RESOURCE

nature neuroscience

An anatomic gene expression atlas of the adult mouse brain

Lydia Ng[1], Amy Bernard[1], Chris Lau[1], Caroline C Overly[1], Hong-Wei Dong[2], Chihchau Kuan[1], Sayan Pathak[1], Susan M Sunkin[1], Chinh Dang[1], Jason W Bohland[3], Hemant Bokil[3], Partha P Mitra[3], Luis Puelles[4,5], John Hohmann[1], David J Anderson[6], Ed S Lein[1], Allan R Jones[1] & Michael Hawrylycz[1]

Studying gene expression provides a powerful means of understanding structure-function relationships in the nervous system. The availability of genome-scale *in situ* hybridization datasets enables new possibilities for understanding brain organization based on gene expression patterns. The Anatomic Gene Expression Atlas (AGEA) is a new relational atlas revealing the genetic architecture of the adult C57Bl/6J mouse brain based on spatial correlations across expression data for thousands of genes in the Allen Brain Atlas (ABA). The AGEA includes three discovery tools for examining neuroanatomical relationships and boundaries: (1) three-dimensional expression-based correlation maps, (2) a hierarchical transcriptome-based parcellation of the brain and (3) a facility to retrieve from the ABA specific genes showing enriched expression in local correlated domains. The utility of this atlas is illustrated by analysis of genetic organization in the thalamus, striatum and cerebral cortex. The AGEA is a publicly accessible online computational tool integrated with the ABA (http://mouse.brain-map.org/agea).

The complexity of the mammalian brain makes it necessary to rely on maps and atlases to analyze and interpret observations effectively[1,2]. Modern neuroanatomic digital atlases[3–9] are based on multiple-modality datasets that include histology, immunohistochemistry, magnetic resonance imaging, positron emission tomography and three-dimensional (3D) reconstruction to describe structures, nuclei and connectivity. With the advent of large-scale gene expression profiling using microarrays and high-throughput *in situ* hybridization[10,11], it is now possible to build an anatomical brain atlas based on the transcriptome alone or using it as a key supporting modality.

The brain exhibits complex and combinatorial gene expression patterns with variations depending on its highly differentiated structure[12]. Expression patterns of individual genes in the cerebral cortex and other brain structures have been shown to highlight useful genetic markers for anatomic regions, boundaries, gradients and cell types[13]. Profiling of larger brain structures using laser-capture microdissection[14] and microarrays has suggested that gene expression patterns established during embryogenesis are largely retained in the adult and are important for regional specificity and for the functional connective relationships between brain regions[15]. Combinatorial gene expression patterns have been found to define a diversity of neural progenitor domains that yield particular functional components in the mature brain[16]. It follows that combinatorial gene expression characteristics should be reflected directly or indirectly in the neuroanatomic organization in the adult. Through the synthesis of *in silico* expression patterns across many genes in a spatially aligned dataset, an enhanced understanding of the relationships between genes expressed and structural and functional neuroanatomy may emerge.

We developed a gene expression–derived (gene-architectonic) atlas of the adult C57Bl/6J mouse brain as an open-access online resource for exploring the structural and functional organization of the brain in new ways. Based on over 4,000 gene expression profiles from *in situ* hybridization (ISH) data from the ABA (http://mouse.brain-map.org), the AGEA (http://mouse.brain-map.org/agea) allows users to navigate a transcriptionally based map of the brain and to interpret these results in the context of a conventional Nissl-based reference atlas, the *Allen Reference Atlas* (ARA)[5]. The AGEA resource consists of three components: browser-accessible 3D correlation maps based on average gene expression profiles; a hierarchical, transcriptome-based and navigable spatial ontology of the mouse brain; and a tool that allows users to retrieve lists of genes that exhibit localized enrichment and then directly access the original high-resolution ISH data in the ABA. The ARA and its ontology are also fully integrated with AGEA, allowing direct comparison with neuroanatomical and ontological content.

RESULTS

Our goals were to visualize the gene expression–derived spatial organization of the adult mouse brain based on the extensive *in situ* hybridization (ISH) dataset in the Allen Brain Atlas (ABA) and to present the resulting atlas (AGEA) as an interactive online resource,

[1]Allen Institute for Brain Science, Seattle, Washington 98103, USA. [2]Laboratory for NeuroImaging, University of California Los Angeles, Los Angeles, California 90095, USA. [3]Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. [4]Department of Human Anatomy and Psychobiology and [5]CIBER en Enfermedades Raras, U736, Faculty of Medicine, University of Murcia, Murcia, Spain. [6]Division of Biology, California Institute of Technology, Pasadena, California 91125, USA. Correspondence should be addressed to M.H. (mikeh@alleninstitute.org).

Received 6 November 2008; accepted 22 January 2009; published online 15 February; doi:10.1038/nn.2281

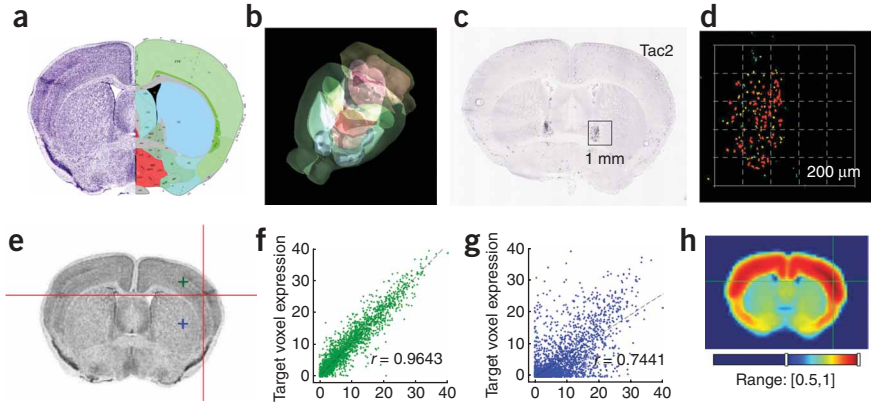

Figure 1 Construction and representation of the Anatomic Gene Expression Atlas (AGEA). (**a**) Level 53 coronal plate (bregma 0.145 mm) from the *The Allen Reference Atlas* (ARA) delineating 2D anatomic boundaries of a Nissl-stained mouse brain section. (**b**) 3D assembly of high-level ARA structures formed by 3D reconstruction of the Nissl sections. The 3D ARA space is partitioned into 200-μm^3 voxels forming the smallest spatial unit for analysis. (**c**) Coronal plane *in situ* hybridization (ISH) image of gene tachykinin 2 (*Tac2*) from the Allen Brain Atlas showing enriched expression in the bed nucleus of the stria terminalis (BST). The box represents a 1-mm^2 square. (**d**) Enlarged expression mask view of boxed area in **c** depicting gene expression levels color coded by ISH signal intensity (red, higher expression level; green/blue, lower expression level). Outer square corresponds to the box in **c**, and the internal dashed lines mark 200-μm subdivisions illustrating the level of resolution in the AGEA. (**e**) Seed voxel in somatosensory cortex in the Nissl atlas view (red cross-hair) and two distinct target voxels in cortex (green) and striatum (blue). (**f**) A scatter plot of expression level of each gene in the seed voxel (*x* axis) against the green cortical target voxel (*y* axis). The correlation of these gene expression pairs is 0.9643. (**g**) Similar scatter plot for the seed voxel and the blue striatal target voxel showing lower correlation of 0.7441. (**h**) 2D colorized coronal plane of a 3D AGEA correlation volume computed by correlating all voxels against the seed voxel. A scale bar enables users to change the dynamic range of the color map. This example shows strong correlation between the seed and other cortical areas and lower correlation with subcortical areas, consistent with **f** and **g**, respectively.

fully integrated with the ABA and the annotated ARA (**Supplementary Fig. 1** online). The AGEA was created by a correlation-based analysis of the expression patterns of all 4,376 genes examined in the coronal plane in the ABA. All ISH data were spatially registered to a common 3D atlas space derived from the ARA (**Fig. 1a**,**b**) using previously described image processing algorithms[12,17]. Applying a 3D grid to that reference atlas space subdivides the brain volume into 51,533 cubic grid cells, or voxels, each 200 μm on a side, to create discrete, positionally defined subunit volumes for spatially mapped gene expression analysis. Using ISH signal quantification algorithms (**Fig. 1c**,**d**)[12,17], a composite gene expression profile or 'transcriptome profile' can be derived for each voxel. This profile is represented by a vector in which each component contains the expression level for one of the 4,376 genes.

The AGEA consists of three functional tools each based on per-voxel transcriptome profile analysis: Correlation maps, Clusters analysis and Gene Finder. First, Correlation maps were created to present the degree of transcriptome profile similarity between a user-specified reference point of interest ('seed voxel') and other areas of the brain (**Fig. 1e–h**). A 3D correlation map was generated for each possible seed voxel (51,533 in total) by computing the Pearson correlation between transcriptome profile expression vectors for the reference seed voxel and each other voxel in the brain volume (**Fig. 1f**,**g**). For example, in **Figure 1e**, the seed voxel was chosen at the red cross-hairs. Gene expression values at this location are correlated with the corresponding values at a second cortical location indicated in green. A scatter plot of these 4,376 values is shown in **Figure 1f** having correlation 0.9643. The correlation of the seed expression values with a striatal location, in blue (**Fig. 1e**), is lower, 0.7441 (**Fig. 1g**). For a chosen seed, the complete set of resulting correlation coefficients are displayed simultaneously within a 3D brain volume as 24-bit color values (**Fig. 1h**) that can be thresholded for significance by the user (**Supplementary Fig. 2** online). Second, Clusters were generated by applying hierarchical clustering techniques to delineate apparent anatomical subdivisions comprised of voxels with similar composite gene expression profiles across all 4,376 genes. The cluster maps, which can be thresholded by a number of user-defined distinct correlation classes (**Supplementary Methods**), expose exclusively differential gene expression–driven subdivisions of the brain and may be useful for investigating conventional neuroanatomical boundaries and regional functional organization (**Supplementary Fig. 3** online). Finally, a Gene Finder search tool was created to informatically retrieve lists of genes that are locally enriched around user-defined points of interest and to provide a path from the AGEA directly to the original, high-resolution ISH data in the ABA (**Supplementary Fig. 4** online). In all three modes, users can navigate between the AGEA and the ARA and its ontology (**Supplementary Fig. 1**).

Using the AGEA: comparison of two thalamic nuclei

To illustrate the AGEA, two thalamic nuclei that differ in position, connections and function are considered here: the ventral posterior complex (VP) and the parafascicular nucleus (PF). VP is the primary relay nucleus for somatosensory information from the head and body to the cortex, and projects primarily to layer 4 of somatosensory cortex[18,19]. In contrast, PF has diverse projections to both cortical and subcortical structures—consistent with the role of the intralaminar nuclei, of which it is a part—as relays for various motor, limbic and somatosensory functions. Projections to striatum, putamen, olfactory tubercle, amygdala and layer 5 pyramidal cells in the cortex are known[20,21].

To examine VP (**Fig. 2**), a seed in VP was selected by navigating the red cross-hair cursor in the Nissl reference images (**Fig. 2a**, upper panels) to relevant bregma coordinates ((coronal, sagittal, horizontal) = (–1.855, 2.975, –1.575)). A regionally restricted red cloud reflecting tight correlation around the seed point, as well as one covering the corresponding region in the opposing hemisphere, is clearly evident (**Fig. 2a**, lower panels) and is consistent with the cytoarchitectonic delineation of VP. A pronounced lateral-to-medial correlation gradient across the thalamus is also observed, visualized as a spectrum of increasingly cooler colors. Examination of the Correlation map beyond the thalamus provides insight into the global organization of the transcriptome across the brain, including potential gene expression relationships between distant but synaptically connected brain regions. Clear correlation between the seed voxel in VP and cortical areas are observed (**Fig. 2a**, lower panel), with discrete zones of higher correlation in the middle layers of the somatosensory cortex and the retrosplenial cortex. A more intriguing correlation pattern is observed when the seed is placed in PF (−2.275, 3.375, −0.175). The red cloud is more medial than that seen for VP, reflecting a localized area of correlation

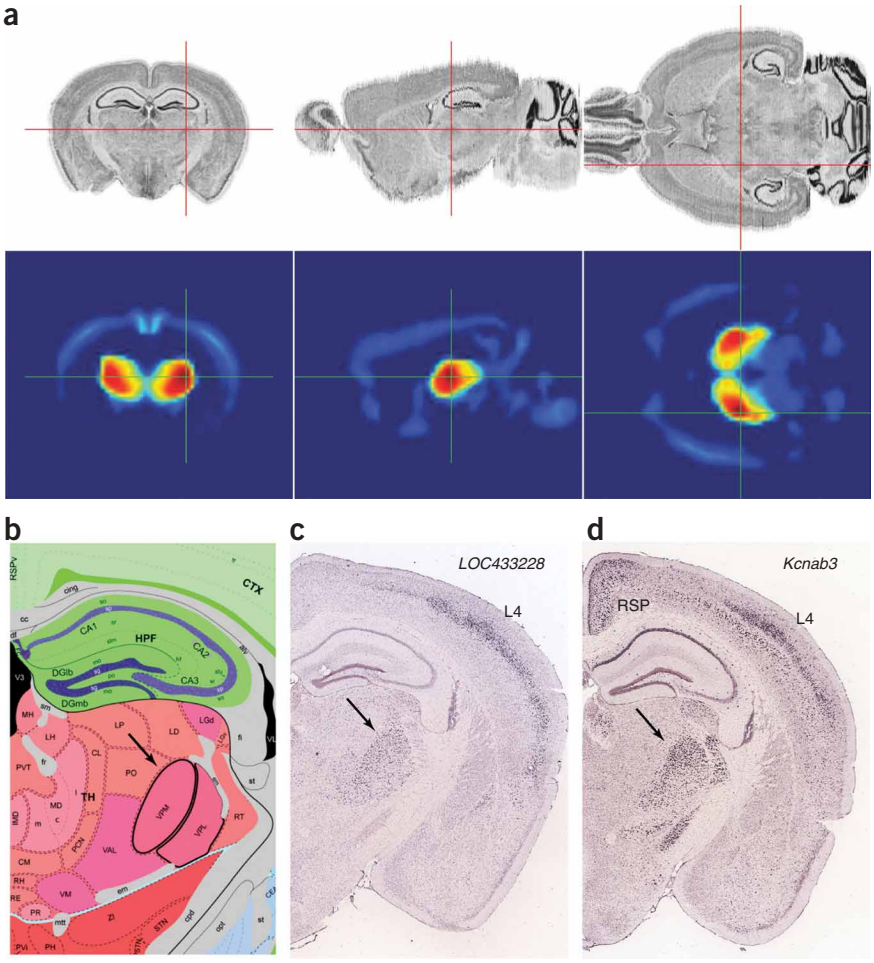

Figure 2 The ventral posterior complex (VP) of the thalamus. (**a**) Above, AGEA interface for a seed voxel located in VP (bregma −1.855, 2.975, −1.575), selected by moving the red cross-hairs on the 2D views of the 3D Nissl reference atlas displayed in the upper panels. Below, the 2D views of the 3D Correlation map showing the correlation of the transcriptome profile of the seed voxel with every other voxel. The computed correlation values are displayed as 24-bit false-color images using a blue-to-red color scale with the threshold interval set to [0.7, 1.0]. (**b**) *The Allen Reference Atlas* plate (coronal level 73, bregma –1.855 mm) delineating VP, comprising lateral (VPL) and medial (VPM) regions. (**c**,**d**) ISH images of two example genes returned with the Gene Finder search at the seed voxel used in **a**: hypothetical gene supported by AK082257 (*LOC433228*; **c**) and potassium voltage-gated channel, shaker-related subfamily, beta member (*Kcnab3*; **d**). Arrows in **b**–**d** indicate VP. L4, cerebral cortex layer 4; RSP, retrosplenial areas. See *The Allen Reference Atlas* or http://mouse.brain-map.org/atlas/index.html for other abbreviations in **b**.

defining PF in both hemispheres (**Fig. 3a**). In contrast to VP, the Correlation map for PF reveals that genes expressing in PF tend also to express with interlaminar thalamic nuclei and subcortical areas in the hypothalamus and midbrain, even more so than with the majority of the thalamus.

For more detailed analysis of these patterns, the AGEA Gene Finder retrieves genes exhibiting enrichment in the local area of the seed voxel and guides users to the original high-resolution ISH data. In this mode, the AGEA fixes a threshold that has high correlation in the local volume of a seed in which to search the ABA for genes with overlapping expression pattern (**Supplementary Methods**). Returned genes include those expressed in the seed-containing brain structure as well as in more distant correlated areas highlighted by the Correlation map. *LOC43328* (hypothetical gene supported by AK082257) and *Kcnab3* (encoding potassium voltage-gated channel, shaker-related subfamily, beta member) are highly expressed in VP and in correlated cortical areas (**Fig. 2b–d**). Similarly, *Ddc8* (encoding testis-specific protein, Dc8) and *Rprm* (encoding reprimo, TP53-dependent G2 arrest mediator candidate) are expressed in PF and in other correlated areas (**Fig. 3b–d**). These ISH gene expression patterns correlate well with the AGEA Correlation maps (**Figs. 2a** and **3a**).

Subdivisions of the striatum

Although at a gross structural level the striatum shows a degree of internal homogeneity, a focused analysis of transcriptome profile correlations provides support for a nonhomogeneous genetic architecture. Correlation maps based on a reference seed voxel in the caudoputamen reveal a high level of intrastructure transcriptome profile correlation at a threshold level that clearly distinguishes the striatum from other brain regions, consistent with classical anatomic delineation (**Fig. 4a**). This apparent homogeneity is also revealed in AGEA Clusters mode (Methods and **Supplementary Methods**). Visualized at a relatively low threshold level (Tree Depth 0), the striatal voxels exhibiting strong correlation in **Figure 4a** are grouped together as a natural cohesive whole that clearly includes the ventral striatum (**Fig. 4b**, blue-green) while remaining distinct from other, nonstriatal groupings of correlated voxels.

More detailed Cluster analysis of the striatal region (**Fig. 4c**) reveals a deeper organization. Using a higher threshold for intracluster correlation, further down the hierarchical tree of transcriptome profile similarity (Tree Depth 4), several striatal subdivisions emerge. These reflect a dorsal-ventral parcellation and support a 45° oblique dorsolateral-to-ventromedial banding pattern[22] (**Fig. 4c**, left) as well as a rostral-caudal gradient (**Fig. 4c**, right). Interestingly, these striatal subdivisions correlate with the estimated spatial distribution of the derivatives of the diverse striatal progenitor domains recently characterized molecularly by previous workers[16].

The AGEA's Gene Finder can be used to search for marker genes corroborating such structural delineations (**Fig. 5**). The caudoputamen (CP) exhibits a medial-dorsal region reflected by expression of *BC053994* (cDNA sequence BC053994) (**Fig. 5a**,**i**, dark blue). In contrast, *Ddit4l* (encoding DNA-damage-inducible transcript 4-like) is expressed in a ventral medial band (**Fig. 5b**,**i**, yellow), and *Ace* (angiotensin I converting enzyme (peptidyl-dipeptidase A) 1) transcripts appear in a dorsal lateral band (**Fig. 5c**,**i**, blue-green). The marker *Wfs1* (Wolfram syndrome 1 homolog, human) is present in the core of the nucleus accumbens and nearby ventro-medial band of CP (**Fig. 5d**,**i**, red and blue), whereas *Btg1* (B-cell translocation gene 1, anti-proliferative) expression is enriched only in the accumbens core domain (**Fig. 5e**,**i**, red). Moreover, *Wfs1* (**Fig. 5d**) and *Npy1r*

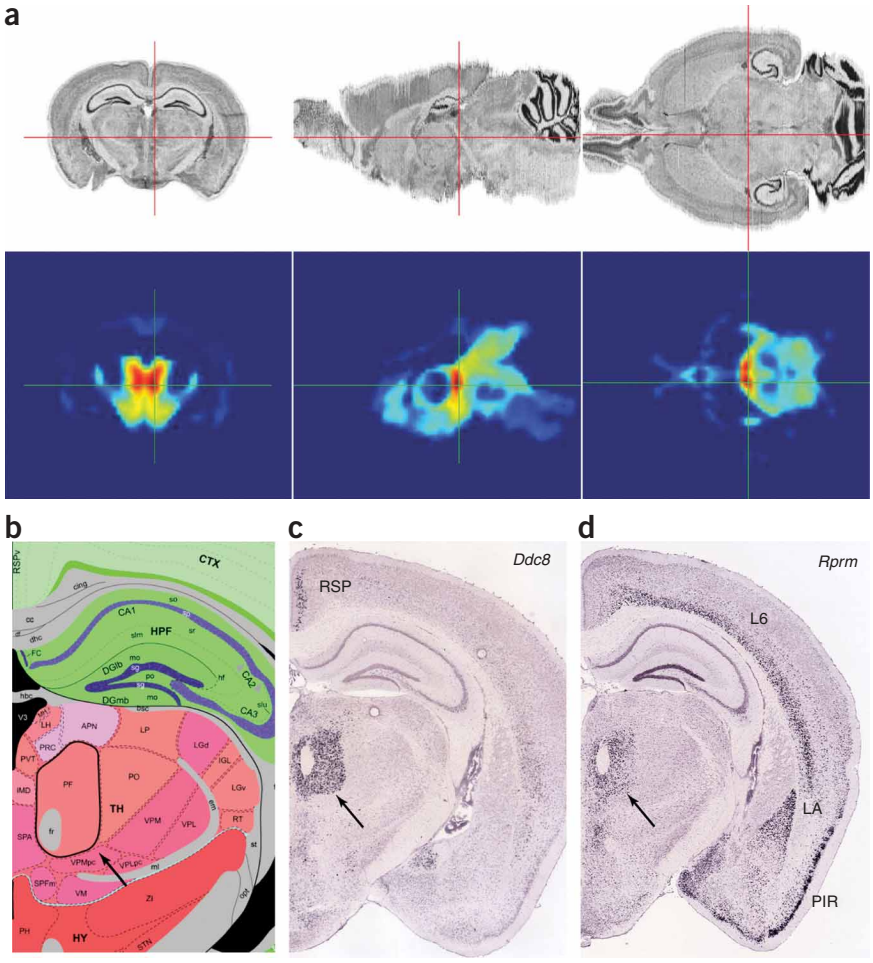

Figure 3 The parafascicular nucleus of the thalamus. (**a**) AGEA interface for a seed voxel located in PF (bregma –2.275, 3.375, –0.175), showing Nissl reference views (above) and the correlation map with a threshold range of [0.8, 1] (below). (**b**) *The Allen Reference Atlas* plate (coronal level 77, bregma –2.255 mm) delineating PF. (**c**,**d**) ISH images of two example genes returned with the Gene Finder search at the seed voxel used in **a**: testis-specific protein, Ddc8 (*Ddc8*; **c**), and reprimo, TP53-dependent G2 arrest–mediator candidate (*Rprm*; **d**). Arrows in **b**–**d** indicate PF. L6, cerebral cortex layer 6 (L6); RSP, retrosplenial areas; LA, lateral amygdalar nucleus; PIR, piriform area. See *The Allen Reference Atlas* or http://mouse.brain-map.org/atlas/index.html for other abbreviations in **b**.

(neuropeptide Y receptor Y1) (**Fig. 5f**) expression show, respectively, relative enrichment in the pyramidal (**Fig. 5i**, dark red) and polymorph (**Fig. 5i**, orange) layers of the olfactory tubercle. Expression of the genes *mCG145872* (ABmCG145872) (**Fig. 5g**) and *Thbs3* (thrombospondin 3) characterize differentially a ventriculo-pial gradient across the striatum in sagittal sections (**Fig. 5h**). Genes identified using this approach often represent novel local anatomic markers for these structures.

Gene expression delineates cortical topography

We investigated to what extent correlation of gene expression profiles reveals laminar and areal organization in the mouse neocortex (**Supplementary Fig. 5** online)[23,24]. Seed points were selected at radial intervals through the somatosensory cortex, based upon a Nissl-stained reference atlas (**Fig. 6a–e**). Analyzing Correlation maps at correlation range [0.89, 1.0] reveals neuroanatomical boundaries corresponding to dominant laminar architecture[25]. Genes expressed in superficial layers have a sharp decline in correlation as one examines deeper layers, but the transition is not smooth, and a distinction between layer 5 (L5) and layer 6 (L6) is evident (**Fig. 6**, column **a**). Similarly, the presence of expression in deeper layers diminishes the correlation with superficial layers, although there are laminar-specific effects. In particular, seeds in somatosensory L6 have comparably lower L4 correlation (**Fig. 6**, column **d**) than seeds in L2/3 (**Fig. 6**, column **b**). Intriguingly, a seed in somatosensory cortex layer 6b (**Fig. 6**, column **e**) results in a restricted cluster of gene expression, with closest correlation appearing in the adjacent L6 (but not throughout other layers) and expanding to neighboring parts of the claustrum. This suggests that genes expressed in L6b comprise a unique cluster and are less likely to be expressed throughout other cortical lamina (**Supplementary Fig. 6** online).

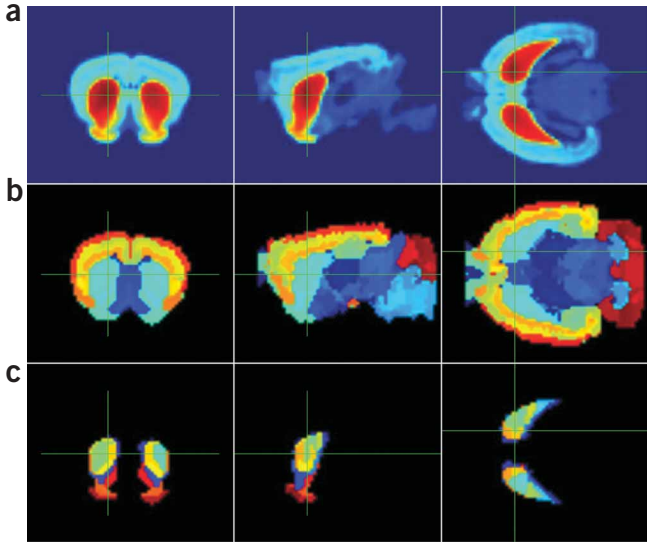

Figure 4 Genetic organization of the striatum. (**a**) Correlation map based on a seed voxel in the caudoputamen (bregma 0.7, 2.985, 1.55), with threshold 0.5938, showing a high level of intrastructure gene expression correlation across the striatum and ventral striatum and lower correlation with other brain regions. (**b**) Clusters mode view of a high-level, global clustering analysis (Tree Depth 0) showing the strongly correlated striatal voxels grouped together (blue-green area). (**c**) View of a deeper, more detailed cluster analysis (Tree Depth 4) of the striatal voxel group revealing several striatal subdivisions reflecting a dorsal-ventral parcellation and supporting a 45° dorsolateral-ventromedial banding pattern and a rostral-caudal gradient. All three parts of the figure show 2D views of the 3D volume, with coronal, sagittal and horizontal planes presented left to right.

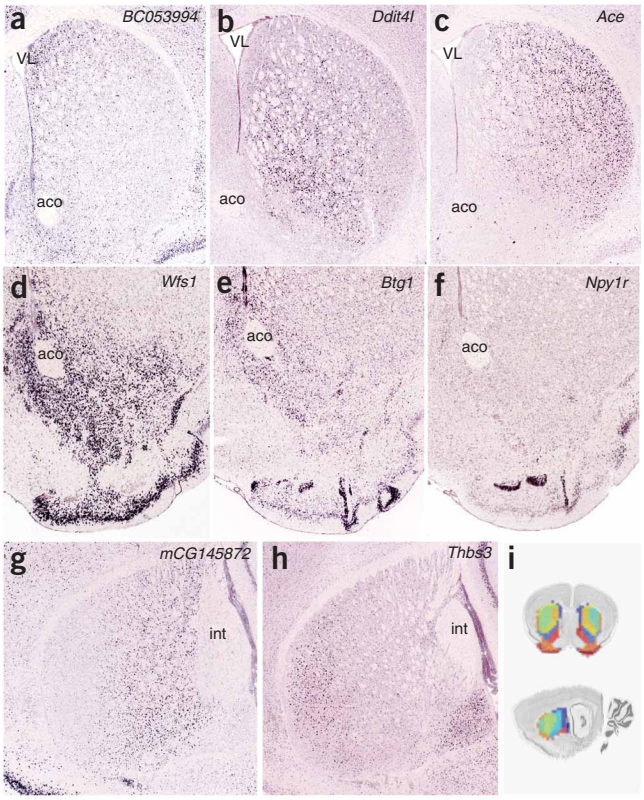

Figure 5 Anatomic gene markers for the dorsal-ventral striatum. (**a**–**f**) Coronal-plane and (**g**,**h**) sagittal-plane ISH images of the expression of genes reflecting AGEA striatal decomposition shown in **i** and **Figure 4c**. (**a**) cDNA sequence BC053994 (*BC053994*), medio-dorsal caudoputamen (CP) region (dark blue). (**b**) DNA-damage-inducible transcript 4-like (*Ddit4l*), ventro-medial CP band (yellow). (**c**) Angiotensin I–converting enzyme (peptidyl-dipeptidase A) 1 (*Ace*), dorsal-lateral CP band (blue-green). (**d**) Wolfram syndrome 1 homolog, human (*Wfs1*), nucleus accumbens (ACB) core and ventromedial CP (red, blue), plus pyramidal layer of olfactory tubercle (OT; dark red). (**e**) B-cell translocation gene 1, anti-proliferative (*Btg1*), ACB core (red). (**f**) Neuropeptide Y receptor Y1 (*Npy1r*), polymorph layers of OT (orange). (**g**,**h**) AB mCG145872 (*mCG145872*) and thrombospondin 3 (*Thbs3*), illustrating the ventriculo-pial gradient in sagittal sections. VL, lateral ventricle; aco, anterior commisure, olfactory limb (aco); int, internal capsule.

To quantitatively investigate cortical correlation structure, the 3D reference atlas (**Fig. 1b**) was used to circumscribe four intersecting layers and regions. The cortical regions chosen were the auditory (AUD), primary somatosensory (SSp), secondary somatosensory (SSs) and primary visual (VISp) areas, and we cross-correlated therein layers (L) 2/3, 4, 5 and 6 (**Supplementary Fig. 7** online). Each of the four specified regions transects the layers, and we refer to the intersection of a given layer and area as a 'domain'. For a given domain, we compute the mean of all Correlation maps having seeds in that domain (**Supplementary Table 1** online) and examine mean correlation structure with the 15 other domains. A convenient representation for this data is shown graphically in **Figure 6f**, whose mean layer-region correlations are summarized in each 4 × 4 submatrix. For example, data in the upper left 4 × 4 submatrix (**Fig. 6f**) represent average correlation values in each domain for seeds in the L2/3-AUD cortex. The dark blue square in this 4 × 4 submatrix corresponds to L6-SSp, indicating the lowest correlation with seeds in L2/3-AUD occurs in this region.

The representation in **Figure 6f** allows a rich interpretation of the relative correlations across layers and regions. The mean correlation is highest, as expected, in the domain containing the seed (to within statistical significance). We may pool data from **Figure 6f** according to dominant area (columns) or layer (rows) to show that adjacent layers

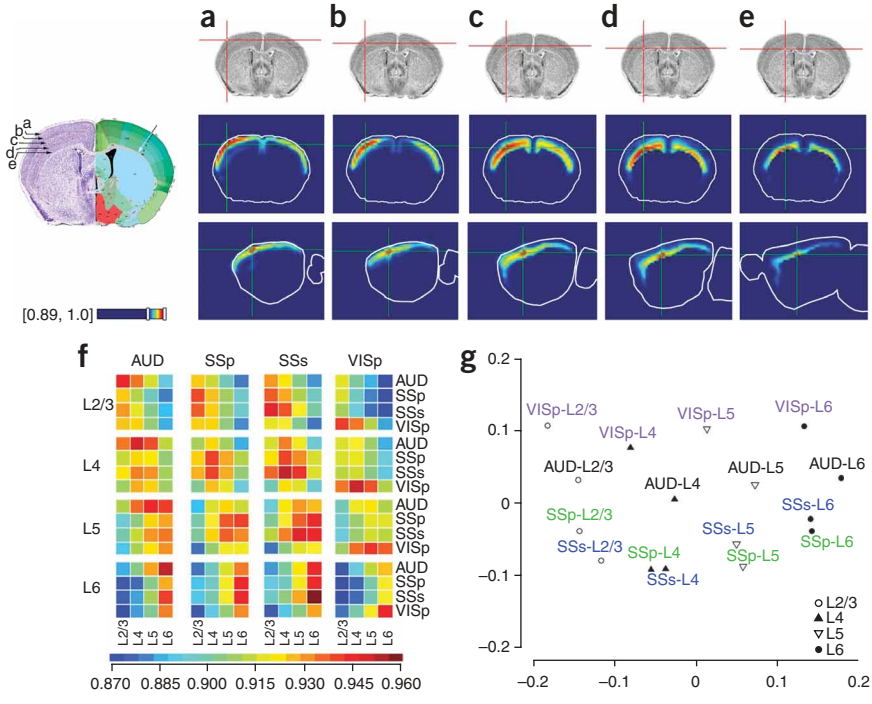

Figure 6 AGEA and cortical topography. (**a**–**e**) Seed points selected on a Nissl-stained reference section shown in upper left. Selected seeds are shown at red cross-hairs from somatosensory cortex layers 1/2 (**a**), layers 2/3 and 4 (**b**), layer 5 (**c**), layer 6 (**d**) and layer 6b (**e**). The corresponding AGEA map for each seed is shown in coronal (row 2) and sagittal (row 3) views. Threshold is restricted to [0.89, 1.0]. (**f**) Each 4 × 4 sub-matrix is a heat-map representation of average layer (L2/3, L4, L5 and L6) and area (AUD, SSp, SSs and VISp) correlations for seeds in the intersection of layer and area labeled on left and top, respectively. (**g**) Multidimensional scaling (MDS) embedding of domain (layer intersect area) correlations in cortex (**Supplementary Methods**). The axes represent implied relative correlation distances based on correlation from **Figure 6f**. In MDS clustering, the relative correlation distance relationships between domains is captured, whereas the absolute scale of distances is not directly interpretable[31]. Domain point symbols are coded for cortical layer, and associated labels are color coded for cortical region. AUD, auditory; SSp, primary somatosensory; SSs, secondary somatosensory; VISp, primary visual; L2/3, layers 2/3; L4, layer 4; L5, layer 5; L6, layer 6. All correlation data of **Figure 6f** are available in **Supplementary Table 7**, including means and variances of correlations and their normally transformed variates by Fisher ρ-to-z transformation[32].

have positive expression correlation, with the strongest concordance between L5 and L6 (**Supplementary Table 2** and **Supplementary Fig. 8** online). For non-adjacent layers, there is a negative correlation that inversely correlates with anatomic proximity, indicating that physically distant layers are less likely to exhibit gene coexpression.

If domain-to-domain correlations are regarded as a measure of similarity, the data in **Figure 6f** may be visualized by multidimensional scaling (MDS)[26] (**Fig. 6g**). This clustering method uses values from pairwise domain correlations (from **Fig. 6f**) to graphically represent relative distance relationships, where distance between points (domains) is proportional to their correlation (that is, shorter distance reflects greater similarity). MDS analysis of gene expression recapitulates the basic laminar and areal relationships of the neocortex. The horizontal-vertical spread in **Figure 6g** is striking, giving a relative measurement of laminar versus areal gene expression in the brain. The proximal and functional relationship of SSp and SSs is also evident, as is the lower concordance of VISp with other regions.

DISCUSSION

High-throughput data production and large-scale data digitization and storage, together with robust image processing algorithms, have enabled the creation of a novel, *in silico*, transcriptome-based atlas of the adult C57Bl/6J mouse brain. The AGEA provides rich information from mesoscale coexpression between brain regions and a suite of data-mining tools, enabling researchers to gain insight into how brain areas are delineated by their underlying constellated gene expression constitution. Although the primary focus of the work described here is on anatomic delineations that can be uncovered from transcription comparisons, such a resource should also be valuable for studying the organization of the genome. Further insight into the underlying biochemistry of various brain regions, driven by the expression of genes in specific functional classes, should contribute to a deeper understanding of the organization of both the brain and the genome.

The analyses presented here highlight the discovery potential of the AGEA, illustrating opportunities not only to investigate regional boundaries and how a particular brain area is organized with respect to gene expression but also to elucidate similarities between a local region of interest and more distant areas of the brain. For example, these data demonstrate that the AGEA enables discovery and characterization of gene expression relationships between synaptically connected brain regions. The Correlation map for VP shows marked transcriptome profile correlation between VP and a discrete domain in the middle layers of the somatosensory cortex (**Fig. 2a**), consistent with the known termination of VP projection neurons in layer 4 of the somatosensory cortex[27]. Use of Gene Finder provides additional detail, identifying specific VP-enriched genes expressed in both domains (**Fig. 2c**,**d**).

Additionally, the AGEA has potential to circumscribe domains not well addressed by traditional atlas-based delineation, such as neurogenic domains. The rostral migratory stream (RMS)[28] can be distinguished for the first time as a coherent structure, consistent with numerous anatomical models of the RMS[29], also based upon gene expression (**Supplementary Fig. 9** online). Gene ontology (GO) analysis of returned genes using Gene Finder reveals a profile enriched in cell differentiation, cell migration, CNS development, neurogenesis, and other differentiation and development processes (**Supplementary Fig. 10** and **Supplementary Tables 3** and **4** online). With further software development, it is also entirely feasible to produce AGEA models using functional subclasses of genes such as transcription factors, ion channels or G protein–coupled receptors. Additionally, a multitude of more sophisticated statistical techniques could be applied in AGEA analysis.

To maximize spatial coverage, registration integrity and correlation data quality, the AGEA was developed based on the coronal ISH image series in the Allen Brain Atlas rather than the full genomic sagittal image data. Because of tissue-sectioning requirements, however, the anatomic mapping resolution of the AGEA remains restricted to a minimum 3D grid voxel size of 200 μm on a side. An argument may be made that this grid voxel size is too large to distinguish smaller nuclei in the mouse brain and that volume averaging of correlation will reduce the value of the resulting atlas. Although there is limitation in spatial resolution based on the present mapping accuracy, many interesting domains can be resolved at this resolution, as we have shown in the thalamus, striatum and cortex. More importantly, access to the original ISH image data in the ABA via Gene Finder provides users a seamlessly integrated resource for higher-resolution analysis.

Selection of a gene for coronal mode in the ABA was made based on classes of known neuroscientific interest, such as ion channels and G protein–coupled receptors, or through *post hoc* identification of a marked non-ubiquitous expression pattern as a part of the ABA project. There was no quantitative rule based either on gene class function, specificity of pattern, coefficient of variation or any other parameter to determine which genes were made available in coronal plane, although each gene in the mouse genome was considered. Ultimately, our goal for the AGEA was to illustrate the fundamental power of the pooled genetic approach in relation to neuroanatomy. Although a larger dataset could be used, it is salient that, even using less than the entire genome, the basic anatomy of the brain is recapitulated.

There is increasing interest in multimodal atlases, and several digital atlases of the C57Bl/6J mouse brain (adult and neonate) exist, providing frameworks for storing and accessing multiple types of information[8,28,29]. As large genomic-related datasets in neuroscience are increasingly being made publicly available, the potential for cross-modality neuroanatomic atlases increases. The integration and cross-connectivity of these large scale databases will continue to drive the quest for more potent computational approaches, undoubtedly favoring the emergence of new insights on the brain[1,30]. It is our hope that the present approach will complement these methods, and that with the availability of large-scale spatial gene expression databases, the AGEA becomes a key addition to the available tools of the neuroscientist.

METHODS

Image informatics. Each ISH image is processed through an automated pipeline that identifies signal and reconstructs each image series as a 3D volume while registering that volume to a Nissl series of the ARA using methods of morphological image processing and signal detection[12,17,30]. Details on expression statistics, correlation analysis and spatial clustering are in the **Supplementary Methods** online.

AGEA dataset. The AGEA dataset consists of all 4,376 genes of the coronal plane ABA. These genes show marked regional expression patterns in the C57Bl/6J mouse brain. The AGEA dataset generally lacks genes with widespread expression such as those represented by various metabolic, biosynthetic, and basal or homeostatic gene functions[11]. There is also some selection bias toward genes enriched for cortical and/or hippocampal regional expression patterns, owing to our specific research interest in these areas. A gene ontology analysis for the AGEA gene set reveals strong enrichment in categories related to neuronal cell process and function, GABA receptor activity, ion and potassium channel binding, neuron differentiation, axon guidance, synaptic transmission and long-term potentiation (see **Supplementary Tables 5** and **6** online).

AGEA application. The 200-μm 3D grid consists of 67 coronal × 41 horizontal × 58 sagittal voxels (sum = 159,326), of which 51,533 voxels are masked to intersect the ABA data. For each voxel, a 3D correlation map is computed,

resulting in an equivalent 51,533 correlation maps, stored as raw byte images and consisting of approximately 31 gigabytes of online data. The AGEA application is multiuser and has no plug-ins. It is fully linked with the ARA and ontology and does two-dimensional (2D) slicing of volumes and threshold computation at the time of user access.

All ABA image data, ARA plates, expression statistics, AGEA 3D correlation volumes and associated metadata are available for online public access. Specific URLs for the manuscript figures are provided in the **Supplementary Results** online. All correlation data of **Figure 6f** are available in **Supplementary Table 7** online.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

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. The authors also thank D. Haynor of the University of Washington, Department of Neuroradiology, and C. Thompson of the Allen Institute.

AUTHOR CONTRIBUTIONS

L.N., CL. and C.D. built the AGEA application; LN., A.B, J.W.B., H.B., P.P.M. and M.H. performed the analyses; H.-W.D., L.P. and J.H. interpreted neuroanatomy; L.K. and S.P. provided informatics support; E.S.L., D.J.A., A.J.R., M.H. provided overall guidance; and L.N., C.C.O., A.B., S.M.S. and M.H. wrote the manuscript.

Published online at http://www.nature.com/natureneuroscience/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. MacKenzie-Graham, A. *et al.* The informatics of a C57BL/6J mouse brain atlas. *Neuroinformatics* **1**, 397–410 (2003).
2. Toga, A.W. & Thompson, P.M. in *Brain Mapping: The Systems* (eds. Toga, A.W. & Mazziotta, J.C.): 4–5 (Academic Press, San Diego, 2000).
3. Swanson, L.W. *Structure of the Rat Brain* (Academic Press, San Diego, 2004).
4. Paxinos, G. & Franklin, K.B.J. *The Mouse Brain in Stereotaxic Coordinates* (Academic Press, San Diego, 2004).
5. Dong, H.W. *The Allen Reference Atlas: A Digital Color Brain Atlas of the C57BL/6J Male Mouse* (John Wiley & Sons, Hoboken, New Jersey, USA, 2008).
6. Ma, Y. *et al.* A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. *Neuroscience* **135**, 1203–1215 (2005).
7. Toga, A.W., Thompson, P.M., Mori, S., Amunts, K. & Zilles, K. Towards multimodal atlases of the human brain. *Nat. Rev. Neurosci.* **7**, 952–966 (2006).
8. MacKenzie-Graham, A. *et al.* A multimodal, multidimensional atlas of the C57BL/6J mouse brain. *J. Anat.* **204**, 93–102 (2004).
9. Van Essen, D.C., Drury, H.A., Joshi, S. & Miller, M.I. Functional and structural mapping of human cerebral cortex: solutions are in the surfaces. *Proc. Natl. Acad. Sci. USA* **95**, 788–795 (1998).
10. Gee, C.E. & Roberts, J.L. In situ hybridization histochemistry: a technique for the study of gene expression in single cells. *DNA* **2**, 157–163 (1983).
11. 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).
12. Lein, E.S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
13. Sunkin, S.M. & Hohmann, J. G. Insights from spatially mapped gene expression in the mouse brain. *Hum. Mol. Genet.* **16** (Spec. No. 2): R209–R219 (2007).
14. Bonner, R.F. *et al.* Laser capture microdissection: molecular analysis of tissue. *Science* **278**, 1481–1483 (1997).
15. Zapala, M.A. *et al.* Adult mouse brain gene expression patterns bear an embryologic imprint. *Proc. Natl. Acad. Sci. USA* **102**, 10357–10362 (2005).
16. Flames, N. *et al.* Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J. Neurosci.* **27**, 9682–9695 (2007).
17. Ng, L.L. *et al.* Neuroinformatics for genome-wide 3D gene expression mapping in the mouse brain. *IEEE Trans. Comput. Biol. Bioinformatics* **4**, 382–393 (2007).
18. Price, J.L. in *The Rat Nervous System* (ed. Paxinos, G.) 632–633 (Academic Press, Sydney, 1995).
19. Sherman, S.M. & Guillery, R.W. *Exploring the Thalamus and its Role in Cortical Function* (MIT Press, Cambridge, Massachusetts, USA, 2006).
20. Nakamura, Y., Otake, K. & Tokuno, H. The parafascicular nucleus relays spinal inputs to the striatum: an electron microscope study in the rat. *Neurosci. Res.* **56**, 73–79 (2006).
21. Marini, G., Pianca, L. & Tredici, G. Thalamocortical projection from the parafascicular nucleus to layer V pyramidal cells in frontal and cingulate areas of the rat. *Neurosci. Lett.* **203**, 81–84 (1996).
22. Voorn, P., Vanderschuren, L.J., Groenewegen, H.J., Robbins, T.W. & Pennartz, C.M. Putting a spin on the dorsal-ventral divide of the striatum. *Trends Neurosci.* **27**, 468–474 (2004).
23. Brodmann, K. in *Some Papers on the Cerebral Cortex* [originally published 1909; translated as On the comparative localization of the cortex] 201–230 (Thomas, Springfield, Illinois, USA, 1960).
24. Yamamori, T. & Rockland, K.S. Neocortical areas, layers, connections, and gene expression. *Neurosci. Res.* **55**, 11–27 (2006).
25. Mountcastle, V.B. *Perceptual Neuroscience: The Cerebral Cortex* (Harvard University Press, Cambridge, Massachusetts, USA, 1998).
26. Kruskal, J.B. Non-metric multidimensional scaling: a numerical method. *Psychometrika* **29**, 115–129 (1964).
27. Paxinos, G. *The Rat Nervous System* (Academic Press, New York, 1995).
28. Alvarez-Buylla, A. & Garcia-Verdugo, J.M. Neurogenesis in adult subventricular zone. *J. Neurosci.* **22**, 629–634 (2002).
29. Abrous, D.N., Koehl, M. & Le Moal, M. Adult neurogenesis: from precursors to network and physiology. *Physiol. Rev.* **85**, 523–569 (2005).
30. Yushkevich, P.A. *et al.* Using MRI to build a 3D reference atlas of the mouse brain from histological images. *Proc. Intl. Soc. Magn. Reson. Med. Annu. Mtg.* **13**, 2809 (2005).
31. Kruskal, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27 (1964).
32. DeGroot, M.H. & Schervish, M.J. *Probability and Statistics* (Addison-Wesley, Boston, 2002).