset for every gene in the major cell classes of the brain. Although a systematic and deep analysis of this splicing data is beyond the scope of this study, a preliminary exploration of the data has already identified interesting splicing patterns of considerable biological significance. For example, the neural cell adhesion molecule (NCAM), encoded by the *Ncam1* gene, is a homophilic cell adhesion molecule involved in synaptic plasticity, learning, and memory (Becker et al., 1996; Senkov et al., 2006; Stoenica et al., 2006). Multiple splicing isoforms of *Ncam1* have been reported previously (Owens et al., 1987; Krushel et al., 1998). We found that *Ncam1* is spliced uniquely in neurons and glia. The neuronal isoform includes an exon that encodes a transmembrane domain, whereas the glial form excludes this exon. Both isoforms contain a signal peptide sequence. Therefore, neuronal NCAM is likely located on the plasma membrane, whereas glial NCAM is likely secreted or glycoprophosphatidylinositol anchored. This discovery raises interesting new possibilities of the cellular mechanism by which NCAM controls synaptic plasticity, learning, and memory.

Another finding from our alternative splicing dataset is the differential expression of splicing isoforms of the *H2afy* gene in oligodendrocyte lineage. *H2afy* encodes an atypical histone variant macroH2A1 enriched in heterochromatin. The gene *H2afy* contains two mutually exclusive exons, and alternative splicing generates two isoforms, mH2A1.1 and mH2A1.2 (Kustatscher et al., 2005). Embryonic stem cells exclusively express mH2A1.2, and as development progresses, the expression of the mH2A1.1 isoform increases. In addition, cancer cells downregulate the mH2A1.1 isoform, which inhibits cell proliferation (Novikov et al., 2011). We found that OPCs predominantly express mH2A1.2, NFOs express approximately equal amounts of mH2A1.1 and mH2A1.2, and MOs predominantly express mH2A1.1. This developmental progression of splicing isoform change correlates with the reduced proliferation capacity as oligodendrocytes differentiate. Our data are consistent with a role of *mH2A1* in regulating oligodendrocyte differentiation through chromosomal changes.

RNA-Seq is highly reproducible and reference independent. Thus, this database of wild-type mouse neurons, glia, and vascular cells of the brain provides a valuable reference dataset for future studies that compare gene expression patterns of diseased versus healthy tissues. Our cell purification protocols can be used to similarly purify cell types from brains of mutant mice. Comparing these transcriptome studies at the resolution of individual cell types will help elucidate mechanisms of neurological diseases. Accumulating evidence has demonstrated that glia are involved in a variety of neurological diseases, including schizophrenia, autism spectrum disorders, epilepsy, Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis, and stroke (Matute et al., 2005; Tian et al., 2005; Nagai et al., 2007; Chen et al., 2009; Li et al., 2011; Liyo et al., 2011; Molofsky et al., 2012; Zamanian et al., 2012). As the research community continues to look beyond neurons to understand the pathogenesis of neurologic dysfunction, the comprehensive transcriptome dataset of the brain presented here will be a valuable resource.

## Notes

Supplemental material for this article is available at [http://web.stanford.edu/group/barres\\_lab/brain\\_rnaseq.html](http://web.stanford.edu/group/barres_lab/brain_rnaseq.html). Searchable RNA-Seq database for purified cell types. This material has not been peer reviewed.

## References

- Allen NJ, Barres BA (2009) Neuroscience: glia—more than just brain glue. *Nature* 457:675–677. [CrossRef](#) [Medline](#)
- Almeida A, Bolaños JP, Moncada S (2010) E3 ubiquitin ligase APC/C-Cdh1 accounts for the Warburg effect by linking glycolysis to cell proliferation. *Proc Natl Acad Sci U S A* 107:738–741. [CrossRef](#) [Medline](#)
- Ashurst JL, Chen CK, Gilbert JGR, Jekosch K, Keenan S, Meidl P, Searle SM, Stalker J, Storey R, Trevanion S, Wilming L, Hubbard T (2005) The Vertebrate Genome Annotation (Vega) database. *Nucleic Acids Res* 33:D459–D465. [CrossRef](#) [Medline](#)
- Bainbridge MN, Warren RL, Hirst M, Romanuk T, Zeng T, Go A, Delaney A, Griffith M, Hickenbotham M, Magrini V, Mardis ER, Sadar MD, Siddiqui AS, Marra MA, Jones SJ (2006) Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics* 7:246. [CrossRef](#) [Medline](#)
- Bandyopadhyay U, Cotney J, Nagy M, Oh S, Leng J, Mahajan M, Mane S, Fenton WA, Noonan JP, Horwich AL (2013) RNA-Seq profiling of spinal cord motor neurons from a presymptomatic SOD1 ALS mouse. *PLoS One* 8:e53575. [CrossRef](#) [Medline](#)
- Bansal R, Warrington AE, Gard AL, Ranscht B, Pfeiffer SE (1989) Multiple and novel specificities of monoclonal antibodies O1, O4, and R-mAb used in the analysis of oligodendrocyte development. *J Neurosci Res* 24:548–557. [Medline](#)
- Becker CG, Artola A, Gerardy-Schahn R, Becker T, Welzl H, Schachner M (1996) The polysialic acid modification of the neural cell adhesion molecule is involved in spatial learning and hippocampal long-term potentiation. *J Neurosci Res* 45:143–152. [CrossRef](#) [Medline](#)
- Bélanger M, Allaman I, Magistretti PJ (2011) Brain energy metabolism: focus on astrocyte-neuron metabolic cooperation. *Cell Metab* 14:724–738. [CrossRef](#) [Medline](#)
- Beutner C, Linnartz-Gerlach B, Schmidt SV, Beyer M, Mallmann MR, Staratschek-Jox A, Schultze JL, Neumann H (2013) Unique transcriptome signature of mouse microglia. *Glia* 61:1429–1442. [CrossRef](#) [Medline](#)
- Bouzier AK, Thiaudiere E, Biran M, Rouland R, Canioni P, Merle M (2000) The metabolism of [3-(13)C]lactate in the rat brain is specific of a pyruvate carboxylase-deprived compartment. *J Neurochem* 75:480–486. [CrossRef](#) [Medline](#)
- Bouzier-Sore AK, Pellerin L (2013) Unraveling the complex metabolic nature of astrocytes. *Front Cell Neurosci* 7:179. [CrossRef](#) [Medline](#)
- Bracko O, Singer T, Aigner S, Knobloch M, Winner B, Ray J, Clemenson GD Jr, Suh H, Couillard-Despres S, Aigner L, Gage FH, Jessberger S (2012) Gene expression profiling of neural stem cells and their neuronal progeny reveals IGF2 as a regulator of adult hippocampal neurogenesis. *J Neurosci* 32:3376–3387. [CrossRef](#) [Medline](#)
- Buckanovich RJ, Yang YY, Darnell RB (1996) The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *J Neurosci* 16:1114–1122. [Medline](#)
- Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA, Thompson WJ, Barres BA (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 28:264–278. [CrossRef](#) [Medline](#)
- Chen K, Deng S, Lu H, Zheng Y, Yang G, Kim D, Cao Q, Wu JQ (2013) RNA-seq characterization of spinal cord injury transcriptome in acute/subacute phases: a resource for understanding the pathology at the systems level. *PLoS One* 8:e72567. [CrossRef](#) [Medline](#)
- Chen PC, Vargas MR, Pani AK, Smeyne RJ, Johnson DA, Kan YW, Johnson JA (2009) Nrf2-mediated neuroprotection in the MPTP mouse model of Parkinson's disease: critical role for the astrocyte. *Proc Natl Acad Sci U S A* 106:2933–2938. [CrossRef](#) [Medline](#)
- Chih B, Gollan L, Scheiffele P (2006) Alternative splicing controls selective trans-synaptic interactions of the neuroligin-neurexin complex. *Neuron* 51:171–178. [CrossRef](#) [Medline](#)
- Chiu IM, Morimoto ET, Goodarzi H, Liao JT, O'Keeffe S, Phatnani HP, Muratet M, Carroll MC, Levy S, Tavazoie S, Myers RM, Maniatis T

- (2013) A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. *Cell Rep* 4:385–401. CrossRef Medline
- Christopherson KS, Ullian EM, Stokes CC, Mullaney CE, Hell JW, Agah A, Lawler J, Mosher DF, Bornstein P, Barres BA (2005) Thrombospondins are astrocyte-secreted proteins that promote CNS synaptogenesis. *Cell* 120:421–433. CrossRef Medline
- Citri A, Pang ZP, Südhof TC, Wernig M, Malenka RC (2012) Comprehensive qPCR profiling of gene expression in single neuronal cells. *Nat Protoc* 7:118–127. CrossRef Medline
- Clonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5:613–619. CrossRef Medline
- Daneman R, Zhou L, Agalliu D, Cahoy JD, Kaushal A, Barres BA (2010) The mouse blood-brain barrier transcriptome: a new resource for understanding the development and function of brain endothelial cells. *PLoS One* 5:e13741. CrossRef Medline
- Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Dougherty ML, Gong S, Greengard P, Heintz N (2008) Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* 135:749–762. CrossRef Medline
- Dugas JC, Tai YC, Speed TP, Ngai J, Barres BA (2006) Functional genomic analysis of oligodendrocyte differentiation. *J Neurosci* 26:10967–10983. CrossRef Medline
- Emery B, Agalliu D, Cahoy JD, Watkins TA, Dugas JC, Mulinyawe SB, Ibrahim A, Ligon KL, Rowitch DH, Barres BA (2009) Myelin gene regulatory factor is a critical transcriptional regulator required for CNS myelination. *Cell* 138:172–185. CrossRef Medline
- Eroglu C, Allen NJ, Susman MW, O'Rourke NA, Park CY, Ozkan E, Chakraborty C, Mulinyawe SB, Annis DS, Huberman AD, Green EM, Lawler J, Dolmetsch R, Garcia KC, Smith SJ, Luo ZD, Rosenthal A, Mosher DF, Barres BA (2009) Gabapentin receptor alpha2delta-1 is a neuronal thrombospondin receptor responsible for excitatory CNS synaptogenesis. *Cell* 139:380–392. CrossRef Medline
- Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, Gil L, García-Girón C, Gordon L, Hourlier T, Hunt S, Juettemann T, Kähäri AK, Keenan S, Komorowska M, et al. (2012) Ensembl 2013. *Nucleic Acids Res* 41:D48–D55. CrossRef Medline
- Foo LC, Allen NJ, Bushong EA, Ventura PB, Chung WS, Zhou L, Cahoy JD, Daneman R, Zong H, Ellisman MH, Barres BA (2011) Development of a method for the purification and culture of rodent astrocytes. *Neuron* 71:799–811. CrossRef Medline
- Friedrich B, Euler P, Ziegler R, Kuhn A, Landwehrmeyer BG, Luthi-Carter R, Weiller C, Hellwig S, Zucker B (2012) Comparative analyses of Purkinje cell gene expression profiles reveal shared molecular abnormalities in models of different polyglutamine diseases. *Brain Res* 1481:37–48. CrossRef Medline
- Gehman LT, Stoilov P, Maguire J, Damianov A, Lin CH, Shiue L, Ares M Jr, Mody I, Black DL (2011) The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat Genet* 43:706–711. CrossRef Medline
- Gong X, Carmon KS, Lin Q, Thomas A, Yi J, Liu Q (2012) LGR6 is a high affinity receptor of R-spondins and potentially functions as a tumor suppressor. *PLoS One* 7:e37137. CrossRef Medline
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabilio MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223–227. CrossRef Medline
- Halim ND, McFate T, Mohyeldin A, Okagaki P, Korotchkina LG, Patel MS, Jeoung NH, Harris RA, Schell MJ, Verma A (2010) Phosphorylation status of pyruvate dehydrogenase distinguishes metabolic phenotypes of cultured rat brain astrocytes and neurons. *Glia* 58:1168–1176. CrossRef Medline
- Heintz N (2004) Gene expression nervous system atlas (GENSAT). *Nat Neurosci* 7:483. CrossRef Medline
- Herrero-Mendez A, Almeida A, Fernández E, Maestre C, Moncada S, Bolaños JP (2009) The bioenergetic and antioxidant status of neurons is controlled by continuous degradation of a key glycolytic enzyme by APC/C-Cdh1. *Nat Cell Biol* 11:747–752. CrossRef Medline
- Itoh Y, Esaki T, Shimoji K, Cook M, Law MJ, Kaufman E, Sokoloff L (2003) Dichloroacetate effects on glucose and lactate oxidation by neurons and astroglia in vitro and on glucose utilization by brain in vivo. *Proc Natl Acad Sci U S A* 100:4879–4884. CrossRef Medline
- Johnson MB, Kawasawa YI, Mason CE, Krnsik Z, Coppola G, Bogdanović D, Geschwind DH, Mane SM, State MW, Sestan N (2009) Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* 62:494–509. CrossRef Medline
- Krushel LA, Tai MH, Cunningham BA, Edelman GM, Crossin KL (1998) Neural cell adhesion molecule (N-CAM) domains and intracellular signaling pathways involved in the inhibition of astrocyte proliferation. *Proc Natl Acad Sci U S A* 95:2592–2596. CrossRef Medline
- Kustatscher G, Hothorn M, Pugieux C, Scheffzek K, Ladurner AG (2005) Splicing regulates NAD metabolite binding to histone macroH2A. *Nat Struct Mol Biol* 12:624–625. CrossRef Medline
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25. CrossRef Medline
- Lau P, Verrier JD, Nielsen JA, Johnson KR, Notterpek L, Hudson LD (2008) Identification of dynamically regulated microRNA and mRNA networks in developing oligodendrocytes. *J Neurosci* 28:11720–11730. CrossRef Medline
- Lebon V, Petersen KF, Cline GW, Shen J, Mason GF, Dufour S, Behar KL, Shulman GI, Rothman DL (2002) Astroglial contribution to brain energy metabolism in humans revealed by <sup>13</sup>C nuclear magnetic resonance spectroscopy: elucidation of the dominant pathway for neurotransmitter glutamate repletion and measurement of astrocytic oxidative metabolism. *J Neurosci* 22:1523–1531. Medline
- Lee JE, Hollenberg SM, Snider L, Turner DL, Lipnick N, Weintraub H (1995) Conversion of *Xenopus* ectoderm into neurons by NeuroD, a basic helix-loop-helix protein. *Science* 268:836–844. CrossRef Medline
- Lerch JK, Kuo F, Motti D, Morris R, Bixby JL, Lemmon VP (2012) Isoform diversity and regulation in peripheral and central neurons revealed through RNA-Seq. *PLoS One* 7:e30417. CrossRef Medline
- Li C, Zhao R, Gao K, Wei Z, Yin MY, Lau LT, Chui D, Hoi Yu AC (2011) Astrocytes: implications for neuroinflammatory pathogenesis of Alzheimer's disease. *Curr Alzheimer Res* 8:67–80. CrossRef Medline
- Ling SC, Polymenidou M, Cleveland DW (2013) Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* 79:416–438. CrossRef Medline
- Lioy DT, Garg SK, Monaghan CE, Raber J, Foust KD, Kaspar BK, Hirrlinger PG, Kirchhoff F, Bissonnette JM, Ballas N, Mandel G (2011) A role for glia in the progression of Rett's syndrome. *Nature* 475:497–500. CrossRef Medline
- Lovatt D, Sonnewald U, Waagepetersen HS, Schousboe A, He W, Lin JH, Han X, Takano T, Wang S, Sim FJ, Goldman SA, Nedergaard M (2007) The transcriptome and metabolic gene signature of protoplasmic astrocytes in the adult murine cortex. *J Neurosci* 27:12255–12266. CrossRef Medline
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18:1509–1517. CrossRef Medline
- Matute C, Melone M, Vallejo-Illarramendi A, Conti F (2005) Increased expression of the astrocytic glutamate transporter GLT-1 in the prefrontal cortex of schizophrenics. *Glia* 49:451–455. CrossRef Medline
- Mercer TR, Dinger ME, Mattick JS (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10:155–159. CrossRef Medline
- Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenblom KR, Roe G, Rhead B, Raney BJ, Pohl A, Malladi VS, Li CH, Lee BT, Learned K, Kirkup V, Hsu F, Heitner S, Harte RA, et al. (2013) The UCSC Genome Browser database: extensions and updates 2013. *Nucleic Acids Res* 41:D64–D69. CrossRef Medline
- Molofsky AV, Krenick R, Ullian EM, Tsai HH, Deneen B, Richardson WD, Barres BA, Rowitch DH (2012) Astrocytes and disease: a neurodevelopmental perspective. *Genes Dev* 26:891–907. CrossRef Medline
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621–628. CrossRef Medline
- Motoike T, Loughna S, Perens E, Roman BL, Liao W, Chau TC, Richardson CD, Kawate T, Kuno J, Weinstein BM, Stainier DY, Sato TN (2000) Universal GFP reporter for the study of vascular development. *Genesis* 28:75–81. CrossRef Medline
- Nagai M, Re DB, Nagata T, Chalazonitis A, Jessell TM, Wichterle H, Przedborski S (2007) Astrocytes expressing ALS-linked mutated SOD1 release

- factors selectively toxic to motor neurons. *Nat Neurosci* 10:615–622. CrossRef Medline
- Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320:1344–1349. CrossRef Medline
- Novikov L, Park JW, Chen H, Klerman H, Jalloh AS, Gamble MJ (2011) QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation. *Mol Cell Biol* 31:4244–4255. CrossRef Medline
- Orre M, Kamphuis W, Osborn LM, Melief J, Kooijman L, Huitinga I, Klooster J, Bossers K, Hol EM (2014) Acute isolation and transcriptome characterization of cortical astrocytes and microglia from young and aged mice. *Neurobiol Aging* 35:1–14. CrossRef Medline
- Owens GC, Edelman GM, Cunningham BA (1987) Organization of the neural cell adhesion molecule (N-CAM) gene: alternative exon usage as the basis for different membrane-associated domains. *Proc Natl Acad Sci U S A* 84:294–298. CrossRef Medline
- Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 40:1413–1415. CrossRef Medline
- Quackenbush J (2002) Microarray data normalization and transformation. *Nat Genet* 32[Suppl]:496–501. CrossRef Medline
- Qu H, Eloqayli H, Unsgård G, Sonnewald U (2001) Glutamate decreases pyruvate carboxylase activity and spares glucose as energy substrate in cultured cerebellar astrocytes. *J Neurosci Res* 66:1127–1132. CrossRef Medline
- Ramskold D, Wang ET, Burge CB, Sandberg R (2009) An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLOS Comput Biol* 12:e1000598. CrossRef Medline
- Ritz K, van Schaik BDC, Jakobs ME, van Kampen AH, Aronica E, Tijssen MA, Baas F (2011) SGCE isoform characterization and expression in human brain: implications for myoclonus-dystonia pathogenesis? *Eur J Hum Genet* 19:438–444. CrossRef Medline
- Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12:R22. CrossRef Medline
- Rossner MJ, Hirrlinger J, Wichert SP, Boehm C, Newrzella D, Hiemisch H, Eisenhardt G, Stuenkel C, von Ahsen O, Nave KA (2006) Global transcriptome analysis of genetically identified neurons in the adult cortex. *J Neurosci* 26:9956–9966. CrossRef Medline
- Schaeren-Wiemers N, Gerfin-Moser A (1993) A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: in situ hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* 100:431–440. CrossRef Medline
- Schurr A, Payne RS, Miller JJ, Rigor BM (1997) Brain lactate is an obligatory aerobic energy substrate for functional recovery after hypoxia: further in vitro validation. *J Neurochem* 69:423–426. CrossRef Medline
- Senkov O, Sun M, Weinhold B, Gerardy-Schahn R, Schachner M, Dityatev A (2006) Polysialylated neural cell adhesion molecule is involved in induction of long-term potentiation and memory acquisition and consolidation in a fear-conditioning paradigm. *J Neurosci* 26:10888–10989. CrossRef Medline
- Stoenica L, Senkov O, Gerardy-Schahn R, Weinhold B, Schachner M, Dityatev A (2006) In vivo synaptic plasticity in the dentate gyrus of mice deficient in the neural cell adhesion molecule NCAM or its polysialic acid. *Eur J Neurosci* 23:2255–2264. CrossRef Medline
- Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, Schmidt D, O'Keeffe S, Haas S, Vingron M, Lehrach H, Yaspo ML (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321:956–960. CrossRef Medline
- Tani K, Yoshida MC, Satoh H, Mitamura K, Noguchi T, Tanaka T, Fujii H, Miwa S (1988) Human M2-type pyruvate kinase: cDNA cloning, chro-mosomal assignment and expression in hepatoma. *Gene* 73:509–516. CrossRef Medline
- Thierry-Mieg D, Thierry-Mieg J (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* 7 [Suppl 1]: S12.1–S12.14. CrossRef Medline
- Tian GF, Azmi H, Takano T, Xu Q, Peng W, Lin J, Oberheim N, Lou N, Wang X, Zielke HR, Kang J, Nederaard M (2005) An astrocytic basis of epilepsy. *Nat Med* 11:973–981. CrossRef Medline
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28:511–515. CrossRef Medline
- Walz W, Mukerji S (1988) Lactate release from cultured astrocytes and neurons: a comparison. *Glia* 1:366–370. CrossRef Medline
- Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456:470–476. CrossRef Medline
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10:57–63. CrossRef Medline
- Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bähler J (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453:1239–1243. CrossRef Medline
- Wu JQ, Habegger L, Noisa P, Szekely A, Qiu C, Hutchison S, Raha D, Egholm M, Lin H, Weissman S, Cui W, Gerstein M, Snyder M (2010) Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. *Proc Natl Acad Sci U S A* 107:5254–5259. CrossRef Medline
- Wu J, Anczukow O, Krainer AR, Zhang MQ, Zhang C (2013) OLego: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. *Nucleic Acids Res* 41:5149–5163. CrossRef Medline
- Wylie CJ, Hendricks TJ, Zhang B, Wang L, Lu P, Leahy P, Fox S, Maeno H, Deneris ES (2010) Distinct transcriptomes define rostral and caudal serotonin neurons. *J Neurosci* 30:670–684. CrossRef Medline
- Xu Q, Modrek B, Lee C (2002) Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res* 30: 3754–3766. CrossRef Medline
- Yang YY, Yin GL, Darnell RB (1998) The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. *Proc Natl Acad Sci U S A* 95:13254–13259. CrossRef Medline
- Yeo G, Holste D, Kreiman G, Burge CB (2004) Variation in alternative splicing across human tissues. *Genome Biol* 5:R74. CrossRef Medline
- Zamanian JL, Xu L, Foo LC, Nouri N, Zhou L, Giffard RG, Barres BA (2012) Genomic analysis of reactive astrogliosis. *J Neurosci* 32:6391–6410. CrossRef Medline
- Zhang C, Frias MA, Mele A, Ruggiu M, Eom T, Marney CB, Wang H, Licatalosi DD, Fak JJ, Darnell RB (2010) Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* 329:439–443. CrossRef Medline
- Zhang HM, Chen H, Liu W, Liu H, Gong J, Wang H, Guo AY (2012) AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res* 40:D144–D149. CrossRef Medline
- Zhou Q, Anderson DJ (2002) The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. *Cell* 109:61–73. CrossRef Medline
- Zonta B, Tait S, Melrose S, Anderson H, Harroch S, Higginson J, Sherman DL, Brophy PJ (2008) Glial and neuronal isoforms of neurofascin have distinct roles in the assembly of nodes of Ranvier in the central nervous system. *J Cell Biol* 181:1169–1177. CrossRef Medline