

Cell type-specific genes show striking and distinct patterns of spatial expression in the mouse brain

Younhee Ko^{a,1}, Seth A. Ament^{b,1}, James A. Eddy^{b,c}, Juan Caballero^b, John C. Earls^{b,d}, Leroy Hood^{b,2}, and Nathan D. Price^{a,b,c,d,2}

^aInstitute for Genomic Biology and Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801; ^bInstitute for Systems Biology, Seattle, WA 98109; ^cDepartment of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801; and ^dDepartment of Computer Science and Engineering, University of Washington, Seattle, WA 98195

Contributed by Leroy Hood, January 2, 2013 (sent for review December 26, 2012)

To characterize gene expression patterns in the regional subdivisions of the mammalian brain, we integrated spatial gene expression patterns from the Allen Brain Atlas for the adult mouse with panels of cell type-specific genes for neurons, astrocytes, and oligodendrocytes from previously published transcriptome profiling experiments. We found that the combined spatial expression patterns of 170 neuron-specific transcripts revealed strikingly clear and symmetrical signatures for most of the brain's major subdivisions. Moreover, the brain expression spatial signatures correspond to anatomical structures and may even reflect developmental ontogeny. Spatial expression profiles of astrocyte- and oligodendrocyte-specific genes also revealed regional differences; these defined fewer regions and were less distinct but still symmetrical in the coronal plane. Follow-up analysis suggested that region-based clustering of neuron-specific genes was related to (i) a combination of individual genes with restricted expression patterns, (ii) region-specific differences in the relative expression of functional groups of genes, and (iii) regional differences in neuronal density. Products from some of these neuron-specific genes are present in peripheral blood, raising the possibility that they could reflect the activities of disease- or injury-perturbed networks and collectively function as biomarkers for clinical disease diagnostics.

The mammalian brain can be subdivided into more than 100 anatomically and functionally distinct regions, each containing multiple cell types, including various classes of neurons and glia (1). Despite several decades of modern neuroscience research, we lack a complete understanding of how these brain compartments are specified and maintained or of how structural differences are translated into the diverse functions performed within the brain. Understanding the transcriptional correlates of brain region and cell type diversity holds promise for elucidating brain function and development. Moreover, unique transcriptional signatures for brain regions have potential clinical uses as biomarkers for disease diagnostics (2).

Recent technological innovations have enabled researchers to begin systematically characterizing regional differences in brain gene expression. Transcriptome profiling of isolated neurons and glia has revealed that certain genes are specifically expressed in neurons, astrocytes, or oligodendrocytes (3), or even within specific classes of cortical neurons (4). A complementary approach, high-throughput *in situ* hybridization of more than 20,000 mouse genes (5, 6), allows visualization of expression patterns across the entire brain at the single-cell level, revealing tremendous diversity in the expression profiles of individual genes.

The diversity of spatial expression patterns for genes in the adult brain (as well as their distinct functions) suggests that many regional differences have gene expression correlates. Indeed, clustering analysis using 3,041 genes from the Mouse Brain Atlas produced by the Allen Institute for Brain Science (Allen Brain Atlas, ABA) (5), the most comprehensive *in situ* hybridization database, revealed 30 transcriptionally distinct spatial units, which had 70% overlap with a standardized reference atlas (7). However, the specific genes responsible for these regional differences have not been identified, leading to a number of intriguing questions.

What genes are most strongly correlated with regional specification in the adult brain? How does gene expression at the level of brain regions relate to the cell type-specific expression diversity revealed by microarray experiments?

We addressed these questions by studying spatial expression profiles for panels of genes expressed in only one of the brain's major cell types. We analyzed ABA data for cell type-specific genes characterized by Cahoy et al. (3), who used microarrays to profile expression patterns in purified populations of neurons, astrocytes, and oligodendrocytes (3). The presumption is that these cell type-specific transcripts will at least in part specify the distinct cell types (states) that characterize functionally distinct neurons and glia. The combined expression of 170 neuron-specific genes in the coronal plane revealed strikingly distinct and symmetrical spatial expression patterns, delineating regions of the mouse brain that correlated both with major brain compartments and with small subregions. Similar analyses with 50 astrocyte- or 44 oligodendrocyte-specific genes also revealed symmetrical regional differences, but these patterns were less distinct. We discuss the relevance of our results to understanding the function of different brain regions, to studying brain development, and to discovering disease biomarkers.

Results

Spatial Expression Patterns of Neuron-Specific Genes Correspond Precisely to Regional Differences in the Brain. To what extent do subdivisions in the mammalian brain correspond to regional differences in the expression of genes within major cell types? To address this question, we analyzed highly spatially resolved brain gene expression data from the ABA for panels of cell type-specific genes. We broke the coronal and sagittal sections into small (coronal: 20 × 30 μm; sagittal: 20 × 40 μm) "patches" (pixels) reflecting local gene expression. We then performed *k*-means clustering (with *k* between 2 and 100) to find subsets of patches with similar gene expression across all of the neuron-specific, astrocyte-specific, or oligodendrocyte-specific genes (i.e., the *k*-means clustering grouped patches based on their similarity in the cell type-specific gene expression space). For neurons, the clusters even up to *k* = 100 were highly spatially coordinated (Fig. 1 and Fig. S1), although the size of the added regions became much smaller above *ca.* *k* = 60, which is the *k* we will use for a number of examples herein. The resulting transcriptionally defined clusters were then transformed

Author contributions: Y.K., S.A.A., J.A.E., J.C., L.H., and N.D.P. designed research; Y.K., S.A.A., J.A.E., J.C., and J.C.E. performed research; Y.K., S.A.A., and J.A.E. contributed new reagents/analytic tools; Y.K., S.A.A., J.A.E., J.C., and N.D.P. analyzed data; and Y.K., S.A.A., J.A.E., L.H., and N.D.P. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

¹Y.K. and S.A.A. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: lhood@systemsbiology.org or nprice@systemsbiology.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1222897110/-DCSupplemental.

back into a color-coded brain image to assess their overlap with known brain regions. Our goal was then to systematically characterize the transcriptionally distinct, spatially contiguous neuronal and glial subtypes based on their gene expression patterns across the brain.

We defined cell type-specific genes as those that were previously shown to be >10-fold enriched in each of these cell types (Dataset S1). Cahoy et al. (3) found 320 genes that were >10-fold enriched in neurons, 185 in astrocytes, and 131 in oligodendrocytes, but ABA data were available for only a subset of these genes. We downloaded all available in situ hybridization images from the ABA showing expression on coronal or sagittal slices through the centermost part of the brain (one to three images per gene on each plane). The resulting dataset included images for 170 neuron-, 44 oligodendrocyte-, and 50 astrocyte-specific genes for the coronal plane and for 250 neuron-, 101 oligodendrocyte-, and 154 astrocyte-specific genes for the sagittal plane (where ABA measured more genes).

The 170 neuron-specific genes with coronal data were enriched for genes annotated in Gene Ontology (GO) as part of the synapse (28 genes, $P = 7.1 \times 10^{-18}$) and for biological processes such as synaptic transmission (18 genes, $P = 4.1 \times 10^{-13}$), regulation of neurotransmitter levels (11 genes, $P = 1.8 \times 10^{-10}$), and neuron projection development (14 genes, $P = 8.5 \times 10^{-8}$). Many of the 50 astrocyte-specific genes with coronal data encode proteins that are secreted into the extracellular region (11 genes, $P = 0.02$) or that are integral to membranes (19 genes, $P > 0.05$), and some of these genes are involved in biological processes such as blood vessel development (5 genes, $P = 4.6 \times 10^{-3}$), steroid metabolism (4 genes, $P = 0.01$), and response to oxidative stress (3 genes, $P = 0.025$). The 44 oligodendrocyte-specific genes with coronal data included genes with functions in the ensheathment of neurons (4 genes, $P = 6.8 \times 10^{-5}$), as well as components of cell surface receptor linked signaling pathways (8 genes, $P > 0.05$) and genes intrinsic to the plasma membrane (20 genes, $P > 0.05$).

k-Means clustering of brain patches on the coronal plane based on the expression of neuron-specific genes revealed remarkably clear and symmetrical spatial patterning, with most clusters corresponding tightly to known anatomical subdivisions in the brain (Fig. 1A, C, and E). When we subdivided the brain into small numbers of clusters ($k < 10$), we observed transcriptionally distinct units corresponding to major brain compartments (Fig. 1C), as indicated by manual comparison of clusters to the ABA reference atlas (Fig. 1A and B). The cerebral cortex and striatum were the first brain regions to appear as distinct clusters ($k = 3$). As k increased to 8, we observed distinct clusters for major midbrain and hindbrain regions such as thalamus, hypothalamus, and pallidum, as well as the corpus callosum and other white-matter regions. Similar results were obtained using images of neuron-specific genes on the sagittal plane (Fig. 1B, D, and F). These results suggest that large brain regions correspond to the most transcriptionally distinct, spatially defined units in the brain.

Considering larger numbers of *k*-means clusters revealed finer-scale brain structures. For instance, as we increased the number of clusters to $k = 20$, we found clusters corresponding to four distinct cortical layers (Fig. 1C and Fig. S2). Increasing the number of clusters still more, at $k = 60$, we observed unique transcriptionally defined and spatially contiguous clusters corresponding to smaller structures such as the dentate gyrus of the hippocampus and small nuclei within the amygdala and thalamus (Fig. 1E and F). These results suggest that relatively fine structures in the mammalian brain are associated with transcriptionally distinct classes of neurons.

Importantly, clusters derived from neuron-specific genes showed strong mirror symmetry between the left and right hemispheres of the brain at both small and large k (Fig. 2A). Major brain structures are almost completely symmetrical between hemispheres, despite a few minor structural and functional differences (8, 9). Therefore,

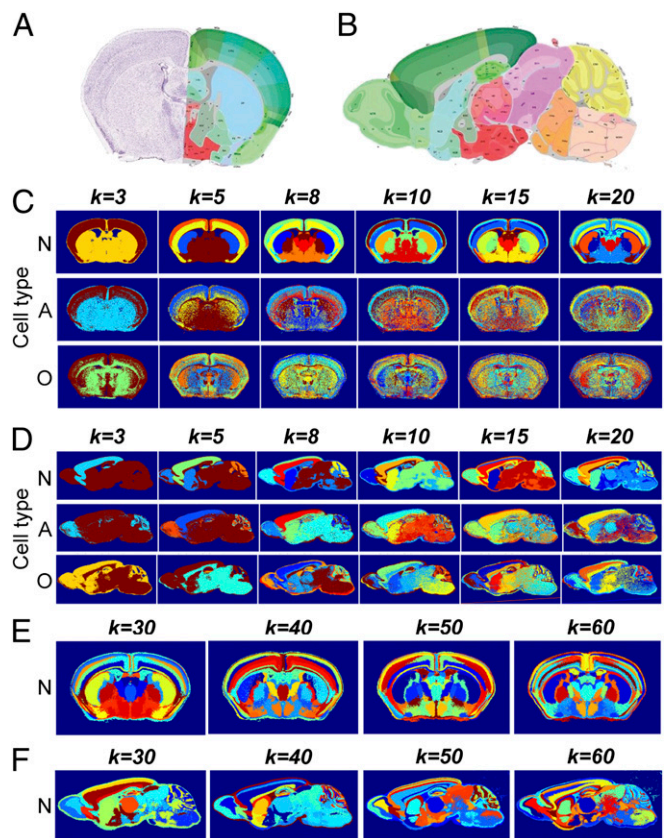


Fig. 1. Transcriptionally distinct, spatially contiguous regions in the adult mouse brain revealed by clustering of neuron-, astrocyte-, and oligodendrocyte-specific genes. Expression data were assembled for sections through the center of the mouse brain. Reference atlases for regions visible on the coronal (A) and sagittal (B) planes. (C) *k*-Means clusters derived from the expression of neuron-, astrocyte-, or oligodendrocyte-specific genes (labeled N, A, and O, respectively) on the coronal plane ($k = 3$ –20). (D) *k*-Means clustering based on the expression of these genes on the sagittal plane. Clustering of neuron-specific genes for larger numbers of clusters ($k = 30$ –60), using expression on the coronal (E) and sagittal (F) planes.

the strong symmetry in cluster assignments provides yet more evidence for the robustness of region-based clusters of neuron-specific genes. In addition, the result suggests that there may be relatively few hemispheric differences in gene expression within the major cell types in the adult mouse brain—at least with respect to the cell type-specific transcripts.

Clusters derived from neuron-specific genes were highly spatially contiguous; adjacent patches in the brain were very frequently assigned to the same cluster (Fig. 2B). For $k = 60$, 59 of the 60 clusters were spatially contiguous with >59% of the patches adjacent to cluster patches belonging to that same cluster (Fig. 2B). All 59 of these spatially contiguous clusters had significantly greater ($P < 0.001$) contiguity by this measure than we found in a distribution generated from randomly permuting the spatial coordinates for the patches within the mouse brain (while keeping the same number of patches per cluster). Known brain regions are spatially contiguous within the brain. As such, the high levels of spatial contiguity we observed for *k*-means clusters suggests that the vast majority of these transcriptionally defined regions correspond to functionally relevant brain regions, even with high numbers of clusters.

The sequence in which brain structures emerged as distinct *k*-means clusters (as we increased k) paralleled not only functional relationships between brain regions but also seemed to reflect developmental stages. For instance, at low k , we observed a single

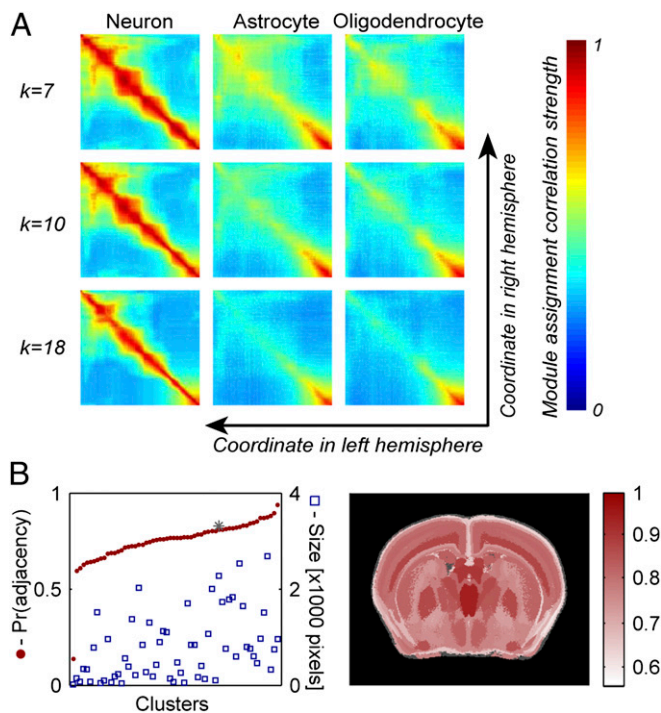


Fig. 2. Symmetry and transcriptional distinctiveness of k -means clusters derived from the spatial expression patterns of neuron-, astrocyte-, and oligodendrocyte-specific genes. (A) Heat maps depicting correlations between cluster assignments in the left vs. right hemisphere, relative to the distance of each point from the midline. Correlations are based on k -means clusters for neuron-, astrocyte-, or oligodendrocyte-specific genes at three representative numbers of clusters, $k = 7, 10,$ and 18 . (B) Probabilities (for $k = 60$ clusters) that adjacent neighbors of a patch lie in the same cluster. Probabilities for each cluster mapped to corresponding brain regions. Stars and grayed-out regions denote predominantly background clusters that could not be excluded a priori based on gene expression alone.

cluster corresponding to several hindbrain regions; these regions were resolved into distinct clusters for thalamic and hypothalamic nuclei at higher k . Similarly, we observed a single cluster for cortex at low k , which separated into clusters for each cortical layer at higher k . This trajectory mirrors the developmental sequence in which hindbrain and forebrain fates of neural progenitors are determined, before their assignment to specific hindbrain nuclei or cortical layers (10). This result provides further support for the biological relevance of the expression-defined brain region clusters. Similar relationships between developmental origins and adult gene expression have been reported previously using lower-resolution gene expression data (11).

The Allen Reference Atlas includes coordinates for >100 distinct anatomical regions. Our results suggest that many of these can be distinguished based on the expression of 170 neuron-specific transcripts. These observations raise fascinating questions: How many specific types of neurons exist in the brain, how are they differentiated from one another, and what are their functions? These data provide the means for beginning to analyze the number of transcriptionally distinct neuron-specific cell types.

Glia-Specific Genes also Reveal Spatially Clustered and Symmetrical Expression in the Brain, but with Fewer Distinct Clusters and Less-Precise Boundaries. Astrocyte- and oligodendrocyte-specific transcripts also displayed spatially clustered and symmetrical expression patterns, but they delineated fewer regions and the subregion boundaries were less clearly defined. As with neurons, astrocyte-specific genes displayed transcriptionally distinct clusters that corresponded to cortex, thalamus, and striatum, among other struc-

tures (Fig. 1 C and D). Many of the same structures were revealed from the expression of both neuron- and astrocyte-specific genes, which may relate to functional pairing of specific classes of astrocytes and neurons (defined by distinct expression patterns) in the distinct brain regions. This in turn could imply that region-specific transcriptional regulation within neurons and astrocytes encodes synergistic functions.

Oligodendrocyte-specific genes revealed transcriptionally distinct subdivisions that displayed a different architecture from what we observed with neurons and astrocytes. That is, they corresponded primarily to distinctions between white matter and gray matter. For instance, at $k = 3$, we observed a single cluster that mapped to most of the corpus callosum and third and fourth ventricles. These results are consistent with the primary function of oligodendrocytes in myelination of fiber tracts.

Glia-derived clusters were generally less distinct than those for neurons and revealed fewer known brain regions. Clustering became increasingly spatially noncontiguous (Fig. 1 C and D and Fig. S3) and less symmetrical (Fig. 2A) when we considered more than five to eight clusters of astrocytes or oligodendrocytes. The smaller number of glia-specific genes in our analysis (170 neuron-specific genes vs. 44 astrocyte-specific genes or 50 oligodendrocyte-specific genes) did not explain the weaker clustering; a follow-up analysis showed that randomly generated subsets of neuron-specific genes with a similar number of images to the glial gene sets still outperformed glial-specific genes in revealing brain regions (Fig. S4). These results suggest that whereas there are transcriptionally distinct subtypes of astrocytes and oligodendrocytes they are less distinct and explain far fewer brain region differences than do similar patterns in neurons. There seem to be significantly more transcriptionally defined neuron cell types than oligodendrocyte and astrocyte cell types.

Transcriptomic Signatures Distinguishing Cortical and Noncortical Brain Regions Are Characterized by Single Genes, Gene Sets, and Neuronal Density. What features of neuronal and glial gene expression differ between brain regions, accounting for these distinct, symmetric, spatially contiguous clusters? We addressed three (non-mutually exclusive) hypotheses: (i) single gene “markers” with dramatically different expression between regions, (ii) functionally related groups of genes with regionally different patterns of expression, and (iii) differences in neuronal cell density between brain regions. We focused on the major brain subdivisions that emerged as distinct transcriptional units based on k -means clustering with $k = 8$.

A neighbor-joining (hierarchical) tree comparing the characteristic expression patterns for each k -means cluster (derived from $k = 8$) revealed a primary division between cortical and noncortical brain regions, followed by further subdivisions among both cortical and noncortical brain regions (Fig. 3A). This result is consistent with our finding (above) that cortex vs. noncortex is the subdivision observed at the smallest k ($k = 3$).

The primary subdivision between cortical and noncortical regions corresponded to dramatically higher expression of many neuron-specific genes in the cortex (Fig. 3B). These differences could relate either to distinctions between the kinds of neurons in cortical vs. noncortical areas or to differences in cell density. To evaluate the latter contribution, we estimated the density of neurons, astrocytes, and oligodendrocytes within each brain region by calculating the ratio of positively stained vs. unstained patches, summed over all of the genes for each cell type. These ratios differed strongly across the brain (Fig. S5). For instance, oligodendrocyte density was very high in parts of the cerebral cortex, presumably reflecting the large number of myelinated fibers in the lower cortical layers. By contrast, certain gray matter nuclei in the mid- and hindbrain seemed depleted of oligodendrocyte-specific gene expression. These results, combined with the global differences in gene expression between cortical and noncortical regions, suggest that brain region differ-

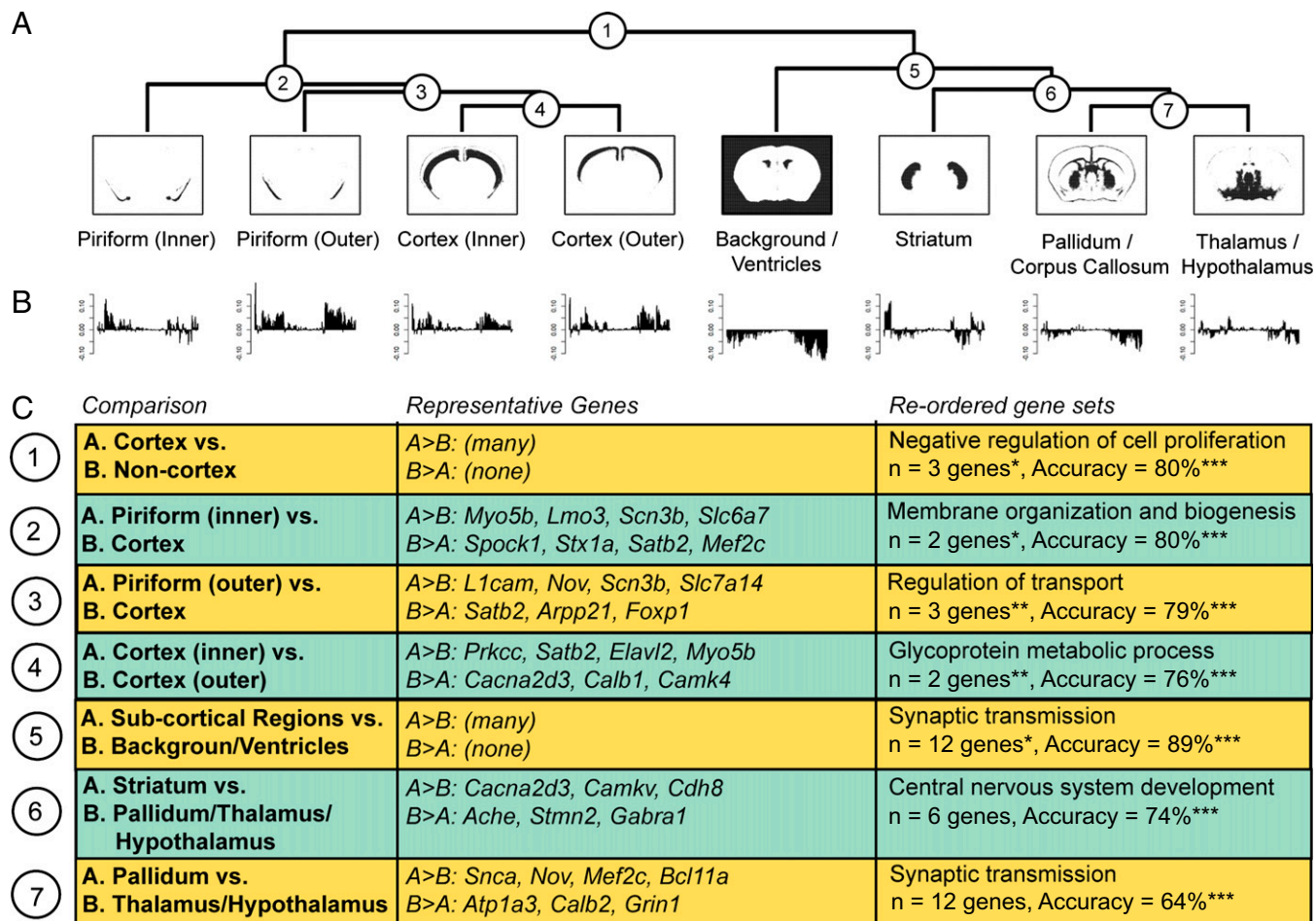


Fig. 3. Clusters corresponding to major subdivisions in the mammalian brain are characterized by expression differences for individual marker genes and reordering of functionally related gene sets. (A) Images of individual clusters for neuron-specific genes at $k = 8$ and hierarchical clustering of their characteristic expression patterns. (B) Characteristic expression patterns for 170 neuron-specific genes in each of the eight clusters. (C) Selected marker genes with >5-fold expression differences between the clusters shown. Gene Ontology gene sets with consistent differences in relative gene expression (reordering of gene set components) between clusters. P values for added-value of GO term vs. random sets of neuron-specific genes (genes, size n), and separability of clusters (accuracy) vs. randomly distributed pixels, calculated by permuting gene and pixel cluster labels, respectively. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

ences in cell type density contribute in part to the transcriptional distinctiveness of brain regions.

By contrast, examination of clusters defining somewhat finer subdivisions in the brain showed differences in the expression of subsets of neuron-specific genes, with high expression for different sets of genes in each brain region. We characterized statistically up- and down-regulated neuron-specific genes at each of the “branches” in the neighbor-joining tree (Dataset S2). These differentially expressed genes included known markers for some brain regions. For instance, changes in the expression of *Satb2*—a marker for layer 3/4 cortical neurons (12, 13)—distinguished k -means clusters for piriform cortex vs. other cortical regions and for inner vs. outer layers of cortex (Fig. 3C). Known and less-well-known markers with greater than fivefold differences in expression between each of the k -means clusters ($k = 8$) are shown in Fig. 3C.

The results above suggest that individual genes serve as markers for some regional differences identified through clustering. However, the robust clustering we observed using random subsets of just 40 neuron-specific genes (Fig. S4) suggests that no single gene is unique in defining the characteristic expression patterns of these brain regions. We used Differential Rank Conservation (DIRAC) (14) to evaluate cluster-specific differences in the relative expression levels of genes within functionally related categories (defined by GO). Such “shuffled” pathways may be differently regulated in

different parts of the brain (Fig. 3C). Intriguingly, DIRAC analysis showed that the k -means cluster for striatum differed from the clusters for the pallidum, thalamus, and hypothalamus in the relative expression levels for genes involved in central nervous system development. The clusters corresponding to the pallidum vs. thalamus/hypothalamus could then be further distinguished by changes in the expression of genes related to synaptic transmission. These data are consistent with the idea that neurons in different regions of the brain exhibit distinct states (encoding different functions), specified by a multiplicity of transcripts, and that there are quite a large number of distinct neuronal states. Thus, the relative expression differences between a broad range of neuron-specific genes is the main contributor to the observed spatially contiguous, transcriptionally defined clusters.

Neuron-Specific Genes Are Potential Blood Biomarkers. Blood contains organ-specific gene products that are released into circulation through tissue damage, enzymatic cleavage from the cell membrane, or normal blood secretion, providing a source of clinically useful biomarkers for differentiating normal from diseased organs (15). A comparison with published human tissue expression atlases (data on deep comparative transcriptome analyses of multiple organs) (16–18) showed that 130 of the 320 neuron-specific genes from Cahoy et al. (3), of which our 170 genes are a subset, are

produced specifically in the human brain. Cross-referencing these lists with lists of proteins previously reported as being present in the blood of healthy individuals (19) revealed that 29 neuron-specific genes were present in blood; 13 of these were brain-specific as well (Dataset S3). We believe that concentrations of the brain-specific blood proteins could change during disease or injury (presumably reflecting the activity of disease-perturbed biological networks), or that disease or injury could cause brain-specific proteins not normally found in the blood to be secreted. We anticipate that the distinct spatial expression patterns of many of these genes, singly or collectively, could potentially be used to help indicate particular subregions affected by injury or disease.

Discussion

We have shown that subsets of neuron-specific genes exhibit clear and symmetrical spatial expression patterns across the adult mouse brain. We obtained similar results for glia-specific transcripts, although patterns are less complex (e.g., fewer compartments) and less well defined. Our results suggest that both major brain compartments and smaller subregions are defined by specific transcriptionally distinct classes of neurons and glia. Moreover, they reveal insights into the transcriptional organization of the brain and possibly brain biology and development and raise fascinating questions about the numbers of discrete types of neurons and glia present in the brain. There is also a possibility to develop biomarker panels that would be relevant in either tissue (brain) or blood analyses of normal and diseased (injured) brains.

Our results for neurons make sense given the well-known anatomical and physiological diversity of neurons across the brain and the gene expression markers for some neuronal classes that have been characterized previously. The less-precise regional clustering from glia-specific genes suggests that astrocytes and oligodendrocytes, like neurons, differ between brain regions, but that these regional differences in expression are less pronounced. Studies over the last few decades have transformed our understanding of the function of glial cells, from passive “support” cells to highly active participants in brain development, brain plasticity, and synapse function. Our results are consistent with the idea that both distinct neurons and distinct glia play diverse roles in brain functions that differ between regions, and suggest that glia are less regionally distinct than neurons. Our data also suggest anatomical and functional associations between some types of neurons and some types of glia.

Spatial clustering from neuron-specific genes revealed that both major brain compartments and remarkably fine-scale brain anatomical structures have distinct transcriptional signatures. Overall, we characterized transcriptionally distinct clusters corresponding to more than 50 brain regions. This suggests that there are at least this many classes of transcriptionally distinct neurons. By contrast, we found evidence for fewer than 10–15 regionally distinct classes of astrocytes and of oligodendrocytes, suggesting that these cell types are less diverse across the brain and/or that the subtypes are less spatially structured.

We found more transcriptionally distinct brain regions than were discovered in a previous study of ABA data (7), which reported transcriptional signatures for ~30 brain regions based on *k*-means clustering of the expression patterns for 3,041 genes. Several factors may account for the larger number of distinct regions revealed in our analysis. First, our analysis was performed using a finer-resolution grid of the mouse brain (i.e., pixels or patches in our analysis were smaller), which may retain more information from the original *in situ* hybridization images. Likely more important, the focus on neuron-specific genes herein removed noise coming from less well spatially differentiating sets of genes, such as those specific to oligodendrocytes and astrocytes.

The vast diversity of anatomically distinct neuronal subtypes was described by Cajal more than a century ago. We now know that neurons in the brain display innumerable differences in morphol-

ogy, connectivity, and physiology that contribute to their diverse functions. It is not surprising to find that presumably all of these differences reflect different patterns of gene expression. It is remarkable, however, that these transcriptionally distinct, spatially contiguous, symmetrical subdivisions in the brain can be seen so clearly from analysis of the coordinated gene expression patterns of just 170 transcripts, suggesting tight control over neuronal anatomy and function at the level of transcription.

Whereas our approach emphasizes the transcriptional differences between distinct brain regions, others have focused on patterns of transcriptional coexpression shared between multiple classes of neurons and brain regions (20, 21). Presumably, neuronal classes are specified by a combination of subtype-specific transcriptional regulatory mechanisms and their resulting distinct patterns of gene expression. These regulatory networks likely are first enacted during development. Consistent with the developmental origin of adult gene expression patterns, the order in which brain regions emerged as distinct clusters roughly paralleled the order in which developmentally distinct regions of the brain emerge. Developmental gene networks, as well as adult-specific transcriptional regulation, may contribute to the maintenance of distinct cell states in the adult brain. Thus, development builds in additional specificity as it proceeds—and this is similar to the pattern followed as we increase the number of *k*-means clusters. We may be able to follow sequentially the changing patterns of gene expression that specify development by deconvoluting the regulatory control that governs gene expression as we move from a few clusters to, for example, 60 clusters. Obviously, this approach will not reveal all of the transcripts governing development because it can only assess those that are present in the adult brain (and not those turned off during development).

Our results imply that some aspects of brain region-specific gene expression are highly stable from development to adulthood and between individual mice (the ABA was constructed using many different individual mice), but the brain transcriptome is also known to be highly dynamic. An individual's experience, time of day, and other environmental factors each influence the expression of hundreds of thousands of transcripts (e.g., refs. 22–24). The ABA is a static map of gene expression in the adult mouse brain, and different relationships between genes and brain regions would likely emerge if comparable data sets could be analyzed for mice in a variety of distinct states. Integrating spatial and temporal analyses of gene expression patterns in the brain will be an important goal of future research.

A long-term goal will be to use spatial gene expression patterns and related information to develop brain region biomarkers either for tissue or blood analyses. Such biomarkers could then be used in disease diagnosis, particularly when compartment-defining gene products are released into the blood. In disease (or injury), biological networks in the brain are perturbed. This alters the patterns of protein expression that the disease-perturbed networks encode, and if these proteins are secreted into the blood, disease would be reflected in concentrations of the proteins encoded by the perturbed networks. Another possibility is that disease or injury could cause brain proteins to be released in the blood that are normally not there—from brain cell death, promoting cleavage of membrane proteins or abnormal secretion patterns. In addition, biomarkers that are expressed on the cell surfaces of diseased tissues could potentially be used to target therapeutic treatments to particular cell types. Our results extend previous demonstrations that brain regions have distinct gene expression profiles (4, 7) and show that small subsets of genes—especially those that are neuron specific—can serve as biomarkers for identifying individual brain regions. Future studies should characterize optimized subsets of neuron-specific genes that retain diagnostic information when measured at the periphery.

A second long-term goal will be to use spatial gene expression patterns and related information to understand how these patterns relate to the unique biological functions of distinct anatomical

features. These studies might also provide insights into the developmental processes that lead to the immense complexities of the mammalian brain. An extension of these studies might in time let us begin to describe many of the distinct types of neurons that populate the brain. It may even be possible to use such descriptions of neuronal subtypes to predict specific transcriptional regulators that specify these cell types during development. Such hypotheses might have tremendous value in developing stem cell-derived neurons of various types for clinical and experimental uses.

Our analyses suggest a large number of hypotheses about the genes and gene networks that drive the development and functions of specific brain regions. Some genes within these networks have already been validated as regulators of brain development (12, 13), whereas other predictions remain to be tested. Hypotheses about the specific regulators of brain region identity could be refined by integration with time course expression data for each brain regions through development and adulthood. Ultimately, a refined set of predictions should be tested in vivo; the functions of specific genes within particular brain compartments can be assessed through the generation of conditional knockout mice using brain region-specific drivers of gene expression.

In conclusion, we implemented a unique strategy for delineating regional subdivisions in the mammalian brain through the integration of spatial gene expression patterns from the ABA with panels of cell type-specific genes from transcriptome profiling experiments. We identified neuron-specific transcripts as particularly information-rich in that their diverse gene expression patterns reflected precise, symmetrical and well-delineated spatial compartments—reflecting both the structure and development of the brain. These data suggested that there are many transcriptionally distinct types of neurons. Our studies of oligodendrocyte-specific and astrocyte-specific transcripts also delineate symmetrical, spatially distinct compartments, which are fewer in number and with less-well-defined edges. The overlaps of some of these compartments suggested that certain subtypes of neurons and glial cell associated with one another in the distinct compartments. Our results suggest both known and previously undescribed relationships among genes and brain regions. We anticipate that panels of genes derived from studies of this type will serve as useful biomarkers for clinical diagnosis.

- Kandel ER (2000) *Principles of Neural Science* (McGraw-Hill, New York), 4th Ed.
- Wang K, Lee I, Carlson G, Hood L, Galas D (2010) Systems biology and the discovery of diagnostic biomarkers. *Dis Markers* 28(4):199–207.
- Cahoy JD, et al. (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *J Neurosci* 28(11):264–278.
- Sugino K, et al. (2006) Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat Neurosci* 9(1):99–107.
- Lein ES, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445(7124):168–176.
- Visel A, Thaller C, Eichele G (2004) GenePaint.org: An atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res* 32(Database issue):D552–D556.
- Bohland JW, et al. (2010) Clustering of spatial gene expression patterns in the mouse brain and comparison with classical neuroanatomy. *Methods* 50(2):105–112.
- Galaburda AM, LeMay M, Kemper TL, Geschwind N (1978) Right-left asymmetries in the brain. *Science* 199(4331):852–856.
- Sun T, Walsh CA (2006) Molecular approaches to brain asymmetry and handedness. *Nat Rev Neurosci* 7(8):655–662.
- Gage FH (2000) Mammalian neural stem cells. *Science* 287(5457):1433–1438.
- Zapala MA, et al. (2005) Adult mouse brain gene expression patterns bear an embryologic imprint. *Proc Natl Acad Sci USA* 102(29):10357–10362.
- Alcamo EA, et al. (2008) Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. *Neuron* 57(3):364–377.
- Britanova O, et al. (2008) Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron* 57(3):378–392.

Methods

Image Processing. In situ hybridization images for cell type-specific genes were downloaded from the Allen Institute for Brain Science web site (www.brain-map.org). We used all available images within a 200- μ m range in the centermost part of the brain, one to three images per gene on the coronal plane and one to three images on the sagittal plane. Downloaded images were registered to a standard size and shape, following the protocol of Jagalur et al. (25). We then performed bicubic interpolation to generate average expression levels for patches within a 300 \times 300 grid across each image. Each patch represents the averaged expression values for \sim 600 pixels on a coronal section or for \sim 1,000 pixels on a sagittal section.

k-Means Clustering. We applied the *k*-means clustering algorithm to group brain patches based on their expression profiles across all neuron-, astrocyte-, or oligodendrocyte-specific genes. The starting point for this analysis was a matrix of *p* images \times 90,000 patches, where *p* is the number of images for each cell type. The *k*-means algorithm clustered subsets of brain patches by partitioning the 90,000 patches in the grid into *k* groups, while minimizing the sum of squares between patches within the clusters. The *k*-means algorithm rarely converges on a global optimum, so we ran it multiple times with different initial seeds. We report the clustering result with the lowest value for the within-cluster sum of point-to-centroid distance. We performed this clustering approach for *k* ranging from 3 to 100.

Correlation Between Left and Right Hemispheres. We defined the correlation coefficient for cluster assignments between patches at each position to the left and right of the midline by comparing cluster indices, as follows:

$$\text{Corr}(C_{RH(m)}, C_{LH(n)}) = \sum_i \text{Coff}(C_{RH(m,i)}, C_{LH(n,i)}),$$

where $\text{Coff}() = 1$ if $C_{RH(m,i)} = C_{LH(n,i)}$, otherwise $\text{Coff}() = 0$. $C_{RH(m,i)}$, $C_{LH(n,i)}$ are the cluster indices of position (*m, i*) and (*n, i*) in the right and left hemisphere, respectively, where *m* and *n* represent column position index and *i* represents the row position index. To identify the exact location of the brain's midline, we searched for the position that maximizes the correlation coefficient between the two hemispheres; this position was then used to divide the brain into its two hemispheres.

ACKNOWLEDGMENTS. We thank V. Cassen for computational assistance; G. Aldridge and M. Mustroph for consultation on mouse neuroanatomy; and M. Hawrylycz, K. Koch, C. Milne, and J. Bletz for critical reading of the manuscript. This work was supported by Department of Defense Grant WX1XWH-08-1-0420 (to L.H. and N.D.P.), a National Cancer Institute Howard Temin Pathway to Independence Award in Cancer Research (to N.D.P.), and an Institute for Systems Biology, University of Luxembourg Strategic Partnership.

- Eddy JA, Hood L, Price ND, Geman D (2010) Identifying tightly regulated and variably expressed networks by Differential Rank Conservation (DIRAC). *PLoS Comput Biol* 6(5):e1000792.
- Qin S, et al. (2012) SRM targeted proteomics in search for biomarkers of HCV-induced progression of fibrosis to cirrhosis in HALT-C patients. *Proteomics* 12(8):1244–1252.
- Su AI, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 101(16):6062–6067.
- Ge X, et al. (2005) Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* 86(2):127–141.
- Roth RB, et al. (2006) Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics* 7(2):67–80.
- Farrar T, et al. (2011) A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol Cell Proteomics* 10(9):M110 006353.
- Oldham MC, et al. (2008) Functional organization of the transcriptome in human brain. *Nat Neurosci* 11(11):1271–1282.
- Winden KD, et al. (2009) The organization of the transcriptional network in specific neuronal classes. *Mol Syst Biol* 5:291.
- Dong S, et al. (2009) Discrete molecular states in the brain accompany changing responses to a vocal signal. *Proc Natl Acad Sci USA* 106(27):11364–11369.
- Panda S, et al. (2002) Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* 109(3):307–320.
- Chandrasekaran S, et al. (2011) Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. *Proc Natl Acad Sci USA* 108(44):18020–18025.
- Jagalur M, Pal C, Learned-Miller E, Zoeller RT, Kulp D (2007) Analyzing in situ gene expression in the mouse brain with image registration, feature extraction and block clustering. *BMC Bioinformatics* 8(Suppl 10):S5.