

ARTICLES

Genome-wide atlas of gene expression in the adult mouse brain

Ed S. Lein^{1*}, Michael J. Hawrylycz^{1*}, Nancy Ao², Mikael Ayres¹, Amy Bensinger¹, Amy Bernard¹, Andrew F. Boe¹, Mark S. Boguski^{1†}, Kevin S. Brockway¹, Emi J. Byrnes¹, Lin Chen¹, Li Chen², Tsuey-Ming Chen², Mei Chi Chin¹, Jimmy Chong¹, Brian E. Crook¹, Aneta Czaplinska², Chinh N. Dang¹, Suvro Datta¹, Nick R. Dee¹, Aimee L. Desaki¹, Tsega Desta¹, Ellen Diep¹, Tim A. Dolbeare¹, Matthew J. Donelan¹, Hong-Wei Dong¹, Jennifer G. Dougherty¹, Ben J. Duncan¹, Amanda J. Ebbert¹, Gregor Eichele³, Lili K. Estin¹, Casey Faber¹, Benjamin A. Facer¹, Rick Fields², Shanna R. Fischer¹, Tim P. Fliss¹, Cliff Frensley¹, Sabrina N. Gates¹, Katie J. Glattfelder¹, Kevin R. Halverson¹, Matthew R. Hart¹, John G. Hohmann¹, Maureen P. Howell¹, Darren P. Jeung¹, Rebecca A. Johnson¹, Patrick T. Karr¹, Reena Kawal¹, Jolene M. Kidney¹, Rachel H. Knapik¹, Chihchau L. Kuan¹, James H. Lake¹, Annabel R. Laramee¹, Kirk D. Larsen¹, Christopher Lau¹, Tracy A. Lemon¹, Agnes J. Liang², Ying Liu², Lon T. Luong¹, Jesse Michaels¹, Judith J. Morgan¹, Rebecca J. Morgan¹, Marty T. Mortrud¹, Nerick F. Mosqueda¹, Lydia L. Ng¹, Randy Ng¹, Geralyn J. Orta¹, Caroline C. Overly¹, Tu H. Pak¹, Sheana E. Parry¹, Sayan D. Pathak¹, Owen C. Pearson¹, Ralph B. Puchalski¹, Zackery L. Riley¹, Hannah R. Rockett¹, Stephen A. Rowland¹, Joshua J. Royall¹, Marcos J. Ruiz², Nadia R. Sarno¹, Katherine Schaffnit¹, Nadiya V. Shapovalova¹, Taz Sivisay¹, Clifford R. Slaughterbeck¹, Simon C. Smith¹, Kimberly A. Smith¹, Bryan I. Smith¹, Andy J. Sodt¹, Nick N. Stewart¹, Kenda-Ruth Stumpf¹, Susan M. Sunkin¹, Madhavi Sutram¹, Angelene Tam², Carey D. Teemer¹, Christina Thaller², Carol L. Thompson¹, Lee R. Varnam¹, Axel Visel^{3†}, Ray M. Whitlock¹, Paul E. Wohnoutka¹, Crissa K. Wolkey¹, Victoria Y. Wong¹, Matthew Wood², Murat B. Yaylaoglu², Rob C. Young¹, Brian L. Youngstrom¹, Xu Feng Yuan¹, Bin Zhang², Theresa A. Zwingman¹ & Allan R. Jones¹

Molecular approaches to understanding the functional circuitry of the nervous system promise new insights into the relationship between genes, brain and behaviour. The cellular diversity of the brain necessitates a cellular resolution approach towards understanding the functional genomics of the nervous system. We describe here an anatomically comprehensive digital atlas containing the expression patterns of ~20,000 genes in the adult mouse brain. Data were generated using automated high-throughput procedures for *in situ* hybridization and data acquisition, and are publicly accessible online. Newly developed image-based informatics tools allow global genome-scale structural analysis and cross-correlation, as well as identification of regionally enriched genes. Unbiased fine-resolution analysis has identified highly specific cellular markers as well as extensive evidence of cellular heterogeneity not evident in classical neuroanatomical atlases. This highly standardized atlas provides an open, primary data resource for a wide variety of further studies concerning brain organization and function.

The mammalian central nervous system (CNS) contains an enormous variety of cell types, each with unique morphology, connectivity, physiology and function. A characterization of the full complement of neural cell types is essential to understand functional circuit properties and their relation to higher cognitive functions and behaviours. The phenotypic properties of different neuronal and non-neuronal cells are largely the product of unique combinations of expressed gene products; therefore, gene expression profiles provide an informative modality to define cellular diversity in the brain. However, histological data are typically generated for one gene at a time, and data are neither systematically produced and analysed nor consolidated in an easily accessible format. Consequently, a limited set of

established cellular markers dominates the current literature, and expression patterns of many genes remain uncharacterized.

The recent availability of genomic sequence information and high-throughput advances in experimental techniques has allowed an expansion from single-gene to genome-wide transcriptional analyses. These approaches have been used to investigate the relationship between brain structure and function in normal adult brain, development and disease, most notably using DNA microarrays and serial analysis of gene expression (SAGE)^{1–4}. However, these techniques have typically been applied to large brain regions, yielding data that are difficult to interpret because they do not resolve cellular diversity within those structures (although see refs 5 and 6). This diversity is

¹Allen Institute for Brain Science, Seattle, Washington 98103, USA. ²Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. ³Department of Genes and Behavior, Max Planck Institute of Biophysical Chemistry, 37077 Goettingen, Germany. †Present addresses: Division of Genome and Proteome Sciences, Novartis Institutes for Biomedical Research, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA (M.S.B.); Lawrence Berkeley National Laboratory, Genomics Division, MS84-171, 1 Cyclotron Road, Berkeley, California 94720, USA (A.V.).

*These authors contributed equally to this work.

exemplified by single neuron studies comparing gene expression with anatomical and physiological properties, demonstrating complex relationships between gene expression and neuronal subtypes, frequently with very different expression profiles in adjacent neurons^{7–9}. The fine (cellular but not single cell as above^{7–9}) resolution of scalable techniques such as *in situ* hybridization (ISH) is needed to begin to make meaningful global correlations between gene expression, gene regulation, CNS function and cellular phenotype. An increasing number of large-scale projects are now determining messenger RNA and protein localization in adult and developing brain^{10–14}. The marriage of the unbiased genomic approach with semi-quantitative histological methods promises to open many new research avenues, particularly in a core model for mammalian brain development and behavioural genetics, the mouse¹⁵.

The Allen Brain Atlas project has taken a global approach to understanding the genetic structural and cellular architecture of the mouse brain by generating a genome-scale collection of cellular resolution gene expression profiles using ISH. Highly methodical data production methods and comprehensive anatomical coverage via dense, uniformly spaced sampling facilitate data consistency and comparability across >20,000 genes. The use of an inbred mouse strain with minimal animal-to-animal variance allows one to treat the brain

essentially as a complex but highly reproducible three-dimensional tissue array. Image-based informatics methods for signal detection and three-dimensional registration of ISH data to a *de novo* age-matched annotated reference atlas have been developed to allow automated signal quantification across the anatomical structures in the reference atlas, or its associated grid-based coordinate system. These methods enable global analysis and mining for detailed expression patterns in the brain. The entire Allen Brain Atlas data set and associated informatics tools are available through an unrestricted web-based viewing application (<http://www.brain-map.org>).

Global analysis of gene expression

Quantification and mapping strategies. To create a searchable, anatomy-centred database of gene expression, ISH data for each gene is aligned in the same three-dimensional coordinate space through registration with a reference atlas created for the project (Supplementary Methods 3). Three-dimensional volumetric representations are created from the reference atlas and the series of ISH images for each gene (Fig. 1). The atlas volume is registered in three dimensions to the ISH volume, thereby alleviating animal-to-animal differences in cutting plane. This process applies a standardized three-dimensional grid coordinate frame to the ISH data for each gene (Fig. 1c), with mapping to corresponding anatomical structures provided by the annotated reference atlas (Fig. 1d). Although there are caveats associated with handling image artefacts and consistent registration across different brains, the result of this registration system is a genome-scale ISH data set searchable either by anatomical structure or using a finer (100 μm to 300 μm) resolution geometric grid.

Informatics tools have also been developed to identify pixels corresponding to true ISH signal and to identify (segment) expressing objects (contiguous clusters of pixels corresponding to cells) on the two-dimensional images for each gene. This process allows relative quantification based on either the fraction of expressing cells (relative density) or on expression level (approximating total transcript count),

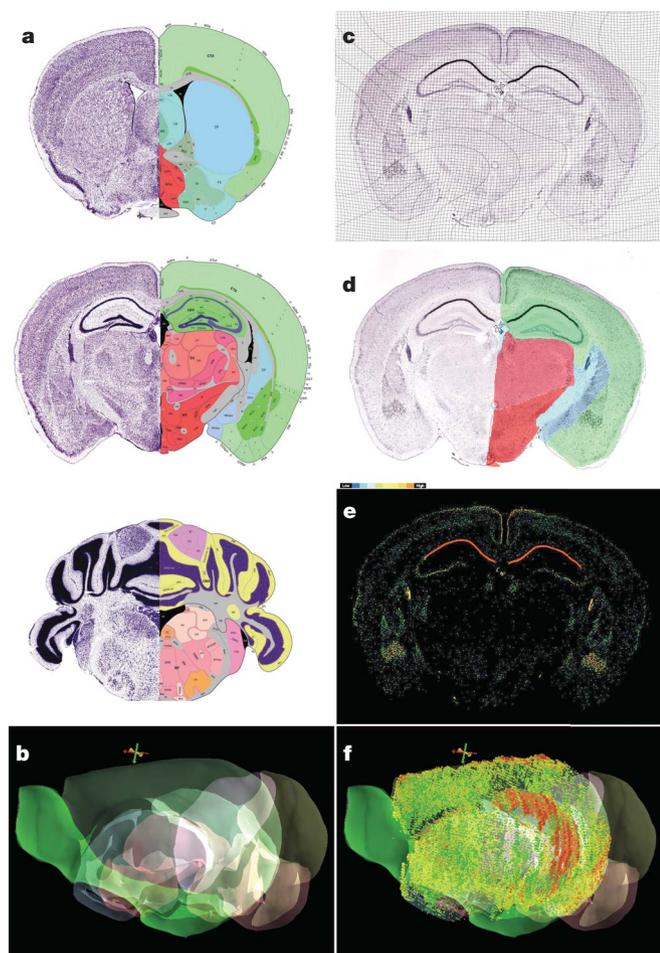


Figure 1 | Global analysis strategy. **a**, Panels from the Allen Reference Atlas, annotated and colour-coded using an anatomical nomenclature modified from ref. 55. **b**, These images are rendered into a three-dimensional annotated reference volume. ISH images are aligned in three dimensions with the reference volume (*Man1a* shown in **c–f**), allowing overlay of a fine grid framework (**c**) or a plane-matched slice through the reference volume to delineate structural boundaries (**d**) onto the two-dimensional ISH images. **e**, **f**, Image segmentation and false-colouring by object pixel intensity (**e**) allows quantification, assignment of signal to anatomical structure and visualization on the three-dimensional reference framework (**f**).

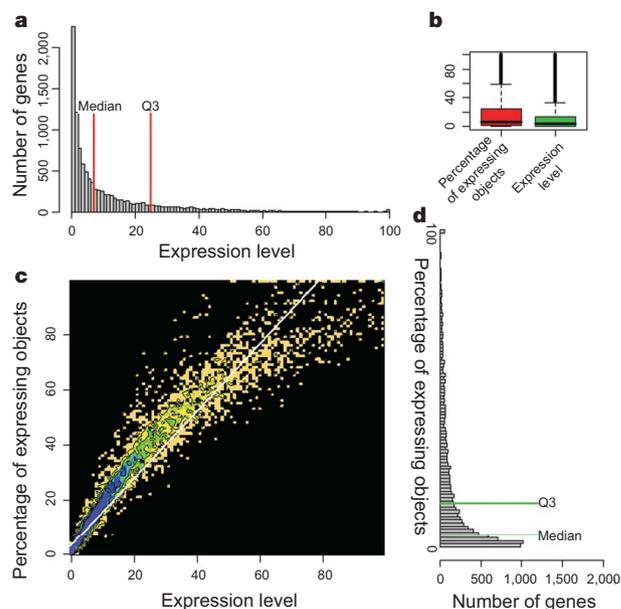


Figure 2 | Genome-wide analysis of expression level versus percentage of expressing cells across the entire brain. **a**, **d**, Histograms of genes versus expression level (**a**, 100 is maximum expression level) or percentage of expressing cells (**d**, 100 is ubiquitous expression). **b**, Box plot indicating median percentage of expressing objects and expression level and distribution across the entire set of expressed genes. Expression scores are normalized to [0,100] with median and third quartiles indicated. **c**, Expression level and percentage of expressing cells are highly correlated among expressing genes, as shown by a regression line ($R^2 = 0.92$). Similar plots for individual structures are available in Supplementary Fig. 1.

binned either by anatomical structure or into three-dimensional voxels of arbitrary size (see Supplementary Methods 2 for mathematical descriptions of variables). To allow visual differentiation of signal from noise, 'heat maps' representing signal intensity (average pixel intensity/expressing object) are generated for each two-dimensional image. Pseudo-colouring of these maps by signal intensity provides a visually accessible quantification of ISH signal in individual cells within individual sections (Fig. 1e). An interactive application, Brain Explorer, was also developed to allow visualization of three-dimensional representations of gene expression superimposed on the reference atlas (Fig. 1f).

Expressed versus non-expressed genes. Approximately 80% of total genes assayed display some cellular expression above background in the brain. This number was equivalently derived through manual scoring and informatics-based analysis using expression thresholds calibrated through expert confirmation. This percentage is significantly higher than predicted from microarray analysis of brain tissues³, indicating that many genes are expressed at low levels and/or in small numbers of cells for which expression levels fall below the sensitivity of arrays applied to large brain regions. The distribution of expression level and percentage of expressing cells, and the correlation between these two variables across the entire set of expressers, is plotted in Fig. 2. Notably, most genes are expressed in a relatively small percentage of cells (70.5% of genes are expressed in less than 20% of total cells). Although expression level correlates strongly with the percentage of expressing cells ($R^2 = 0.92$), the distributions of these variables differ from one another (chi-squared, $P < 2.2 \times 10^{-16}$). Similar distributions are observed for individual brain regions (Supplementary Fig. 1). Although expression level can vary as a function of probe length, the distribution is unchanged between the entire set of expressing genes and a subset with probe lengths ± 0.5 standard deviations (s.d.) from the mean (649–826 nucleotides, $P = 0.318$).

To search for 'housekeeping' genes necessary for the basal function of all cells, we examined the genes with the greatest percentage of expressing cells. Although true ubiquity cannot be established without a nuclear counter stain, genes with 'near ubiquity' can be validated by visual confirmation of expression in essentially all cells relative to Nissl staining. A total of 186 of the top 500 genes showed detectable expression in all cell classes, although very few genes (for

example, *2610002F03Rik* and *Fbxo22*; Fig. 3) are expressed at high levels in all easily definable cell types (neurons, astrocytes, oligodendrocytes, choroid plexus, ependymal lining of ventricles and pial surface). A much larger set of genes appears to be expressed in all cells but at very low levels in some cellular populations. As expected, gene ontology (GO)¹⁶ categories involving cellular metabolism are highly represented in apparently ubiquitous genes.

Enriched expression in major cell types. A targeted search for genes enriched in major cell classes with distinctive anatomical localization is enabled by the atlas grid coordinate system superimposed on each gene's expression pattern. These coordinates enable searches for genes having similar expression patterns to genes of interest. Seeding with oligodendrocyte-specific genes (*Mbp*, *Mobp*, *Cnp1*) returns a large set of oligodendrocyte-enriched genes; seeding with choroid-plexus-enriched genes (*Col8a2*, *Lbp*, *Msx1*) returns a large set of choroid-enriched genes. Where spatial distributions of cell classes overlap greatly, such as for neurons and astrocytes, this strategy is less robust but can still be effective. Essentially all well-established markers for different cell types were identified with these informatics-based methods for prediction followed by expert confirmation (Supplementary Table 1). GO analysis of genes enriched in major cell classes reflects the different functions of these cells (Fig. 3 and Supplementary Table 2). For example, many oligodendrocyte-enriched genes are involved in myelin production. Certain functional ontologies are over-represented in the non-expressed set as well, including immune responses, meiosis and sensory organ development. Many previously uncharacterized genes are enriched in each cell class, providing novel gene candidates for cell-type-specific function.

Regionally enriched gene expression. Genes with regionalized expression patterns provide potential substrates for functional differences between brain regions, whereas correlations between regions may predict functional similarities. To identify the most specific genes for each of 12 major brain regions, the area occupied by signal in each structure was divided by the area occupied by signal in the entire brain for each gene. These ranked lists were visually confirmed to identify the top 100 most specific genes for each region. These lists of enriched genes (Supplementary Table 3) are representative but not comprehensive, as the search missed several well-known genes with high specificity (for example, *Agrp* in the hypothalamus) owing to

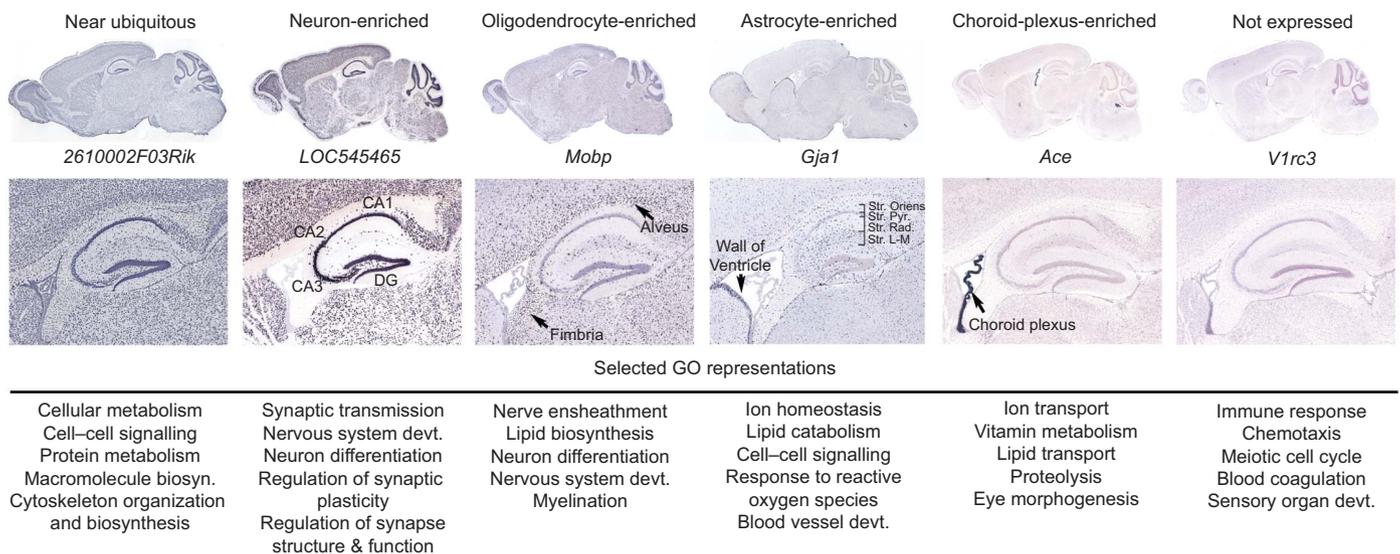


Figure 3 | Representative cell-type-specific genes and corresponding molecular functions. Examples of genes expressed in all cell types, neurons, oligodendrocytes, astrocytes and the choroid plexus, and a representative non-expressed gene. Selected GO representations¹³ are shown for each class (see Supplementary Tables 1 and 2 for complete gene lists and GO terms). *Mbp*, myelin-associated oligodendrocytic basic protein; *Gjal*, gap junction

membrane channel protein alpha 1; *Ace*, angiotensin I converting enzyme 1; *V1rc3*, vomeronasal 1 receptor. C3, Hippocampal subfields: CA1, CA2, CA3, DG (dentate gyrus). Layers of CA1: Str. Oriens, stratum oriens; Str. Pyr., stratum pyramidale; Str. Rad., stratum radiatum; Str. L-M, stratum lacunosum-moleculare.

artefacts interfering with accurate registration and/or signal detection. The top candidates for each structure show highly enriched expression, typically restricted to a subset of nuclei/cells within that structure, although few genes are expressed only in a single structure. Some structures have a greater degree of enriched gene expression than others, easily visualized by plotting the average expression of the top 100 genes for a given structure on a representative brain section (Fig. 4). For example, the hippocampus, thalamus, cerebellum and olfactory bulb display a great deal of enriched gene expression. In contrast, the pallidum, hypothalamus, midbrain, pons and medulla exhibit highly overlapping expression patterns.

The complexity of structural combinations is more striking when plotted on a gene-by-gene basis (Fig. 4, bottom row, left column of each panel). Hierarchical clustering of enriched gene sets based on correlations between expression level for each structure reveals unique patterns of correlated gene expression between different structures (Fig. 4, bottom row, right column of each panel). For example, within the cortically enriched gene set, strong correlations are largely with other cortical structures, with relatively distinct gene clusters showing co-expression between combinations of cortical structures. The retrohippocampal formation and hippocampus show strongly correlated expression, as do the pallidum and hypothalamus and the pons and medulla.

Clustering of correlated gene expression. Classical definition of brain regions is based on a combination of overall morphology, cellular cytoarchitecture, ontological development and functional connectivity. Grid-level correlations across the brain provide the means to assess whether gene expression in the adult mirrors these organizational principles. To approach this question, correlations were performed between evenly spaced voxels throughout the volume of a brain hemisphere across a set of >5,000 genes chosen to contain a large proportion of genes with regionalized expression but excluding non-expressers or very widely expressed genes. Organizing voxels by inclusion in large anatomical structures (Fig. 5b, d) reveals high correlations between voxels in the same gross anatomical region, and complex but distinct correlations between brain regions. Alternatively, unbiased hierarchical clustering of voxels based on their overall correlations with other voxels produces a different result (Fig. 5c, e). Large anatomical divisions with relatively uniform internal structure, such as the striatum and cerebral cortex, are represented by single tight clusters. Other structures that consist of discrete nuclei, such as the hypothalamus, pons, medulla and midbrain, have local clusters interleaved with one another (Fig. 5e). On the basis of these data, organization of gene expression more closely mirrors the smaller nuclear structure, with stronger relationships between nuclei than between gross anatomical regions.

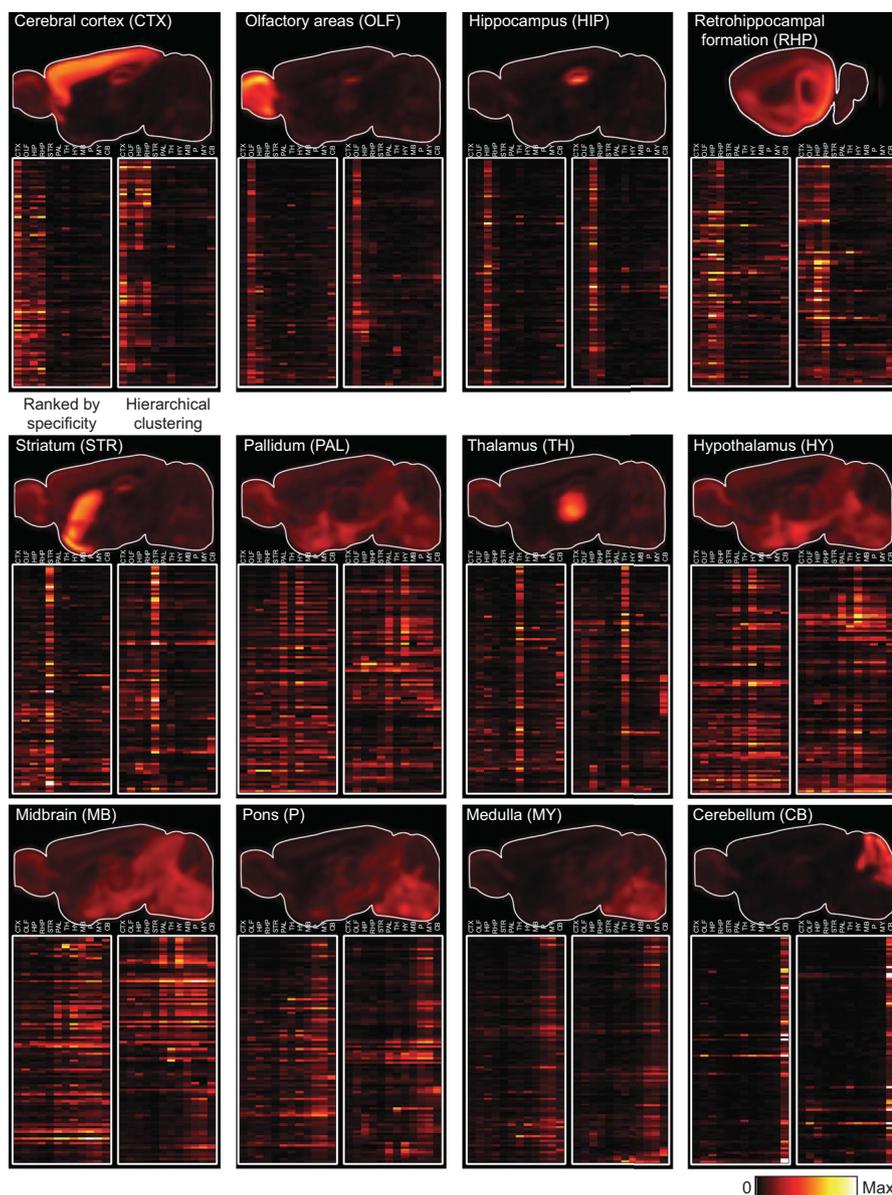


Figure 4 | Hierarchical ranking and clustering of the most specific genes in 12 major brain regions. Top row: average expression of the top 100 enriched genes for each region, plotted as the fractional percentage of expressing pixels/300 μm^3 voxel on a representative sagittal section containing that region. Bottom two rows: expression of the same genes across all 12 structures ranked by structural specificity (left column of each panel) or ordered by hierarchical clustering based on correlations between structures (right column of each panel). 'Hot metal' colour scale for percentage of expressing pixels in panels of the top row: dark red (0) to white (≥ 0.06); expression level in panels of the bottom two rows: dark red (0) to white (> 80).

Nuclear- and cellular-level analysis of gene expression

A genome-wide ISH data set allows a variety of analyses to understand the fine molecular anatomy of brain regions. One approach is to generate complete expression profiles for gene families of obvious functional relevance, such as the ligand-gated ion channels (Supplementary Data 4 and 5). Alternatively, unbiased analysis of cellular expression patterns in specific brain regions can identify the most specific gene markers for those regions and (putative) cell types, as well as reveal fine, presumably functional subdivisions that are indistinguishable by conventional cytoarchitecture alone.

Laminar and regional cortical specificity. Systematic analysis of different cortical cell types has been hampered by the lack of specific markers defining neuronal classes with consistent phenotypic properties. Numerous genes have been identified with highly specific cortical gene expression restricted to discrete neuronal populations in different cortical layers, many of which have no known function. For example, *A930038C07Rik* (Fig. 6b) is expressed nearly exclusively in layer 1 cells, potentially Cajal–Retzius cells. Pyramidal neurons in layers 2/3 express the regulator of G-protein signalling family member 8 (*Rgs8*, Fig. 6d), consistent with observations that RGS proteins are involved in higher-level synaptic plasticity, as persistent plasticity is observed in superficial cortical layers¹⁷. Cocaine-amphetamine regulated transcript (*Cart*; also called *Cartpt*), a neuropeptide involved in reward/reinforcement circuitry¹⁸, is quite specifically expressed in deep layer 3 of somatosensory cortex (Fig. 6e). Layer 4 is predominantly labelled by *LOC433228* (Fig. 6f). Pyramidal cells in layer 5 are delineated by expression of *C030003D03Rik* (Fig. 6g), whereas layer 6 neurons specifically express the immunoglobulin heavy chain (TIGR accession TC1460681; Fig. 6h). Layer 6b neurons

express the early growth-response gene *Ctgf* (Fig. 6i), as previously described in rat¹⁹.

Most genes expressed in the neocortex display relatively uniform laminar expression across all cortical areas, consistent with the idea that the basic (canonical) microcircuit is conserved across the entire neocortex. However, functionally discrete cortical regions are also delineated by gene expression, albeit infrequently. For example, the cytochrome P450 family member *Cyp39a1* (Fig. 6j) is expressed almost exclusively in somatosensory cortex, and Marcks-like protein (*Mlp*; also known as *Marcksl1*) is expressed in layer 5 of all neocortical regions except somatosensory cortex (Fig. 6k). These findings suggest that cellular properties vary between different regions of the adult neocortex.

Heterogeneity in hippocampal subregions. The hippocampus has been extensively studied owing to its essential role in certain types of learning and memory^{20,21}. Lorente de No²² divided the pyramidal cell layer of the hippocampus proper into CA1, CA2 and CA3 subfields, and there is extensive anatomical evidence for regional specialization within these subfields^{22–24}.

As described previously, the main subregions of the hippocampus are specifically delineated by gene expression^{25,26}, and a number of additional markers have been identified (Fig. 7a–c). Heterogeneous expression within hippocampal subregions is also common, and is found in all subregions in all cardinal axes. For example, within CA3, *Col15a1* is expressed in a gradient with higher expression proximal to the dentate gyrus (Fig. 7d), and *Crym* is expressed in a reciprocal gradient (Fig. 7e). Heterogeneity is occasionally observed across the depth of the pyramidal cell layer. *Col6a1* is preferentially expressed in a band of cells on the outer border of CA2 and CA3 (Fig. 7f).

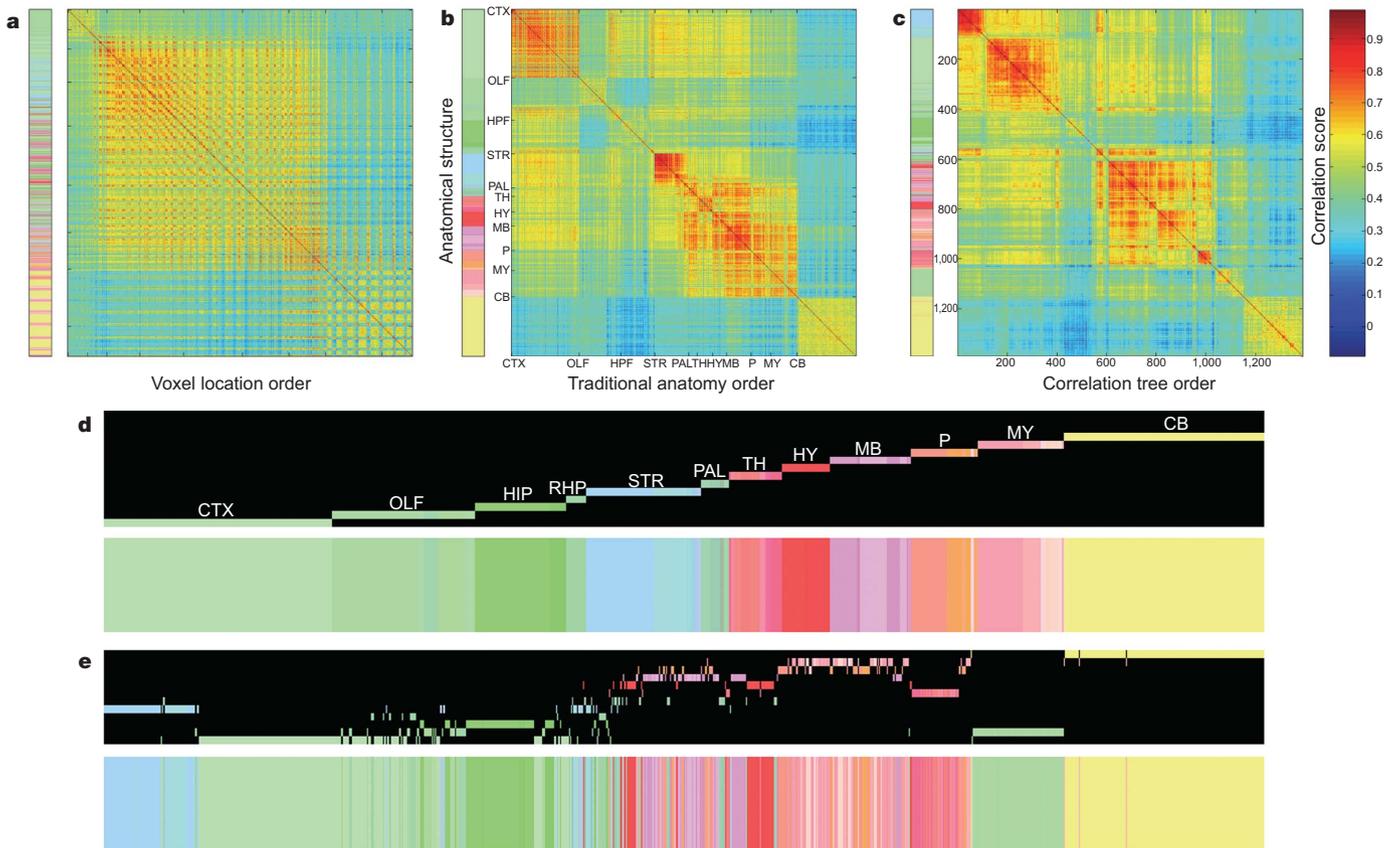


Figure 5 | Correlative analysis of gene expression. **a**, Correlations between voxels ordered in grid space where colour-coding is based on correlation scores of expression in each voxel to all others across 5,195 genes. **b**, Voxels ordered by structures defined in the Allen Reference Atlas. Colour bar (left) indicates inclusion in listed structures. Most structures display high internal correlations, with differential cross-structural correlations. **c**, Hierarchical

clustering of voxels independent of anatomical location. *x-y* axes in **a–c** represent voxel locations, with the colour scale on the left indicating the anatomical structure in which each voxel is contained. Correlation score bar (range [0,1]) on the right applies to **a–c**. **d, e**, Magnification of colour bars indicating structural identity from **b** and **c**, respectively. Voxel inclusion in the 12 main structures is indicated by a line in the row for each structure.

Dorsal/ventral (septotemporal) heterogeneity is very common (Fig. 7g–m). Individual genes often vary across more than one dimension, defining complex three-dimensional compartments within a given subfield. Five of the ten genes shown in Fig. 7 (selected solely based on specificity) are involved in cell adhesion, and 56 out of 188 cell-adhesion-related genes in the Allen Brain Atlas demonstrate spatially restricted hippocampal expression (data not shown). From a different perspective, 1,137 out of 5,099 analysed genes displayed some form of regionality in the hippocampus. The top over-represented functional category within the regional gene set is cell adhesion ($P < 1.79 \times 10^{-10}$). Differential cell adhesion may be important for establishment and maintenance of topographic connectivity, or, as described recently, different forms of synaptic plasticity and remodelling²⁷.

Although there is increasing experimental evidence suggesting functional differentiation along the septotemporal axis of the hippocampus^{28–30}, the precise cellular substrate has not been defined. The observation that different neuropeptides (*Grp* and *Nmb*; Fig. 7g–j) are restricted to either dorsal or ventral hippocampus provides evidence for functional differentiation involving discrete signalling pathways. Identification of these regionally restricted markers allows anatomical delineation of novel hippocampal compartments

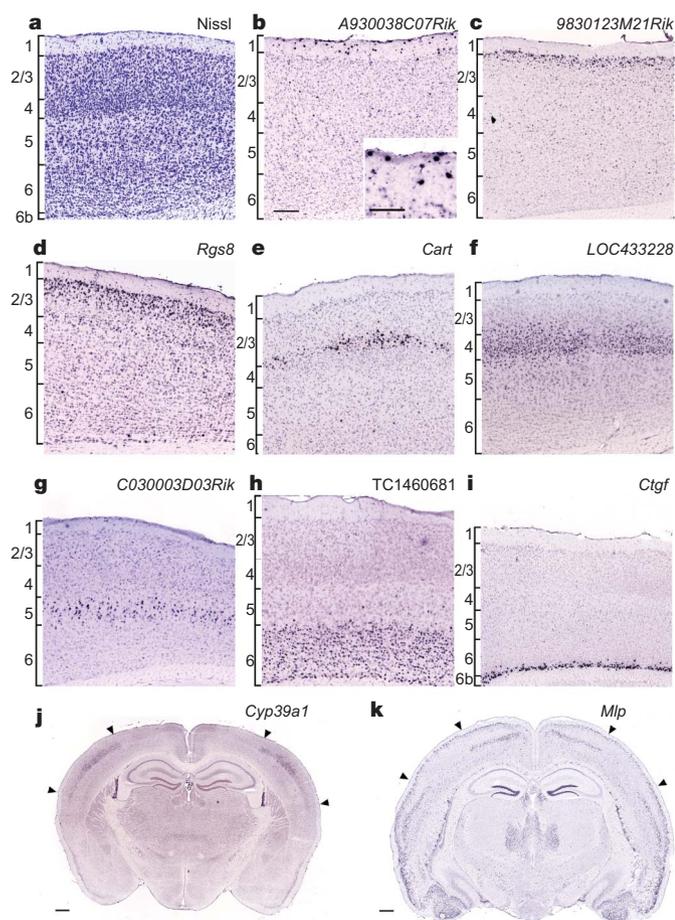


Figure 6 | Laminar and region-specific neocortical gene expression. **a**, Nissl-stained section with cortical layer boundaries. **b–i**, Layer-specific expression in sagittal sections: *A930038C07Rik* in layer 1 (**b**, inset shows higher magnification view); *9830123M21Rik* in layer 2 (**c**); *Rgs8* in layer 2/3 (**d**); *Cart* in deep layer 3 (**e**); *LOC433228* in layer 4 (**f**); *C030003D03Rik* in layer 5 (**g**); *TC1460681/IgM* in layer 6 (**h**); *Ctgf* in layer 6b (**i**). **j, k**, Selective *Cyp39a1* expression in layer 4 (**j**) and lack of *Mlp* expression in layer 5 (**k**) delineate somatosensory cortex (arrowheads). Scale bars: **b–i**, 200 μm ; inset in **b**, 50 μm ; **j, k**, 500 μm .

and provides the means for future experimental manipulation to assess function.

Novel cerebellar compartments. The basic structure of the cerebellum is well known and consists of several functionally discrete gross divisions. Additionally, the cerebellar cortex exhibits a bilaterally symmetric series of sagittally oriented bands³¹ mirrored by a number of genes, most notably zebrin^{32,33}. Strong (although not complete) correlation between patterns of cerebellar afferent segregation and zebrin expression³⁴ indicate that molecular markers can delineate functionally discrete regions in the cerebellum.

A number of genes display heterogeneity within cerebellar granule and Purkinje cell populations. For example, *Rasgrf1* defines a previously unrecognized large, contiguous domain with sharp boundaries in the granule cell layer of the rostral (Fig. 8a) and dorsal cerebellum (Fig. 8b). More complex regional patterns are observed in the Purkinje cell layer, such as that of *Opn3*, a non-canonical opsin (Fig. 8c, d) whose expression in the cerebellum has been described as a rostro-caudal gradient with radial stripes³⁵. Rather than a gradient, three-dimensional reconstruction of ISH data for *Opn3* reveals a more coherent pattern (Fig. 8e) involving a sharply delineated diagonal band lacking expression extending across the entire cerebellum. The overall pattern of *Opn3* is both complex and discrete, with regionalized expression in distinct lobules and sagittal banding in the posterior vermis.

Subcellular mRNA targeting. Subcellular localization and translation of mRNA transcripts in dendrites is increasingly recognized as a widespread phenomenon³⁶, and is thought to be involved in certain forms of synaptic plasticity^{37,38}. Dendritic mRNA targeting is particularly obvious in the hippocampus (Fig. 9a–f) and cerebellum (Fig. 9g, h), where clear distinctions can be made between cell-dense layers, dendritic molecular layers and white matter. Targeting throughout the entire dendritic field is exemplified by the well-characterized patterns of *Camk2a* and *Dnd1* in the hippocampus (Fig. 9b, c)³⁹. Although often subtle, this distribution is independent of expression level, thereby distinguishing targeting from passive diffusion of mRNA into dendrites. For example, labelling for the highly expressed gene *Nptx1* is confined to the soma of CA3 pyramidal cells (Fig. 9d), whereas microtubule associated proteins 1A (*Mtap1a*) and 2 (*Mtap2*) also label proximal or proximal and distal dendrites (Fig. 9e, f). Similar forms of targeting are seen in cerebellar Purkinje cells (Fig. 9h, i), as well as in oligodendrocytes (*Mbp*; Fig. 9k)⁴⁰.

Many genes exhibiting dendritic targeting are involved in cytoskeletal organization and biogenesis, as well as in regulating synaptic plasticity. There appear to be multiple *cis*-acting sequence elements that mediate mRNA targeting⁴¹. Identification of sets of dendritically targeted mRNAs with shared features (Supplementary Table 4), such as regional specificity and distribution in dendrites, may aid in the identification of conserved sequence elements that correlate with cellular and intracellular transport specificity.

Discussion

A global analysis of the genetic underpinnings of the brain's structural and cellular complexity requires a genome-wide, cellular resolution, anatomically comprehensive data set. The Allen Brain Atlas has taken a genomics-style approach to understanding this complexity by creating an integrated set of data production and analysis methodologies to systematically produce and analyse a comprehensive atlas of gene expression in the adult C57BL/6J mouse brain. Accurate and comprehensive analysis and annotation of data from the Allen Brain Atlas and other similar projects are difficult but essential hurdles to realizing the full potential of these data. New suites of image analysis tools are required to apply analysis methods developed by the genomics community to cellular-resolution gene expression data in the brain. Ongoing efforts aim to allow correlative cross-gene and cross-structural analysis across the entire Allen Brain Atlas data set by improving automated and semi-automated

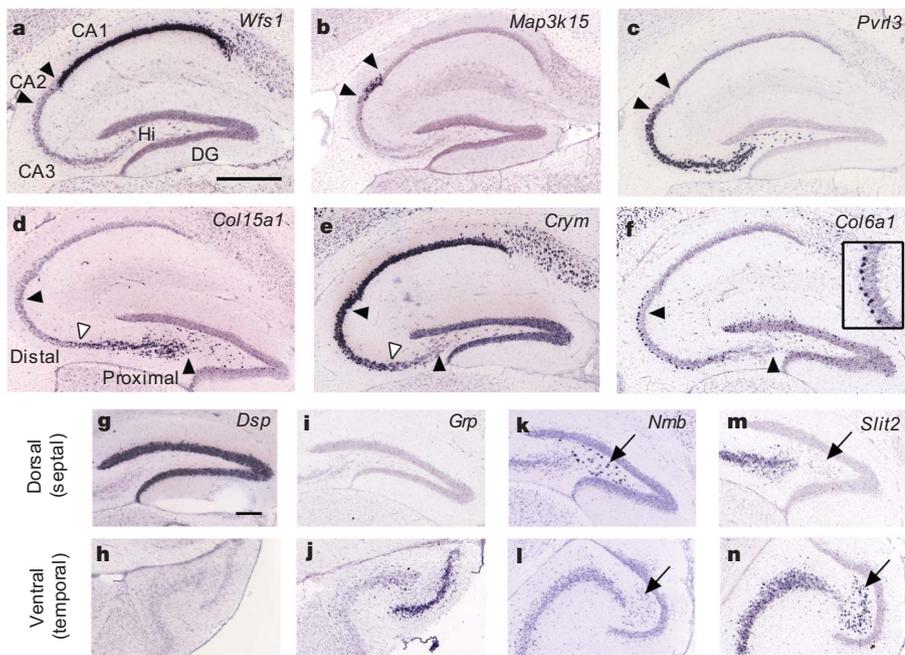


Figure 7 | Heterogeneity of hippocampal gene expression. Classically defined hippocampal subregions (black arrowheads) are delineated by gene expression: **a**, *Wfs1* in CA1; **b**, *Map3k15* in CA2; **c**, *Pvr13* in CA3. Heterogeneous expression within hippocampal subfields: proximal–distal (*Col15a1*, **d**) and distal–proximal (*Crym*, **e**) gradients in CA3 (white arrowheads mark approximate expression boundaries); selective expression of *Col6a1* in the outer band of cells in CA3 (**f**); differential expression in dorsal (*Dsp*, **g**, **h**) and ventral dentate gyrus granule cells (*Grp*, **i**, **j**), and dorsal (*Nmb*, **k**, **l**) and ventral hilar neurons (*Slit2*, **m**, **n**). Scale bars: **a–f**, 500 μm; **g–n**, 200 μm. Hi, hilus (arrows in **k–n**).

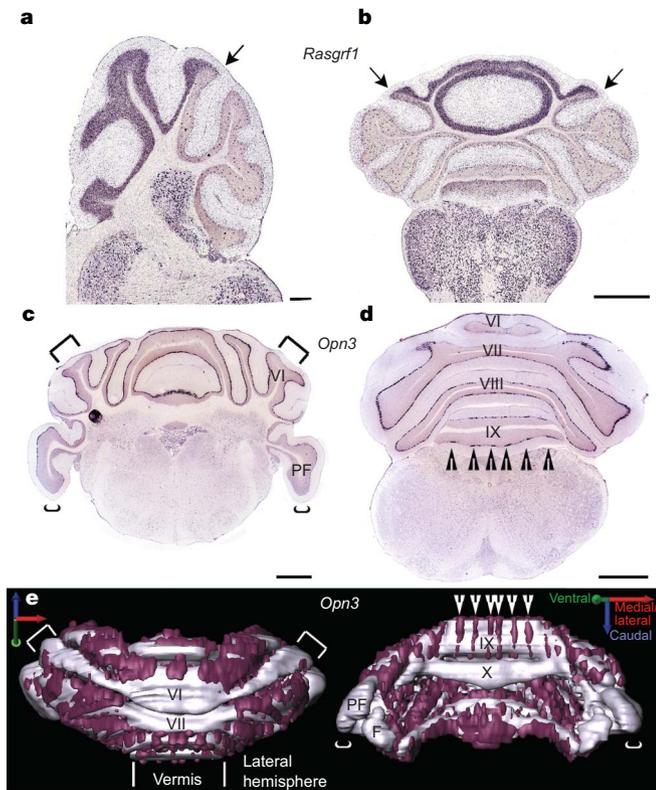


Figure 8 | Cerebellar compartments revealed by gene expression. **a**, **b**, Sagittal and coronal views of regional *Rasgrf1* expression in the rostral and dorsal granule cell layer (arrows). **c–e**, Complex regional *Opn3* expression in Purkinje cells in anterior (**c**) and posterior (**d**) sections. **e**, Three-dimensional reconstruction of *Opn3* expression reveals sagittal and diagonal bands. Surface rendering of *Opn3* expression (purple) is superimposed on a model of the granule cell layer (silver) generated from the same brain. Arrowheads mark sagittal bands of *Opn3* expression, whereas brackets distinguish areas lacking *Opn3* expression. Scale bars: **a**, 200 μm; **b–d**, 1 mm. F, flocculus; PF, paraflocculus.

methodologies in order to quantify and map gene expression at increasingly finer anatomical resolution.

It should be noted that the expression data produced by the current project provide an incomplete snapshot of the molecular complexity of the brain. There is increasing evidence of complex, often cell-type-specific transcriptional splice variation that significantly increases the complexity and size of the transcriptome⁴², whereas a significant proportion of the transcriptome is also dynamically regulated, for example by circadian rhythms⁴³. It will also be important to complement these transcriptional data with protein data in the future. In these respects the Allen Brain Atlas serves as a comprehensive baseline data set with which to compare gene expression in other tissues, species, developmental time points, behavioural and disease states.

The identification of large numbers of genes with restricted expression in specific cell populations yields insights into the fine structural and functional organization of specific brain regions and provides the basis for precise functional manipulations. Although soma location and cell density are strong predictors of cell class, important future goals will be to determine the relationship between transcriptional profiles and specific cellular phenotypes by correlating gene expression with more salient characteristics such as morphology, connectivity and physiology⁸. Cell-type-specific transgenic manipulation using promoters derived from highly specific genes is an obvious next step towards these goals, which will be facilitated by systematic promoter analysis of co-expressed genes to identify the necessary *cis*-acting regulatory elements⁴⁴ driving gene expression in particular cell types. This powerful approach towards cellular characterization has demonstrated that promoter-driven transgenes can define cell ‘classes’ with consistent morphology and patterns of connectivity^{6,11}. Advances in transgenic methodologies to produce spatially and temporally regulated transgene expression promise to revolutionize modern neuroscience, allowing a wide variety of cell-type-specific manipulations including lesions⁴⁵, silencing of neuronal activity⁴⁶, trans-synaptic labelling^{47,48} and real-time imaging of neuronal morphology and activity⁴⁹. Region- and cell-type-specific markers identified using the Allen Brain Atlas should broadly facilitate these approaches, allowing specific manipulations to reveal the contributions of specific cell types to higher-level brain function and disease-related dysfunction.

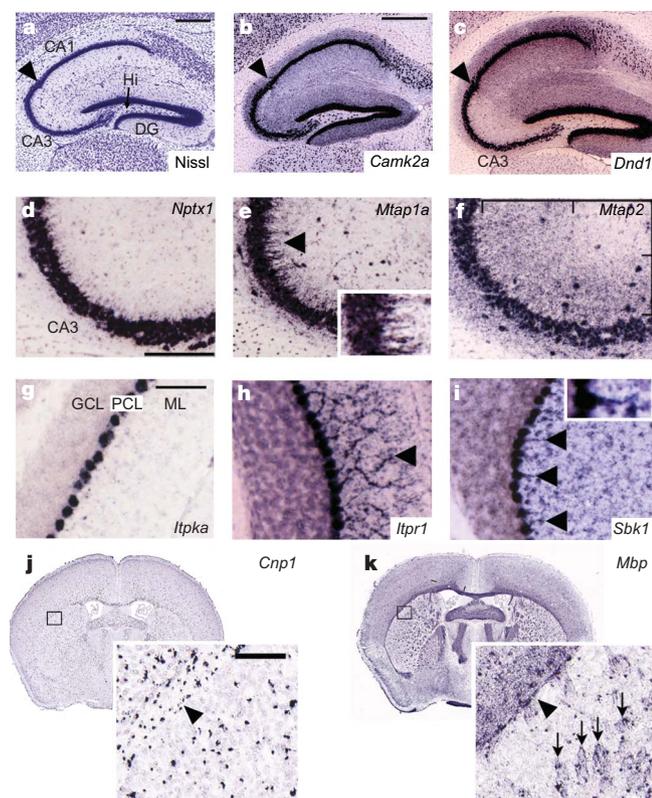


Figure 9 | Subcellular mRNA targeting. **a**, Nissl-stained section through the hippocampus. **b**, **c**, Dendritic targeting in apical and basal dendritic zones of CA1–CA3 and DG (*Camk2a*, **b**) or just CA1 and DG (*Dnd1*, **c**). **d–f**, Differential CA3 mRNA targeting: soma only (*Nptx1*, **d**), proximal apical dendrite only (*Mtap1a*, **e**), and proximal and distal dendrites (*Mtap2*, **f**). **g–i**, Cerebellar Purkinje cell targeting: soma only (*Itpka*, **g**), entire dendritic field (*Itpr1*, **h**) or proximal dendrites only (*Sbk1*, **i**). **j**, **k**, Oligodendrocyte targeting: soma only (*Cnp1*, **j**) versus white matter (myelin) labelling in external capsule (arrowhead) and striatum (arrows) (*Mbp*, **k**). Scale bars: **a–c**, **g–i**, 200 μm ; **d–f**, 600 μm ; **j**, **k**, 100 μm . GCL, granule cell layer; ML, molecular layer; PCL, Purkinje cell layer.

METHODS

High-throughput data generation was achieved by industrializing and automating ISH data production, as well as subsequent image capture for analysis and web-based viewing. This involved creating streamlined pipelines for gene selection, probe manufacturing, tissue sectioning, ISH and *in silico* image processing, and data analysis⁵⁰. Brief descriptions are presented below, and extensive methodological details are supplied as Supplementary Information (Supplementary Methods 1 and 2).

Animals and tissue processing. Male, 56-day-old C57BL/6J mice (Jackson Labs West) were maintained on a 12 h light/dark cycle. Brain tissue was collected between 3 and 6 h after light onset. Serial 25- μm fresh-frozen cryostat sections were systematically collected starting at a standardized plane of section to ensure reproducible anatomical coverage.

Riboprobe generation. A semi-automated process was used to design gene-specific probes based on sequences from multiple sources including Refseq, MGC, Celera, TIGR, Riken and Unigene. Preliminary sequence data were obtained from The Institute for Genomic Research (<http://www.tigr.org>). Probes were designed against unique regions of transcripts to avoid cross-reactivity.

In situ hybridization. The semi-automated non-isotopic digoxigenin (DIG)-based ISH platform of ref. 12 was used with minor modifications. Many controls were performed to establish the reproducibility of the platform and to demonstrate both qualitative and quantitative concordance to conventional radioactive ISH methodologies (see Supplementary Data 1–3). Each riboprobe was initially processed on 200- μm spaced sagittal sections spanning an entire hemisphere, and a coronal replicate was generated for a subset (~3,500) of genes with restricted expression patterns. An automated image capture platform was developed to digitize ISH data.

ISH anatomic mapping and quantification. Mapping of ISH data to a standard coordinate frame was achieved through three-dimensional registration of ISH

data to an annotated volume generated from a Nissl-stained reference atlas created for this project (Supplementary Methods 2 and 3). Rigid registration methods⁵¹ were used to construct the three-dimensional reference atlas, using a 1.5-T, low-resolution, three-dimensional averaged magnetic resonance imaging (MRI) volume as a template to ensure that reconstruction resulted in a realistic volume. Registration of each ISH data set to the Nissl volume involves iterative application of global rigid and subsequently deformable methods⁵².

Signal detection algorithms were developed to identify labelled pixels and segment expressing cells using methods of adaptive image morphology⁵³ to recognize objects of the correct shape and intensity distribution while ignoring tissue artefacts. The expression mask for each ISH section is constructed by combining tissue boundary detection with small and isolated object recognition (dense or compact nuclei require special treatment). The number of labelled cells proved to be the best predictor of whether or not a gene is expressed in the brain. A minimum cutoff of $>1.5\%$ of all 8×8 pixel regions containing some segmented cell body was used to ascertain brain expression, a threshold derived through extensive manual calibration. The two most salient variables for quantitative analysis proved to be relative density (D), defined as the proportion of labelled cells, and expression level (L), a measure correlated with total transcript count incorporating both area occupied by expressing pixels as well as pixel intensity. **Global correlation analysis.** Cell-type-enriched markers were identified by searching for genes with highly correlated expression to specific markers for those cell types. A Pearson correlation-based metric was used to compare expression levels between seed genes and the entire Allen Brain Atlas data set in $100\text{-}\mu\text{m}^3$ voxels across the entire brain. *Gjal* was used to search for astrocyte-enriched genes, *Col8a2* for the choroid plexus, *Cnp1* for oligodendrocytes, and *Snrpn* for neurons.

To assess global correlations between different brain regions, this same technique was applied to compare expression levels in a pairwise fashion between each of 1,500 uniformly distributed voxels across the brain. Display of these voxel correlations grouped by (1) inclusion in anatomic structures, or (2) hierarchically clustered by expression correlations between voxels, was computed using Matlab (The MathWorks). Details are provided in Supplementary Methods 4.

We searched for genes enriched in a specific structure by dividing expression in that structure (in this case the fraction of expressing pixels) by expression in the brain as a whole. The resulting ranked gene lists were manually curated to identify the 100 most specific genes. The average expression across the gene lists for each structure was calculated by computing the average expression level L on a voxel-by-voxel basis in three-dimensions throughout the brain, and then plotting these values onto a selected plane of section through that structure. Hierarchical clustering based on (Pearson-based) correlations of L across 12 brain regions was performed using Matlab.

Three-dimensional reconstruction of cerebellar gene expression. Three-dimensional reconstruction of *Opn3* expression was performed using Adobe Photoshop (Adobe Systems) and the 3D Constructor module of Image-Pro Plus (v.5.0; Media Cybernetics) to align and surface render gene expression from uniformly spaced ISH and Nissl images, as described previously⁵⁴.

Received 28 July; accepted 15 November 2006.

Published online 6 December 2006.

- Zapala, M. A. et al. Adult mouse brain gene expression patterns bear an embryologic imprint. *Proc. Natl Acad. Sci. USA* **102**, 10357–10362 (2005).
- Datson, N. A., van der Perk, J., de Kloet, E. R. & Vreugdenhil, E. Expression profile of 30,000 genes in rat hippocampus using SAGE. *Hippocampus* **11**, 430–444 (2001).
- Sandberg, R. et al. Regional and strain-specific gene expression mapping in the adult mouse brain. *Proc. Natl Acad. Sci. USA* **97**, 11038–11043 (2000).
- Siddiqui, A. S. et al. A mouse atlas of gene expression: large-scale digital gene-expression profiles from precisely defined developing C57BL/6J mouse tissues and cells. *Proc. Natl Acad. Sci. USA* **102**, 18485–18490 (2005).
- Kamme, F. et al. Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. *J. Neurosci.* **23**, 3607–3615 (2003).
- Sugino, K. et al. Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature Neurosci.* **9**, 99–107 (2006).
- Markram, H. et al. Interneurons of the neocortical inhibitory system. *Nature Rev. Neurosci.* **5**, 793–807 (2004).
- Toledo-Rodriguez, M., Goodman, P., Wu, C. & Markram, H. Neuropeptide and calcium-binding protein gene expression profiles predict neuronal anatomical type in the juvenile rat. *J. Physiol. (Lond.)* **567**, 401–413 (2005).
- Monyer, H. & Markram, H. Interneuron diversity series: Molecular and genetic tools to study GABAergic interneuron diversity and function. *Trends Neurosci.* **27**, 90–97 (2004).
- Christiansen, J. H. et al. EMAGE: a spatial database of gene expression patterns during mouse embryo development. *Nucleic Acids Res.* **34**, D637–D641 (2006).

11. Gong, S. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).
12. Visel, A., Thaller, C. & Eichele, G. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.* **32**, D552–D556 (2004).
13. Gray, P. A. *et al.* Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* **306**, 2255–2257 (2004).
14. Magdaleno, S. *et al.* BGEM: an *in situ* hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol.* **4**, e86 (2006).
15. Boguski, M. S. & Jones, A. R. Neurogenomics: at the intersection of neurobiology and genome sciences. *Nature Neurosci.* **7**, 429–433 (2004).
16. Dennis, G. Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, P3 (2003).
17. Kirkwood, A., Rioult, M. C. & Bear, M. F. Experience-dependent modification of synaptic plasticity in visual cortex. *Nature* **381**, 526–528 (1996).
18. Elias, C. F. *et al.* Characterization of CART neurons in the rat and human hypothalamus. *J. Comp. Neurol.* **432**, 1–19 (2001).
19. Heuer, H. *et al.* Connective tissue growth factor: a novel marker of layer VII neurons in the rat cerebral cortex. *Neuroscience* **119**, 43–52 (2003).
20. Zola-Morgan, S., Squire, L. R. & Amaral, D. G. Human amnesia and the medial temporal region: enduring memory impairment following a bilateral lesion limited to field CA1 of the hippocampus. *J. Neurosci.* **6**, 2950–2967 (1986).
21. Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. 1957. *J. Neuropsychiatry Clin. Neurosci.* **12**, 103–113 (2000).
22. Lorente de N6, R. Studies on the structure of the cerebral cortex. II. Continuation of the study of the ammonic system. *J. Psychol. Neurol. (Lpz.)* **46**, 113–177 (1934).
23. Amaral, D. G. & Witter, M. P. The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* **31**, 571–591 (1989).
24. Ishizuka, N., Weber, J. & Amaral, D. G. Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J. Comp. Neurol.* **295**, 580–623 (1990).
25. Tole, S., Christian, C. & Grove, E. A. Early specification and autonomous development of cortical fields in the mouse hippocampus. *Development* **124**, 4959–4970 (1997).
26. Lein, E. S., Zhao, X. & Gage, F. H. Defining a molecular atlas of the hippocampus using DNA microarrays and high-throughput *in situ* hybridization. *J. Neurosci.* **24**, 3879–3889 (2004).
27. Yamagata, M., Sanes, J. R. & Weiner, J. A. Synaptic adhesion molecules. *Curr. Opin. Cell Biol.* **15**, 621–632 (2003).
28. Small, S. A. The longitudinal axis of the hippocampal formation: its anatomy, circuitry, and role in cognitive function. *Rev. Neurosci.* **13**, 183–194 (2002).
29. Moser, M. B. & Moser, E. I. Functional differentiation in the hippocampus. *Hippocampus* **8**, 608–619 (1998).
30. Bannerman, D. M. *et al.* Regional dissociations within the hippocampus—memory and anxiety. *Neurosci. Biobehav. Rev.* **28**, 273–283 (2004).
31. Voogd, J., Hess, D. & Marani, E. *The Parasagittal Zonation of the Cerebellar Cortex in Cat and Monkey: Topography, Distribution of Acetylcholinesterase, and Development* (ed. King, E. S.) (Liss, New York, 1987).
32. Herrup, K. & Kuemerle, B. The compartmentalization of the cerebellum. *Annu. Rev. Neurosci.* **20**, 61–90 (1997).
33. Hawkes, R. & Herrup, K. Aldolase C/zebrin II and the regionalization of the cerebellum. *J. Mol. Neurosci.* **6**, 147–158 (1995).
34. Gravel, C. & Hawkes, R. Parasagittal organization of the rat cerebellar cortex: direct comparison of Purkinje cell compartments and the organization of the spinocerebellar projection. *J. Comp. Neurol.* **291**, 79–102 (1990).
35. Blackshaw, S. & Snyder, S. H. Enkephalin: a novel mammalian extraretinal opsin discretely localized in the brain. *J. Neurosci.* **19**, 3681–3690 (1999).
36. Eberwine, J., Belt, B., Kacharmina, J. E. & Miyashiro, K. Analysis of subcellularly localized mRNAs using *in situ* hybridization, mRNA amplification, and expression profiling. *Neurochem. Res.* **27**, 1065–1077 (2002).
37. Kang, H. & Schuman, E. M. A requirement for local protein synthesis in neurotrophin-induced hippocampal synaptic plasticity. *Science* **273**, 1402–1406 (1996).
38. Huber, K. M., Kayser, M. S. & Bear, M. F. Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. *Science* **288**, 1254–1257 (2000).
39. Burgin, K. E. *et al.* *In situ* hybridization histochemistry of Ca²⁺/calmodulin-dependent protein kinase in developing rat brain. *J. Neurosci.* **10**, 1788–1798 (1990).
40. Trapp, B. D. *et al.* Spatial segregation of mRNA encoding myelin-specific proteins. *Proc. Natl Acad. Sci. USA* **84**, 7773–7777 (1987).
41. Kindler, S., Wang, H., Richter, D. & Tiedge, H. RNA transport and local control of translation. *Annu. Rev. Cell Dev. Biol.* **21**, 223–245 (2005).
42. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
43. Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320 (2002).
44. Vavouri, T. & Elgar, G. Prediction of cis-regulatory elements using binding site matrices—the successes, the failures and the reasons for both. *Curr. Opin. Genet. Dev.* **15**, 395–402 (2005).
45. Luquet, S., Perez, F. A., Hnasko, T. S. & Palmiter, R. D. NPY/AgRP neurons are essential for feeding in adult mice but can be ablated in neonates. *Science* **310**, 683–685 (2005).
46. Lechner, H. A., Lein, E. S. & Callaway, E. M. A genetic method for selective and quickly reversible silencing of mammalian neurons. *J. Neurosci.* **22**, 5287–5290 (2002).
47. Yoshihara, Y. *et al.* A genetic approach to visualization of multisynaptic neural pathways using plant lectin transgene. *Neuron* **22**, 33–41 (1999).
48. DeFalco, J. *et al.* Virus-assisted mapping of neural inputs to a feeding center in the hypothalamus. *Science* **291**, 2608–2613 (2001).
49. Brecht, M. *et al.* Novel approaches to monitor and manipulate single neurons *in vivo*. *J. Neurosci.* **24**, 9223–9227 (2004).
50. Ng, L. *et al.* Neuroinformatics for genome-wide 3-D gene expression mapping in the mouse brain. *IEEE Trans. Comput. Biol. Bioinform.* (in the press).
51. Thompson, P. M. & Toga, A. W. in *Handbook of Medical Imaging: Processing and Analysis* (ed. Bankman, I. N.) (Academic Press, San Diego, 2000).
52. Viola, P. & Wells, W. M. Alignment by maximization of mutual information. *Int. J. Comput. Vis.* **24**, 137–154 (1997).
53. Chan, T. F. & Shen, J. *Image Processing and Analysis: Variation, PDE, Wavelet, and Stochastic Methods* (Society for Industrial and Applied Mathematics, Philadelphia, 2005).
54. Lein, E. S., Callaway, E. M., Albright, T. D. & Gage, F. H. Redefining the boundaries of the hippocampal CA2 subfield in the mouse using gene expression and 3-dimensional reconstruction. *J. Comp. Neurol.* **485**, 1–10 (2005).
55. Swanson, L. W. *Brain Maps: Structure of the Rat Brain* (Elsevier, Amsterdam, 2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was sponsored by the Allen Institute for Brain Science. The authors wish to thank the Allen Institute founders, P. G. Allen and J. Patton, for their vision, encouragement and support. We also wish to thank key Institute advisors, K. Dooley and S. Coliton, as well as the Scientific Advisory Board for the Atlas project, M. Tessier-Lavigne, D. Anderson, C. Dulac, R. Gibbs, S. Paul, G. Schuler, A. W. Toga and J. Takahashi, for their scientific guidance and dedication to the successful execution of the Atlas project. We would particularly like to acknowledge D. Anderson for his role in the conceptual genesis and continual refinement of Atlas goals, as well as M. Tessier-Lavigne for key scientific and organizational leadership throughout the project. We also thank C. Jennings for his critical reading of the manuscript.

Author Contributions Neuroscience Group: E.S.L. (group leader), A.B., L.C., M.P.H., M.T.M., C.L.T., T.A.Z. Informatics Group: M.J.H. (group leader), C.L.K., C.L., L.L.N., S.D.P. Production Groups: Allen Institute for Brain Science: P.E.W. (group leader), S.M.S. (group leader), R.A.J. (group leader), M.A., A.F.B., E.J.B., S.D., N.R.D., A.L.D., T.D., E.D., M.J.D., J.G.D., A.J.E., L.K.E., S.R.F., S.N.G., K.J.G., K.R.H., M.R.H., J.M.K., R.H.K., J.H.L., T.A.L., L.T.L., R.J.M., N.F.M., R.N., G.J.O., T.H.P., S.E.P., O.C.P., R.B.P., Z.L.R., H.R.R., S.A.R., J.J.R., N.R.S., K.S., N.V.S., T.S., C.R.S., S.C.S., K.A.S., N.N.S., K.-R.S., L.R.V., R.M.W., C.K.W., V.Y.W., X.F.Y.; Baylor College of Medicine: C.T. (group leader), N.A., L.C. (Li Chen), T.-M.C., A.C., R.F., A.J.L., Y.L., M.J.R., A.T., M.W., M.B.Y., B.Z.; Max Planck Institute: G.E. (group leader), A.V. Technology Group: C.N.D. (group leader), C.D.T. (group leader), A.B. (Amy Bensinger), K.S.B., M.C.C., J.C., B.E.C., T.A.D., B.J.D., T.P.F., C.F. (Cliff Frensky), D.P.J., P.T.K., R.K., A.R.L., K.D.L., J.M., B.I.S., A.J.S., M.S., R.C.Y., B.L.Y. Other: H.-W.D., B.A.F., C.F. and J.J.M., Allen Reference Atlas generation; J.G.H., data annotation; C.C.O., critical review and manuscript preparation; M.S.B., overall project leadership 2003–2004; A.R.J., overall project leadership 2004–present.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.S.L. (edl@alleninstitute.org) or A.R.J. (allanj@alleninstitute.org).

Disclaimer The Nature Publishing Group has a business collaboration with the Allen Institute for the creation and maintenance of the Neuroscience gateway (<http://www.brainatlas.org>), but has no role in generating or curating the Allen Brain Atlas database content. As always, *Nature* Editors have been fully independent and solely responsible for the editorial content and peer review of this research article.