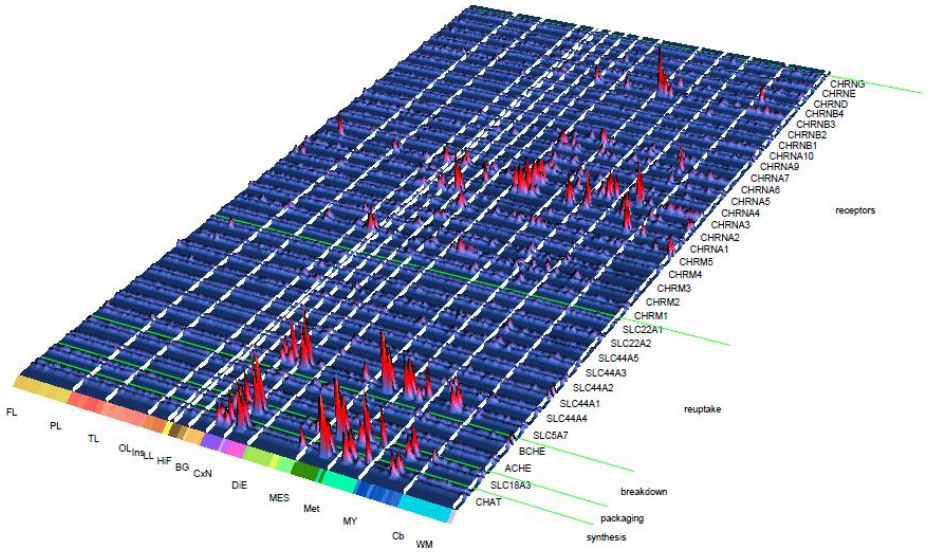
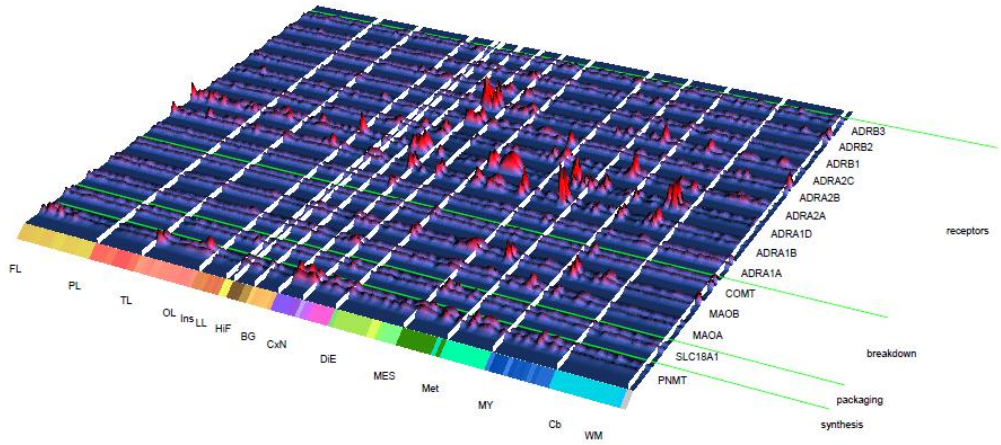
doi:10.1038/nature11405

Supplementary Figures

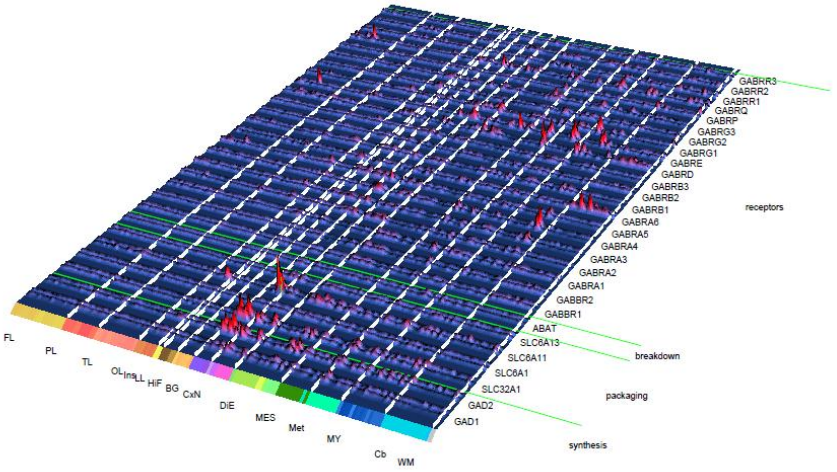
Acetylcholine



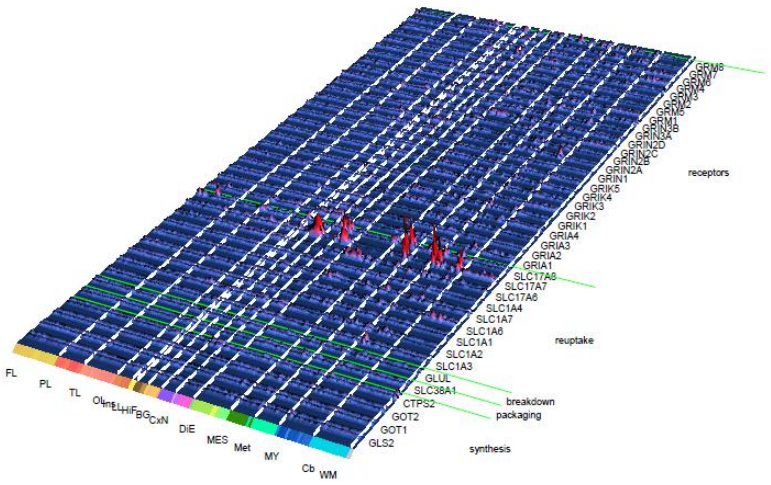
Epinephrine



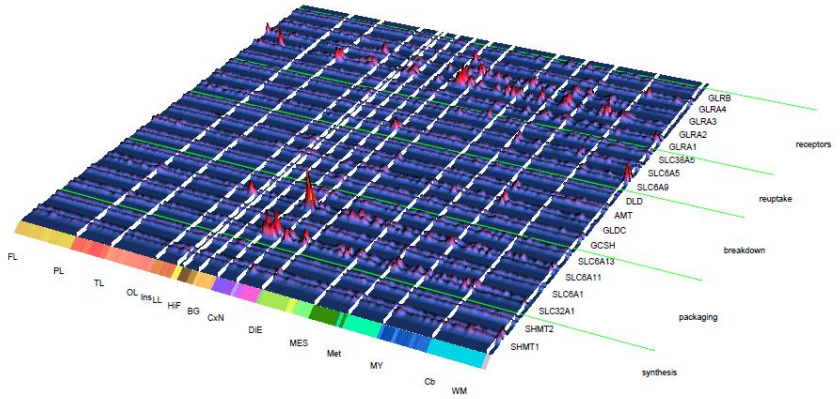
GABA



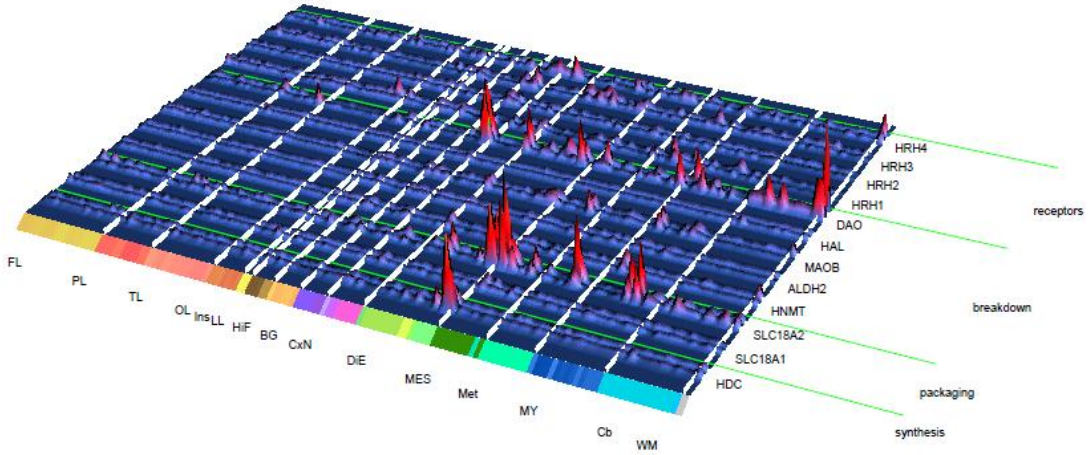
Glutamate

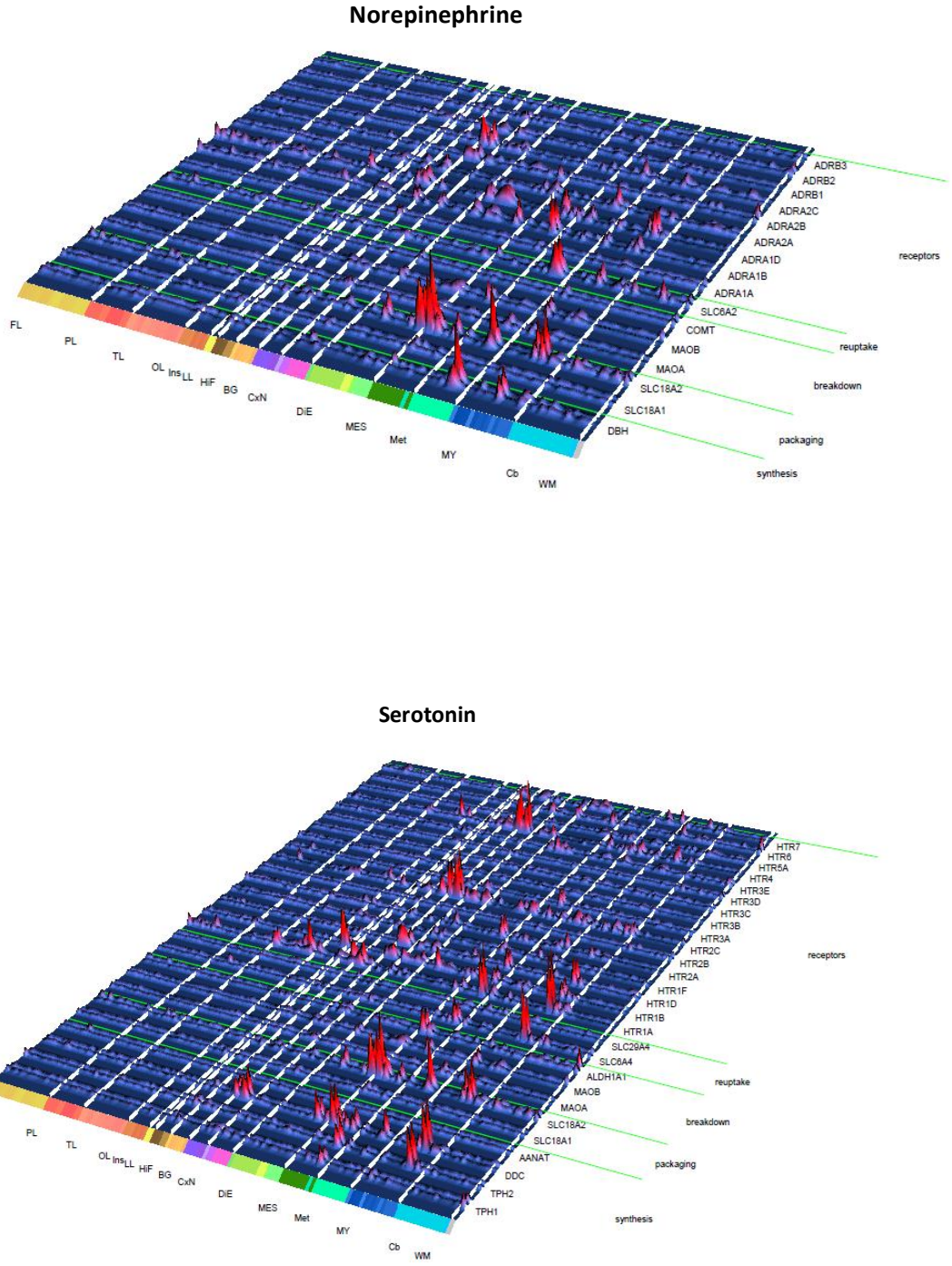


Glycine

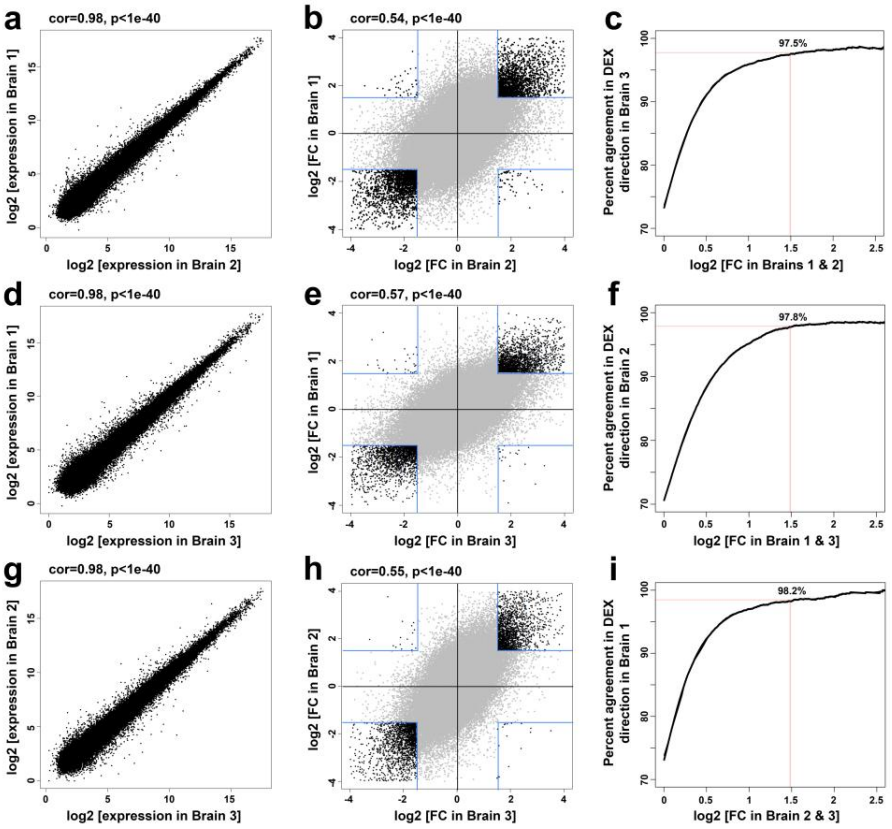


Histamine

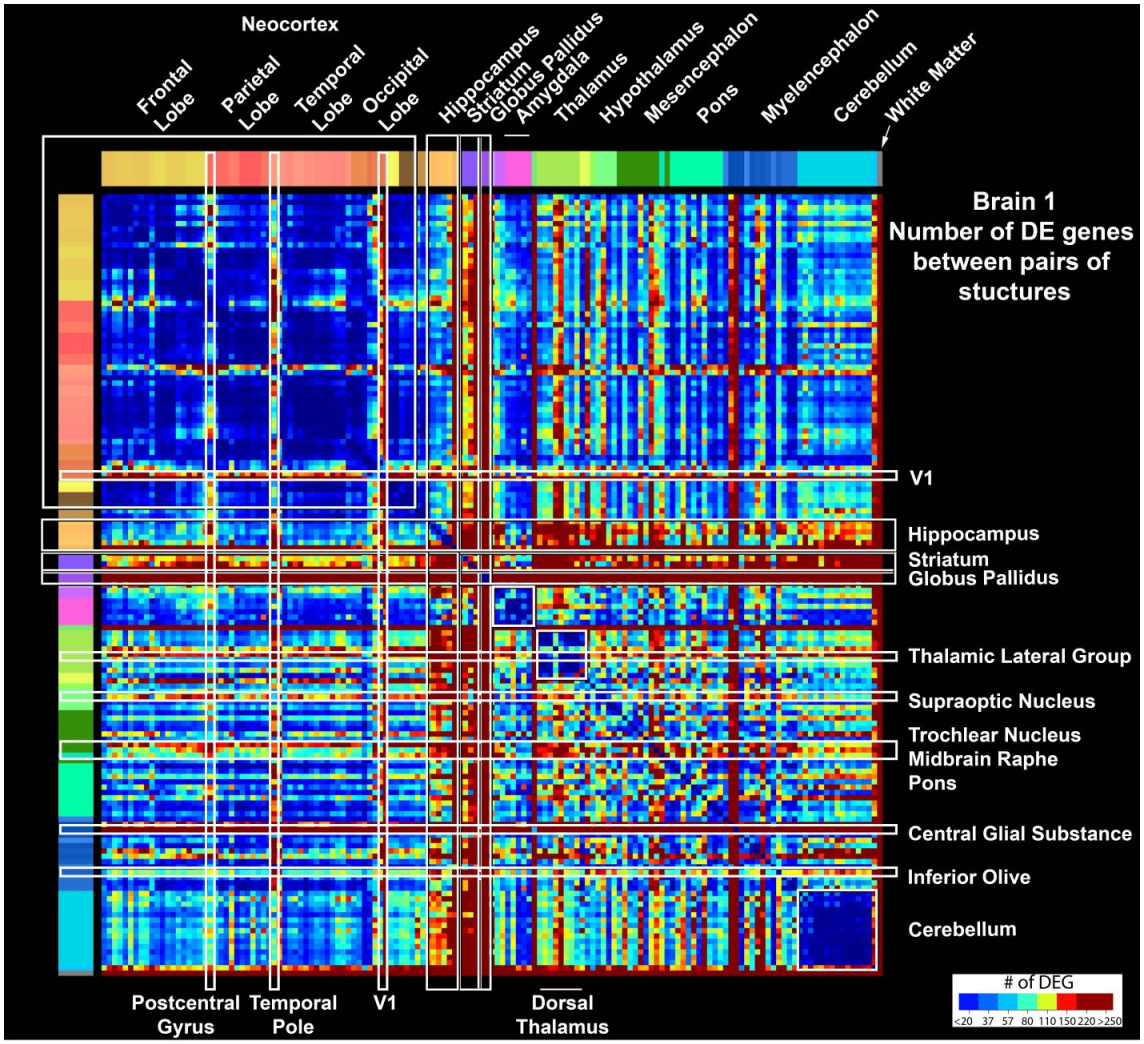


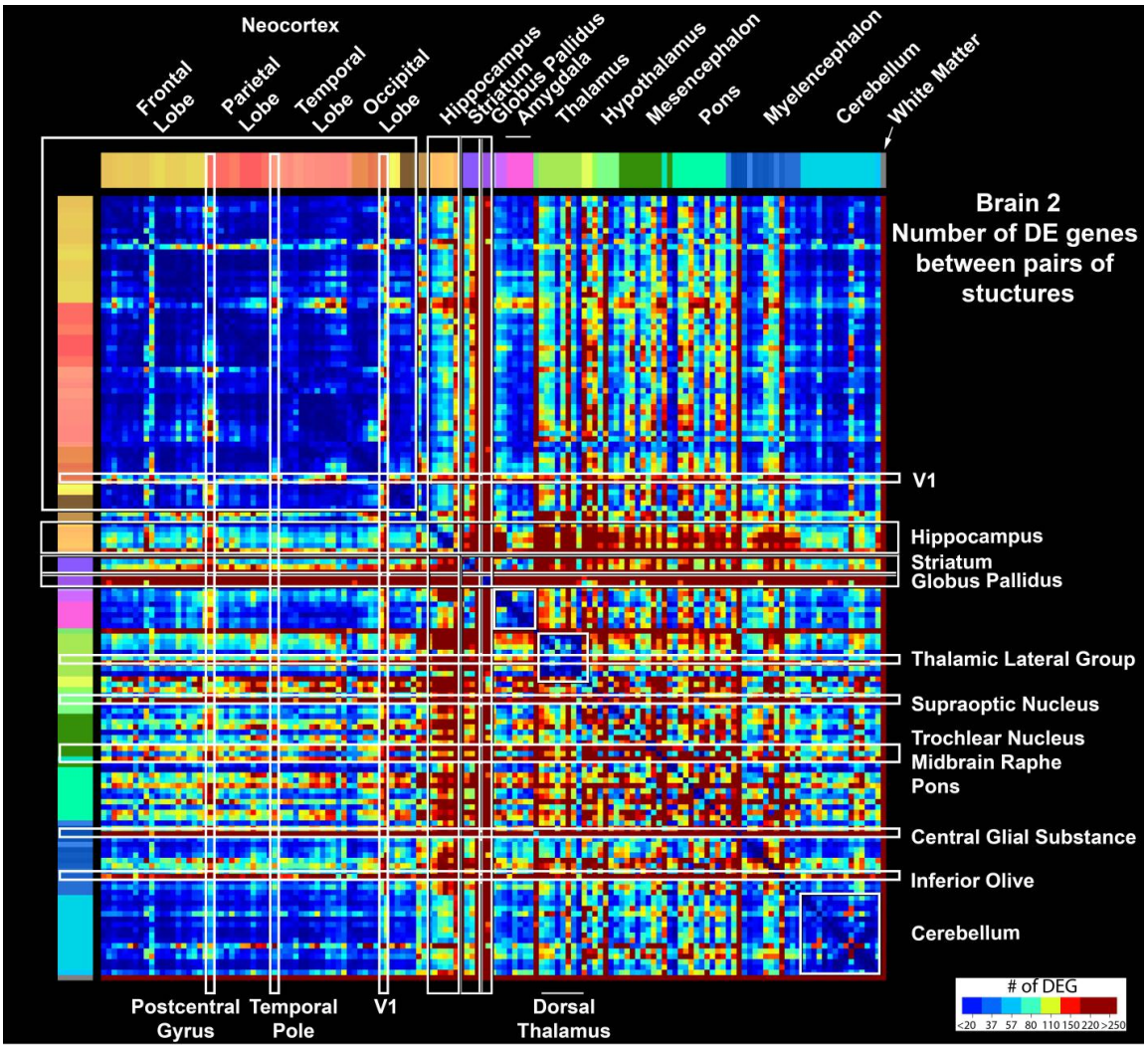


Supplementary Figure 1. Brain-wide distribution of genes involved in synthesis, packaging, breakdown, reuptake, and signaling through postsynaptic receptors for major neurotransmitter systems. Plots for acetylcholine, epinephrine, GABA, glutamate, glycine, histamine, norepinephrine and serotonin are shown. Raw $\log_2$ expression values (60k probes) were quantile-normalized across all 1799 samples in both brains. Subregions represented by at least two samples in each brain (n = 146) were identified, and averaged in each brain. To transform expression values into the same scale across genes, the median expression value for a given gene was subtracted. As most genes were represented by more than one probe, the best probe for each gene was identified by best correlation across 146 regions between the two brains. Expression plots for the neurotransmitter pathways were generated based on these expression values. See **Supplementary Table 2** for structure abbreviations along x-axes.



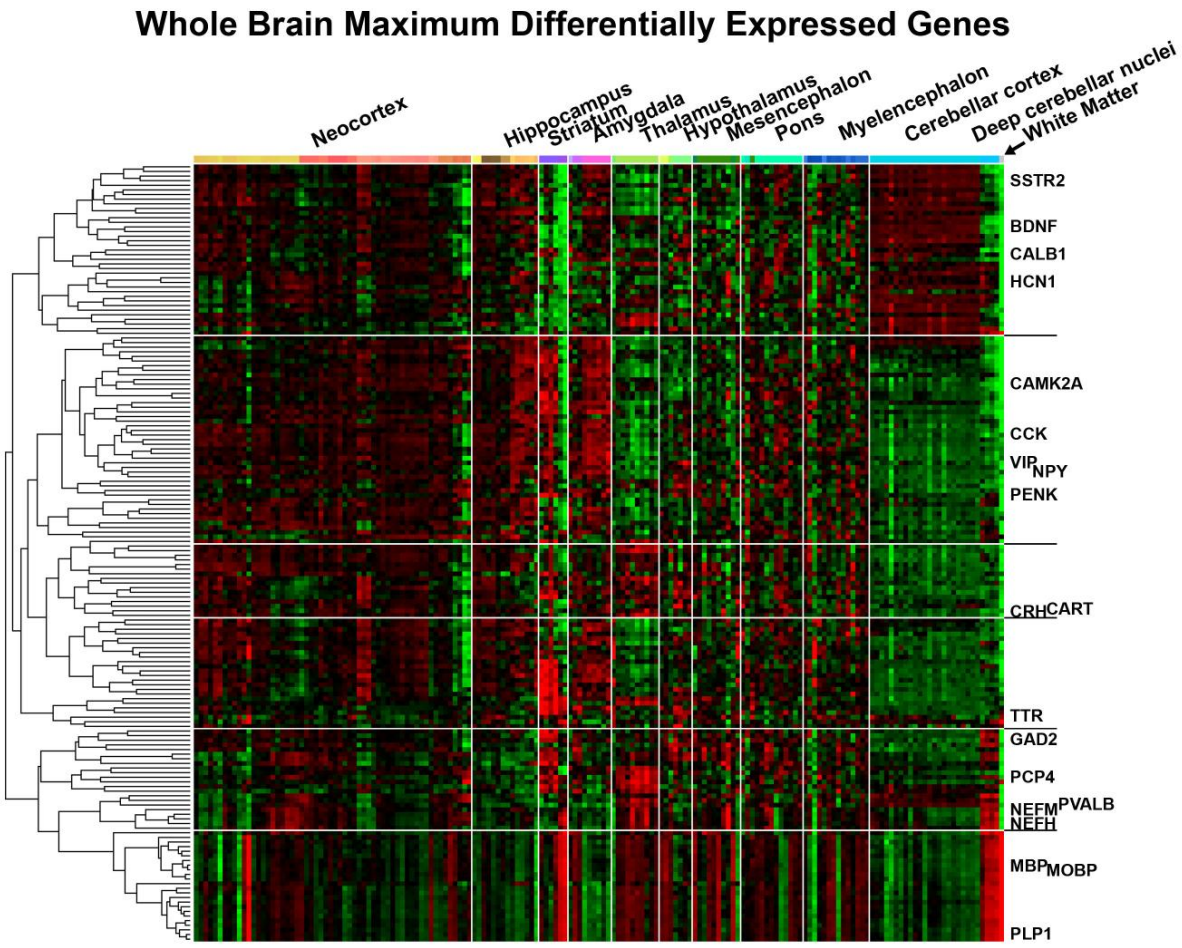
Supplementary Figure 2. Expression patterns in three individual human brains are highly consistent. *Left column*: Expression levels across regions between Brains 1 and 2 (**a**), Brains 1 and 3 (**d**) and Brains 2 and 3 (**g**) are strongly correlated. Each point corresponds to the average log2 [expression level] for one of 10,000 randomly sampled genes x structure data points (with 79 structures and 29,412 genes) in the pairs of brains listed. *Middle column*: Fold change in expression level between pairs of structures in Brains 1 and 2 (**b**), Brains 1 and 3 (**e**) and Brains 2 and 3 (**h**) for a given gene. 500,000 random gene x (structure pair) data points were chosen out of ~$1x10^8$ possible combinations. Differential expression relationships are correlated between brains. Darker black points in upper right and lower left quadrants represent structure pairs displaying at least log2 [FC] > 1.5 in both brains. *Right column*: Directionality of differential relationships is highly conserved between brains. Plot of the percent agreement in directionality of differential expression between structures in Brain 3 (**c**), 2 (**f**), or 1 (**i**) (y-axis) as a function of the fold change observed between those structures in both other brains (x-axis). For example, 97.5% of sampled gene x structure pairs exhibiting log2 [FC]>1.5 in both Brains 1 and 2 (corresponding to the upper right and lower left shaded quadrants in (**b**)) showed an expression gradient in the same direction in Brain 3.


Neocortex
Frontal Lobe
Parietal Lobe
Temporal Lobe
Occipital Lobe
Hippocampus
Striatum
Globus Pallidus
Amygdala
Thalamus
Hypothalamus
Mesencephalon
Pons
Myelencephalon
Cerebellum
White Matter
Brain 1
Number of DE genes between pairs of stuctures
V1
Hippocampus
Striatum
Globus Pallidus
Thalamic Lateral Group
Supraoptic Nucleus
Trochlear Nucleus
Midbrain Raphe
Pons
Central Glial Substance
Inferior Olive
Cerebellum
Postcentral Gyrus
Temporal Pole
V1
Dorsal Thalamus
# of DEG
<20 37 57 80 110 150 220 >250

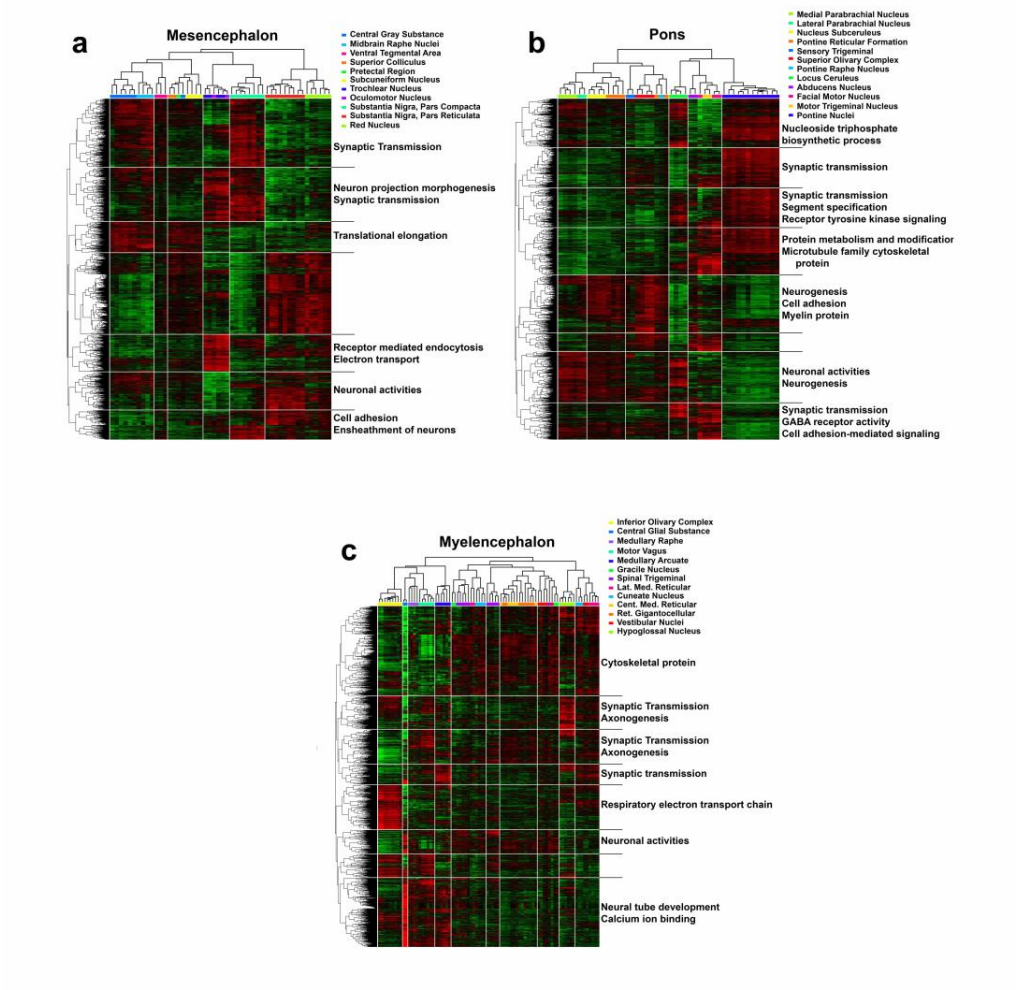


Supplementary Figure 3. Brain 1 and Brain 2 cross-structural variation in gene expression. Matrix of differential expression relationships between 146 brain regions, with each point representing the number of genes differentially expressed between pairs of structures (Benjamini-Hochberg corrected $p<0.01$, $\log_2$ [FC] > 1.5). Differential relationships are tabulated into large preliminary files and subsequently corrected for multiple hypothesis comparisons. A similar signature is apparent in both brains, represented in **Figure 4a** of the main manuscript by extracting common genes exhibiting the same differential relationships in both brains.

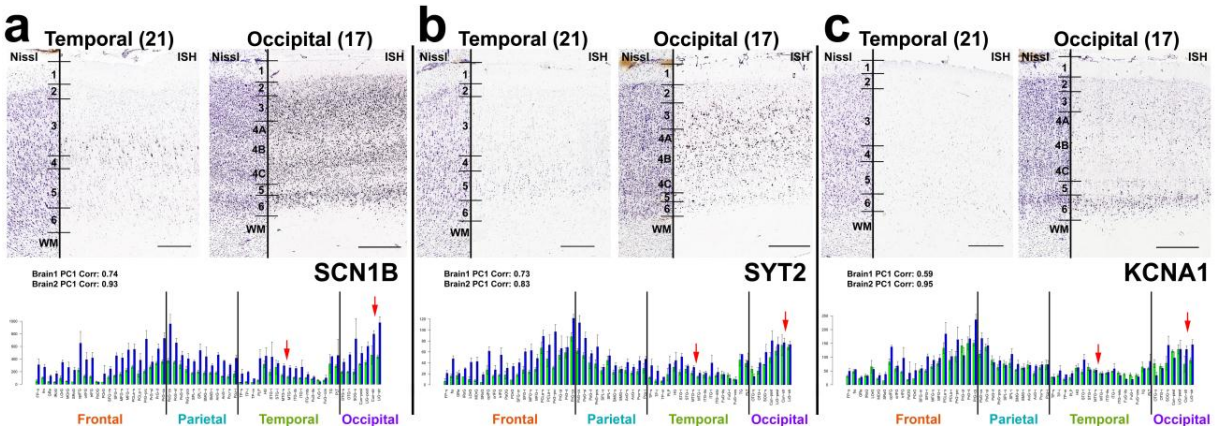
Supplementary Figure 4. Matrix of relationships between 146 brain regions, with each point representing the fold change of the top differentially expressed genes between regions common to Brain 1 and Brain 2. The average $\log_2$ fold change magnitude of the top 20 differentially expressing genes is calculated between each pair of structures using genes common to both brains as in **Figure 4a** of the main manuscript. There is a strong correlation (Pearson r = 0.62) between the number of differentially expressing genes and their maximum magnitude fold change.



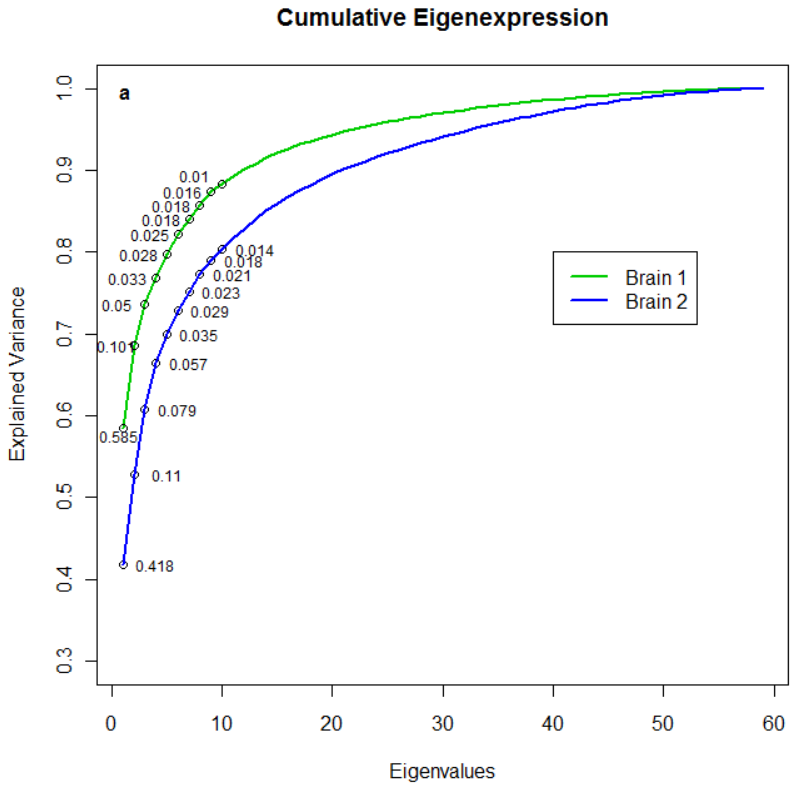
Supplementary Figure 5. Anatomical distribution of genes displaying maximal differences between major brain regions. Genes were selected as the most significantly differential across major brain regions by ANOVA. Expression profiles for the 300 genes with lowest p-values are clustered and represented in heatmap form with fixed anatomical ordering as in **Figure 4** of the main manuscript. While these genes cluster by common anatomical distributions, these complex combinatorial patterns typically span portions of multiple regions, and few if any genes are selective for a single region. Certain known genes are labeled along the y-axis. A subset of these genes is non-neuronal, for example the gene cluster at the bottom containing the oligodendrocyte gene MBP.



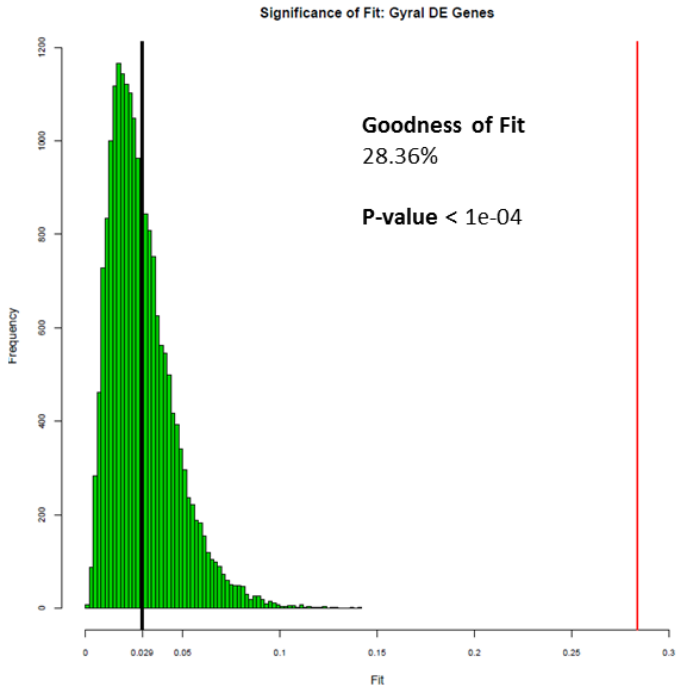
Supplementary Figure 6. Transcriptional signatures reflect subcortical nuclear anatomical parcellation. Clustering and heatmap representation of genes with differential expression within the mesencephalon (**a**) pons (**b**) and myelencephalon **(c).** 2D clustering of individual microarray samples and differentially expressed genes (ANOVA, $p < 0.01$ BH-corrected, top 5000 genes), demonstrates robust region-specific gene signatures, consistency across individual samples and combinatorial expression within structures. Samples from individual substructures generally cluster together. Selected GO categories enriched in particular gene clusters are listed to the right of each heatmap, generally showing enrichment in processes associated with neuronal phenotype and function (full GO analysis in **Suppl. Table 9**).



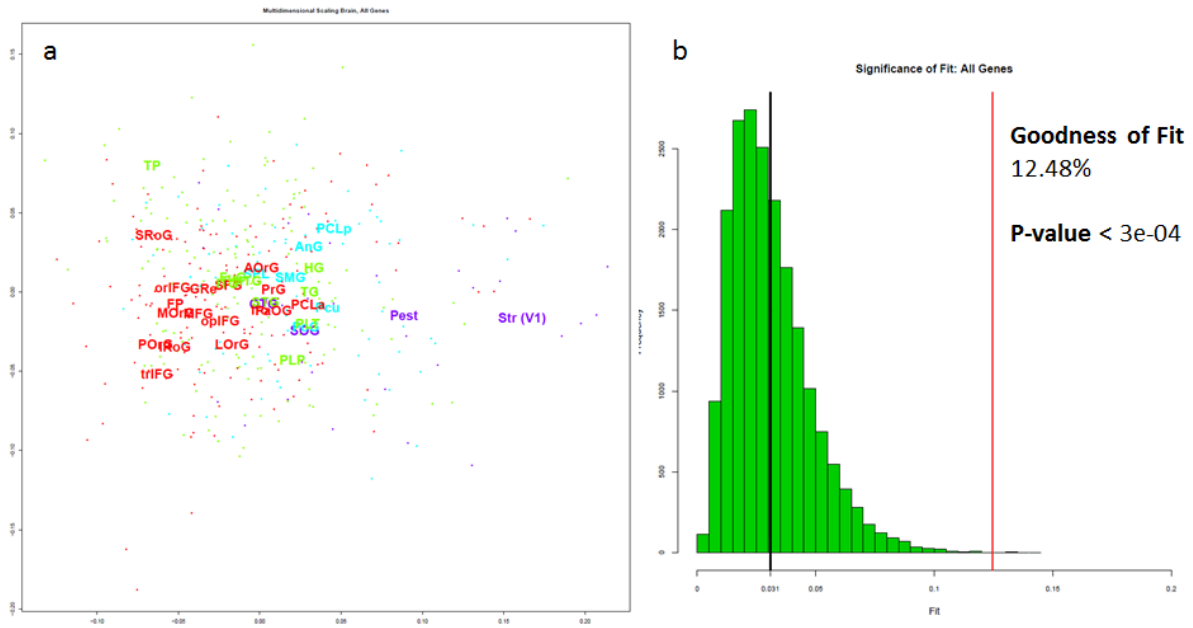
Supplementary Figure 7. ISH confirmation of enriched gene expression in human primary visual cortex. Spatial microarray expression patterns for genes highly correlated to PC1 in **Figure 6** of the main manuscript in the two brains (lower panels) for three neuronal function genes: SCN1B (**a**), SYT2 (**b**) and KCNA1 (**c**). ISH validation (upper panels) of predicted enrichment in occipital lobe (area 17) versus temporal lobe (area 21) for these three genes (corresponding to red arrows in lower panels). Scale bars: 500μm in **c**-**e**.



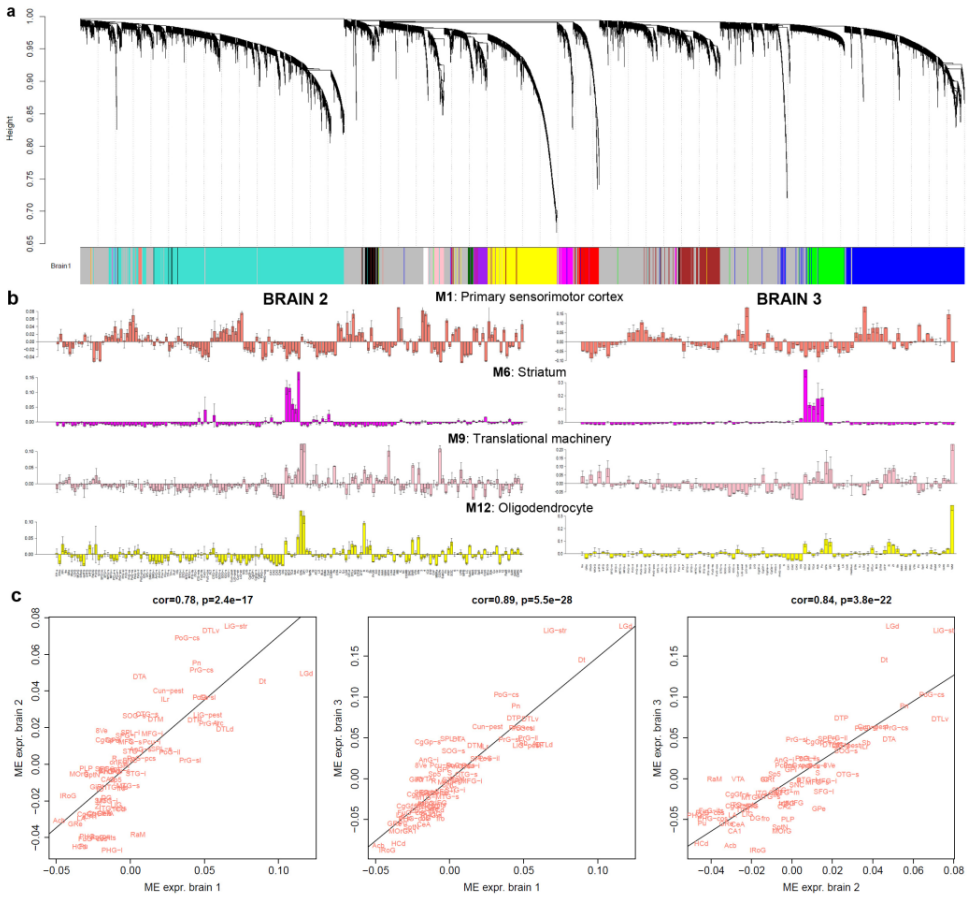
Supplementary Figure 8. Cortical gene expression and representative modes. The 1000 most differentially expressed genes (lowest Benjamini-Hochberg corrected p-value in *t*-tests) between pairs of gyri in the neocortex were selected for analysis. Singular value decomposition shows that a substantial component of variance in gene expression is explained in a small number of eigenvectors for both brains.



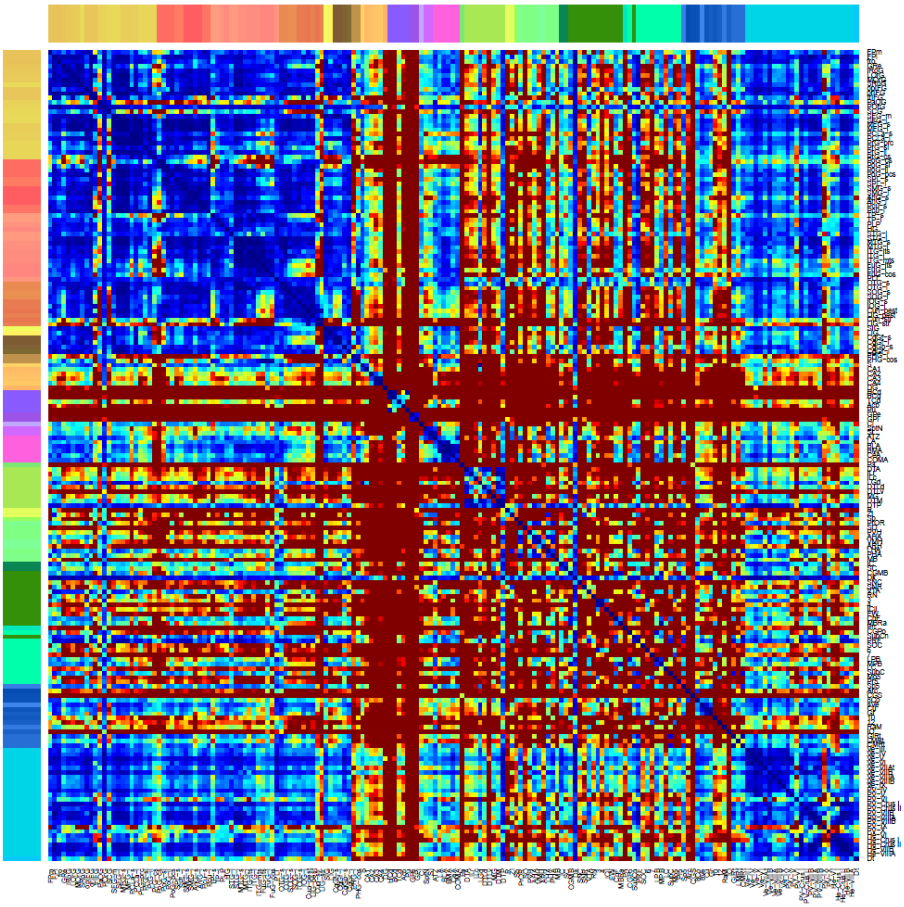
Supplementary Figure 9. Bootstrap permutation test for cortical reconstruction by gene expression. Projecting gene expression vectors for samples into the first three principal components using a multidimensional scaling (MDS) approach yields a representation of the cortex. This figure presents this result projected in the 2D plane for visualization. To measure the accuracy of MDS with the native coordinate representation we use a three-dimensional MDS, projecting onto the first three principal coordinates (**Supplementary Methods 3**). For every sample point there are two directions in 3D space: the actual MNI coordinates direction and the PCA model direction. The projection of the model in the direction of the true coordinates is a measure of the amount of explained variance accounted for by the model. The selection of genes in this figure is made only on the basis of pairwise independently tested differential expression relationships between the gyri. To access the accuracy of the reconstruction based on this gene set compared to the original MNI mean spatial gyral location we retain samples assigned to each given gyrus and permute the gyral location labels. For each permutation, after applying the translation, rotation, and scaling operations described in **Supplementary Methods 3** a goodness of fit measure is derived. The actual number of gyral permutations is large (7.11e+74), but to approximate a p-value we form N = 9999 bootstrap permutations and perform the scaling/rotation transformation on each. The histogram shows the number of occurrences of quality of fit metric with mean 0.029 shown by the black line. The observed reconstruction is far better (28.36%, $p < e\text{-}04$) than would be obtained by this subsample of possible permutations indicating that native structure amongst gyral relationships is well preserved.



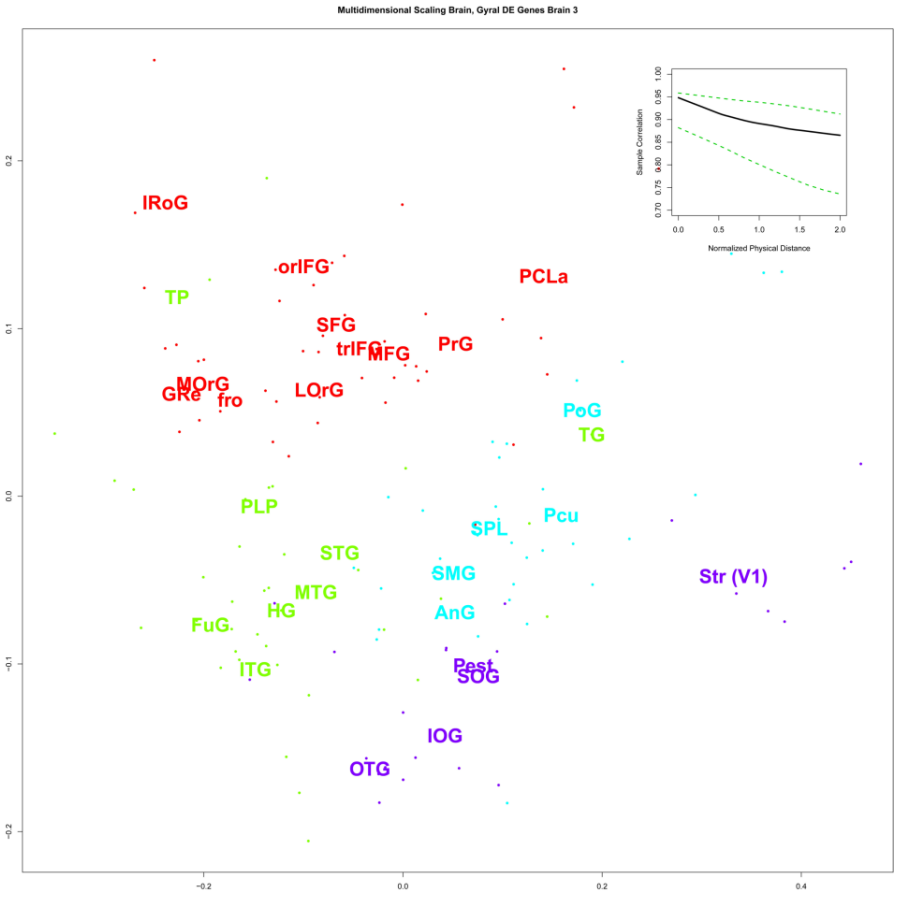
Supplementary Figure 10. Cortical reconstruction using all genes. The effect of spatial correlation of gene expression can be seen across the entire genome. Using the identical PCA scaling and multi-dimensional scaling (MDS) approach the effect of proximal gene expression relationships can be seen in panel (a). Frontal and parietal lobe samples are consistently grouping as is the strongly distinct occipital visual cortex signal. Gyri of the temporal cortex are more intermingled and less well differentiated using the entire gene set. This dataset still yields a meaningful genetic based reconstruction of the cortical topography. The goodness of fit metric shown in (b) is 12.48% with p < 3e-04. This result is obtained by allowing permutations among all samples independently as now the gene set is unconstrained.



Supplementary Figure 11. Global gene networks are preserved across brains. (**a**) Cluster dendrogram groups genes into distinct modules using all samples in all three brains. Y-axis corresponds to co-expression distance between genes, as derived using a consensus network across all three brains (**Supp. Methods 2**). X-axis corresponds to genes. Color bar below dendrogram shows module assignments from network created using data in Brain 1 (**Figure 3** of main manuscript). Note that the consensus network identifies nearly identical modules as brain 1 alone, suggesting strong agreement between brains. (**b**) Module eigengene (ME) expression (y-axis) for four modules across all subregions with N=2 samples or more for brains 2 and 3. Comparison of these plots to those from **Figure 3c** of the main manuscript shows consistent ME expression patterns between brains. Labeling as in **Figure 3c** of main manuscript. (**c**) Primary sensorimotor cortex module (M1) shows highly consistent ME expression levels across all three brains. X- and y-axes correspond to ME expression levels for all three pairs of brains (1 vs. 2, 1 vs. 3, and 2 vs. 3, respectively) across the 79 common subregions with N=2 or more samples in each brain. Each point is the average ME expression level for the named subregion. Note the high correlations (R>=0.78 for all comparisons).



Supplementary Figure 12. Structural variation in gene expression for brain specimen 3. Matrix of differential expression between 179 regions in single left hemisphere from a third brain specimen , with each point representing the number of genes more highly expressed in one structure than another (BH corrected $p<0.01$, $\log_2$ [FC]>1.5). As similarly found with the first two brain specimens, the globus pallidus and striatum have the most distinct molecular signatures, with lowest overall variability in the neocortex and cerebellum.



Supplementary Figure 13. Cortical Topography of brain specimen 3. Multidimensional scaling (MDS) based on top 1000 genes that display significant variation between cortical gyri reproduces the principal cortical topography as in **Figure 6** of the main manuscript. The first principal component associated with sensorimotor cortices is very highly conserved in brain 3 (Pearson correlation brain 1-brain 3 is 87.6% and brain 1-brain 2 is 78.1%). MDS-based cortical mapping and reconstruction achieves 32.54% of the left hemisphere positional alignment in MNI space (similar to the 28.56% reported in the main manuscript). Inset panel again plots the relationship between 3D MDS-based similarity and 3D *in vivo* sample distance, demonstrating correlations that are stronger between proximal samples and decrease with distance.

Supplementary Table 1. **Case qualification and donor profiles**

Detailed descriptions of each case and donor profile are provided. The information provided includes tissue receipt date, sex, age, race/ethnicity, handedness, postmortem interval, serology, toxicology, tissue pH, RNA quality for several regions, neuropathology, tissue received, additional medical information (if available), and available datasets.

Supplementary Table 2. Neuroanatomical sampling overview, complete neuroanatomical ontology and microarray sampling totals in the Allen Human Brain Atlas

A hierarchical ontology spanning all major architectural subdivisions was created to support the microarray sampling strategy. Each structure in this tree is designated a specific RGB color used throughout this paper. The number of samples isolated from both hemispheres and from the (left) hemisphere of each brain region is shown in the first worksheet, along with a summary of the specific subdivisions sampled and the sample isolation method used (Macro = scalpel macrodissection, LMD = laser microdissection). The complete hierarchical ontology and fine structure sampling for each brain is provided in the second worksheet. The hierarchical organization of the neuroanatomical ontology that was used for the microarray sampling is provided. The table contains the database ID, acronym, hemisphere, sample number for Brain 1, sample number for Brain 2, and the structures.

Supplementary Table 3. Differentially expressed postsynaptic density protein genes

The postsynaptic density genes identified in Bayes, van de Lagemaat et al. 2011 most differentially expressed between neuroanatomical structures were identified, and the top 10% of these genes are contained here. The genes plotted in **Figure 2b** of the main manuscript are at the top of this list.

Supplementary Table 4. **Whole brain WGCNA module gene assignment and module characterization**

Module membership (correlation between each gene and each module eigengene) and corresponding p-value for each probe in the whole brain network. Gene symbol and Entrez ID are provided for each probe (columns 2-3), along with the assigned module color/label (column 4). Note that the "grey" module represents all genes not assigned to any module. The "Enrichments" sheet includes a summary of regional enrichments for each module (in bold next to the module name), as well as select pathway enrichments (found using EASE), and enrichments for "Brain Lists" (found using the R function userListEnrichment). Columns A and B list the specific reference list (System) in which the Enrichment Category was found, while column C lists the P-value of enrichment.

Supplementary Table 5. Genes differentially expressed between all pairs of neuroanatomical structures

Differentially expressing genes between pairs of structures, that are common to both brains, are given in the first two columns. The number of genes that are reproducibly differential for the given structures in

both brains are reported in column 3. These genes (to a maximum of 20 ranked by highest differential fold change) are given together with fold change and p-value in Brain 1, then for the same gene in Brain 2 are given for each gene in subsequent columns.

Supplementary Table 6. Unannotated non-coding genes

A subset of unannotated/noncoding genes show high correlation with whole brain module eigengenes. Column A lists all probes to an unannotated/noncoding gene whose average Pearson correlation with any module eigengene is at least 0.7. Columns B-C show the highest correlated WGCNA module, as well as the corresponding Pearson correlation for each probe. Columns D-K reproduce several descriptive columns about the unannotated/noncoding probe from the Agilent annotation file, and are included for completeness.

Supplementary Table 7. Genes locally enriched in subdivisions of major brain regions and GO analysis

All genes showing at least a twofold enrichment in a given subdivision (Subregion) relative to all other subdivisions of that major brain region (Toplevel) are listed, along with the corresponding fold changes in each brain (FC1 and FC2). Column B (Probe) lists the probe for each gene used to measure the fold change, while column G (Type) either lists the glial module from **Figure 3** in the main manuscript to which the corresponding gene was assigned, or is labeled as "Non-glial" if the marker gene was not assigned to a glial-associated module. Gene ontology (GO) analysis was carried out for the genes from each subregion using DAVID with GO biological process 5, GO molecular function 5, PANTHER biological process all and PANTHER molecular function all. For each subregion the worksheet lists the top level structure followed by the subregion. A summary of any DAVID result with BH (Benjamini-Hochberg) p-value < 0.1 is provided on the GO significant summary worksheet for all subregions.

Supplementary Table 8. Genes ubiquitously expressed in brain and GO analysis.

UbiqutiousAcrossBrain sheet lists all genes with expression greater than the mean expression and coefficient of variation less than the predicted coefficient of variation in brains 1 and 2 (i.e., these genes are highly and consistently expressed). The genes and corresponding probes are listed in the first two columns. GO_Enrichment sheet shows the gene ontology (and other category) enrichments for the UbiquitousAcrossBrain list relative to all genes present in the brain using EASE. Note that genes associated with basic cellular processes (i.e., ribosome, cytoplasm, intracellular, mitochondrion, metabolism) tend to be ubiquitous. UbiquitousInSpecificRegions sheet lists additional genes / probes that are ubiquitous ("1") or not (" ") across specific brain regions with sufficient sample number to test (N>=10 in both brains), but that are ubiquitous not across the whole brain. Abbreviations: Cx = Neocortex; MY = Myelencephalon; HiF = Hippocampal formation; DT = Dorsal Thalamus; Str = Striatum; MES = Mesencephalon; CbCx = Cerebellar cortex; PTg = Pontine Tegmentum; CgG = Cingulate Gyrus; Bpons = Basal Pons; Cl = Claustrum; GP = Globus Pallidus.

Supplementary Table 9. Gene ontology results from differential expression analysis of hippocampus, mesencephalon, pons, and myelencephalon

For the results of ANOVA and unsupervised hierarchical clustering of genes differentially expressed between subfields of hippocampus (HiF), mesencephalon (MES), myencephalon (MY), and pons (Pons), the genes are provided along with the microarray probe ID and cluster number (1-8 for HiF, MY, and Pons and 1-7 for MES). Gene ontology (GO) analysis was carried out using DAVID with GO biological process 5, GO molecular function 5, PANTHER biological process all and PANTHER molecular function all. For each structure, a GO significant summary is provided containing any DAVID result with BH (Benjamini-Hochberg) p-value < 0.1.

Supplementary Table 10. Top 1000 differentially expressed genes between cortical gyri and correlation to principal components

Cortical samples were pooled at the level of cortical gyri, and the 1000 genes displaying the lowest p-values from *t*-tests for differential expression between these gyri are shown. For each brain the table contains gene name, probe, mean expression gyrus 1, mean expression gyrus 2, fold change, Benjamini-Hochberg corrected p-value, and the two gyral structure names as well as correlation with the principal components PC 1-3 described in the main manuscript.