ARTICLE

doi:10.1038/nature11405

An anatomically comprehensive atlas of the adult human brain transcriptome

Michael J. Hawrylycz[1]*, Ed S. Lein[1]*, Angela L. Guillozet-Bongaarts[1], Elaine H. Shen[1], Lydia Ng[1], Jeremy A. Miller[1], Louie N. van de Lagemaat[2], Kimberly A. Smith[1], Amanda Ebbert[1], Zackery L. Riley[1], Chris Abajian[1], Christian F. Beckmann[3], Amy Bernard[1], Darren Bertagnolli[1], Andrew F. Boe[1], Preston M. Cartagena[4], M. Mallar Chakravarty[1,5], Mike Chapin[1], Jimmy Chong[1], Rachel A. Dalley[1], Barry David Daly[6], Chinh Dang[1], Suvro Datta[1], Nick Dee[1], Tim A. Dolbeare[1], Vance Faber[1], David Feng[1], David R. Fowler[7], Jeff Goldy[1], Benjamin W. Gregor[1], Zeb Haradon[1], David R. Haynor[8], John G. Hohmann[1], Steve Horvath[9], Robert E. Howard[1], Andreas Jeromin[10], Jayson M. Jochim[1], Marty Kinnunen[1], Christopher Lau[1], Evan T. Lazarz[1], Changkyu Lee[1], Tracy A. Lemon[1], Ling Li[11], Yang Li[1], John A. Morris[1], Caroline C. Overly[1], Patrick D. Parker[1], Sheana E. Parry[1], Melissa Reding[1], Joshua J. Royall[1], Jay Schulkin[12], Pedro Adolfo Sequeira[13], Clifford R. Slaughterbeck[1], Simon C. Smith[14], Andy J. Sodt[1], Susan M. Sunkin[1], Beryl E. Swanson[1], Marquis P. Vawter[13], Derric Williams[1], Paul Wohnoutka[1], H. Ronald Zielke[15], Daniel H. Geschwind[16], Patrick R. Hof[17], Stephen M. Smith[18], Christof Koch[1,19], Seth G. N. Grant[2] & Allan R. Jones[1]

Neuroanatomically precise, genome-wide maps of transcript distributions are critical resources to complement genomic sequence data and to correlate functional and genetic brain architecture. Here we describe the generation and analysis of a transcriptional atlas of the adult human brain, comprising extensive histological analysis and comprehensive microarray profiling of ~900 neuroanatomically precise subdivisions in two individuals. Transcriptional regulation varies enormously by anatomical location, with different regions and their constituent cell types displaying robust molecular signatures that are highly conserved between individuals. Analysis of differential gene expression and gene co-expression relationships demonstrates that brain-wide variation strongly reflects the distributions of major cell classes such as neurons, oligodendrocytes, astrocytes and microglia. Local neighbourhood relationships between fine anatomical subdivisions are associated with discrete neuronal subtypes and genes involved with synaptic transmission. The neocortex displays a relatively homogeneous transcriptional pattern, but with distinct features associated selectively with primary sensorimotor cortices and with enriched frontal lobe expression. Notably, the spatial topography of the neocortex is strongly reflected in its molecular topography—the closer two cortical regions, the more similar their transcriptomes. This freely accessible online data resource forms a high-resolution transcriptional baseline for neurogenetic studies of normal and abnormal human brain function.

The enormous complexity of the human brain is a function of its precise circuitry, its structural and cellular diversity, and, ultimately, the regulation of its underlying transcriptome. In rodents, brain- and transcriptome-wide, cellular-resolution maps of transcript distributions are widely useful resources to complement genomic sequence data[1–3]. However, owing to the challenges of a 1,000-fold increase in size from mouse to human, limitations in post-mortem tissue availability and quality, and the destructive nature of molecular assays, there has been no human counterpart so far. Several important recent studies have begun to analyse transcriptional dynamics during human brain development[4,5], although only in a small number of relatively coarse brain regions. Characterizing the complete transcriptional architecture of the human brain will provide important information for understanding the impact of genetic disorders on different brain regions and functional circuits. Furthermore, conservation and divergence in brain function between humans and other species provide essential information for the understanding of drug action, which is often poorly conserved across species[6].

The goal of the Allen Human Brain Atlas is to create a comprehensive map of transcript usage across the entire adult brain, with the emphasis on anatomically complete coverage at a fine nuclear resolution in a small number of high-quality, clinically unremarkable brains profiled with DNA microarrays for quantitative gene-level transcriptome coverage. Furthermore, structural brain imaging data were obtained from each individual to visualize gene expression data in its native three-dimensional anatomical coordinate space, and to allow correlations between imaging and transcriptome modalities. These data are freely accessible via the Allen Brain Atlas data portal (http://www.brain-map.org).

[1]Allen Institute for Brain Science, Seattle, Washington 98103, USA. [2]Genes to Cognition Programme, Edinburgh University, Edinburgh EH16 4SB, UK. [3]MIRA Institute, University of Twente & Donders Institute, Radboud University Nijmegen, Nijmegen, Netherlands. [4]Department of Psychiatry & Human Behavior, University of California, Irvine, California 92697, USA. [5]Kimel Family Translational Imaging-Genetics Laboratory, Centre for Addiction and Mental Health Toronto, Ontario M5S 2S1, Canada. [6]University of Maryland School of Medicine, Department of Diagnostic Radiology, University of Maryland Medical Center, Baltimore, Maryland 21201, USA. [7]Department of Pathology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA. [8]Department of Radiology, University of Washington, Seattle, Washington 98195, USA. [9]Department of Human Genetics, Gonda Research Center, David Geffen School of Medicine, Los Angeles, California 90095, USA. [10]Banyan Biomarkers, Inc., Alachua, Florida 32615, USA. [11]Office of the Chief Medical Examiner, Baltimore, MD, Department of Pediatrics, University of Maryland, Baltimore, Maryland 21201, USA. [12]Department of Neuroscience, Georgetown University, School of Medicine, Washington DC 20007, USA. [13]Functional Genomics Laboratory, Department of Psychiatry & Human Behavior, School of Medicine, University of California, Irvine, California 92697, USA. [14]Histion LLC, Everett, Washington 98204, USA. [15]The Eunice Kennedy Shriver NICHD Brain and Tissue Bank for Developmental Disorders, University of Maryland, Baltimore, Maryland 21201, USA. [16]Program in Neurogenetics, Department of Neurology and Department of Human Genetics, and Semel Institute, David Geffen School of Medicine-UCLA, Los Angeles, California 90095, USA. [17]Fishberg Department of Neuroscience and Friedman Brain Institute, Mount Sinai School of Medicine, New York, New York 10029, USA. [18]FMRIB, Oxford University, Oxford OX3 9DU, UK. [19]Computation & Neural Systems, California Institute of Technology, Pasadena, California 91125, USA.

*These authors contributed equally to this work.

Global mapping of transcript distributions

A tissue processing and data collection pipeline was established to image the brain and subsequently dissect tissue samples from approximately 900 anatomically defined sites for RNA isolation and microarray analysis (Fig. 1 and Supplementary Methods 1). Two complete normal male brains were analysed from donors aged 24 and 39 years and are referred to here as Brain 1 and Brain 2 (Supplementary Table 1). Briefly, cooled brains underwent *in cranio* magnetic resonance imaging (MRI) followed by embedding, slabbing and freezing. Whole-brain cryosections were made from each slab, after which the slabs were subdivided and sectioned on 2 × 3 inch slides for histological analysis with Nissl and other markers for structure identification. Defined brain regions were isolated either using macrodissection (cortical gyri, other large structures) or laser microdissection (LMD; Leica LMD6000, Leica Microsystems) from tissue sections on polyethylene naphthalate (PEN) membrane slides (Leica Microsystems). Any given anatomical structure was first identified on the basis of histological data, and then sampled in a series of contiguous coronal slabs in both hemispheres. RNA was isolated from each sample and used to generate labelled cRNA probes for hybridization to custom 64K Agilent microarrays. The output of this pipeline was a set of microarrays that sample the entire spatial extent of neocortical gyri that could be reproducibly identified across individuals, as well as subcortical nuclear structures, at the resolution allowed by Nissl staining and sample size requirements for microarray analysis. One-hundred and seventy distinct structures were assayed at least once in both brains, and 146 structures twice or more (Supplementary Table 2). Sample locations were mapped back into the native brain MRI coordinates and subsequently to Montreal Neurological Institute (MNI) coordinate space[7].

These microarray data form the foundation for a publicly accessible online atlas, which includes viewers for microarray data visualization and mining, MRI/histology/sample location, and three-dimensional (3D) visualization of MRI and gene expression. To complement and validate the microarray data, several targeted, large-scale *in situ* hybridization (ISH) data sets were generated using a high-throughput ISH platform[1,8]. All of these data are linked with the other databases available via the Allen Brain Atlas data portal (http://www.brain-map.org) to facilitate comparative analyses with developing and adult mouse, rhesus macaque and human.

The output of the data generation pipeline described above is a detailed quantitative map of transcript distribution across the entire brain. As one example, Fig. 2a depicts the structural distribution of gene expression related to dopaminergic neurotransmission, illustrating the highly localized enrichment of genes associated with dopamine synthesis, packaging, degradation and postsynaptic signalling. Regional enrichments were conserved between the two brains (note similar peaks in paired rows for Brain 1 and 2; Fig. 2a) and were consistent with previous studies[9,10]. For example, tyrosine hydroxylase (*TH*) is enriched in the substantia nigra pars compacta (SNC), ventral tegmental area (VTA), and hypothalamic supraoptic and preoptic nuclei, as well as in the locus ceruleus, the neurons of which use dopamine as a precursor for noradrenaline. Similar brain-wide plots for other neurotransmitter systems are provided in Supplementary Fig. 1.

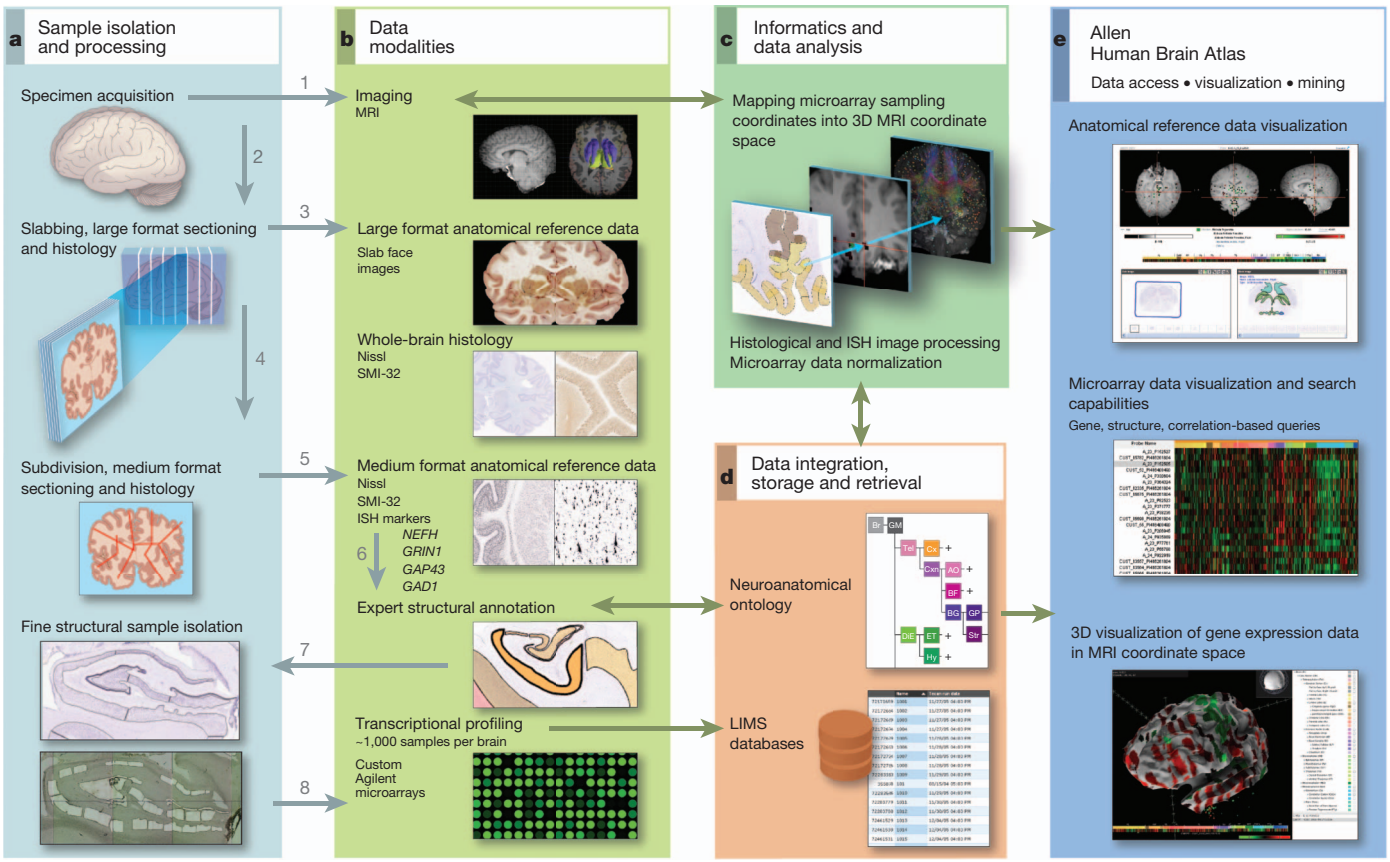

Figure 1 | Data generation and analysis pipeline. **a**, Experimental strategy to subdivide intact brains and isolate precise anatomical samples. **b**, Anatomical reference data are collected at each stage, including whole-brain MRI, large-format slab face and histology, medium (2 × 3-inch slide) format Nissl histology and ISH, and images of dissections. In Brain 2, labelling was performed for additional markers as shown. Histology data are used to identify structures, which are assembled into a database using a formal neuroanatomical ontology (**d**), and to guide laser microdissection of samples (**a**, lower panel). Isolated RNA is used for microarray profiling of ~900 samples per brain (**b**, lower panel). **c**, Microarray data are normalized and sample coordinates mapped to native 3D MRI coordinates. **e**, Data visualization and mining tools underlie the online public data resource. Numbers in **a** and **b** denote the order of sample processing steps leading to microarray data generation.

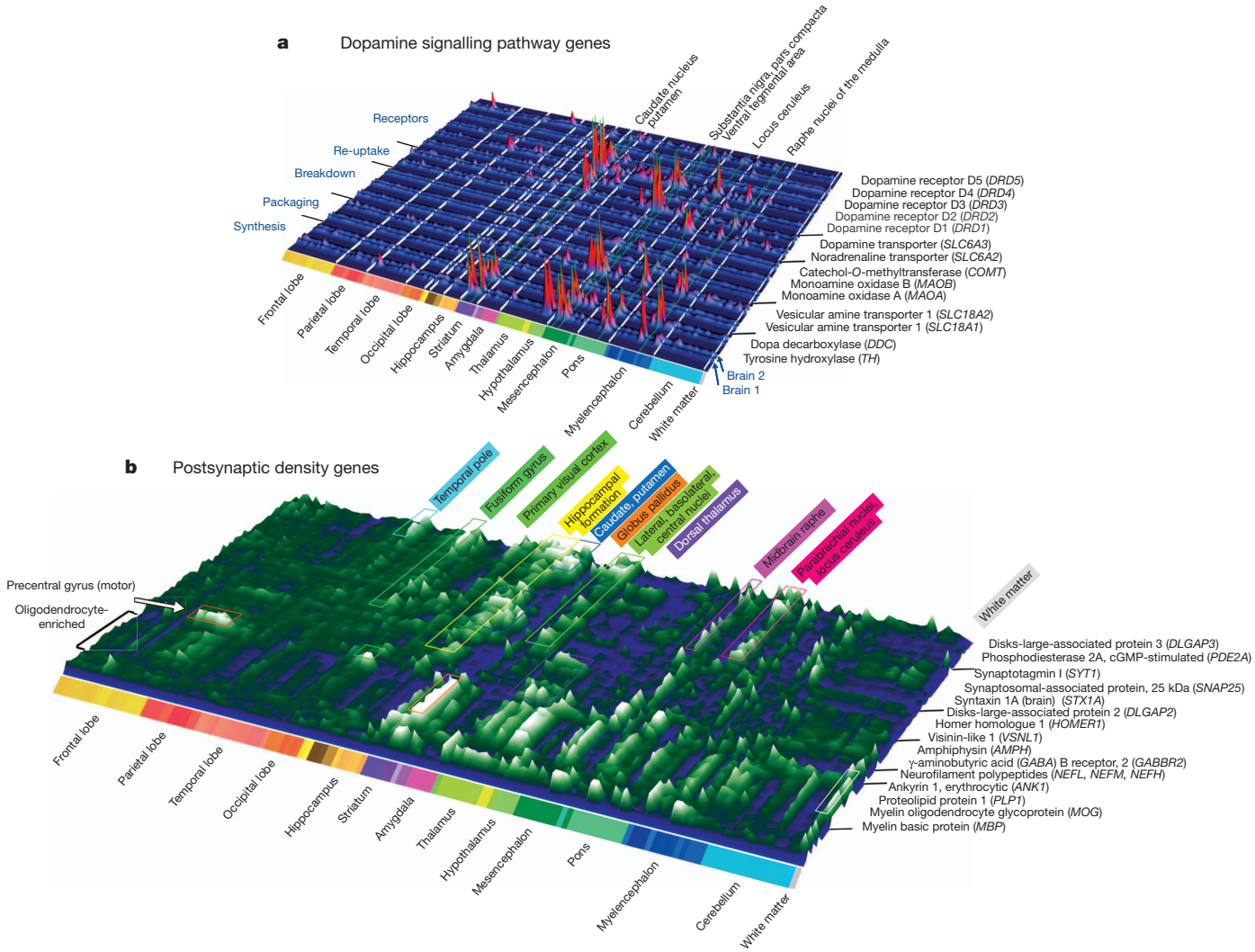

Figure 2 | Topography of transcript distributions for dopamine-signalling- and postsynaptic-density-associated genes. **a**, Gene expression profiles of genes associated with dopamine signalling plotted across 170 brain structures in two brains. Expression profiles for each probe plotted as raw microarray data normalized to mean structural expression, in paired rows to demonstrate consistency between the two brains. **b**, Gene-clustered topographic representation of the 74 most differentially expressed genes in human PSD preparations[12]. Gene profiles represent average expression in each structure between brains, plotted as deviation from the median. Clusters correspond to selective spatial enrichment of genes related to synaptic function, as well as an oligodendrocyte-enriched gene cluster (front cluster).

Interestingly, no statistically significant hemispheric differences could be identified at this fine structural level that were corroborated in both brains (paired one-sided *t*-tests, $P < 0.01$, Benjamini–Hochberg (BH)-corrected). Although surprising given well described lateralization of function, this finding is consistent with a recent study of developing human neocortex that failed to identify hemispheric differences despite extensive efforts using microarrays and quantitative PCR[11]. It may be that the basis for lateralization of function involves more subtle changes in specific cellular components, differences in relative area rather than type of functional domains between hemispheres, or is more related to functional connectivity patterns than molecular differentiation. Given this observation and to increase statistical power, samples from the two hemispheres for each structure were pooled for all subsequent analyses. In each brain independently, 84% of unique transcripts on the microarrays (29,412, referred to as genes for this manuscript) were found to be expressed in at least one structure (91.4% overlap in expressed gene sets between brains), consistent with the percentage of genes expressed in mouse brain by ISH (80%; ref. 1) and fetal human brain by microarrays (76%; ref. 11). Expression levels across anatomical structures were strongly correlated between brains (Pearson $r = 0.98$, $P < 10^{-40}$), with a highly significant correlation in differential expression relationships between structures (Pearson $r = 0.46$, $P < 10^{-40}$). Later in our analysis we completed data generation from a single (left) hemisphere of a third specimen. We found strong corroboration of overall expression levels and fold changes between structures in all three brains (Supplementary Fig. 2).

To illustrate the value of these data in understanding the functional organization of neurotransmission, we examined the 740 genes identified in the human excitatory postsynaptic density (PSD[12]), and in particular those that varied in their neuroanatomical distribution. Thirty-one per cent of PSD genes showed highly regional differential expression (Supplementary Methods 2 and Supplementary Table 3) (fold change >5 between any pair of 170 structures, false discovery rate <0.01), a significantly greater percentage than that observed across all genes (21%, $P < 10^{-6}$, Mann–Whitney *U*-test). As expected, many synapse-associated Gene Ontology (GO) categories[13] were enriched in this gene set, even relative to the PSD genes as a whole, including synapse (GO: 0045202), synaptic vesicle (GO: 0008021), synaptic transmission (GO: 0007268), neurophysiological process (GO: 0050877) and receptor activity (GO: 0004872).

Expression patterns for the most differentially expressed 10% of these PSD genes between any pair of structures are displayed in Fig. 2b

(74 genes with at least a 10.6-fold difference). The synapse-associated genes clustered into groups enriched in specific regions, indicative of a diverse set of excitatory synapse subtypes. For example, the primary motor cortex in the precentral gyrus, the origin of the longest range projection neurons, is delineated by selective enrichment of neurofilament proteins *NEFL*, *NEFM* and *NEFH*, which are frequently enriched in long-range projection neurons[14]. Surprisingly, a number of the most differentially expressed PSD-associated genes seem to be synthesized by glia, an observation made obvious by the stereotyped structural distribution of oligodendrocytes in white matter and other brain regions and the presence of well known myelin-associated genes (for example, myelin oligodendrocyte glycoprotein, *MOG*; myelin basic protein, *MBP*) in this gene cluster (Fig. 2b, front rows). The presence of these proteins in PSD preparations may represent a carry-over of glial fragments. Alternatively, they may be components of glutamatergic synapses between neurons and oligodendrocytes, which have been shown to share many properties with neuronal–neuronal synapses[15]. Overall, these data show remarkable regional variation in synaptic gene expression that probably underlies functional distinctions between regions.

Global transcriptional architecture of the adult brain

We next investigated the dominant features of transcriptional variation across the brain, beginning with global, brain-wide analyses and moving towards targeted local analyses of specific regions. An informative method for identifying biologically relevant patterns in high-dimensional microarray data sets is weighted gene co-expression network analysis (WGCNA)[16,17], which groups genes into modules that have strongly covarying patterns across the sample set. This method can identify gene expression patterns related to specific cell types such as neurons and glia from heterogeneous samples such as whole human cortex[18], due to the highly distinct transcriptional profiles of these cell types and variation in their relative proportions across samples. Each module is represented by an 'eigengene' corresponding to its expression pattern across structures (first left singular vector of the gene × structure matrix[16]), and genes highly correlated with the module eigengene are called 'hub' genes. This unbiased approach allows a module's function or cellular specificity to be imputed based on hub gene function, and allows statistical comparison either across studies to assign function or between brains to examine preservation between individuals.

Applied to the entire 911 sample set from Brain 1, genes were grouped into well-defined co-expression modules with specific anatomical distributions (Fig. 3a, b), consistent with previous studies in brain tissues[18,19]. Gene modules were frequently related to primary neural cell types and molecular functions (Fig. 3b, c). Several modules identify genes with enriched expression in neurons (M1–M2), based on overlap with neural-cell-type-enriched gene sets identified in previous studies[18] (second row in Fig. 3b). Genes in these modules are enriched in the neocortex (fifth row in Fig. 3b), and in particular cortical divisions as shown in eigengene plots (Fig. 3c). Hub genes and enriched GO terms for these modules are associated with neuronal structure and function and energy metabolism, as might be expected given the high metabolic demands of neurons (Supplementary Table 4). Other modules showed subcortical enrichment and correspond to expression in different types of glia (M8–M12), including microglia, astrocytes and oligodendrocytes. Additionally, one module with striking anatomical specificity for the paraventricular thalamus and central glial substance (asterisks in M5 eigengene histogram, Fig. 3c) corresponded to expression in the ventricular ependymal lining and choroid plexus. One highly regionalized neuron-related module (M6) was enriched in the striatum (the dopamine receptor *DRD1* in Fig. 2a is a hub gene). Thus, a major feature of the adult brain transcriptome profiled in this manner is the degree to which anatomical variation reflects the cellular make-up of different brain regions, both neuronal and non-neuronal.

The gene modules identified in Brain 1 were well conserved in Brain 2 as a whole (Fig. 3b), both at the level of regional gene expression patterns (third row) and as measured by a module preservation index (fourth row) using a summary Z-statistic as described previously[20]. Modules corresponding to broad neural cell types also showed highly significant preservation compared to a previous study using human brain samples (ref. 19 and data not shown).

We next took a more direct approach to examine relationships between regions of the brain based on dissimilarity of gene expression, by tabulating genes exhibiting highly differential expression between all pairs of regions. Significant pairwise differential relationships (BH-corrected $P < 0.01$) were independently recorded in each brain and a threshold set for at least a 2.8-fold ratio between structures (Supplementary Table 5). Figure 4a illustrates the resulting neuroanatomical molecular 'blueprint' common to both brains, by plotting the number of genes differentially expressed between each pair of structures based on the 11,414 genes passing these criteria in both brains (individual brain maps in Supplementary Fig. 3).

Many features of the brain transcriptome are apparent with this visualization. Remarkably few differences are seen at this fold change threshold across the neocortex (Fig. 4a, upper left) and cerebellum (lower right), reflecting their stereotyped repetitive cytoarchitecture. Exceptions to this relative cortical homogeneity include the postcentral gyrus (primary sensory cortex), temporal pole (area 38) and primary visual cortex (area 17). In contrast, complex differential relationships were observed between specific nuclei in subcortical structures. The globus pallidus and striatum have highly distinct profiles, as do several specific subcortical regions including the midbrain raphe, pontine nuclei and inferior olivary complex. The magnitude of differential expression between pairs of structures is also strongly correlated with the number of differentially expressed genes between these structures (Pearson $r = 0.62$, $P < 10^{-16}$; Supplementary Fig. 4).

Interestingly, a large percentage of these common differentially expressed transcripts (48%, or 5,500 probes) are poorly annotated, including probes not mapped to the human genome (HG19; http://genome.ucsc.edu/cgi-bin/hgGateway), mapped to contig sequences, or not mapped to known GENCODE genes[21]. Approximately 10% of these transcripts had very high correlation with the co-expression modules identified above (Pearson $r > 0.7$; Supplementary Table 6). For example, 38 transcripts demonstrated high correlation with the striatal module (M6), and 87 transcripts with the oligodendrocyte-associated module (M12; Fig. 3c), providing anatomical 'guilt-by-association' annotation of these genes of previously unknown function for selective roles in striatal and myelin function, respectively.

Most genes with high variation across brain regions are not selective for a single major brain region; rather, they are expressed in multiple regions and non-uniformly within these regions (Supplementary Fig. 5). This suggests that many genes may be quite pleiotropic with respect to brain function, and that local gene regulation in specific cytoarchitectural nuclei is the most important level of resolution. To summarize the complexity of structural variation and examine the extent to which major brain regions display local enrichment in specific fine cytoarchitectural divisions, we created a specificity index for each major region that measures enrichment in subdivisions of that region. This index, defined as the ratio of expression in one subdivision relative to the remaining subdivisions in that region (Supplementary Methods 3 and Supplementary Table 7), measures transcriptional diversity within regions. The results in Fig. 4b bear strong similarity to the plot in Fig. 4a, again with the neocortex and cerebellum displaying the least internal heterogeneity. In contrast, subcortical regions with many well-defined nuclei show the greatest local heterogeneity, including the myelencephalon, mesencephalon, pons, hippocampus and hypothalamus. It is also possible to identify genes with either brain-wide (global) or within structure (local) ubiquity (Supplementary Table 8). Not surprisingly, these gene sets are enriched for cellular

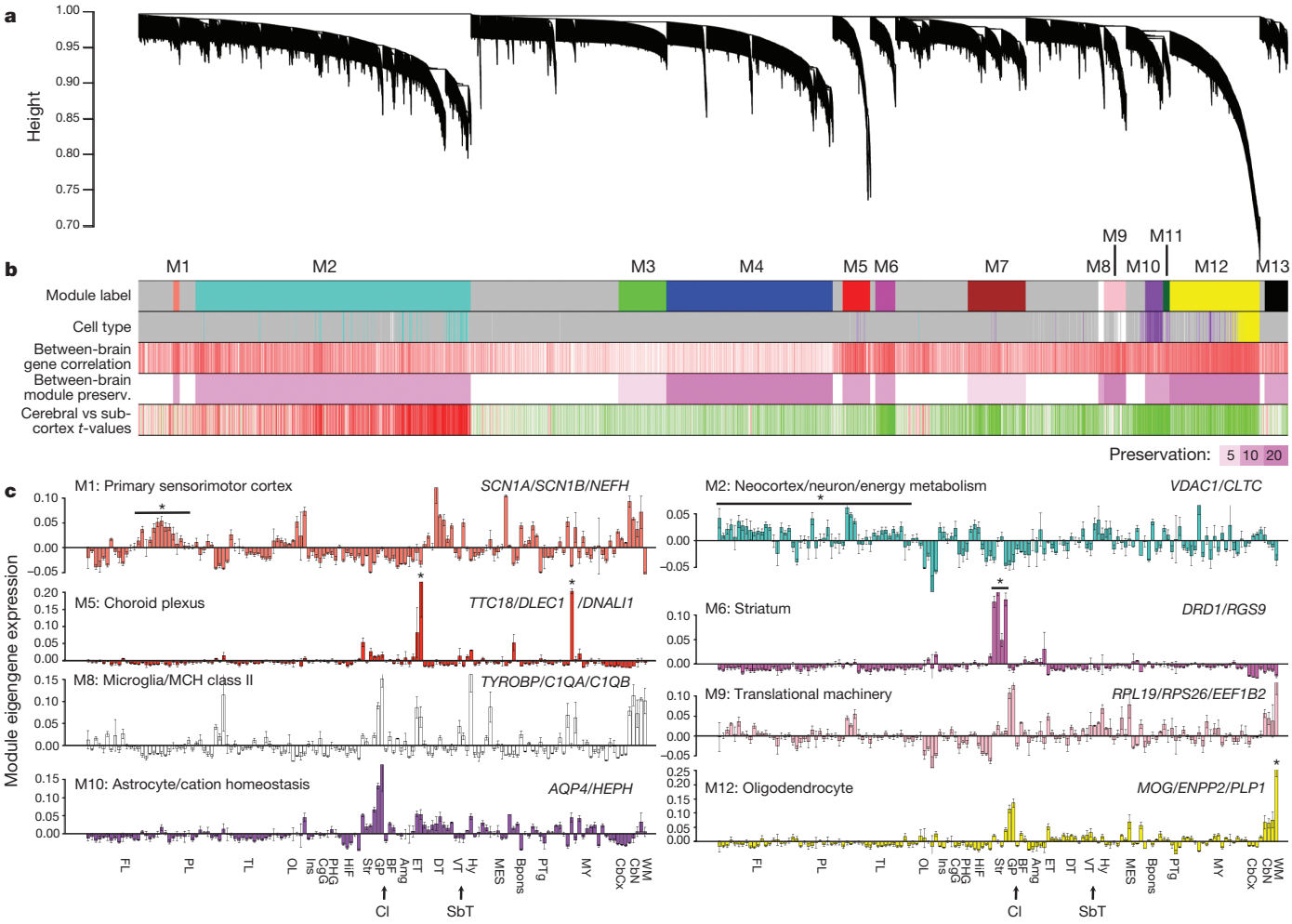

Figure 3 | Global gene networks. **a**, Cluster dendrogram groups genes into distinct modules using all samples in Brain 1, with the *y* axis corresponding to co-expression distance between genes and the *x* axis to genes (Supplementary Methods 2). **b**, Top colour band: colour-coded gene modules. Second band: genes enriched in different cell types (400 genes per cell type[18]) selectively overlap specific modules. Turquoise, neurons; yellow, oligodendrocytes; purple, astrocytes; white, microglia. Third band: correlation of expression across 170 subregions between the two brains. Red corresponds to positive correlations and white to no significant correlation. Fourth band: strong preservation of modules between Brain 1 and Brain 2, measured using a *Z*-score summary ($Z \geq 10$ indicates significant preservation). Fifth band: cortical (red) versus subcortical (green) enrichment (one-side *t*-test). **c**, Module eigengene expression (*y* axis) is shown for eight modules across 170 subregions with standard error. Dotted lines delineate major regions (see Supplementary Table 2 for structure abbreviations). An asterisk marks regions of interest. Module eigengene classifiers are based on structural expression pattern, putative cell type and significant GO terms. Selected hub genes are shown.

organelles and 'housekeeping' functions (for example, ribosome, mitochondrion, metabolism).

Local patterning reflects hippocampal cytoarchitecture

To explore local variation, we identified unique transcriptional signatures by analysis of variance (ANOVA) for the hippocampus. Following unsupervised hierarchical 2D clustering, cytoarchitecturally discrete subdivisions of the hippocampus (dentate gyrus, CA fields and subiculum) showed distinctive expression patterns sufficiently robust to cluster together like-samples while distinguishing subdivisions from one another (Fig. 5a). Interestingly, samples from the CA3 and CA4 subfields were not discriminable (intermixing in Fig. 5a), consistent with the view that CA4 is not a functionally distinct subfield from CA3 (ref. 22). Similarly robust regional clustering was observed in the mesencephalon, pons and myelencephalon (Supplementary Fig. 6 and Supplementary Table 9). Differential expression across hippocampal subfields could be validated by ISH. For example, the calcium-binding protein CALB1 has strong selectivity for the dentate gyrus relative to other hippocampal subdivisions in both brains (Fig. 5b), and cellular specificity for dentate gyrus granule neurons is demonstrated on an independent adult brain specimen by ISH in Fig. 5c. Hippocampal ISH data for CALB1 generated with the same histology platform in adult mouse[1] and rhesus macaque[23] allowed a phyletic comparison. Interestingly, expression in human differs from that in mouse (Fig. 5d) and rhesus monkey (Fig. 5e), where CALB1 is robustly expressed in CA1 and CA2 in addition to dentate gyrus.

Neocortical transcription reflects spatial topography

Our extensive neocortical sampling allowed us to investigate transcriptional variation across the neocortex in relation to spatial position and functional parcellation. Although highly differential expression between cortical regions is much less pronounced than between other brain regions (Fig. 4), many genes show statistically significant variation between lobes or gyri at a lower threshold. We first identified the 1,000 genes displaying the most significant variation in expression between 56 gyri in both brains (ANOVA, $P < 0.01$ BH-corrected, ranked by fold change between gyri; Supplementary Table 10). We then performed principal component analysis (PCA) on the 1,000 (genes) by 56 (sampled gyri) matrices for both brains. As shown in Fig. 6a–c, the first three principal components had striking selectivity for specific cortical regions (samples ordered by lobe and roughly rostral to caudal within each lobe) and were generally

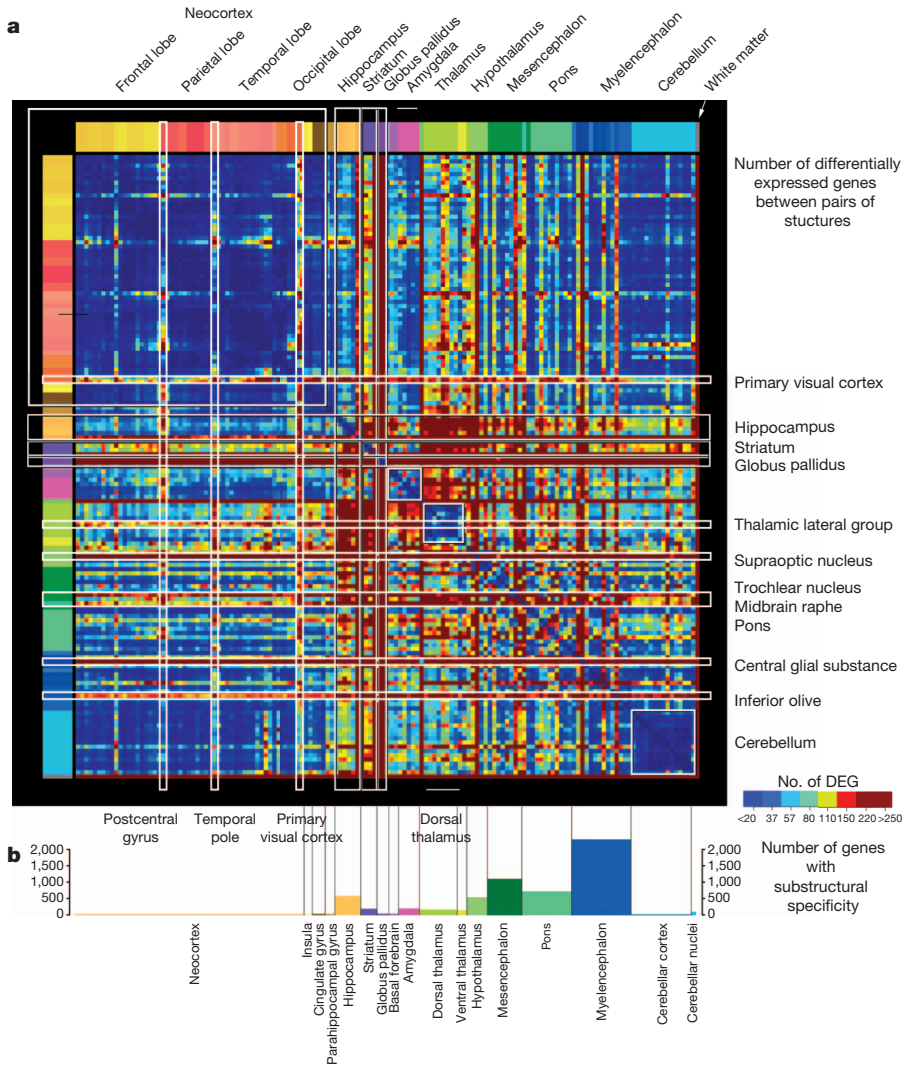

Figure 4 | Structural variation in gene expression. **a**, Matrix of differential expression between 146 regions in both brains. Each point represents the number of common genes enriched in one structure over another in both brains (BH-corrected $P < 0.01$, $\log_2$[fold change] > 1.5). DEG, differentially expressed genes. Several major regions exhibit relatively low internal variation (blue), including the neocortex, cerebellum, dorsal thalamus and amygdala. Subcortical regions show highly complex differential patterns between specific nuclei. **b**, Frequency of marker genes with selective expression in specific subdivisions of major brain regions (greater than twofold enrichment in a particular subdivision compared to the remaining subdivisions).

reproducible across both brains. PC1 is associated with primary sensorimotor cortices, with relative differential expression in precentral (motor) and postcentral (somatosensory) cortex, Heschl's gyrus (primary auditory) and primary and secondary visual areas. Confirmation of the visual cortex enrichment by ISH for several synaptic transmission-associated genes highly correlated to PC1 is shown in Supplementary Fig. 7. PC2 has areal selectivity for posterior orbital, paraolfactory and subcallosal gyri in the frontal lobe, the temporal pole, and the primary visual cortex. PC3 is primarily differential in frontal cortex compared to temporal and occipital cortex. These first three components accounted for a large amount of the variance (PC1: 58% in Brain 1, 42% in Brain 2; PC2: 10% in Brain 1, 11% in Brain 2; PC3: 5% in Brain 1, 8% in Brain 2; Supplementary Fig. 8). The spatial organization of the first three principal components was highly correlated between brains (Pearson $r = 0.71$ for PC1, 0.51 for PC2 and 0.70 for PC3).

To examine molecular relationships between different cortical regions, we applied multi-dimensional scaling (MDS; Supplementary Methods 3) to the samples of Brain 1 to visualize their genetic correlations along the directions of the first two (2D) or three (3D) principal components. Remarkably, the transcriptional relationships between samples recapitulate the spatial topography of the neocortex, as qualitatively illustrated after sample mapping in 2D (Fig. 6e). The relative positions of samples in the MDS plot mirror the actual positions of the gyri in the physical brain, shown in Fig. 6d on the MRI of the brain from which the samples were derived. Not only do samples from each lobe group together, but the relative positions of the lobes are anatomically correct. Furthermore, the relative position of each lobe's samples reflects the cortical topography, with the frontal pole and occipital striate cortex at opposite ends, precentral gyrus near postcentral gyrus, and so on. To provide a quantitative measure of this result we then applied the MDS method in 3D. As the positions of the samples were mapped back into MRI coordinate space, the correlation between 'genetic distance' and physical distance can be calculated after projecting the original cortical samples to a sphere and applying suitable rotation and scaling operations (Supplementary Methods 3). MDS-based sample correlations vary nearly linearly with 3D physical distance (Fig. 6e, inset), with a goodness of fit between native and MDS coordinates of 28.36% ($P < 10^{-4}$, Supplementary Fig. 9). This effect is strongest when limited to genes that are differential between gyri as above, but can also be seen using the entire ~30,000 gene set, achieving a fit of 12.48% between native and MDS coordinates ($P < 10^{-4}$, Supplementary Fig. 10). Therefore, gene expression profiles substantially determine position on the cortical sheet.

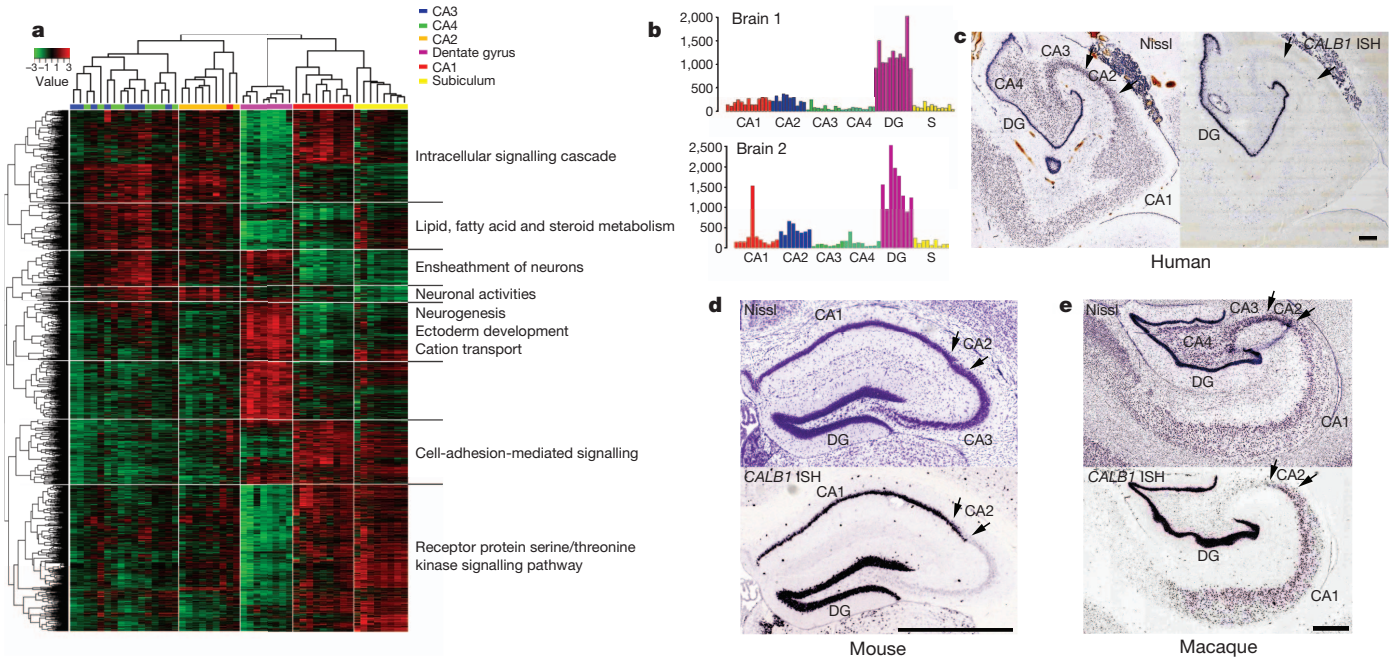

Figure 5 | Distinct transcriptional profiles of hippocampal subfields and human-specific pattern of *CALB1* expression. **a**, 2D clustering of microarray samples and differentially expressed genes across hippocampal subdivisions (ANOVA, $P < 0.01$ BH-corrected, top 5,000 genes), with selected enriched GO terms. **b**, Microarray data for *CALB1* shows enrichment in the dentate gyrus (DG) in both brains (*y* axis shows normalized raw microarray values). S, subiculum. **c**, Nissl (left) and *CALB1* ISH (right) through adult human hippocampus confirms dentate-gyrus-selective expression. **d**, **e**, Unlike human, *CALB1* ISH in the adult mouse (**d**) and rhesus macaque (**e**) show high *CALB1* expression in CA1 and CA2 (arrows) in addition to dentate gyrus. Scale bars: 1 mm.

Discussion

Molecular studies of human tissues are necessary for understanding the details of human brain function in the context of specific pathways and cell types and how they are affected in disease conditions. Here we describe the creation of an anatomically comprehensive transcriptional map in a small number of carefully selected, clinically unremarkable specimens, applying standardized digital molecular brain atlasing methods used in model organisms[3,24,25]. The combination of histology-guided fine neuroanatomical molecular profiling and mapping of gene expression data into MRI coordinate space produced an anatomically accurate quantitative map of transcript distribution across the entire human brain. This strategy was borne out in the robust differential molecular profiles of cytoarchitecturally and functionally distinct nuclei, providing a high-resolution genome-wide map of transcript distribution and the ability to analyse genes underlying the function of specific brain regions. Similar application of RNA sequencing methods[26,27], which were cost-prohibitive and technologically immature when the project was initiated, holds great promise for elucidating finer details of transcriptional regulation in the future.

Regional transcriptional signatures are highly conserved between the two brains assayed. These two individuals were males of similar age and ethnicity and therefore do not capture population or sex diversity; nevertheless, this high degree of similarity is suggestive of a strong underlying common blueprint for the human brain transcriptome and is consistent with other recent studies of human neocortical gene expression[4,5]. The availability of an entire hemisphere of a third brain specimen, as remarked above, enabled several confirmatory analyses to be performed. In particular, Supplementary Figs 11–13 report positively on the network analyses, structural variation of gene expression, and genetic topography of the neocortex. In summary, the high recapitulation of gene expression patterns across all three brains indicates that the basic transcriptional blueprint is robust across individuals. Ongoing work is focused on processing additional brains of both sexes to estimate the consistency of this blueprint.

The primary feature that distinguishes the human brain from that of other species is the enormous expansion of the neocortex relative to total brain volume. Our extensive profiling allowed us to ask directly how transcription varies across the neocortex. Surprisingly, we find a remarkable degree of transcriptional uniformity compared to other brain regions, apparently reflecting the similarity in laminar architecture across the entire neocortex[28]. However, there is significant, albeit less robust, variation in gene expression across cortical areas with two hallmark features. First, individual cortical samples showed such strong transcriptional similarities to neighbouring samples that the topography of the neocortex as a whole can, in part, be reconstructed based on their molecular profiles. One possible explanation is that these proximity relationships mirror lineage relationships of neocortical neurons generated from proximal parts of the developing neuroepithelium. Second, some primary sensory and motor regions do have distinct whole-transcriptome signatures, probably related to their specialized cellular and functional architecture. It is also likely that other more subtle features of cortical parcellation may not have been detected in the current analysis, including those identified using neurotransmitter receptor distributions[29] and functional connectivity[30]. One issue is that gyral patterns do not correlate perfectly with either cytoarchitectural or functional cortical parcellation. Greater regional differences may emerge if the samples can be grouped either by Brodmann area or on the basis of correlation to functional parcellations derived from functional imaging studies, now possible given the mapping of these data to MRI coordinates. Furthermore, it is likely that greater variation across areas will be found when assayed at the level of specific cortical cell types, as the excitatory neuron types in different layers display highly distinct molecular profiles[31] that have been shown to vary significantly across areas in primate neocortex[23]. Finally, higher confidence in consistent regional differences should emerge as more samples are investigated[32]. Nevertheless, the relative homogeneity of the two largest neuronal structures, with ~69 billion (cerebellar cortex) and ~16 billion (cortex) neurons out of the 86 billion neurons in the human brain[33], is striking and suggests an evolutionary expansion of a canonical cortical blueprint[34].

Finally, these data allow comparisons between humans and other animals, with particular relevance for studies of human disease. The

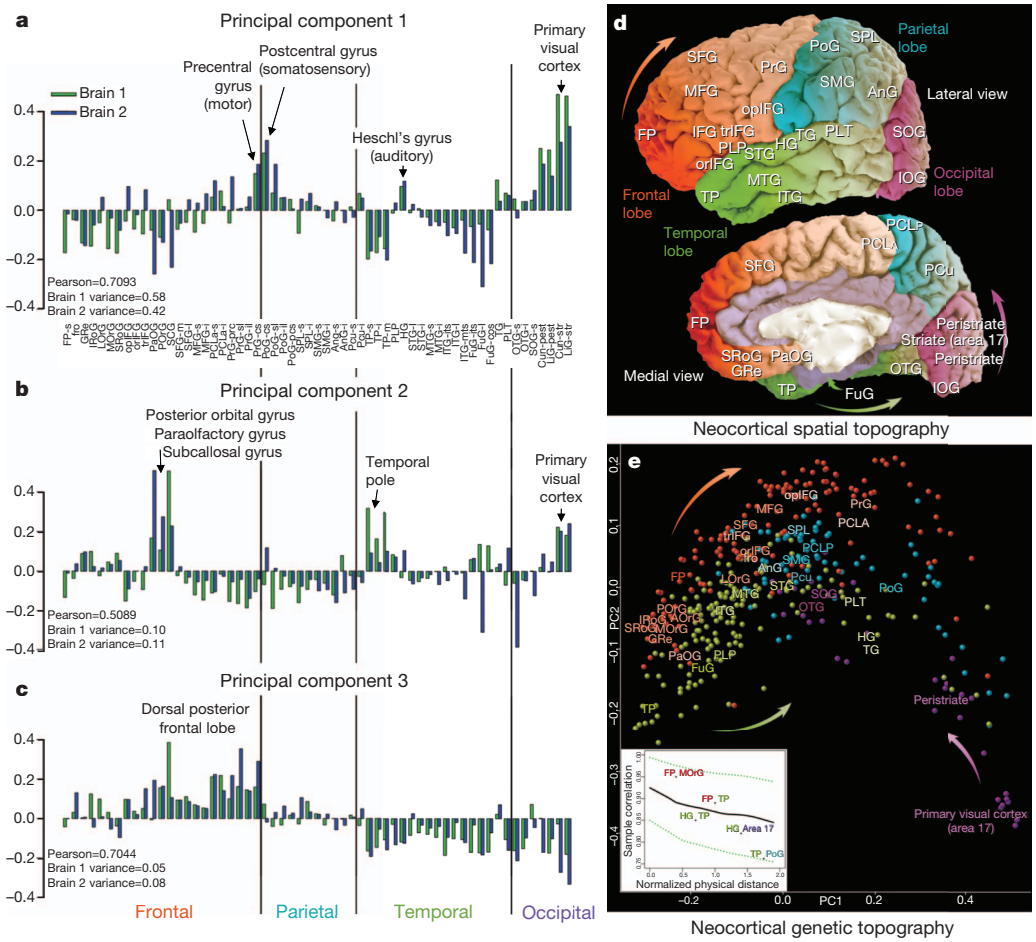

Figure 6 | The neocortical transcriptome reflects primary sensorimotor specialization and *in vivo* spatial topography. **a**–**c**, First three neocortical principal components, plotted across 57 cortical divisions ordered roughly rostral to caudal (frontal to occipital pole), are highly reproducible between brains. PC1 (Pearson $r = 0.71$) is selective for primary sensory and motor areas (**a**). PC2 (Pearson $r = 0.51$) is differential for specific subdivisions of the frontal, temporal and occipital poles (**b**), whereas PC3 (Pearson $r = 0.70$) is selective for the caudal portion of the frontal lobe (**c**). **d**, **e**, Relationship between the (x, y, z) location of sampled cortical gyri and their transcriptional similarities. Native Brain 1 MRI is shown in **d** with major gyri labelled (Supplementary Table 2). **e**, MDS applied to the same cortical samples, where distance between points reflects similarity in gene expression profiles. Median samples for major gyri are labelled. Samples cluster by lobe, and both lobe positions and gyral positions generally mirror the native spatial topography, emphasized by arrows in **d** and **e**. Inset panel in **e** plots the relationship (mean ± 1 s.d.) between 3D MDS-based similarity and 3D *in vivo* sample distance, demonstrating correlations that are stronger between proximal samples and decrease with distance. Selected gyral pairs are labelled. See Supplementary Table 2 for cortical gyrus abbreviations.

current manuscript describes a human-specific pattern for *CALB1* in the hippocampus compared to mouse and rhesus monkey. There are certain to be many such differences. In this light, these data should be extremely valuable from a translational perspective, allowing analysis of candidate genes and functional parcellation derived from genetic and imaging studies, and as a baseline for investigating neurological and neuropsychiatric disease.

METHODS SUMMARY

Anatomically comprehensive transcriptional profiling of adult human brains used high-throughput tissue processing and data generation pipelines for post-mortem brain imaging, anatomical delineation, sample isolation and microarray analysis. Data visualization and mining tools were developed to create a publicly accessible data resource (http://human.brain-map.org/). Extensive methodological details are supplied in Supplementary Methods 1.

Post-mortem tissue acquisition and screening. Tissue was provided by NICHD Brain and Tissue Bank for Developmental Disorders and the University of California, Irvine Psychiatry Brain Donor Program. After obtaining informed consent from decedent next-of-kin, specimens with no known neuropsychiatric or neuropathological history were collected and underwent serology, toxicology and neuropathological screening, and testing for RNA quality (RNA integrity number >6). Tissue collection was approved by Institutional Review Boards of the Maryland Department of Health and Hygiene, University of Maryland Baltimore and University of California Irvine. Specimens for microarray profiling were a 24-year-old African American male (Brain 1), a 39-year-old African American male (Brain 2), and a 57-year old Caucasian male (Brain 3; Supplementary Table 1).

Sample processing. Brains were imaged *in cranio* using MRI, cut into 0.5–1.0-cm-thick slabs and frozen. Slabs were subdivided and sectioned to allow histological staining, anatomical delineation and sample isolation using macro-dissection or laser microdissection. Total RNA was isolated and microarray data were generated by Beckman Coulter Genomics on Agilent 8 × 60K custom-design arrays (AMADID no. 024915). Sample locations were mapped from histology data into MR space using Inkscape (http://www.inkscape.org) and BioImage Suite (http://www.bioimagesuite.org) (Supplementary Methods 1).

Microarray data analysis. Weighted Gene Coexpression Analysis (WGCNA) was performed as described (Supplementary Methods 2)[16,17,20]. Module characterizations used Enrichment Analysis Systematic Explorer[35]. R (http://www.r-project.org/) was used for analysis and visualization (Supplementary Methods 2), principal component analysis (PCA), multidimensional scaling (MDS), and to transform MDS embedding into MNI space (Supplementary Methods 3).

***In situ* hybridization.** *In situ* hybridization used a semi-automated non-isotopic technology platform[1].

Received 22 December 2011; accepted 9 July 2012.

1. Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445,** 168–176 (2007).
2. Diez-Roux, G. *et al.* A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* **9,** e1000582 (2011).

3. Baldock, R. A. *et al.* EMAP and EMAGE: a framework for understanding spatially organized data. *Neuroinformatics* **1,** 309–325 (2003).
4. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478,** 483–489 (2011).
5. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478,** 519–523 (2011).
6. Markou, A., Chiamulera, C., Geyer, M. A., Tricklebank, M. & Steckler, T. Removing obstacles in neuroscience drug discovery: the future path for animal models. *Neuropsychopharmacology* **34,** 74–89 (2009).
7. Evans, A. C. *et al.* Anatomical mapping of functional activation in stereotactic coordinate space. *Neuroimage* **1,** 43–53 (1992).
8. Zeng, H. *et al.* Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149,** 483–496 (2012).
9. Bentivoglio, M. & Morelli, M. in *Handbook of Chemical Neuroanatomy. Dopamine* (eds Dunnett, S. B., Bentivoglio, M., Bjorklund, A. & Hokfelt, T.) Ch. I 1–107 (Elsevier, 2005).
10. Hurd, Y. L. & Hall, H. in *Handbook of Chemical Neuroanatomy. (Dopamine)* (eds S.B. Dunnett, M. Bentivoglio, A. Bjorklund, & T. Hokfelt) Ch. IX 525–571 (Elsevier, 2005).
11. Johnson, M. B. *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62,** 494–509, doi:S0896–6273(09)00286–4 (2009).
12. Bayes, A. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nature Neurosci.* **14,** 19–21 (2011).
13. Huang da, W. Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4,** 44–57 (2009).
14. Hof, P. R., Nimchinsky, E. A. & Morrison, J. H. Neurochemical phenotype of corticocortical connections in the macaque monkey: quantitative analysis of a subset of neurofilament protein-immunoreactive projection neurons in frontal, parietal, temporal, and cingulate cortices. *J. Comp. Neurol.* **362,** 109–133 (1995).
15. Bergles, D. E., Roberts, J. D., Somogyi, P. & Jahr, C. E. Glutamatergic synapses on oligodendrocyte precursor cells in the hippocampus. *Nature* **405,** 187–191 (2000).
16. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4** (2005).
17. Horvath, S. *et al.* Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc. Natl Acad. Sci. USA* **103,** 17402–17407 (2006).
18. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature Neurosci.* **11,** 1271–1282 (2008).
19. Miller, J. A., Horvath, S. & Geschwind, D. H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc. Natl Acad. Sci. USA* **107,** 12698–12703 (2010).
20. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7,** e1001057 (2011).
21. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7** (suppl. 1), 1–9 (2006).
22. Amaral, D. G. & Insausti, R. in *The Human Nervous System* (ed. Paxinos, G.) 771–755 (Academic, 1990).
23. Bernard, A. *et al.* Transcriptional architecture of the primate neocortex. *Neuron* **73,** 1083–1099 (2012).
24. Hawrylycz, M. *et al.* Digital atlasing and standardization in the mouse brain. *PLoS Comput. Biol.* **7,** e1001065 (2011).
25. Shattuck, D. W. *et al.* Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* **39,** 1064–1080 (2008).
26. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5,** 621–628 (2008).
27. Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature Struct. Mol. Biol.* **18,** 1435–1440 (2011).
28. DeFelipe, J. & Jones, E. G. *Cajal on the Cerebral Cortex: an Annotated Translation of the Complete Writings* (Oxford Univ. Press, 1988).
29. Zilles, K. *et al.* Architectonics of the human cerebral cortex and transmitter receptor fingerprints: reconciling functional neuroanatomy and neurochemistry. *Eur. Neuropsychopharmacol.* **12,** 587–599 (2002).
30. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1,** 1–47 (1991).
31. Belgard, T. G. *et al.* A transcriptomic atlas of mouse neocortical layers. *Neuron* **71,** 605–616 (2011).
32. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474,** 380–384 (2011).
33. Herculano-Houzel, S. The human brain in numbers: a linearly scaled-up primate brain. *Front. Hum. Neurosci.* **3,** 31 (2009).
34. Douglas, R. J. & Martin, K. A. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* **27,** 419–451 (2004).
35. Hosack, D. A., Dennis, G. Jr, Sherman, B. T., Lane, H. C. & Lempicki, R. A. Identifying biological themes within lists of genes with EASE. *Genome Biol.* **4,** R70 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We wish to thank the Allen Institute founders, P. G. Allen and J. Allen, for their vision, encouragement, and support. We express our gratitude to past and present Allen Institute staff members R. Adams, K. Aiona, A. Alpisa, J. Arnold, C. Bennet, K. Brouner, S. Butler, E. Byrnes, S. Caldejon, J. Campiche, A. Carey, J. Chen, C. Copeland, C. Cuhaciyan, T. Desta, N. Dotson, S. Faber, T. Fliss, E. Fulfs, G. Gee, T. Gilbert, L. Gourley, G. Gu, J. Heilman, N. Ivanov, K. Keyser, A. Kriedberg, J. Laoenkue, F. Lee, S. Levine, L. Luong, N. Mastan, N. Mosqueda, E. Mott, N. Motz, D. Muzia, K. Ngo, A. Oldre, E. Olson, J. Parente, J. Phillips, L. Potekhina, T. Roberts, K. Roll, D. Rosen, M. Sarreal, S. Shapouri, N. Shapovalova, C. Simpson, D. Simpson, M. Smith, N. Stewart, K. Sweeney, A. Szafer, L. Velasquez, U. Wagley, W. Wakeman, C. White and B. Youngstrom for their technical assistance. We thank C. Long for mechanical engineering contract work. We thank R. Gullapalli, A. McMillan and R. Morales for post-mortem magnetic resonance imaging and radiology interpretation of MR data; J. Cottrell, M. Davis, R. Johnson, K. Moraniec, R. Vigorito, A. Weldon and the NICHD Brain and Tissue Bank for Developmental Disorders for tissue acquisition and processing; J. Davis for donor coordination; F. Mamdani, M. Martin, E. Moon, L. Morgan, B. Rollins and D. Walsh for tissue processing and psychological autopsy (DW); D. Patel for magnetic resonance imaging; and J. Sonnen for consultation on tissue microneuropathology. We also thank the External RNA Controls Consortium (ERCC), the US National Institute of Standards (NIST) and Technology, and M. Salit for access to ERCC transcripts during Phase V testing. We are grateful to Beckman Coulter Genomics (formerly Cogenics) and their staff P. Hurban, E. Lobenhofer, K. Phillips, A. Rouse and S. Beaver for microarray data generation and design of the custom Agilent array. We also wish to thank the Allen Human Brain Atlas Advisory Council members D. Geschwind, R. Gibbs, P. Hof, E. Jones, C. Koch, C. Saper, L. Swanson, A. Toga and D. Van Essen for their scientific guidance and dedication to the successful execution of this project. The project described was supported in part by Grant Numbers 1C76HF15069-01-00 and 1 1C76HF19619-01-00 from the Department of Health and Human Services Health Resources and Services Administration Awards and its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Department of Health and Human Services Health Resources and Services Administration Awards. S.G.N.G and L.V.L. were supported by the MRCD, Wellcome Trust and European Union Seventh Framework Programme under grants 241498 EUROSPIN, 242167 SynSys, and 241995 GENCODYS Projects.

Author Contributions A.R.J., A.L.G.-B., E.H.S. and K.A.S. contributed significantly to overall project design. A.L.G.-B., E.H.S., K.A.S., A.E. and P.W. managed the tissue and sample processing in the laboratory. D.B., A.F.B., R.A.D., J.G., B.W.G., R.E.H., M.K., T.A.L., P.D.P., S.E.P., M.R., J.J.R. and B.E.S. contributed to tissue and sample processing. E.H.S. and Z.L.R. contributed to establishing the tissue acquisition pipeline. P.M.C., B.D.D., D.R.F., L.L., P.A.S., M.P.V. and H.R.Z. contributed to tissue acquisition and MR imaging. Z.L.R., A.B., M.M.C., N.D., A.J., J.M.J., E.T.L., S.C.S. and P.R.H. contributed to protocol development. S.D., J.M.J., C.R.S. and D.W. provided engineering support. A.L.G.-B., R.A.D., P.D.P., J.G.H., J.A.Mo., J.J.R. and B.E.S. contributed to the neuroanatomical design and implementation. L.N. and C.D. managed the creation of the data pipeline, visualization and mining tools. L.N., C.D. and C.C.O. contributed to the overall online product concept. L.N., C.A., M.C., J.C., T.A.D., D.F., Z.H., C.La., Y.L. and A.J.S. contributed to the creation of the data pipeline, visualization and mining tools. M.J.H., E.S.L., J.A.Mi., D.H.G., L.N.L., C.F.B., S.M.Sm., S.G.N.G., A.L.G.-B., E.H.S., K.A.S., A.B., D.B., V.F., J.G., D.R.H., S.H., C.Le., J.S., S.M.Su., P.R.H. and C.K. contributed to data analysis and interpretation. A.R.J. supervised the overall project, and the manuscript was written by M.J.H. and E.S.L. with input from other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.H. (mikeh@alleninstitute.org).