

# Molecular diversity of astrocytes with implications for neurological disorders

Robert M. Bachoo<sup>\*†‡</sup>, Ryung S. Kim<sup>\*§</sup>, Keith L. Ligon<sup>¶</sup>, Elizabeth A. Maher<sup>\*</sup>, Cameron Brennan<sup>\*</sup>, Nathan Billings<sup>¶</sup>, Suzanne Chan<sup>\*</sup>, Cheng Li<sup>§</sup>, David H. Rowitch<sup>¶</sup>, Wing H. Wong<sup>¶</sup>, and Ronald A. DePinho<sup>\*,\*\*\*††</sup>

Departments of <sup>\*</sup>Medical Oncology and <sup>¶</sup>Pediatric Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115; <sup>†</sup>Department of Neurology, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115; Departments of <sup>§</sup>Biostatistics and <sup>¶</sup>Statistics, Harvard University, Boston, MA 02115; <sup>\*\*</sup>Departments of Medicine and Genetics, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115

Communicated by Webster K. Cavenee, University of California at San Diego, La Jolla, CA, March 29, 2004 (received for review January 5, 2004)

The astrocyte represents the most abundant yet least understood cell type of the CNS. Here, we use a stringent experimental strategy to molecularly define the astrocyte lineage by integrating microarray datasets across several *in vitro* model systems of astrocyte differentiation, primary astrocyte cultures, and various astrocyte-rich CNS structures. The intersection of astrocyte data sets, coupled with the application of nonastrocytic exclusion filters, yielded many astrocyte-specific genes possessing strikingly varied patterns of regional CNS expression. Annotation of these astrocyte-specific genes provides direct molecular documentation of the diverse physiological roles of the astrocyte lineage. This global perspective in the normal brain also provides a framework for how astrocytes may participate in the pathogenesis of common neurological disorders like Alzheimer's disease, Parkinson's disease, stroke, epilepsy, and primary brain tumors.

The vertebrate CNS is comprised of three predominant cell types (neurons, oligodendrocytes, and astrocytes) that are thought to arise from multipotent neural stem cells (NSCs). Recent studies (1, 2) have provided considerable insight into the development and diversity of neurons and the oligodendrocyte lineage. In contrast, there is a more limited molecular understanding of the development and diversity of the astrocyte lineage.

Historically, astrocytes have been viewed as a homogenous population of cells functioning to provide passive support by supply of essential substrates and removal of toxic metabolites. This perceived limited functional range of the astrocyte is not consistent with the emerging data that they may retain stem cell-like properties (3, 4) and modulate almost every facet of functional neural networks (5, 6). For instance, astrocytes may express voltage-gated ion channels and neurotransmitter receptors that are coactivated at synapses and then participate in removing potentially toxic excitatory amino acids from synapses by high-affinity transporters (7). Astrocyte involvement in neuron homeostasis may also extend to trophic support (8), antioxidant functions, and production of critical substrates for neuron membrane synthesis. Dysregulation of these and other putative astrocyte functions have been variously implicated in the pathogenesis of numerous developmental, genetic, idiopathic, and acquired neurodegenerative diseases (9).

To date, precise genetic analyses of the astrocyte in normal physiology and disease processes have been limited to *in vitro* studies by using specific glial differentiation model systems (10–12). These important efforts have focused on specialized aspects of early glial differentiation and as such have yielded limited information on the diverse roles of astrocytes in normal brain. Thus, the challenge remains to develop a comprehensive molecular profile of the astrocyte lineage that reflects its apparent developmental complexity, its full range of physiological capacities, and its lineage heterogeneity.

In this study, we have defined the mature astrocyte on the molecular level through an exhaustive and integrated transcriptional analysis of many distinct astrocyte-rich cultures and CNS

tissues by using standard and nonstandard bioinformatic approaches. These data sets were curated further through nonastrocytic subtractive filters. The resulting astrocyte-specific candidate genes were subjected to a rigorous RNA *in situ* hybridization (ISH) validation scheme that revealed distinctive regional patterns of expression. Results presented here serve to define this ubiquitous CNS lineage and to provide a molecular framework within which to explore more precisely its diverse roles in the normal brain and their increasingly recognized role in the pathogenesis of many common neurological disorders.

## Methods

**NSC and Astrocyte Culture Techniques.** Primary NSCs were isolated from murine embryonic day (E)13.5 embryos as described (13). Neurospheres were differentiated into astrocyte predominant cultures by exposure to the following: 10% FBS, ciliary neurotrophic factor (CNTF; 50 ng/ml), bone morphogenetic protein 2 (BMP2; 50 ng/ml) or pituitary adenylate cyclase-activating polypeptide (PACAP; 50 ng/ml). Neuronal cultures were derived from E13.5 hippocampal primordia by previously described methods. Primary cortical astrocytes were isolated from 1- to 2-day-old neonates and prepared according to published methods (14).

The transcriptional profile strategy for the identification of astrocyte-relevant transcripts exploited various primary and induced astrocyte culture systems, primary neuron cultures, NSCs, and regionally defined and developmentally staged CNS tissues, including E13.5 cortex, the corpus callosum (CC) isolated from postnatal day [(P)5 and P10] brain, and neocortical layers II–VI and subpial/glia limitans (GL) of the 6-week-old adult brain. Total RNA was used to prepare *in vitro*-transcribed amplified probes and the hybridized to oligonucleotide microarrays (Affymetrix U74Av2) representing 12,488 probe sets representing 8,585 mouse genes.

**Data Processing and Normalization.** The CEL files are obtained by using Affymetrix MICROARRAY SUITE software. The DNA-Chip Analyzer (DCHIP, VERSION 1.3) is used to normalize all CEL files to the baseline arrays and compute the model-based expression (PM-only model) (15). We normalized arrays only within tissue types by picking the baseline arrays for each tissue type. When calculating the model-based expression, constant array outliers

Abbreviations: NSC, neural stem cell; ISH, *in situ* hybridization; *En*, embryonic day *n*; *Pn*, postnatal day *n*; CNTF, ciliary neurotrophic factor; BMP, bone morphogenetic protein; PACAP, pituitary adenylate cyclase-activating polypeptide; LBFC, lower-bound fold change; SVM, support vector machine; SVZ, subventricular zone; GFAP, glial fibrillary acidic protein; ApoE, apolipoprotein E, UHC, unsupervised hierarchical clustering; R-SVM, recursive-supervised machine; CC, corpus callosum; GL, glial limitans; GM, gray matter; WM, white matter.

<sup>†</sup>R.M.B. and R.S.K. contributed equally to this work.

<sup>††</sup>To whom correspondence should be addressed at: Dana-Farber Cancer Institute, Harvard Medical School, 44 Binney Street (M413), Boston, MA 02115. E-mail: ron.depinho@dfci.harvard.edu.

© 2004 by The National Academy of Sciences of the USA

within each of the six groups; 10% FBS, CNTF/BMP2/PACAP, hippocampal neurons, primary cortical astrocytes, the gray matter (GM) and the CC/GL, were considered as a real biological effect.

**Pooling the Replicates and Lower Bound Fold Change.** We averaged replicate samples by considering model-based SEs of individual expression values by using a resampling method (15). To identify the genes with large change between two groups of replicates, we use the parametric estimation of lower-bound fold change (LBFC) as the measure of change (16). The selection criteria of selecting genes with an LBFC of  $\geq 3$  typically corresponds to a fold change of at least 5 in conventional gene expression assays such as quantitative PCR.

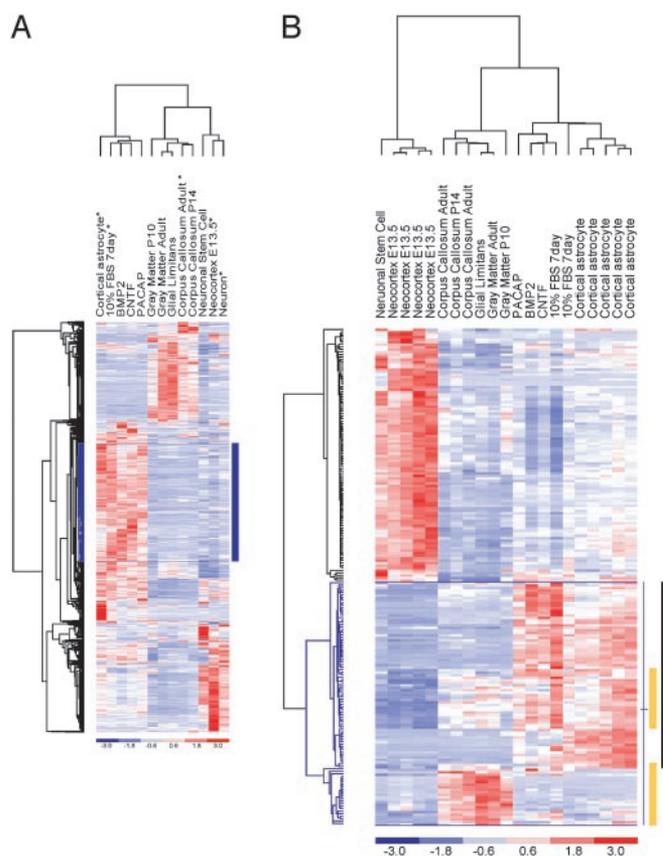
**Selection of Genes That Contribute Most in Distinguishing the Astrocyte Culture Samples and NSC/Early Embryo Brain Samples.** A total of 2,061 genes with large variation and high presence call (by DCHIP) were chosen, and R-SVM, software developed by W.H.W., is used to identify the genes contributing most in distinguishing two sample groups (17). Unlike other gene-wise analysis, this algorithm finds a set of genes that work as a group in separating two sample groups by recursively building linear support vector machines (SVMs) (18). To overcome possible overfitting, we performed the leave-one-out crossvalidation and the error rate was zero. Both the gene selection step and the SVM building step were included in crossvalidation.

**Validation of Astrocyte Candidate Genes by ISH.** Probes were scored for labeling efficacy, CNS expression, and brain region/cell type distribution. Genes with a “glial” expression pattern across neural development [i.e., increasing expression from E13.5 to adult, expression in white matter (WM), the GL, and the subventricular zone (SVZ)] were in most cases easily separated from those with a neuronal-only type pattern (i.e., hippocampus and specific cortical layers) by anatomical assessment alone. Genes suspected to be “astrocytic/glial” by anatomic criteria were stringently validated for cell type specificity by combining ISH with several lineage specific immunohistochemical markers, including glial fibrillary acidic protein (GFAP) (astrocytic), Olig2 (oligodendroglial), or NeuN (neuronal).

**Supporting Information.** *Supporting Methods*, Figs. 5 and 6, Tables 1–8, and CEL files can be found in supporting information, which is published on the PNAS web site.

## Results

**Reproducibility and Distinctiveness of Transcriptional Profiles.** In an effort to identify astrocyte-specific genes, we implemented a multilevel biological prioritization and filtering scheme based on the combined use of several *in vitro* astrocyte differentiation systems, isolated primary astrocytes from the perinatal brain, and various microdissected astrocyte-rich regions of the telencephalon (Fig. 6). The first series of experiments exploited the capacity of various agents (10% FBS, CNTF, BMP2, or PACAP; see *Methods*) to elicit a common astrocytic phenotype from NSCs, reasoning that comparative transcriptional profiles across these protocols should enrich for genes common to most astrocytes, regardless of isolation and induction procedures. Replicates within a given experimental modality or tissue type demonstrated a high degree of reproducibility (correlation coefficient 0.95–0.99) and when analyzed as groups, highlighted the distinctiveness of the astrocyte profile from the profiles of neurons, NSCs, and embryonic cortex (Fig. 5). The reproducibility and fidelity of these microarray data sets serve to document the quality and purity of the samples used, thereby supporting their use in molecularly defining the astrocyte lineage.



**Fig. 1.** Identification of astrocyte candidate genes by UHC and R-SVM. (A) UHC analysis divides the experimental samples into two distinct groups that cluster on separate branches of the dendrogram. All of the astrocyte samples are together. The short vertical distance between the astrocyte samples in the dendrogram indicates statistical similarity between the cortical astrocyte samples and the various differentiated astrocytes. Similarly, the CC, WM, and GL cluster together, suggesting a common transcriptional signal. The NSC, the E13.5 cortex, and neuronal-lineage-committed cells cluster together. The expression level matrix is shown representing standardized values from  $-3$  (blue, below the mean) to  $3$  (red, above the mean). The mean (0 value) is white. Rows correspond to different genes and the columns represent the various experimental samples. When we use all *in vitro* and *in vivo* experimental samples, UHC generates a large cluster of 393 genes (outlined in blue), which are strongly expressed among the *in vitro* astrocyte samples (indicated by blue bar). Although GFAP is among this group of astrocyte-associated genes, there is no obvious GFAP subcluster. (B) R-SVM, a class prediction tool, identified a subset of 85 genes, which contribute most to distinguishing astrocytes from undifferentiated or early lineage-committed cells. A majority (53%) of the astrocyte candidate genes were from only *in vitro* astrocyte experimental samples (indicated by black bar), and the remainder were differentially expressed both in cultured astrocytes and among the brain subregions (yellow bars). Regions of overlap indicate genes, which were differentially expressed in both experimental samples.

**Identification of Astrocyte Markers.** To test our assumption that all astrocytes share a common transcriptional profile, we performed unsupervised hierarchical clustering (UHC) of all experimental samples (Fig. 1A). This unbiased approach organized the experimental samples into two major groups, one consisting of multipotent NSCs and lineage-committed progenitors (NSCs, E13.5 neocortex, and hippocampal neuroblasts) and the other containing differentiated astrocytes. The remarkably similar expression profiles among the various *in vitro*-differentiated astrocytes and their tight association with the primary cortical astrocytes suggests that, despite the distinct signaling pathways engaged by the various agents used to drive commitment to the

astrocyte lineage, they share a similar molecular profile. Similarly, the tight clustering of microdissected brain subregions, CC, GM, and GL, demonstrates a high degree of transcriptional relatedness that is presumably related in part to their astrocyte-rich composition. The lack of significant overlap among the *in vitro* and *in vivo* UHC clusters likely results from a combination of the cell type heterogeneity of the CNS tissues as well as culture-induced stimulation of astrocytes to assume a reactive rather than the resting state typical of astrocytes in the normal brain. Together, these findings support the use of the multiple sample types as a starting experimental approach for the identification of candidate astrocyte genes that are expressed in normal brain.

An in-depth bioinformatic search for astrocyte-specific genes comprised three different methods: (i) a biased search for genes with an expression pattern similar to the best known astrocyte marker, *GFAP*, (ii) an unbiased search by a class prediction tool, recursive-supervised machine (R-SVM) analysis, and (iii) an empirical threshold approach to identify a common set of genes among complementary data sets. In the *GFAP* cluster analysis, *GFAP* captured a group of 393 genes that were differentially expressed by all *in vitro* astrocyte samples (Fig. 1A). Whereas this cluster likely represents a potential source of novel astrocyte genes (see Table 2 for a complete list), it is notable that no clear *GFAP* subcluster emerged, and correspondingly, a random sampling of genes among the *GFAP* cluster revealed their limited expression in CNS astrocytes by RNA ISH (see below). Next, to avoid the bias of so-called signature genes, R-SVM analysis was used to identify genes that “as a group” (unlike genes found by other traditional two-group comparison methods) contribute most to distinguishing the two groups (17). From a total of 2,005 genes, which show sufficient variation in expression over all samples, R-SVM identified a set of 85 genes that most significantly contribute to the astrocyte group (Fig. 1B). The relative contribution of each gene in distinguishing astrocytes from the NSC and early lineage committed cells is presented in Table 3. Strikingly, the gene that contributes most significantly is the main lipid transport protein in the CNS, apolipoprotein E (*apoE*) (8.8%), whereas the *GFAP* is ranked at number 22 (0.43%).

To capture a greater representation of the astrocyte transcriptome in a physiological context, empirical threshold studies were also conducted to compare the data sets of CC, GM, and GL with those derived from E13.5 cortex (a period of development preceding the emergence of astrocytes, ca. E17.5). The application of the expression criteria of more than three LBFCs relative to E13.5 cortex yielded 84 genes in the CC, 103 in the GM, and 100 in the GL. An intersection of these three gene lists generated a list of 47 “common *in vivo*” genes, which could represent candidate astrocyte genes because this cell type is common to these brain subregions. Of these 47 common *in vivo* genes, eight genes were differentially expressed (more than three LBFCs) among the “common *in vitro*” genes, which are defined as genes expressed in at least four of five *in vitro* astrocyte data sets (a complete list of the union of these genes is presented in Table 3). These data support our starting assumption that there may be qualitative and quantitative differences between the genes expressed by astrocytes maintained in cell culture and those found in the normal brain. It is also possible that this limited overlap reflects regionally restricted expression patterns of astrocytes in various brain microenvironments (i.e., only in the CC, GM, or GL). In an attempt to obtain a more complete view of potential molecular diversity of astrocytes in the adult brain, we performed a pairwise comparison of each CC, GM, and GL gene list and union of the four *in vitro* astrocyte data sets. These comparisons yielded 33, 29, and 23 genes for the CC, GM, and GL gene list, respectively (see Table 4).

These data serve to underscore that markedly distinct molecular profiles can emerge through the use of specific tissues,

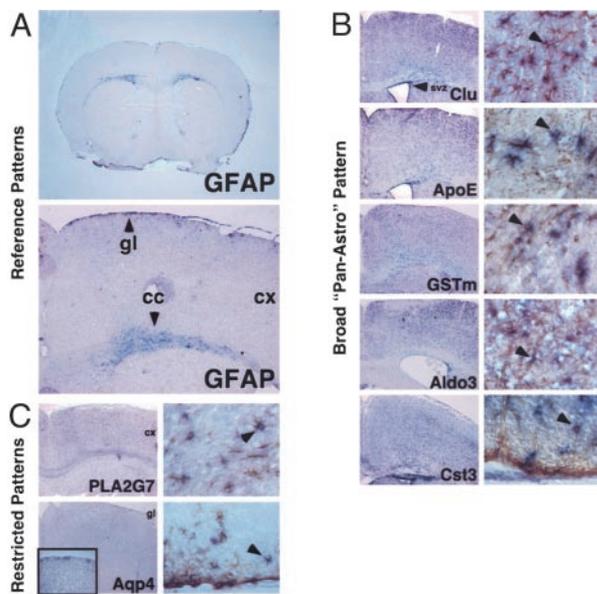
model systems, and experimental conditions and the application of specific bioinformatic methods, further justifying the comprehensive set of comparisons, bioinformatic approaches, model systems, and distinct astrocyte and nonastrocytic tissues and cell types.

**Validation of Candidate Astrocyte-Specific Genes.** We used a combination of ISH and lineage-specific immunohistochemical markers to assess the temporal and spatial patterns of a cross section of our candidate astrocyte-associated genes derived from various data sets.

An astrocyte candidate list derived from up-regulated genes after differentiation of NSC by exposure to serum ( $n = 4$  experiments), was a poor predictor of astrocyte genes. Of the 83 genes that were differentially expressed (more than three LBFCs) by exposure to 10% FBS, the top 19 genes were tested by ISH; of these genes, only two (*GFAP* and *aquaporin 4*) identified astrocytes in adult mouse brain (Table 1). The ability to identify astrocytes by ISH improved moderately when genes were selected from among the large UHC of astrocyte-associated genes (the so-called *GFAP* cocluster, 19 of 37). The R-SVM approach proved to be most effective with 10 of 13 candidate genes identifying astrocytes, whereas the common *in vivo* and common *in vitro* lists identified 9 of 12 genes and 8 of 13 genes, respectively (see Table 1 for summary of results).

Notably, the validated astrocyte genes exhibit widely varying patterns of expression with only a small subset showing coexpression with *GFAP*, yet staining cells with unequivocal astrocyte morphology. Of the genes with a restricted expression pattern, three genes had a *GFAP*-like pattern with cells predominantly labeled in the WM, GL, and SVZ (*aquaporin 4*, *brain glycogen phosphorylase*, and *brevican*). *Aquaporin 4* was predominantly expressed by astrocytes of the GL in the subpial and perivascular locations, as reported (19), whereas five genes were prominently expressed in the SVZ and to a more limited extent in the adjacent GM (*Id3*, *vascular cell adhesion molecule 1*, *N-myc down-regulated 2*, *integral membrane protein 1*, and *endothelial differentiation receptor 1*). In addition, two genes labeled ependymal cells (*diazepam-binding inhibitor* and *IL6 signal transducer*); these genes were included here because ependymal cells label positively with *GFAP* and may arise from a common precursor. Many genes had a heterogeneous expression pattern, labeling scattered populations of cells in the GM, whereas five genes had a broad pattern of expression labeling cells in the SVZ, WM, and throughout the GM of the telencephalon; these genes included: *clusterin*, *cystatin C*, *apoE*, *GST*, and *aldolase 3* (see Fig. 2; see also Table 5, where all genes are listed).

**Tight-Cluster Analysis of Candidate Astrocyte-Specific Genes.** The collection of validated astrocyte genes makes possible an effective prospective bioinformatic identification of additional astrocyte genes and a more comprehensive molecular view of this lineage. To that end, we used a clustering algorithm capable of identifying genes that “tightly cluster” with validated genes (see *Methods* for details). A total of 2,061 genes with sufficient variability over samples were selected for tight-cluster analysis by using both *in vitro* and *in vivo* samples; 51% of the probe sets were assigned to the top 30 tightest clusters, of which four were identified by inclusion of six validated astrocyte genes (see Fig. 3A). These tight clusters are remarkable for identifying differentially expressed genes in both the cultured astrocytes and normal brain, but not in NSCs, neurons, or embryonic brain. For the *in vitro* sample data sets, there were nine tight clusters, all of which contained one or more of our validated astrocyte-associated genes. Data presented in Fig. 3B show three representatives of nine tight clusters from the cultured astrocyte data set, each identified by several ISH-validated genes (a total of 16 *in situ*-validated genes among the nine tight clusters). Each tight



**Fig. 2.** Astrocytic candidate gene validation. (A) Candidate astrocyte genes with glial expression based on similarity to the reference gene expression patterns for GFAP ISH and/or GFAP immunohistochemistry were chosen for further validation. Note the marked abundance of GFAP RNA in GL and CC (arrowheads) and relative absence in cortical GM (cx). (B) The majority of validated genes showed a broad “pan-astrocytic” pattern of expression in GM and WM astrocytes [shown here are: Clusterin (Clu), ApoE, GST (GSTm), Aldolase 3 (Aldo3), and Cystatin 3 (Cst3)]; a subset of each which were GFAP-positive. (C) Several validated astrocyte genes showed a restricted expression pattern in subsets of astrocytes. Phospholipase A, group7 (Pla2g7) was predominantly expressed in cortical GM astrocytes, whereas Aquaporin 4 (Aqp4) was highly abundant in GL regions.

cluster was dominated by a single, statistically significant ( $P < 0.01$ ), functional gene ontology category (see below and Table 6). These results suggest that tight-cluster analysis is not only an efficient means of identifying cell-type-specific genes, it can also identify functionally related astrocyte genes, which reflect normal astrocyte functions.

Among the 46 *GFAP* coclustered genes identified by tight-cluster algorithm (Fig. 3B), which exclude genes with uncertain membership, five of seven genes tested by ISH proved to be astrocyte-specific (*phospholipase A2 group 7*, *gap-junction channel protein 1- $\alpha$* , *aquaporin 4*, *vascular cell adhesion molecule 1*, and *brain glycogen phosphorylase*). These results underscore that the tight-cluster tool represents an efficient means of identifying additional astrocyte markers. Furthermore, the performance of the tight-cluster results continues to improve as more candidate genes are validated by ISH and that information is used to refine the tight-clustering parameters.

**Functional Annotation of Validated Astrocyte Genes and Tight-Cluster Genes.** Functional classification by the Gene Ontology database yielded many significant (see *Methods*) and distinct categories (see Table 6). Among them are categories consistent with known astrocyte function including genes encoding potent antioxidant activity (e.g., *GST* and peroxiredoxin 5), excitatory amino acid uptake (solute carrier 1), immune modulation/chemotaxis (e.g., *CXCR4* and *CX3-C motif 1*), microvascular regulation (*PLAZG7*, *thrombospondin*, and *vascular cell adhesion molecule*), and blood–brain barrier function (*aquaporin 4*). Of special note is a prominently represented category linked to lipid transport and metabolism (e.g., *apoE*, *fatty acid-binding protein*, and *sterol CoA desaturase-2*), providing experimental evidence for the emerging

concept that astrocytes play an obligatory role in cholesterol synthesis and transport to neurons.

## Discussion

In this study, we used combinatorial analysis of microarray data from multiple *in vitro* and *in vivo* astrocyte samples along with relevant biological filters and bioinformatic approaches to identify astrocyte-specific genes that can be validated *in vivo*. This strategy resulted in a collection of genes with robust astrocyte-associated expression in the normal brain. Of significance, these astrocyte markers revealed surprising molecular heterogeneity among astrocytes in the normal adult brain that belies their functional range in CNS physiology. The regional and functional range is illustrated by identification of astrocyte subpopulations that are restricted to the SVZ, the presumed location of astrocytes that retain stem cell properties, and another population of astrocytes that are highly specialized and regulate water transport from the cerebrospinal fluid space into the brain parenchyma. Finally, these validated astrocyte-specific genes enabled the use of tight-cluster analysis that provided a revealing view of the astrocyte lineage and its functional heterogeneity in the normal brain.

The anatomically diverse expression patterns of the tightly clustered astrocyte genes raises the possibility that the molecular profile of astrocyte subgroups may reflect a physiological response related to a unique biological function or, alternatively, may represent lineage subsets encrypted by different developmental subroutines. With increasing evidence that neuronal–glial communication is critically important for local neuronal function, it is possible that astrocytes with a limited expression pattern reflect an adaptation to a restricted neurovascular environment, whereas those with a broad expression pattern serve some global function. A similar heterogeneous expression pattern is seen among glial specific genes recently identified by microarray analysis in *Drosophila* CNS (20).

The molecular identification of astrocyte subtypes in the mammalian brain and their regional restriction will likely prove informative in understanding the pathophysiology of many of the major brain diseases that have been associated with astrocyte dysfunction but without a link to normal astrocyte biology. The body of evidence pointing to astrocytes having a seminal role in modulating primary and acquired neurodegenerative disease processes is substantial. The question that remains unresolved is the extent to which astrocytes play a causal and specific role in these diseases. What is clear from the identification of the various functional categories presented here is that astrocytes may impact on multiple levels in multistep disease processes. For example, the impact of astrocytes in late-onset Alzheimer’s disease not associated with the *apoE4* isoform mutation may be exerted through various mechanisms, as demonstrated by three broadly expressed genes that were among those most significantly contributing to the astrocyte transcriptome in our profiling studies. *Cystatin C*, *clusterin*, and *apoE*, representing extracellular proteases, protein chaperones, and the cholesterol shuttle, as well as the several genes associated with impaired excitatory amino acid transport and inflammatory immune modulators may be important in the pathogenesis of this disease (see functional categories in Table 6), each working through separate mechanisms (21–25). Indeed, a recent study (22) has reported that transgenic mice, which develop Alzheimer’s type pathology, have a dramatically accelerated and more severe clinical course if they are also deficient for both *apoE* and *clusterin*. An effective and systematic means to evaluate the postulated role of astrocytes would be to view them from a global perspective in disease progression.

Finally, many of the genes identified in this study have previously been implicated to play a modulatory role in neurodegenerative diseases, stroke, and epilepsy. For example, *parkin*, a ubiquitin ligase identified from the tight-cluster analysis is associated with



early-onset Parkinson's disease. Although expressed in both neurons and astrocytes in culture, it is up-regulated only in astrocytes after cytotoxic stress (26). Dysfunction of *parkin* in astrocytes may lead to nigrostriatal degeneration by loss of tropic support, free radical damage, and/or release of cytokines by astrocytes (26). A similar mechanism involving astrocytes may explain the lack of amyotrophic lateral sclerosis-like pathology when the *SOD1* gene was targeted to motoneurons (27). The postulated role of astrocytes in primary neuronal pathology is also supported by identification of *Niemann Pick type C1* gene among the tight-cluster genes. This gene has previously been shown to exclusively label astrocytes (28) and is involved in a pediatric neurodegenerative disease by the same name. In stroke and epilepsy, astrocytes may protect neurons by their strong antioxidant mechanisms. One postulated unifying mechanism is the induction of phase II detoxifying enzymes, several of which were identified in this study including GST, peroxiredoxin, steroyl-CoA desaturase, and aldehyde dehydrogenase. Several of these enzymes may be activated by a common transcriptional factor,

*NF-E2-related factor* (29), a gene also identified from the tight-cluster analysis. On the basis of these data, we suggest that this first, to our knowledge, comprehensive and specific view of the astrocyte transcriptome will provide an experimental framework to dissect the complex role of this highly diverse lineage in the normal and diseased brain.

We thank Lynda Chin for critical reading of the manuscript and helpful discussions; the members of the DePinho, Rowitch, and Wong Laboratories for fruitful discussions; and Shakuntala Karki, Xiaoyou Liang, and Dong In Yuk for assistance with cell culture, RNA preparation, and ISH, respectively. This work was supported by National Institutes of Health Grant P01CA95616 (to R.A.D.); grants from the National Multiple Sclerosis Society and the James S. McDonnell Foundation (to D.H.R.); National Institutes of Health Grants 1R01HG02341 and P2OCA96470 (to W.H.W.); and National Institute of Health Fellowships 5K08NS42737 (to R.M.B.) and 5K08CA82241 (to E.A.M.). K.L.L. is supported by National Institutes of Health Training Grant HD007466 to the Department of Pathology, Harvard Medical School. R.A.D. is an American Cancer Society Research Professor.

1. Shirasaki, R. & Pfaff, S. L. (2002) *Annu. Rev. Neurosci.* **25**, 251–281.
2. Miller, R. H. (2002) *Prog. Neurobiol.* **67**, 451–467.
3. Steindler, D. A. & Laywell, E. D. (2003) *Glia* **43**, 62–69.
4. Doetsch, F. (2003) *Nat. Neurosci.* **6**, 1127–1134.
5. Fields, R. D. & Stevens-Graham, B. (2002) *Science* **298**, 556–562.
6. Newman, E. A. (2003) *Trends Neurosci.* **26**, 536–542.
7. Auld, D. S. & Robitaille, R. (2003) *Neuron* **40**, 389–400.
8. Song, H., Stevens, C. F. & Gage, F. H. (2002) *Nature* **417**, 39–44.
9. Nedergaard, M., Ransom, B. & Goldman, S. A. (2003) *Trends Neurosci.* **26**, 523–530.
10. Liu, Y., Wu, Y., Lee, J. C., Xue, H., Pevny, L. H., Kaprielian, Z. & Rao, M. S. (2002) *Glia* **40**, 25–43.
11. De Smet, C., Nishimori, H., Furnari, F. B., Bogler, O., Huang, H. J. & Cavenee, W. K. (2002) *J. Neurochem.* **81**, 575–588.
12. Geschwind, D. H., Ou, J., Easterday, M. C., Dougherty, J. D., Jackson, R. L., Chen, Z., Antoine, H., Terskikh, A., Weissman, I. L., Nelson, S. F. & Kornblum, H. I. (2001) *Neuron* **29**, 325–339.
13. Reynolds, B. A. & Weiss, S. (1992) *Science* **255**, 1707–1710.
14. McCarthy, K. D. & de Vellis, J. (1980) *J. Cell Biol.* **85**, 890–902.
15. Li, C., & Wong, W. H. (2003) in *The Analysis of Gene Expression Data: Methods and Software*, eds. Parmigiani, G., Irizarry, E. S. G. R. & Zeger, S. L. (Springer, Berlin), pp. 120–141.
16. Li, C. & Wong, W. H. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 31–36.
17. Zhang, X., & Wong, W. H. (2001) *Recursive Sample Classification and Gene Selection Based on SVM: Method and Software Description* (Department of Biostatistics, Harvard School of Public Health, Boston), pp. 1–10.
18. Collobert, R. & Bengio, S. (2001) *J. Machine Learn. Res.* **1**, 143–160.
19. Nielsen, S., Nagelhus, E. A., Amiry-Moghaddam, M., Bourque, C., Agre, P. & Ottersen, O. P. (1997) *J. Neurosci.* **17**, 171–180.
20. Freeman, M. R., Delrow, J., Kim, J., Johnson, E. & Doe, C. Q. (2003) *Neuron* **38**, 567–580.
21. Holtzman, D. M., Bales, K. R., Tenkova, T., Fagan, A. M., Parsadonian, M., Sartorius, L. J., Mackey, B., Olney, J., McKeel, D., Wozniak, D. & Paul, S. M. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2892–2897.
22. DeMattos, R. B., O'Dell, M. A., Parsadonian, M., Taylor, J. W., Harmony, J. A., Bales, K. R., Paul, S. M., Aronow, B. J. & Holtzman, D. M. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 10843–10848.
23. Deng, A., Irizarry, M. C., Nitsch, R. M., Growdon, J. H. & Rebeck, G. W. (2001) *Am. J. Pathol.* **159**, 1061–1068.
24. Bezzi, P., Domercq, M., Brambilla, L., Galli, R., Schols, D., De Clercq, E., Vescovi, A., Baretta, G., Kollias, G., Meldolesi, J. & Volterra, A. (2001) *Nat. Neurosci.* **4**, 702–710.
25. Shih, A. Y., Johnson, D. A., Wong, G., Kraft, A. D., Jiang, L., Erb, H., Johnson, J. A. & Murphy, T. H. (2003) *J. Neurosci.* **23**, 3394–3406.
26. Ledesma, M. D., Galvan, C., Hellias, B., Dotti, C. & Jensen, P. H. (2002) *J. Neurochem.* **83**, 1431–1440.
27. Clement, A. M., Nguyen, M. D., Roberts, E. A., Garcia, M. L., Boillee, S., Rule, M., McMahon, A. P., Doucette, W., Siwek, D., Ferrante, R. J., et al. (2003) *Science* **302**, 113–117.
28. Patel, S. C., Suresh, S., Kumar, U., Hu, C. Y., Cooney, A., Blanchette-Mackie, E. J., Neufeld, E. B., Patel, R. C., Brady, R. O., Patel, Y. C., et al. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1657–1662.
29. Lee, J. M., Calkins, M. J., Chan, K., Kan, Y. W. & Johnson, J. A. (2003) *J. Biol. Chem.* **278**, 12029–12038.