Published in final edited form as:

Physiol Genomics. 2007 August 20; 30(3): 313–321. doi:10.1152/physiolgenomics.00287.2006.

A genome-scale map of expression for a mouse brain section obtained using voxelation

Mark H. Chin^{1,2}, Alex B. Geng¹, Arshad H. Khan¹, Wei-Jun Qian³, Vladislav A. Petyuk³, Jyl Boline⁴, Shawn Levy⁵, Arthur W. Toga⁴, Richard D. Smith³, Richard M. Leahy⁶, and Desmond J. Smith¹

¹Department of Molecular and Medical Pharmacology, David Geffen School of Medicine at UCLA, Los Angeles, California

²Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, California

⁴Laboratory of Neuro Imaging, Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, California

³Biological Sciences Division and Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington

⁵Biomedical Informatics Department, Vanderbilt University, Nashville, Tennessee

⁶Signal and Image Processing Institute, University of Southern California, Los Angeles, California

Abstract

Gene expression signatures in the mammalian brain hold the key to understanding neural development and neurological diseases. We have reconstructed two-dimensional images of gene expression for 20,000 genes in a coronal slice of the mouse brain at the level of the striatum by using microarrays in combination with voxelation at a resolution of 1 mm³. Good reliability of the microarray results were confirmed using multiple replicates, subsequent quantitative RT-PCR voxelation, mass spectrometry voxelation, and publicly available in situ hybridization data. Known and novel genes were identified with expression patterns localized to defined substructures within the brain. In addition, genes with unexpected patterns were identified, and cluster analysis identified a set of genes with a gradient of dorsal/ventral expression not restricted to known anatomical boundaries. The genome-scale maps of gene expression obtained using voxelation will be a valuable tool for the neuroscience community.

Keywords

microarrays; genome; gradient of expression

Creating comprehensive gene expression atlases represents an important challenge for understanding the normal mouse brain. Methods for mapping neural gene expression patterns include in situ hybridization (ISH) and analyzing reporter genes in transgenic mice (1, 4, 11, 12, 16). These techniques provide single cell resolution but are labor intensive and costly. Comprehensive analysis of gene expression in the normal and diseased brain using these methods represents a large undertaking.

The Allen atlas used ISH to map expression patterns in the mouse brain for nearly all known genes (1), and the St. Jude's Brain Gene Expression Map (BGEM) has mapped ~2,600 genes (16). Both will be an invaluable resource. The workflow was streamlined and automated, but the handling and analyses of tens of thousands of samples and reactions must necessarily involve some finite error rate. In addition, because the expression patterns are obtained serially rather than in parallel, quantitative comparison of expression between large numbers of genes is difficult. Another large-scale approach for analysis of brain gene expression uses transgenic mice expressing green fluorescent protein fusions (11, 12). Although this method provides vivid and detailed expression images, it is expensive and unlikely to give genome-wide coverage.

To complement ISH and transgenic methods, we developed an approach combining voxelation with microarrays to create genome-wide atlases of expression patterns in the mouse brain (6, 8, 22). Voxelation involves dicing the brain into spatially registered voxels (cubes). Each voxel is then assayed for gene expression levels and images are reconstructed by compiling the expression data back into their original locations. Although the voxelation approach does not give single cell resolution, it does allow acquisition of expression images in parallel, simplifying cross-analysis of multiple genes. In addition, voxelation is much cheaper and faster than traditional approaches.

We have previously used voxelation in combination with microarrays to analyze whole mouse brains at low resolution (11.5 μ l) (6). We have also used a device for semiautomated harvesting of 1 mm³ (1 μ l) voxels, enabling high resolution expression mapping of selected genes using voxelation and quantitative RT-PCR (qRT-PCR) (22). Here we describe a 1-mm³ voxel map for 20,000 genes obtained using microarrays for a single coronal section of the mouse brain at the level of the striatum. The data was found to be of good quality based on multiple independent criteria and provided insights into the molecular architecture of the mammalian brain.

MATERIALS AND METHODS

Mouse voxelation

Adult C57BL/6J male mice (~8 wk old, 25–31 g) were employed. Coronal sections of ~1 mm at the level of the striatum (bregma = 0.02 mm, interaural = 3.82 mm; Ref. 10) were harvested as described (22) with slight modifications producing 68 voxels. To acquire enough total RNA and avoid possible complications with nonlinear amplification of the RNA, we pooled the comparable voxels from ~20 mice. Using a population of mice also diminished the influence of individual outliers. Gene expression levels were determined using cDNA microarrays. This experiment was performed independently, in triplicate, giving three data sets for a total of ~60 mice and 204 arrays. Selected expression patterns were confirmed using qRT-PCR voxelation of single mouse brains.

RNA isolation

Total RNA was extracted from pooled voxels using TRIzol according to the manufacturer's protocol (Gibco-BRL, Gaithersburg, MD). The tissue samples were homogenized by several passages through a 26-gauge needle. Quantification of RNA was performed using spectrophotometry and resulted in ~10 μ g RNA/sample, or ~500 ng RNA/voxel per mouse. Good RNA quality for all of samples was assessed using a spectrophotometer. Additionally, the Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA) was used to assay all of the samples from the first data set and a random 10–20% of the samples in the two subsequent data sets.

Microarray analysis

For each pooled voxel sample in the first and third experiments, 5 µg of Cy3-labeled voxel total RNA and 5 µg Cy5-labeled control RNA (whole brain homogenized samples from >4 mice) were cohybridized to a separate 20,000 gene cDNA microarray. In the second experiment, the dyes were swapped.

Custom cDNA microarrays were made using the Mouse UniGene Set Release I Library from Research Genetics (Vanderbilt Microarray Shared Resource facility). Hybridization was performed at 65°C in 3× SSC, 0.1% SDS, 0.1× Denhardt solution, 0.4 mg/ml poly(dA) (Amersham, Piscataway, NJ), 0.2 mg/ml yeast tRNA, and 0.5 mg/ml *Cot1* DNA (Gibco-BRL, Grand Island, NY) for 16–24 h in humidified hybridization chambers (Corning Glass Works, Corning, NY). Slides were washed for 5 min at room temperature in 0.5× SSC, 0.1% SDS followed by another 5 min wash in 0.06× SSC. Slides were then spun briefly at 500 rpm to dry and scanned using the GenePix 4000B microarray scanner (Axon Instruments, Burlingame, CA). Images were analyzed using the GenePix Pro 3.0 software. Our complete data set is deposited in the Gene Expression Omnibus database (accession no. GSE7480) at the following website: <http://www.ncbi.nlm.nih.gov/geo/>.

The microarray data were processed using standard normalization techniques to remove noise, biases, and outliers. The dye swap for the second experimental set was accounted for, and another nonlinear transformation (LOWESS) employed to compensate for differences in the labeling and chemical properties of Cy3 and Cy5 dyes. The gene expression ratios (experimental/control) were then log₂ transformed. Genes with >10-fold expression differences compared with the mean in a single experiment were given the value of the average expression for the remaining two data sets to minimize their outlier effect.

Averaged data set

Our working data set was produced by averaging the three data sets following normalization and log₂ transformation. To obtain a consensus expression set we removed individual microarrays with poor results across the entire slide (9 in set 1, and 3 in set 3).

Monte Carlo statistics

Permutation testing was used to assess the similarity of correlation matrices (6, 8). To obtain the null distribution, we randomly permuted the columns of the matrices and determined the Frobenius norm by subtracting one permuted matrix from the other. Autocorrelations had negligible effects. The Frobenius norm from the actual results was used to calculate significance based on the null distribution.

Real-time qRT-PCR

Real-time qRT-PCR was performed on voxels from single mouse brains using TaqMan technology (PE Applied Biosystems, Foster City, CA) on an ABI Prism 7900HT Sequence Detection System (PE Applied Biosystems) (22). TaqMan One-Step RT-PCR Master Mix was used following the manufacturer's instructions (PE Applied Biosystems) using ~500 ng of total RNA. Amplification primers were designed using Primer Express 3.0 (PE Applied Biosystems) and were ~20 nucleotides long. The reporter probes employed 6-carboxyfluorescein (6-FAM) as the 5' reporter dye and tetramethyl-6-carboxyrhodamine (TAMRA) as the 3' quencher (Supplementary Table S3) (this article contains supplementary material, which is available at the *Physiological Genomics* web site). Mouse *GAPDH* control (PE Applied Biosystems) was used to normalize variations in cDNA. For relative quantification, a standard curve was constructed for each primer and probe set using total C57BL/6J mouse brain RNA. The data were analyzed using Sequence Detector software (SDS 2.1.1).

Image processing

Images were created using algorithms written in MATLAB (Mathworks, Natick, MA) which displayed gene expression levels using a pseudocolor scale. The voxelation images were mapped to a 1-mm coronal slice of the mouse brain (0.02 mm bregma, 3.82 mm interaural) at the level of the striatum (10). Smoothed gene expression images were constructed using MATLAB as described previously (5–8, 22). The pixelated expression images (Fig. 2) were interpolated before their deformations onto the atlas section (21, 28). Image registration was accomplished via the thin-plate spline warping algorithm. The interpolated and deformed images were then superimposed on the atlas section (5, 7).

In addition to images presented here, mirror images are shown on our voxelation website (<http://vox.pharmacology.ucla.edu/home.html>). These images were created by averaging the data after overlaying one hemisphere on top of the other. The resulting single hemisphere was subsequently reflected about the midline to create a full coronal image from the brain. Since each voxel now results from an averaging of six replicates, the mirror images should have decreased noise.

Correlation matrix clustering

Voxel $\times$ voxel correlation matrices were constructed by calculating the Spearman rho correlation coefficient for each voxel (row) using all transcript or protein data to every other voxel (column) (Figs. 1, B and C; and 5, A–C). The gene $\times$ gene correlation matrix was constructed for each gene (row) using all 68 voxel expression values compared with every other analyzed gene (column) (Fig. 4). This correlation matrix was ordered based on a similarity measure previously described (15). Briefly, the first row of the matrix was chosen to show the strongest contrast between the highest and lowest correlation coefficient for that row. The remaining rows were then arranged in order of decreasing similarity using a least squares metric.

Dorsal/ventral gradient quantitation

Available ISH images for genes from Fig. 4, *cluster B* (7 of 11), were analyzed using ImageJ (2). Coronal images were divided into seven equally thick horizontal slices. Gene expression in each slice was determined by comparing the detected signal to background. For sagittal ISH images, seven stacked equal squares were marked at the level of the striatum corresponding to the coronal section (bregma = 0.02 mm, interaural = 3.82 mm). The signal compared with background was computed for each square in the column. For the microarray voxelation, the expression for each gene was averaged across the voxels of the seven horizontal slices. Error bars ($\pm$ SE) reflect variation between normalized gene expression levels.

Gene Ontology analysis

Analysis for overrepresentation of Gene Ontology categories was conducted using EASE (14). The analysis was performed on genes with significant regional expression in the cortex, striatum, corpus callosum, and hypothalamus (Supplementary Table S2). Three GO categories, “molecular function,” “biological process,” and “cellular component” were used.

Mass spectrometry voxelation

Mass spectrometry voxelation was performed using the equivalent coronal slice of the mouse brain used for the transcriptomics as previously described (20). Briefly, a coronal section from single mouse brains were divided into 1-mm³ voxels, and each voxel was homogenized and digested into peptides using trypsin according to the previously reported protocol (27). The peptide samples from each voxel were individually analyzed by a high-

throughput liquid chromatography (LC) system coupled with high-resolution Fourier transform ion cyclotron resonance (FTICR) mass spectrometer (MS). Detected LC-MS features were confidently identified as peptides by matching to a preestablished database using the accurate mass and time (AMT) tag strategy and quantified-based detected MS intensities (26). The high-throughput proteomic analyses allowed mapping of the protein abundance patterns within the coronal slice for a total of 1,028 proteins.

URLs

The primary microarray data are available at <http://vox.pharmacology.ucla.edu/datadownload.html>.

Access to gene expression images is available from our web site at <http://vox.pharmacology.ucla.edu/home.html> (Supplementary Fig. S5A) and from the Mouse BIRN Atlasing Toolkit (MBAT) web site at <http://www.loni.ucla.edu/twiki/bin/view/MouseBIRN/MBAT> (Supplementary Fig. S5B) from the Biomedical Informatics Research Network (BIRN) at the Laboratory of Neuro Imaging at UCLA (LONI), at <http://www.loni.ucla.edu/BIRN/>.

RESULTS

A 1-mm coronal slice of mouse brain (8 wk old, male C57BL/6 mice) was harvested at the level of striatum (bregma = 0.02 mm, interaural = 3.82 mm). The slice was voxelated into 1-mm³ cubes using a custom-designed device as described (22) resulting in 68 voxels (Fig. 1A). To obtain enough RNA to perform the microarray analysis without amplification, corresponding voxels from 20 mice were pooled together. The variation in voxelation between these inbred mouse brains was <1 voxel at the periphery of the grid. We reasoned that avoiding potential nonlinearities of RNA amplification from single brains outweighed any degradation of image quality caused by slight misregistration of pooled voxels. The RNA from each pooled voxel sample was analyzed using a 20,000 gene cDNA microarray. In addition, using pooled samples should diminish the effects of outlier individuals. The control channel for each microarray consisted of total mouse brain RNA from a pool of ≥ 4 brains. This experiment was performed in triplicate so that a total of 60 mice and 204 microarrays were used. In the first and third data sets, the voxel RNA was labeled with Cy3 and the control RNA was labeled with Cy5. The second data set represented a dye-swap experiment. Following normalization, the three data sets were averaged together to create a consensus. The average data set is used for the presented results.

To assess the overall quality of the combined data set, we took advantage of the inherent bilateral symmetry of the mouse brain. Figure 1, *B* and *C*, shows the correlation coefficients between each of the 34 voxels in the left and right hemispheres, respectively, for all 20,000 genes. The main diagonals are the autocorrelation for each voxel and are 1. Permutation testing showed significant similarity between these two matrices ($P < 10^{-6}$) (Supplementary Fig. S1A). We also compared the omnibus 68 $\times$ 68 voxel matrices for the three separate data sets. All comparisons were significant, *set 1* vs. *set 2* ($P < 10^{-5}$), *set 1* vs. *set 3* ($P < 10^{-5}$), and *set 2* vs. *set 3* ($P < 10^{-6}$) (Supplementary Fig. S1, *B–D*), suggesting a high degree of congruence between the individual data sets.

In addition to doing voxel $\times$ voxel comparisons, we also analyzed gene $\times$ gene correlations. We used the 68 expression measurements from each voxel to compute correlation coefficients for each gene from a filtered subset selected to reduce noise (Supplementary Fig. S2, *A* and *B*). Again, there was a significant similarity between the two hemispheres for the average data set ($P < 10^{-6}$) (Supplementary Fig. S2C) and significant similarities

between the three individual data sets (all comparisons were less than $P < 10^{-6}$) (Supplementary Fig. S3, A–C).

To identify restricted gene expression patterns, we compared voxels from cortex, striatum, hypothalamus, and corpus callosum to the rest of the brain section and found a number of regionally expressed genes (Table 1 and Supplementary Table S1). Genes were deemed to be differentially expressed if significant using Student's *t*-test followed by a conservative Bonferroni correction. Examples of regionally restricted expression include the caldesmon 1 (*Clstn1*) and phosphatidylinositol-4-phosphate 5-kinase, type I, gamma (*Pip5k1c*) genes, which both showed strong expression in the cortex (Fig. 2, A and B). There was good agreement with ISH data for both genes from the Allen atlas and for the available *Clstn1* gene from BGEM. An intragene scale is used, which maximizes the displayed dynamic range for each gene. Striatum-specific genes protein phosphatase 1, regulatory (inhibitor) subunit 1B (*Ppp1r1b*), and myeloid ecotropic viral integration site-related gene 1 (*Mrg1*) were also highly concordant with ISH (Fig. 2, C and D). The hypothalamic region consists of many small nuclei involved in a variety of functions including regulation of the pituitary, body temperature, and appetite. Figure 2, E and F, shows that both neocortical (*Ndn*) and “imprinted and ancient” (*Impact*) have elevated levels of expression in the hypothalamus (3, 17) and match the ISH images.

Transthyretin (*Ttr*) encodes a transport protein specific to the choroid plexus (13) and showed isolated expression in the region corresponding to the lateral ventricles in both microarray voxelation and ISH data (Fig. 2G). Conversely, ubiquitin-conjugating enzyme E2, J1 (*Ube2j1*) was identified as being expressed at a relatively low level throughout the brain except in the dorsal corpus callosum (Fig. 2H). This observations confirms that the ISH image for *Ube2j1* is likely correct and not due to background. However, the expression patterns of *Ttr* and *Ube2j1* show that distinguishing between the choroid plexus and dorsal corpus callosum is difficult using voxelation, a consequence of the limited resolution (1 mm³) of the technique compared with the cellular resolution of ISH. Nevertheless, there was much overall agreement between these two very different techniques. In general, the agreement was over >75% for all comparisons (>100 genes). Voxelated and smoothed images for *Clstn1* and *Ttr* are shown in Fig. 2, I and J.

There were some instances of discrepancies between the Allen atlas and the microarray voxelation data. Figure 3, A and B, shows two examples. Serine peptidase inhibitor (*Serpinb1a*) and “HS loss of heterozygosity” (*HSLOH11*) showed distinct expression in the striatum/fornix and hypothalamic region, respectively, using voxelation. No expression was detected using ISH. Real-time TaqMan qRT-PCR and voxelation of a single mouse brain was used to determine which of the expression patterns was correct. For both genes, the expression pattern obtained using voxelation and microarrays was correct. In general, when we found discrepancies between the voxelation data and the Allen atlas, there was a clear expression pattern from the voxelation data but little discernable expression in the ISH images. This may be because the sensitivity of ISH is not as high as the sensitivity of microarrays and qRT-PCR. Consistent with this idea, *Serpinb1a* was in the bottom 15% of all genes average expression. However, *HSLOH11* had a very high averaged expression.

Another use of the voxelation atlas is to find novel and unexpected expression profiles for genes in the mouse brain. For example, nuclear factor I/X (*Nfix*) was found to be expressed in a gradient pattern in the cortex, more highly at the dorsal surface and decreasing ventrally (Fig. 3C). This finding was confirmed by qRT-PCR voxelation and ISH. Another unusual expression pattern is for the Pre B-cell leukemia transcription factor 3 (*Pbx3*) gene (Fig. 3D). Microarray voxelation data identified *Pbx3* as being expressed in the striatum and adjacent ventral structures, including the anterior commissure, ventral pallidum, and the

magnocellular preoptic nuclei. This expression pattern was confirmed by qRT-PCR voxelation and ISH.

Gene Ontology (GO) analysis using EASE (14) was conducted for groups of genes identified as differentially expressed in the cortex, striatum, hypothalamus, or corpus callosum (Table 1). Categories were identified as significantly overrepresented in the cortex, hypothalamic region, and corpus callosum (Supplementary Table S2). In the cortex, categories synapse ($P < 0.002$), synaptic transmission ($P < 0.009$), clathrin-coated vesicle ($P < 0.02$), coated vesicle ($P < 0.03$), transport ($P < 0.04$), and synaptic vesicle ($P < 0.04$) were all significantly overrepresented. All of these groups are potentially involved in the transmission of nerve signals.

To further explore the reproducibility of the voxelation data, we compared it with the SymAtlas of the Genome Institute of Novartis Foundation (24). There was significant overlap between the two data sets for genes judged as being in the top 10% of those expressed in cortex, striatum, and hypothalamus ($\chi^2 > 42.6$, $df = 1$, $P < 10^{-11}$, all comparisons).

One advantage of microarray data is that clustering tools can be easily used. To minimize noise, we used stringent filtering criteria. A total of 135 genes was selected by identifying those genes in the top 50% of standard deviation across all 68 voxels and that had a Spearman rho correlation coefficient > 0.70 with > 40 other genes. These criteria should identify coregulated genes with marked regional variation in expression. A matrix was constructed showing all pair-wise comparisons between the 135 genes using the correlation coefficient (Fig. 4). Four distinct clusters were identified.

The average expression pattern for each cluster is shown in Fig. 4, A–D. There was good overall agreement between the averaged expression pattern, in situ images from the Allen atlas, and BGEM with the microarray voxelation expression patterns for the genes in each of the clusters.

Cluster A consisted of 45 genes and was expressed in the cortex (Fig. 4A). *Cluster B* consisted of 11 genes expressed throughout the coronal slice at a low level (Fig. 4B). The cluster showed a dorsal/ventral gradient with expression higher in the dorsal half of the brain. Moreover, the expression pattern did not appear to be regionally defined by known brain structures. Removal of the two outlier voxels on the periphery of the brain did not appreciably affect the correlation coefficients of this cluster (Supplementary Fig. S4). There was good agreement for *cluster B* with the results from ISH. This is illustrated in Fig. 4E (graph). There was a statistically significant main effect of dorsal/ventral row coordinates on expression level for this cluster using both ISH ($F_{(6,42)} = 6.3$, $P < 10^{-3}$) and voxelation data ($F_{(6,70)} = 368.3$, $P < 10^{-3}$). Not surprisingly, due to the strikingly different nature of the two methodologies, there was also a significant main effect of data type ($F_{(1,124)} = 51.3$, $P < 10^{-3}$). The low expression level, the subtle nature of the dorsal/ventral gradient, the difficulty of doing cluster analyses, and the qualitative nature of the data would make it hard to identify this cluster using ISH alone or to assign a gradient pattern to any one gene. In contrast, microarray voxelation has the advantage of reliable and simple data clustering. Interestingly, one of the genes in *cluster B*, cystathionine β -synthase, has been implicated in dorsal neural tube defects (25).

The 58 genes that make up *cluster C* (Fig. 4C) are expressed in two bilaterally symmetric voxels in the hypothalamic region. The resolution of voxelation in this experiment is not high enough to discern which of the many nuclei may be responsible for this expression pattern, but possibilities include the anterior and lateral hypothalamic area. The fourth cluster, *cluster D* (Fig. 4D), is a group of 20 genes that show expression in the striatum and

corpus callosum. The averaged expression profiles for the four clusters were bilaterally symmetric, adding credence to their validity. Interestingly, the genes in *cluster A* had a high degree of overlap with those identified as cortex-specific from Table 1 (44/45). However, the genes in *clusters C* and *D* showed no overlap with hypothalamus- and corpus callosum-specific genes, respectively (0/58, 0/20). This may be because clustering was performed using an unbiased mathematical model while region-specific expression was performed using predefined anatomical regions and a stringent Bonferroni correction.

In a parallel study, voxelation was coupled with proteomics (20; 26; 27) by employing liquid chromatography, Fourier transform ion cyclotron resonance mass spectrometry (LC-FTICR-MS) to analyze an equivalent coronal slice from single mouse brains. A total of 347 proteins had expression profiles that could be confidently compared with the microarray voxelation data. The low number of overlapping proteins reflects the use of voxels from individual brains. In contrast, the microarray studies used pooled voxels, increasing sensitivity. Furthermore, the International Protein Index (IPI) gene lists are not exceptionally well annotated, so many of the peptides that were identified could not be confidently paired to its gene product. The correlation between mRNA and protein levels for the left and right hemispheres is shown in Fig. 5A. Comparison of left hemisphere RNA levels with right hemisphere RNA levels gives the expected broad diagonal of positive correlations. There is no diagonal of perfect autocorrelations (correlation coefficient = 1), as we are not comparing left to left hemisphere or right to right. Similarly, there is a broad diagonal band of positive correlations in the comparison of left protein levels with right protein levels. Interestingly, comparison of similarly ordered voxels for left hemisphere RNA levels with right hemisphere protein levels and left hemisphere protein levels with right hemisphere RNA levels reveals the presence of a similar diagonal, although the degree of the positive correlations was understandably less for this cross-modality comparison (RNA vs. protein) than for the within-modality comparisons (RNA vs. RNA and protein vs. protein). Permutation testing showed significant similarity between all four quadrants of the matrix shown in Fig. 5A ($P < 3.4 \times 10^{-4}$). These correlations in RNA and protein levels for equivalent voxels suggest robustness of both the transcriptomic and proteomic maps.

To further illustrate the correspondence between transcript and protein levels in the mouse brain, correlations were plotted for left RNA vs. left RNA and left protein vs. left protein (Fig. 5B). Again, these matrices were significantly similar using permutation testing ($P < 2 \times 10^{-6}$). Corresponding results were obtained when comparing right RNA vs. right RNA and right protein vs. right protein ($P < 2 \times 10^{-6}$) (Fig. 5C). We found good agreement between individual gene expression patterns obtained using microarray voxelation, mass spectrometry, and ISH. Examples are shown in Fig. 5, D–G, for genes expressed in the cortex, striatum, corpus callosum, and hypothalamus.

Gene expression images are available through our web site (Supplementary Fig. S5A) as well as the Mouse BIRN Atlasing Toolkit (MBAT), in which the image is overlaid on top of the corresponding coronal slice of the mouse brain (Supplementary Fig. S5B). MBAT is a three-dimensional visualization and neuroinformatics programs designed by a collaborative group of neuroscientists comprising the Mouse Biomedical Informatics Research Network (BIRN).

DISCUSSION

Existing techniques such as ISH allow gene expression images to be obtained from the brain. Although permitting single cell resolution, these methods are cumbersome, making genome-wide coverage expensive and difficult. Microarray voxelation allowed gene expression profiles for 20,000 genes to be obtained from the mouse at a resolution of 1 mm^3 .

Multiple criteria suggested the data was of good quality including left/right symmetry and comparisons between replicates, qRT-PCR voxelation, mass spectrometry voxelation, and ISH.

We used pooled voxels for the microarray voxelation to avoid potential problems with RNA amplification and outlier individuals. The use of an inbred mouse strain led to high reproducibility in the voxelation process and hence minimal degradation of image quality from pooling. This is confirmed by the similarity of images from microarray voxelation, qRT-PCR voxelation (Fig. 3), and mass spectrometry voxelation (Fig. 5D), in which the last two image sets each used single mouse brains.

There was good agreement between the voxelation data and the high-throughput ISH data from the Allen and BGEM atlases. A survey of ~100 genes from the Allen atlas yielded a concordance of >75%. In most cases of disagreement, there was no clear signal from the ISH data, whereas there were defined patterns from the microarray voxelation. We further investigated two cases of discrepancy using qRT-PCR voxelation and the microarray voxelation results appeared to be correct. Because ISH is a serial process with multiple steps, it is susceptible to errors such as reaction and hybridization failures. The high-parallelism of microarrays renders voxelation less prone to these types of error.

Another advantage of microarray voxelation is that the expression data for different genes are acquired on an equal footing. This greatly simplifies quantitative analysis and comparisons of many genes. In contrast, individual ISH images have idiosyncratic brightness and contrast characteristics. ISH also introduces random spatial deformations into individual images, further complicating multi-gene comparisons.

Comparison of expression patterns in microarray voxelation can be done using cluster analysis, a powerful and unbiased discovery-driven method. Clusters can highlight the average expression pattern of many correlated genes, decreasing noise and giving greater clarity of the consensus expression images. We identified four clusters with distinct expression patterns. *Clusters A, C, and D* defined the cortex, hypothalamus, and white matter. *Cluster B* consisted of 11 genes with a subtle dorsal/ventral gradient that did not appear to respect traditional anatomic boundaries. In retrospect, this gradient was visible in the ISH patterns but the pattern would have been difficult to discern in individual genes. The dorsal/ventral genes may be involved in global patterning of the mammalian brain (9, 18). One of the genes in this cluster, cystathionine β -synthase, has been implicated in dorsal neural tube defects in mammals (25).

Using the voxelation data, we also used hypothesis-driven approaches to find genes expressed at high levels in various conventionally defined brain regions such as cortex, striatum, white matter, and hypothalamus. A number of structures showed functional enrichment of GO categories, including the cortex which had elevation of genes involved in synaptic transmission (Supplementary Table S1). In addition to standard expression patterns, we found genes with more unusual expression characteristics such as the striatum plus ventral structures (*Pbx3*) (Fig. 2D) and a gradient in the cortex (*Nfix*) (Fig. 4C). Additional in-depth analysis of the voxelation data set may give further insights into the molecular anatomy of the mammalian brain.

Classical methods such as ISH and immunohistochemistry can give single cell resolution. In contrast, the resolution of voxelation is limited by the voxel size, here 1 mm³ or 1 μ l. The finite spatial resolution of voxelation could result in a decrease in the signal-to-noise ratio for each gene's expression image. However, it seems unlikely that this caused loss of much gene expression information, since a gene cluster was detected with expression in only one voxel in the left and right hypothalamus (Fig. 4, *cluster C*).

Another advantage of voxelation is its modality independence. In this report, we used qRT-PCR to confirm microarray voxelation at the transcript level and LC-FTICR-MS to obtain proteomic expression maps of the mouse brain. The high degree of concordance between the transcriptomic and proteomic data bolsters confidence in these two very different maps. Further investigation of mouse models of neurological disease might provide interesting examples of translational control.

Compared to traditional techniques, voxelation is relatively inexpensive. A complete survey of the expression pattern for all known genes in the whole mouse brain in three dimensions would be thousands of times cheaper than ISH. Continued developments in microelectromechanical systems (MEMS) and nanotechnology may improve the resolution of voxelation (23). Furthermore, advances in profiling technology such as arrays of arrays (29) and improved sensitivity and reproducibility should help drive costs further down.

In summary, voxelation is a relatively affordable approach for global mapping of gene expression in the mouse brain, with complementary strengths and weaknesses to traditional methods. At a resolution of 1 mm³, it is feasible to use voxelation to investigate mouse models of brain disease at a genomic scale, which would be beyond the reach of ISH or immunohistochemistry. Thus, voxelation should provide future insights into the genomic mechanisms of neurological disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Steve Anderson, Allan Mackenzie-Graham, and Daniel Sforza at LONI for help in importing the voxelation data into MBAT. We thank Rita Cantor for advice and Leah Hutnick and Keri Martinowich for reading the manuscript.

GRANTS

This study was supported by National Institutes of Health Grants RO1-DA-015802 (to D. J. Smith) and RO1-NS-050148 (to D. J. Smith, R. M. Leahy, and R. D. Smith), by the Staglin Music Festival and NARSAD Young Investigator Award, by the Tobacco-Related Disease Research Program (11RT-0172), and by the Alzheimer's Association (IIRG-02-3609) (D. J. Smith). Pacific Northwest National Laboratory (PNNL) efforts were supported by NIH National Center for Research Resources Grant RR-18522 (to R. D. Smith), and proteomics measurements were performed in the Environmental Molecular Sciences Laboratory, a US Department of Energy national scientific user facility located at PNNL in Richland, Washington. PNNL is a multi-program national laboratory operated by Battelle Memorial Institute for the DOE under Contract DE-AC05-76RL01830.

References

1. ABA (Allen Brain Atlas). Allen Brain Atlas. Allen Institute for Brain Science; Seattle (WA): 2004. cited July 2006. Available from: <http://www.brain-map.org>
2. Abramoff MD, Magelhaes PJ, Ram SJ. Image processing with ImageJ. *Biophotonics Int.* 2004; 11:36–42.
3. Bischof JM, Wevrick R. Genome-wide analysis of gene transcription in the hypothalamus. *Physiol Genomics.* 2005; 22:191–196. [PubMed: 15870395]
4. Boguski MS, Jones AR. Neurogenomics: at the intersection of neurobiology and genome sciences. *Nat Neurosci.* 2004; 7:429–433. [PubMed: 15114353]
5. Bookstein F. Principal warps: thin-plate splines and decomposition of deformations. *IEEE Trans Patt Anal Machine Intell.* 1989; 11:567–585.
6. Brown VM, Ossaditchi A, Khan AH, Cherry SR, Leahy RM, Smith DJ. High-throughput imaging of brain gene expression. *Genome Res.* 2002; 12:244–254. [PubMed: 11827944]

7. Brown VM, Ossadtchi A, Khan AH, Gambhir SS, Cherry SR, Leahy RM, Smith DJ. Gene expression tomography. *Physiol Genomics*. 2002; 8:159–167. [PubMed: 11875194]
8. Brown VM, Ossadtchi A, Khan AH, Yee S, Lacan G, Melega WP, Cherry SR, Leahy RM, Smith DJ. Multiplex three-dimensional brain gene expression mapping in a mouse model of Parkinson's disease. *Genome Res*. 2002; 12:868–884. [PubMed: 12045141]
9. Flanagan JG, Vanderhaeghen P. The ephrins and Eph receptors in neural development. *Annu Rev Neurosci*. 1998; 21:309–345. [PubMed: 9530499]
10. Franklin, KBJ.; Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates*. San Diego: Academic; 1997. p. xxii186 of plates
11. Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, Heintz N. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*. 2003; 425:917–925. [PubMed: 14586460]
12. Heintz N. BAC to the future: the use of bac transgenic mice for neuroscience research. *Nat Rev Neurosci*. 2001; 2:861–870. [PubMed: 11733793]
13. Herbert J, Wilcox JN, Pham KT, Freneau RT Jr, Zeviani M, Dwork A, Soprano DR, Makover A, Goodman DS, Zimmerman EA, Roberts JL, Schon EA. Transthyretin: a choroid plexus-specific transport protein in human brain The 1986 S Weir Mitchell award. *Neurology*. 1986; 36:900–911. [PubMed: 3714052]
14. Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. *Genome Biol*. 2003; 4:R70. [PubMed: 14519205]
15. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 2005; 21:3448–3449. [PubMed: 15972284]
16. Magdaleno S, Jensen P, Brumwell CL, Seal A, Lehman K, Asbury A, Cheung T, Cornelius T, Batten DM, Eden C, Norland SM, Rice DS, Dosoooye N, Shakya S, Mehta P, Curran T. BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol*. 2006; 4:e86. [PubMed: 16602821]
17. Niinobe M, Koyama K, Yoshikawa K. Cellular and subcellular localization of necdin in fetal and adult mouse brain. *Dev Neurosci*. 2000; 22:310–319. [PubMed: 10965153]
18. Peel AD, Chipman AD, Akam M. Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat Rev Genet*. 2005; 6:905–916. [PubMed: 16341071]
20. Petyuk VA, Qian WJ, Chin MH, Wang H, Livesay EA, Monroe ME, Adkins JN, Jaitly N, Anderson DJ, Camp DG 2nd, Smith DJ, Smith RD. Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography-mass spectrometry. *Genome Res*. 2007; 17:328–336. [PubMed: 17255552]
21. Rosen GD, Williams AG, Capra JA, Connolly MT, Cruz B, Lu L. The Mouse Brain Library at www.mbl.org. *Int Mouse Genome Conference*. 2000; 14:166.
22. Singh RP, Brown VM, Chaudhari A, Khan AH, Ossadtchi A, Sforza DM, Meadors AK, Cherry SR, Leahy RM, Smith DJ. High-resolution voxelation mapping of human and rodent brain gene expression. *J Neurosci Methods*. 2003; 125:93–101. [PubMed: 12763235]
23. Staples M, Daniel K, Cima MJ, Langer R. Application of micro- and nano-electromechanical devices to drug delivery. *Pharm Res*. 2006; 23:847–863. [PubMed: 16715375]
24. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A, Patapoutian A, Hampton GM, Schultz PG, Hogenesch JB. Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci USA*. 2002; 99:4465–4470. [PubMed: 11904358]
25. van der Put NM, van Straaten HW, Trijbels FJ, Blom HJ. Folate, homocysteine and neural tube defects: an overview. *Exp Biol Med* (May-wood). 2001; 226:243–270.
26. Wang H, Qian WJ, Chin MH, Petyuk VA, Barry RC, Liu T, Gritsenko MA, Mottaz HM, Moore RJ, Camp DG 2nd, Khan AH, Smith DJ, Smith RD. Characterization of the mouse brain proteome using global proteomic analysis complemented with cysteinyl-peptide enrichment. *J Proteome Res*. 2006; 5:361–369. [PubMed: 16457602]
27. Wang H, Qian WJ, Mottaz HM, Clauss TR, Anderson DJ, Moore RJ, Camp DG 2nd, Khan AH, Sforza DM, Pallavicini M, Smith DJ, Smith RD. Development and evaluation of a micro- and

- nanoscale proteomic sample preparation method. *J Proteome Res.* 2005; 4:2397–2403. [PubMed: 16335993]
28. Williams, RW. Mapping genes that modulate mouse brain development: a quantitative genetic approach. In: Goffinet, AF.; Rakic, P., editors. *Mouse Brain Development*. New York: Springer; 2000. p. 21-49.
29. Zarrinkar PP, Mainquist JK, Zamora M, Stern D, Welsh JB, Sapinoso LM, Hampton GM, Lockhart DJ. Arrays of arrays for high-throughput gene expression profiling. *Genome Res.* 2001; 11:1256–1261. [PubMed: 11435408]

Fig. 1. Correlation matrices. A: voxelation scheme. Spatial (voxel $\times$ voxel) correlation analysis using all genes from the 34 left hemisphere voxels (B) and the 34 right hemisphere voxels (C). The Spearman rho correlation coefficient between each pair of voxels is read by looking along the relevant row and column and finding the intersection. The diagonal is the autocorrelation in each voxel and is equal to 1. The observed difference between the two matrices is measured using the Frobenius norm and is significantly different ($P < 10^{-6}$) compared with a distribution of randomly permuted matrices.

Fig. 2.

Mouse brain gene expression images. Genes were selected as having restricted expression patterns based on the microarray voxelation data. The nonradioactive in situ hybridization (ISH) images (*left* for each gene) are reproduced with permission from the Allen Brain Atlas. The voxelation images use a pseudocolor scale individualized for each gene. The dark-field ISH images are reproduced from St. Jude's BGEM database. Cortically expressed genes included calyntenin 1 (*Clstn1*; A) and phosphatidylinositol-4-phosphate 5-kinase, type I, gamma (*Pip5k1c*; B). Genes with spatially restricted expression in the striatum included protein phosphatase 1, regulatory (inhibitor) subunit 1B (*Ppp1r1b*; C) and myeloid ecotropic viral integration site-related gene 1 (*Mrg1*; D). Two genes, necdin (*Ndn*; E) and imprinted and ancient (*Impact*; F), show expression restricted to the hypothalamic region. The transthyretin (*Ttr*; G) gene was expressed in the choroid plexus, and the ubiquitin-conjugating enzyme E2 J1 (*Ube2j1*; H) gene showed expression almost everywhere except for the corpus callosum. The arrow in the sagittal ISH images indicates the level at which the coronal slice was taken. Voxelated images for *Clstn1* (I) and *Ttr* (J) are shown with their smoothed images.

Fig. 3.

Gene expression patterns confirmed using quantitative RT-PCR (qRT-PCR). *A*: serine (or cysteine) peptidase inhibitor, clade B, member 1a (*Serpinb1a*) was identified as a striatally expressed gene from the microarray voxelation data, while ISH (Allen Atlas) suggested there was no specific expression in the brain. The coronal slice taken for voxelation was taken at the level of the striatum (arrow). qRT-PCR voxelation performed on an identical slice from the mouse brain agreed with the microarray voxelation data. *B*: microarray voxelation identified the *Homo sapiens* loss of heterozygosity 11 (*HSLOH11*) gene as being expressed in the hypothalamus and was confirmed using voxelation qRT-PCR. In contrast, the ISH pattern suggested little or no expression in the hypothalamus. *C*: a gradient of gene expression is identified for nuclear factor I/X (*Nfix*) in the cortex where highest expression is detected in the most dorsomedial surface, decreasing toward the ventrolateral surfaces. There was broad agreement between the pattern discovered by microarray voxelation and those found by qRT-PCR voxelation and ISH. *D*: pre-B-cell leukemia transcription factor 3 (*Pbx3*) exhibits expression in the striatum and also high levels ventral to the striatum when assayed using microarray voxelation. This was confirmed by qRT-PCR voxelation and ISH.

Fig. 4.

Clustering of genes using a gene $\times$ gene correlation matrix. Genes were selected in the top 50% standard deviation across all voxels with a correlation coefficient >0.70 with >40 other genes. The genes were then ordered using an unsupervised clustering algorithm, resulting in four distinct clusters. The expression was averaged across all the genes for each cluster and spatially reconstructed (A–D). *Cluster A* was expressed in cortex, *cluster B* in a shallow dorsal/ventral gradient, *cluster C* in the hypothalamus, and *cluster D* in the striatum and corpus callosum. ISH images for representative genes from each cluster are shown. The arrowheads indicate discrete regions of expression. *E*: quantitative measurement of the dorsal/ventral gradient in *cluster B*. The graph shows averaged gene expression using ISH (blue line) and microarray voxelation data (red line). The coronal sections were divided into rows from dorsal to ventral. The ordinate shows the normalized relative level of expression $\pm$ SE.

Fig. 5. Correlation matrices comparing transcript and protein voxelation data. *A*: a correlation matrix (voxel $\times$ voxel) comparing transcript and protein data for whole brain. *B*: left hemisphere data. *C*: right hemisphere data. *D*: patterns for genes expressed in the cortex, striatum, hypothalamus, and corpus callosum obtained using ISH, microarray voxelation, and mass spectrometry voxelation.

Table 1

Regional gene expression in the brain

Brain Structure	No. of Genes Differentially Expressed
Cortex	152
Striatum	39
Corpus callosum	33
Hypothalamus	63

Genes were significant ($P < 0.05$) after applying a Bonferroni correction.