# Supporting Information

## Miller et al. 10.1073/pnas.0914257107

### SI Materials and Methods

**Data Set Acquisition and Filtering.** All data sets were downloaded from the GEO database, and consisted of experiments run on either mouse or human brain tissue (Fig. 1*A*). We filtered out all but a core collection of data sets that were similar enough for useful bioinformatic comparison. First, we removed all data sets that were not run on an Affymetrix platform, leaving three platforms in human (HG-U95A, HG-U133A, HG-U133 Plus 2) and two in mouse (MG-U74A, MG-U430A). Second, we excluded all samples in each data set that were not taken from brain tissue (for example, in one expression atlas study, more than 80% of the samples were excluded). Third, to make the correlations between genes more comparable across studies, we omitted all data sets with fewer than 20 samples and split data sets with more than 40 samples into subsets when a biologically meaningful splitting parameter was available (i.e., brain region, disease state, or mouse strain). Finally, data sets were preprocessed identically (as detailed later) and all data sets with average within-species expression correlation (correlation between expression ranks of genes in two studies) and connectivity correlation (correlation between connectivity ranks of genes in two studies) that were disproportionately low were excluded. For the connectivity correlation assay, test networks were made for each data set with a power of five using WGCNA (see refs. 1 and 2 for more details). After filtering, a total of 18 data sets in human and 20 in mouse remained for our analysis (Table S1).

**Preprocessing and Network Formation.** An initial expression matrix was either downloaded from GEO and scaled such that the average intensity was 200 (if no .cel files were available), or created from Affymetrix .cel files. These .cel files were downloaded, read into R, and preprocessed using the "expresso" function and the MAS5 method of preprocessing. We chose MAS5 based on a study by Lim et al. (3), which benchmarked four commonly used normalization procedures (MAS5, RMA, GCRMA and Li-Wong) in the context of established algorithms for the reverse engineering of protein-DNA and protein-protein interactions (PPIs). Using replicate sample, randomized, and human B-cell data sets as input, their study suggests that MAS5 provides the most faithful cellular network reconstruction. We then calculated the correlation of gene expression between samples, and outliers with mean sample correlations more than two to three SDs below average (exact value specific to each study) were omitted until no outliers remained (as described in ref. 2). After performing quantile normalization on the filtered data, probe sets that were not present were excluded from the analysis either by using the "pma" function in R and excluding probe sets that were called "absent" in more than 90% of the samples (in the data sets where .cel files were available), or by removing a comparable number of probe sets (approximately 40%) with the lowest 90% quantile of expression. To allow comparison across Affymetrix platforms, only a single probe set for each gene was kept—for genes with two corresponding probe sets we chose for each sample the probe set with highest expression, whereas for genes with three or more probe sets we chose the probe set with the highest connectivity across samples. To make the final expression file for each data set, all probe sets without associated genes were omitted and all remaining probe sets were reassigned the name of their corresponding gene symbol. In mouse network, mouse gene symbols were converted to human orthologues using data from Jackson Laboratory Mouse Genome Informatics (August 2006). The result of this step was 18 human expression files with 20 to 40 samples and 5,629 to 9,731 genes each and 20 mouse expression files with 18 and 44 samples and 5,176 to 6,157 genes each. All preprocessed data files (as well as the resulting network data and some associated code and support files) are available at the WGCNA group Web site (www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/MouseHumanBrain).

From these preprocessed expression files we created a human and a mouse consensus network (method modified from ref. 4). For each consensus network we first created correlation matrices from each data set (obtained by calculating the Pearson correlations between all variable probe sets across all subjects in each data set), and then weighted them based on the number of samples used in that data set. Each data set was weighted as follows:

$$W_k = 0.5^* \left[ \sqrt{N_k} / \sum \left( \sqrt{N_a} \right) + 1/S \right] \qquad \textbf{[S1]}$$

where $W_k$ is the weight of the kth expression matrix, $N_k$ is the number of samples in the kth data set, and S is the number of data sets. Expression matrices where a given gene was not considered present were omitted from this calculation and genes that were present in fewer than 50% of data sets were excluded from the consensus network, leaving a total of 9,778 genes in the human network and 6,368 genes in the mouse network. Networks were formed from the weighted correlation matrices following the protocols of WGCNA, as previously described (1, 2). In short, the adjacency matrices were calculated by raising the absolute values of the weighted correlation matrices to a power of five. Finally, topological overlap (TO), a measure of node similarity (i.e., how close the neighbors of gene 1 are to the neighbors of gene 2) that has proven biologically meaningful, was then calculated as described previously (1, 5). We compared TO with known PPIs for both human and mouse in two different databases (6, 7) by placing all of the TO values into 100 bins representing the percentile of ranked values from largest to smallest, and determined what percentage of known PPIs were present in each bin (Fig. S2).

**Module Formation, Characterization, and Preservation.** For the initial module identification, all but 5,000 of the most connected genes in the human network (3,000 in mouse) were excluded to decrease noise. This filtering step left genes with high intramodular connectivity in any module as part of the network while omitting genes with weak membership in all modules. Omitted genes are included in the gene by eigengene table, which is used for many of the comparisons in this paper (as detailed later). Genes were hierarchically clustered using 1 − TO as the distance measure and modules were determined by using a dynamic tree-cutting algorithm (8). Module identifiers in the mouse network were then changed to match the modules in the human network with the most significant gene overlap (Fig. 1 *B* and *C*) (3). Each gene's module membership (MM) for a given module was then estimated as the average Pearson correlation between that gene and the five genes in that module with the highest within-module connectivity ($k_{in}$), which has been shown to be a good approximation of the module eigengene (ME) (2, 9). *P* values were obtained by (*i*) averaging the T-score from gene–eigengene Pearson correlation across data sets, (*ii*) scaling to the square root of the number of data sets per gene, and (*iii*) calculating a *P* value from the T-distribution of the resulting scaled T-score. For the final module characterizations, all genes with MM values of $R > 0.2$ and with $P < 10^{-13}$ were assigned to that module, leaving an average of approximately 5% of all of the genes in each

module, with some genes assigned to multiple modules. Although these values represent one of many possible sets of module-thresholding parameters, the results were relatively robust to changes in module size. We also used enrichment analysis by way of Expression Analysis Systematic Explorer (10) and IPA to characterize modules based on gene ontology (GO). Finally, modules were graphically depicted using VisANT (11) as previously described (12, 13).

For unbiased disease gene (DG) characterization we used an annotated list of approximately 5,000 orthologous mouse and human genes, for which mutations in the gene were known to produce disease phenotypes for any human or mouse disease (Jackson Laboratory) (14). This list of "all DGs" was curated from the literature in an unbiased manner and distinguishes genes causing human-specific disease phenotypes from those causing similar phenotypes in both mouse and human. We created a list of "dementia DGs" by taking the subset of DGs that was associated with dementia-related disorders. Dementia-related disorders were defined as the list of all disorders returned in a search of Online Mendelian Inheritance in Man (ncbi.nlm.nih.gov/omim) for the term "brain AND ("dementia" OR "neurodegenerative" OR "neurodegeneration")". Overall, we found that approximately 20% of all genes in our networks were DGs and approximately 3% of the genes in our network were dementia DGs.

Finally, we used a variety of strategies to measure module preservation. First, we used a permutation test procedure implemented in the WGCNA R package, which produces a summary preservation Z-score (Table 1) (15). Second, we assessed the significance of module overlap between genes in corresponding mouse and human modules (Table 1). Third, we estimated the similarity of module annotation both by showing that most corresponding human and mouse modules show significant overlap with the same module from ref. 2 (Table 1), as well as by showing that these module pairs share similar annotations as measured by GO and Ingenuity Pathway Analysis (IPA; www.ingenuity.com) (Table S2).

## SI Results and Discussion

This article provides a case study regarding metaanalysis on the level of coexpression networks. Among other things, the article shows (*i*) how to successfully reduce potential biases from individual studies, (*ii*) how to weigh the information from different data sets, and (*iii*) how to compare the resulting networks between species. Our results illustrate that our method for compiling multiple data sets into a single correlation matrix allows across-experiment and between-species comparisons.

**Linear Relationship Between TO and PPIs in Both Mouse and Human.** We and others have previously shown that our measure of gene coexpression, TO, predicts many biologically meaningful relationships. For example, across multiple species such as yeast (16) and human (2, 17), highly coexpressed genes are more likely to interact on the protein level than genes with low coexpression. Therefore, to provide another level of systematic network validation, we determined the relative likelihood that gene pairs of specific TO would also have PPIs (Fig. S2) (6, 7). For positively correlated genes, there was a strong, positive linear relationship between TO and PPIs in both networks. In contrast, we observed a similar negative linear relationship between negatively correlated gene pairs, which fits expectations (18). We also find strong relationships in the mouse network, despite the fact that these genes are actually mouse homologues of the genes in both PPI databases.

**Networks Are Robust to Choice of Data Sets.** We next determined the robustness of our networks to choice of data set. To do this, we randomly split the mouse and human data into two groups of nine data sets (approximately 250 arrays), creating new networks for each of these grouping as described earlier in the *SI Text*. For each network we then calculated both the average expression rank across data sets, as well as the connectivity rank for each gene in the network, and correlated these two measures within species (Fig. S1). There was nearly perfect expression correlation between both the human ($R = 0.94$; $P < 10^{-1,900}$) and the mouse ($R = 0.98$; $P < 10^{-2700}$) group pairs, suggesting that our choice of data sets has minimal impact on gene expression ranking or levels. There was lower, but still highly significant, connectivity correlation between the human ($R = 0.76$; $P < 10^{-700}$) and the mouse ($R = 0.62$; $P < 10^{-400}$) group pairs, suggesting a high preservation of the gene coexpression relationships in networks made from different data sets, consistent with previous studies (2, 18, 19). Finally, to ensure that module selection is robust to our choice of data set, we performed WGCNA following the procedure outlines in Fig. 1*A* on each of these human and mouse subnetworks, defining modules by using the same characterization as in Fig. 1 *B* and *C*. This network formation schema resulted in networks highly overlapping with those created using all data sets. Taken together, these results show that there is no significant within-species bias in the data and suggest that the composition of brain-specific data sets used in this analysis does not significantly impact the results.

**Cortex/Control Networks and Overall Brain Networks Are Equally Comparable.** One key idea in our analysis is that we include only arrays from brain, to reduce noise generated from samples in which we are not interested. One possible problem with this approach is that there are relatively few data sets in mouse and human brain that both have data publicly available and also include enough samples to perform viable coexpression analyses. As such, our mouse and human data sets might not be completely matched; for example, most human samples are quite region-specific, with several data sets acquired via laser capture microdissection, whereas many mouse data sets are from larger regions or even whole brain. It is therefore possible that resulting network differences could be a result of differences in data set selection rather than differences between species. To address this issue, we created separate "C/C" networks in mouse and human that included only cortex samples from control subjects, removing bias caused by brain area, disease, or treatment state, thus making our networks more directly comparable. We included whole brain samples along with cortex samples in mouse because (*i*) the mouse brain largely consists of cortex and (*ii*) we would not have had enough samples to perform such a comparison using only mouse control samples from cortex. This filtering step left three mouse data sets (57 arrays) and five human data sets (137 arrays) in our analysis (Table S1), which we compared following the same procedures outlined in the text for the original networks (e.g., Table 1 and Fig. S1).

We find that the two analyses (C/C and all brain) produce highly similar networks, suggesting that our between-species analysis using all available brain tissue is not biased by networks specific to brain region, disease states, or to medications that may be taken by human subjects. At the global level, we find similar preservation between mouse and human expression in the all-brain analysis compared with the C/C analysis ($R = 0.60$ vs. $R = 0.64$), whereas node connectivity preservation is actually better in the all-brain analysis ($R = 0.27$ vs. $R = 0.15$), consistent with the notion that connectivity measures are more sensitive to the amount of data than to the precise matching of data sets. In the case of module comparison, although we find highly similar results at the level of module overlap (nearly all modules in both networks show gene overlap between analyses with significance levels of $P < 10^{-40}$), in the case of module preservation, most modules show lower preservation Z-scores in the C/C networks compared with the overall networks. In fact, modules for oligo-

dendrocytes (M2h) and multiple cellular components (M5h, M8h, M14h) that were significantly preserved in the overall network no longer show significant preservation between the C/C networks. In short, we find that networks made using data sets from all reliable brain samples are more comparable than networks made using fewer, but better-matched data sets.

These C/C networks can also be used to further address the issue of how sensitive networks are to data set selection. To do so, we measured within-species module preservation in our original networks compared with the C/C networks. For all modules, we found similar average preservation Z-scores (as a typical example, module M1h in the original and C/C analyses have $Z = 6.08$ and $Z = 5.77$, respectively), suggesting that both networks are robust to removal of a large percentage of the data.

**Minimal Effects of Agonal State on Between-Species Transcriptional Changes.** Many genes are known to change expression levels with death. It is therefore important to rule out agonal state as a prominent cause of between-species coexpression differences. First, gene expression levels show high preservation between human and mouse, which points to minimal effects of agonal state in general between species. Furthermore, others have shown that, despite the relatively large number of affected genes, agonal state has a minimal (if any) effect on differential expression analyses (20)—in other words, changes in expression with death are disease and region blind. Finally, we used the hypergeometric distribution to measure overlap between each module in both networks and a core set of human genes previously shown to be related to agonal state (either showing increased or decreased expression in autopsy vs. biopsy tissue) (20). Consistent with results from this group, we found that modules associated with mitochondria and ribosome were enriched with genes showing decreased expression after death (for M4h, $P < 10^{-43}$; for M5h, $P < 10^{-9}$; for M8h, $P < 10^{-20}$), whereas M14h (nucleus) contained excess genes showing increased expression with death ($P < 10^{-9}$). As these are among the most preserved modules between the species, and the corresponding mouse modules showed comparable enrichments (for M4m, $P < 10^{-48}$; for M5m, $P < 10^{-38}$; for M8m, $P < 10^{-16}$; for M14m, $P < 10^{-12}$), agonal state does not appear to play different roles between mouse and human. Although there is weak enrichment for genes increasing with agonal state in M9h ($P < 10^{-3}$), this enrichment is much less significant than in the highly preserved cellular component modules and is almost entirely caused by genes not present in the mouse network. Thus, although we cannot rule out the fact that agonal state may partially underlie some species differences in this poorly preserved module, it is highly unlikely that this effect is substantial.

**Preservation of Modules Associated with General Cellular Components.** Although the thrust of our research is brain-specific, confirmation of basic cellular biology is a key validation of our method. Orthologous ribosomal, mitochondrial, and other ubiquitous cellular components are found in nearly all known species, with high conservation between species as distant as yeast, fly, and human (19, 21). Furthermore, previous studies of general transcriptional similarities between mouse and human have found multiple common modules of coexpressed genes—including the ribosomal subunits (22, 23). Our metaanalyses uncovered similarities between many global network properties, including general measures of gene expression, connectivity, and module preservation, with genes involved in basic cellular components showing the highest level of preservation. In our human network, within-species preservation can be most clearly seen in module M12h, which contains 64 of the 71 ribosomal subunits present in the human network (Table S2), whereas between-species preservation is most obvious in M4h, which shows the most highly significant module preservation Z-score ($Z = 17.21$). Replication of these relatively well established coexpression links is an important step toward confirming the validity of our methods, and demonstrating the reliability of our results.

To further flesh out the mitochondrial result, we compared M4 and M5 with validated lists of genes transcribed in somatic versus synaptic mitochondria (24). M4 showed enrichment over M5 for somatic mitochondria in both mouse ($P < 10^{-6}$ vs. $P < 10^{-3}$) and human ($P < 10^{-12}$ vs. $P < 10^{-7}$). Conversely, although there was equal enrichment for synaptic mitochondria in both modules, we found higher significance for glutamatergic synapse genes (M10 from the CTX network in ref. 2) in M5 relative to M4 for both species ($P < 10^{-26}$ vs. $P = 10^{-3}$ in mouse; $P < 10^{-240}$ vs. $P < 10^{-83}$ in human). As a specific example, we find that cyclin-dependent kinase 5 (CDK5) is the top interspecies marker for M5 (Table S4). Although cytoplasmic, CDK5 inhibition has been shown to rescue mitochondrial damage occurring from neurotoxic insults (25); therefore, the role of CDK5 as a mitochondrial hub, particularly one also highly coexpressed with glutamatergic synapse genes, is not unreasonable. Furthermore, unlike most other AD-related genes, CDK5 overexpression can result in similar disease phenotypes in mouse and human, causing neurodegeneration in mice (26) and playing a role in a number of neurodegenerative human diseases. These results suggest that, although the two mitochondrial modules are highly overlapping in both species, they represent separate, evolutionarily conserved biological components.

**Networks Correctly Sort Known Marker Genes by Cell Type.** In addition to being a useful resource, our list of interspecies marker genes (Table S3) provides an important validation for our methods. For example, although there may be some biological differences between cell types in mouse and human, we should find that a majority of known highly specific marker genes for cell type show strong coexpression with modules corresponding to relevant cell types. To address this issue, we measured the correlation of 40 highly cell type–specific markers for neurons, astrocytes, and oligodendrocytes in mouse (figure *3* in ref. 27) and 10 in human (figure *4* in ref. 2) against each ME in both network. We then calculated for which module each marker gene showed the highest correlation and where each gene ranked in that module by significance of MM (Table S5). In the case of oligodendrocytes, our network precisely matched expectations: nearly all known oligodendrocyte markers were reproduced in both our mouse and human networks, with known mouse markers tending to be hubs in our mouse network and known human markers tending to be hubs in our human network. Similarly, we found that the majority of known neuronal markers showed the highest correlation with neuron-associated modules, although not necessarily the module chosen for Table S3 (M13), consistent with the diverse neuronal populations throughout the brain. Although the results for astrocyte markers were less clear, the hubs between the network of Oldham et al. (2) and our human network were highly reproducible. Overall, these results suggest that our method can sort genes by cell type in both species to a relatively high degree of accuracy, and that lists of marker genes from our networks may provide valuable biological insight.

**Mouse Models in the Study of Human Disease.** Animal models are essential tools in the study of human disease, and have led to breakthroughs in nearly every area of medical science. As a result of their relative ease of genetic manipulation and short life spans, mice present especially useful animal models (28–30). Given the widely varying success of mouse models at mimicking human disease phenotypes, having a method to more accurately predict the effectiveness of model systems would be extremely useful. One possible strategy for predicting the effectiveness of specific mouse models in disease is to compare and contrast the human and mouse transcriptome. For example, coexpression preservation has been successfully used as a restrictive filter for predicting

which are the relevant human genes in disease loci (31). Furthermore, transcriptional analyses in human have found that genes causing the same disease tend to have shared expression patterns, a finding that is enhanced when data from other species are also included (32). Finally, phenotypic information from mouse homologues has been shown to improve DG prioritization beyond that which can be obtained using human expression data and associated categorization databases (e.g., GO and Kyoto Encyclopedia of Genes and Genomes) (33). By restricting our analysis to brain data and including the aforementioned modifications to standard transcriptional analysis, we believe that our results will be particularly useful in DG prioritization for neurological and neurodegenerative disorders, such as AD.

**DGs Show Differential Expression and Connectivity Patterns in Mouse and Human Brain.** Comparison of transcriptional programs across species has previously demonstrated that core metabolic functions and cellular structures are highly conserved between human and species as distant as *Escherichia coli* (19, 21). Similar gene expression preservation studies have been used to gain significant insight into disease in myriad cases, including in sleep and circadian rhythm biology (34, 35), in neurodegenerative disease (36), and as a filter for predicting DGs under a linkage peak (31), and is often a basic assumption of such work. We hypothesize that the opposite is also true. That is, we expect genes showing poor expression and connectivity preservation between species to be enriched for DGs associated with human-specific disease phenotypes. To assess the viability of this hypothesis, we first identified the sets of genes showing significantly high expression or node connectivity in mouse or human, but not both (genes in the upper left and lower right of Fig. S1 *A* and *B*). Unbiased GO annotation found enrichment for such disease-related genes: the top GO hit for genes with high connectivity in mouse, but not human was "disease mutation" (uncorrected $P < 0.005$) (10), whereas three of the main players in AD—apolipoprotein E, mitochondrial associated protein tau (MAPT), and PSEN1—have high connectivity in human, but not mouse.

Similar between-species differences in gene expression patterns can be seen at the level of modules. Apolipoprotein E shows high expression correlation with the astrocyte module (M3h) in the human network ($R = 0.43$), but not the mouse network ($R = -0.01$). As mentioned in the main text, glycogen synthase kinase-3β (GSK3B), a key protein involved in abnormal tau phosphorylation (37), is a hub gene for the poorly characterized, yet highly human-specific module M7h, which also contains MAPT (Fig. 3*E*). In contrast, CDK5—another AD-related kinase—shows different expression patterns (as described earlier). This is especially interesting given the recent evidence that GSK3B plays a dominant role in overall tau phosphorylation, whereas the main effects of CDK5 in AD progression are in the regulation of amyloidogenic APP processing (38) in addition to tau phosphorylation. Finally, there several other modules that are human specific, including one related to AD progression in humans (M9h; as detailed later). Given the human predilection for this disease, such genes and modules with divergent expression patterns become important candidates for studying the pathophysiology of AD in humans, and suggest that there may be a lot of information in control transcriptional networks regarding AD—as well as other neurodegenerative disorders—that has yet to be uncovered.

**Module M9h Is of Particular Interest in AD and Aging.** Interestingly, we found replication of the red module (12) in the human network in our current analysis. Not only does module M9h show very high within-species module preservation in human, in the sense that four of the hubs are replicated in both modules and further confirmed in the Celsius database (Fig. 3 *A* and *B*), but M9h also shows low between-species module preservation (Table 1), increasing the plausibility of its role in human-specific brain disease. To follow up on this finding, we performed WGCNA (*Materials and Methods*) on two additional large data sets in human, which were recently deposited in GEO (39, 40). The first study compared the relationship between gene expression and genomics in AD, finding that relative transcript levels are a good endophenotype for disease (40). From these data, which were run on the Illumina Human Refseq-8 microarray platform, we used 118 control and 95 AD samples from temporal cortex in our analysis. The second study compared gene expression changes between male and female across a wide range of ages in four different brain regions (hippocampus, entorhinal cortex, superior frontal gyrus, and postcentral gyrus) (39). This study on aging was run using the Affymetrix HG-U133 Plus 2.0 array and had 32 to 43 control samples for each brain region, all of which we used in our analysis.

In the AD study (40), we performed WGCNA using only the 118 control samples, and found a total of 24 modules, most of which showed either a significant increase or decrease in expression between control and AD. The black module showed significant overlap with M9h ($P < 10^{-4}$), including the common hub gene CXXC1 (Fig. S3*A*). Furthermore, when we compared the ME values between control and AD, we find that this module showed a significant increase in expression with AD progression ($P < 10^{-11}$; Fig. 3*C*). Thus we find modules from two separate data sets (40, 41) that both show increased expression in AD as well as significant overlap with M9h. Also, taking into consideration the fact that these studies were run in different laboratories, on different microarray platforms, and using tissue from different areas of the brain, we are confident that our result is biologically meaningful. To assess whether this module is unique to AD or whether it may also play a role in normal aging, we performed a second WGCNA analysis using aging data (39). In this analysis, one of the 10 modules we found (the yellow module) showed significant overlap with M9h ($P < 10^{-14}$; Fig. S3*B*) as well as positive correlation with age across all four brain regions ($P < 10^{-5}$; Fig. 3*D*). Therefore, in addition to the role of M9h in AD progression, M9h may also be involved in normal aging and possibly other neurodegenerative disorders. Finally, to assist other groups who may wish to follow up the results from this analysis, we have complied a ranked list of 50 genes that show high MM in M9h and all comparative modules (Table S7).

**Glossary of WGCNA and Comparative Network Terms.** Between-species preservation is any measure of preservation comparing data from human to data from mouse (e.g., Fig. S1 *A* and *B*)

A (coexpression) network is an undirected, weighted network with nodes corresponding to genes and edges based upon gene–gene coexpression levels. To evaluate coexpression levels between genes, Pearson correlations are taken and then weighted by raising their absolute value to a power. This weighting emphasizes strong correlations at the expense of weak ones.

A consensus network is a single network defined from multiple sources of data (in this case created from the weighted average of the correlation matrices from each human or mouse data set).

Expression preservation is the Pearson correlation between the ranked average expression value of genes across two sets of studies. Although this type of preservation is independent of network formation, it is useful in assessing the comparability of data sets across species (e.g., between the 18 human and 20 mouse data sets used in this analysis).

A hub is any highly connected gene. A hub can be characterized by high MM, high intramodular connectivity, or a strong presence in network depictions (e.g., circled genes in Fig. 3*A*).

Intramodular connectivity ($k_{in}$) is a measurement of network position that reflects how connected a given gene is with respect to the genes of a particular module. The higher the $k_{in}$, the more central a gene is to the network.

A module is a group of genes with strong sharing of coexpression relationships as measured by high TO. Modules are identified via hierarchical clustering (Fig. 1 *B* and *C*) using a measure of dissimilarity (i.e., 1 – TO). Genes in a module show much higher coexpression with each other (either positive or negative) than with genes outside the module.

Module characterization is a short descriptive term characterizing genes in a module based on GO or IPA annotation, module overlap with experimentally derived gene lists, and module overlap with modules previously characterized and published.

The ME is the first principal component of a module. The ME summarizes the characteristic expression pattern of a module.

MM is the Pearson correlation between the expression level of a given gene and a given ME. This quantity describes the extent to which a gene "belongs" to a module, and is used in the final module definitions.

Module (or list) overlap is the number of common genes between one module (or list) and a different module (or list). The significance of module overlap can be measured using the hypergeometric test.

Module preservation is any of a number of tests that measure how well characteristics of a module in one network are reproduced in another network. As the particulars of module preservation are beyond the scope of this article, we present a single Z-score that summarizes a variety of preservation measures (15).

(Node) connectivity preservation is the Pearson correlation between the ranked connectivities of genes common to two networks (e.g., the mouse and human networks).

(Overall) connectivity is the sum of connection strengths (adjacency matrix values) with all other network genes. The connectivity measures how correlated a gene is with all other genes in a network.

TO is a quantity describing gene pair similarity by comparing the weighted correlation of these genes with all other genes in the network.

(Weighted) adjacency matrix is a symmetric matrix whose off-diagonal elements lie between 0 and 1. The adjacencies measure the connection strength between pairs of nodes. In correlation networks, the adjacency between two genes is a power of the Pearson correlation between their expression profiles.

Within-species preservation is any measure of preservation comparing two sets of data in the same species (e.g., Fig. S1 *C–F*).


1. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4:e17.
2. Oldham MC, et al. (2008) Functional organization of the transcriptome in human brain. *Nat Neurosci* 11:1271–1282.
3. Lim WK, Wang K, Lefebvre C, Califano A (2007) Comparative analysis of microarray normalization procedures: effects on reverse engineering gene networks. *Bioinformatics* 23:i282–i288.
4. Langfelder P, Horvath S (2007) Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* 1:54.
5. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297:1551–1555.
6. Hermjakob H, et al. (2004) IntAct: An open source molecular interaction database. *Nucleic Acids Res* 32 (database issue):D452–D455.
7. Keshava Prasad TS, et al. (2009) Human Protein Reference Database—2009 update. *Nucleic Acids Res* 37 (database issue):D767–D772.
8. Langfelder P, Zhang B, Horvath S (2008) Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* 24:719–720.
9. Horvath S, Dong J (2008) Geometric interpretation of gene coexpression network analysis. *PLOS Comput Biol* 4:e1000117.
10. Hosack DA, Dennis G, Sherman BT, Lane HC, Lempicki RA (2003) Identifying biological themes within lists of genes with EASE. *Genome Biol* 4:R70.
11. Hu Z, Mellor J, Wu J, DeLisi C (2004) VisANT: An online visualization and analysis tool for biological interaction data. *BMC Bioinformatics* 5:17.
12. Miller JA, Oldham MC, Geschwind DH (2008) A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci* 28:1410–1420.
13. Oldham MC, Horvath S, Geschwind DH (2006) Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci USA* 103: 17973–17978.
14. Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA; Mouse Genome Database Group (2008) The Mouse Genome Database (MGD): Mouse biology and model systems. *Nucleic Acids Res* 36 (database issue):D724–D728.
15. Langfelder P, Horvath S (2008) WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559.
16. Ge H, Liu Z, Church GM, Vidal M (2001) Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. *Nat Genet* 29: 482–486.
17. Huang Y, et al. (2007) Systematic discovery of functional modules and context-specific functional annotation of human genome. *Bioinformatics* 23:i222–i229.
18. Lee HK, Hsu AK, Sajdak J, Qin J, Pavlidis P (2004) Coexpression analysis of human genes across many microarray data sets. *Genome Res* 14:1085–1094.
19. Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302:249–255.
20. Franz H, et al. (2005) Systematic analysis of gene expression in human brains before and after death. *Genome Biol* 6:R112.
21. Bergmann S, Ihmels J, Barkai N (2004) Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol* 2:E9.
22. Berg J, Lässig M (2006) Cross-species analysis of biological networks by Bayesian alignment. *Proc Natl Acad Sci USA* 103:10967–10972.
23. Tsaparas P, Mariño-Ramírez L, Bodenreider O, Koonin EV, Jordan IK (2006) Global similarity and local divergence in human and mouse gene co-expression networks. *BMC Evol Biol* 6:70.
24. Winden KD, et al. (2009) The organization of the transcriptional network in specific neuronal classes. *Mol Syst Biol* 5:291.
25. Sun KH, de Pablo Y, Vincent F, Shah K (2008) Deregulated Cdk5 promotes oxidative stress and mitochondrial dysfunction. *J Neurochem* 107:265–278.
26. Cruz JC, Tseng HC, Goldman JA, Shih H, Tsai LH (2003) Aberrant Cdk5 activation by p25 triggers pathological events leading to neurodegeneration and neurofibrillary tangles. *Neuron* 40:471–483.
27. Cahoy JD, et al. (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *J Neurosci* 28: 264–278.
28. Ahmad-Annuar A, Tabrizi SJ, Fisher EM (2003) Mouse models as a tool for understanding neurodegenerative diseases. *Curr Opin Neurol* 16:451–458.
29. Lin CH, et al. (2001) Neurological abnormalities in a knock-in mouse model of Huntington's disease. *Hum Mol Genet* 10:137–144.
30. Strand AD, et al. (2007) Conservation of regional gene expression in mouse and human brain. *PLoS Genet* 3:e59.
31. Ala U, et al. (2008) Prediction of human disease genes by human-mouse conserved coexpression analysis. *PLOS Comput Biol* 4:e1000043.
32. Oti M, van Reeuwijk J, Huynen MA, Brunner HG (2008) Conserved co-expression for candidate disease gene prioritization. *BMC Bioinformatics* 9:208.
33. Chen J, Xu H, Aronow BJ, Jegga AG (2007) Improved human disease candidate gene prioritization using mouse phenotype. *BMC Bioinformatics* 8:392.
34. Allada R, Emery P, Takahashi JS, Rosbash M (2001) Stopping time: The genetics of fly and mouse circadian clocks. *Annu Rev Neurosci* 24:1091–1119.
35. Ptácek LJ, Jones CR, Fu YH (2007) Novel insights from genetic and molecular characterization of the human clock. *Cold Spring Harb Symp Quant Biol* 72:273–277.
36. Karsten SL, et al. (2006) A genomic screen for modifiers of tauopathy identifies puromycin-sensitive aminopeptidase as an inhibitor of tau-induced neurodegeneration. *Neuron* 51:549–560.
37. Ishiguro K, et al. (1993) Glycogen synthase kinase 3 beta is identical to tau protein kinase I generating several epitopes of paired helical filaments. *FEBS Lett* 325: 167–172.
38. Wen Y, et al. (2008) Interplay between cyclin-dependent kinase 5 and glycogen synthase kinase 3 beta mediated by neuregulin signaling leads to differential effects on tau phosphorylation and amyloid precursor protein processing. *J Neurosci* 28: 2624–2632.
39. Berchtold NC, et al. (2008) Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc Natl Acad Sci USA* 105:15605–15610.
40. Webster JA, et al.; NACC-Neuropathology Group (2009) Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet* 84:445–458.
41. Blalock EM, et al. (2004) Incipient Alzheimer's disease: Microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proc Natl Acad Sci USA* 101:2173–2178.
42. Su AI, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 101:6062–6067.
43. Lu T, et al. (2004) Gene regulation and DNA damage in the ageing human brain. *Nature* 429:883–891.
44. Liang WS, et al. (2008) Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. *Physiol Genomics* 33:240–256.
45. Liang WS, et al. (2007) Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. *Physiol Genomics* 28:311–322.
46. Ryan MM, et al. (2006) Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol Psychiatry* 11:965–978.
47. Vawter MP, et al. (2004) Gender-specific gene expression in post-mortem human brain: localization to sex chromosomes. *Neuropsychopharmacology* 29:373–384.
48. Hodges A, et al. (2006) Regional and cellular gene expression changes in human Huntington's disease brain. *Hum Mol Genet* 15:965–977.

49. Lesnick TG, et al. (2007) A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet* 3:e98.
50. Moran L, et al. (2006) Whole genome expression profiling of the medial and lateral substantia nigra in Parkinson's disease. *Neurogenetics* 7:1–11.
51. Levenson JM, et al. (2004) A bioinformatics analysis of memory consolidation reveals involvement of the transcription factor c-rel. *J Neurosci* 24:3933–3943.
52. Natale JE, Ahmed F, Cernak I, Stoica B, Faden AI (2003) Gene expression profile changes are commonly modulated across models and species after traumatic brain injury. *J Neurotrauma* 20:907–927.
53. Benn CL, et al. (2005) Contribution of nuclear and extranuclear polyQ to neurological phenotypes in mouse models of Huntington's disease. *Hum Mol Genet* 14:3065–3078.
54. Hovatta I, et al. (2005) Glyoxalase 1 and glutathione reductase 1 regulate anxiety in mice. *Nature* 438:662–666.
55. Zapala MA, et al. (2005) Adult mouse brain gene expression patterns bear an embryologic imprint. *Proc Natl Acad Sci USA* 102:10357–10362.
56. Keeley MB, et al. (2006) Differential transcriptional response to nonassociative and associative components of classical fear conditioning in the amygdala and hippocampus. *Learn Mem* 13:135–142.
57. Majdan M, Shatz CJ (2006) Effects of visual experience on activity-dependent gene regulation in cortex. *Nat Neurosci* 9:650–659.
58. Hovatta I, et al. (2007) DNA variation and brain region-specific expression profiles exhibit different relationships between inbred mouse strains: implications for eQTL mapping studies. *Genome Biol* 8:R25.
59. Fernandes C, et al. (2004) Hippocampal gene expression profiling across eight mouse inbred strains: Towards understanding the molecular basis for behaviour. *Eur J Neurosci* 19:2576–2582.
60. Somel M, et al. (2008) Human and chimpanzee gene expression differences replicated in mice fed different diets. *PLoS ONE* 3:e1504.
61. Mackiewicz M, et al. (2007) Macromolecule biosynthesis: a key function of sleep. *Physiol Genomics* 31:441–457.
62. Maret S, et al. (2007) Homer1a is a core brain molecular correlate of sleep loss. *Proc Natl Acad Sci USA* 104:20090–20095.
63. Kuhn A, et al. (2007) Mutant huntingtin's effects on striatal gene expression in mice recapitulate changes observed in human Huntington's disease brain and do not differ with mutant huntingtin length or wild-type huntingtin dosage. *Hum Mol Genet* 16: 1845–1861.
64. Lein ES, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168–176.
65. Gan L, et al. (2004) Identification of cathepsin B as a mediator of neuronal death induced by Abeta-activated microglial cells using a functional genomics approach. *J Biol Chem* 279:5565–5572.
66. Albright AV, González-Scarano F (2004) Microarray analysis of activated mixed glial (microglia) and monocyte-derived macrophage gene expression. *J Neuroimmunol* 157: 27–38.
67. Day A, Carlson MR, Dong J, O'Connor BD, Nelson SF (2007) Celsius: A community resource for Affymetrix microarray data. *Genome Biol* 8:R112.

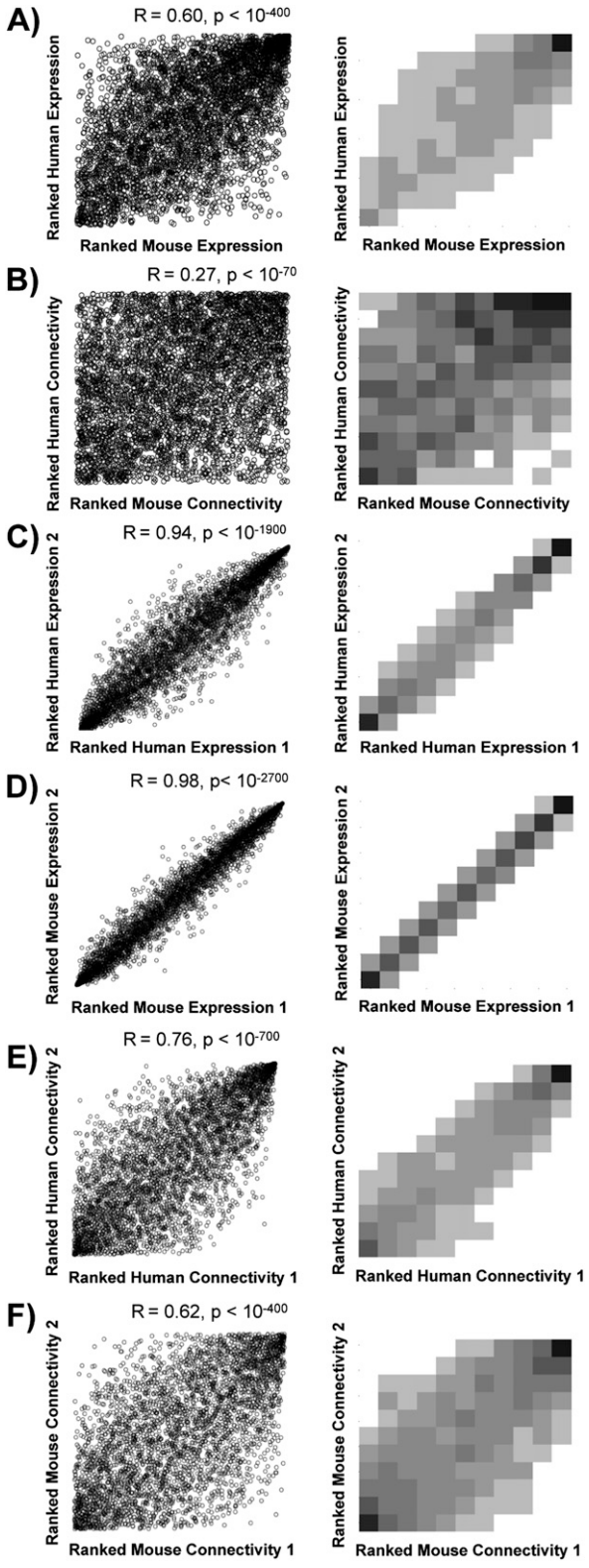


**Fig. S1.** Networks show significant between- and within-species expression and connectivity preservation. Gene expression (*A*) and connectivity (*B*) show significant preservation between mouse and human. The *x* and *y* axes represent the average expression rank (*A*) and rank of overall connectivity (*B*) across studies in the mouse and human data sets, respectively. Genes with high expression/connectivity in both mouse and human are in the upper right of each plot. Gene expression is preserved within-species both for human (*C*) and mouse (*D*). The *x* and *y* axes both represent the average expression rank across a subset of nine random mouse/human studies. Gene connectivity is preserved within-species both for human (*E*) and mouse (*F*). The *x* and *y* axes both represent the rank of overall connectivity for each gene in a network built using the same subset of nine random mouse/human studies. The randomization of studies was performed twice with comparable results. Dot plots (*Left*) and density plots (*Right*) present the same information in different ways.

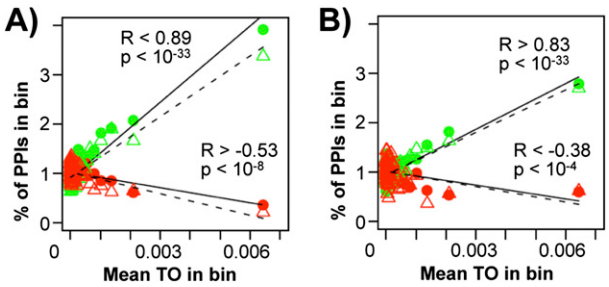


**Fig. S2.** TO reflects PPIs in both the human (*A*) and mouse (*B*) networks, with positively correlated (green) and negatively correlated (red) genes showing opposite effects. This result was replicated in both the HPRD (solid points and lines) (1) as well as the IntAct (hollow points and dashed lines) databases (2). The *y* axis represents the percent of all interactions from a given comparison contained in each of 100 bins of gene pairs sorted based on TO, whereas the *x* axis represents the average TO of each bin.

1. Keshava Prasad TS, et al. (2009) Human Protein Reference Database—2009 update. *Nucleic Acids Res 37*(database issue):D767–D772.
2. Hermjakob H, et al. (2004) IntAct: An open source molecular interaction database. *Nucleic Acids Res 32*(database issue):D452–D455.

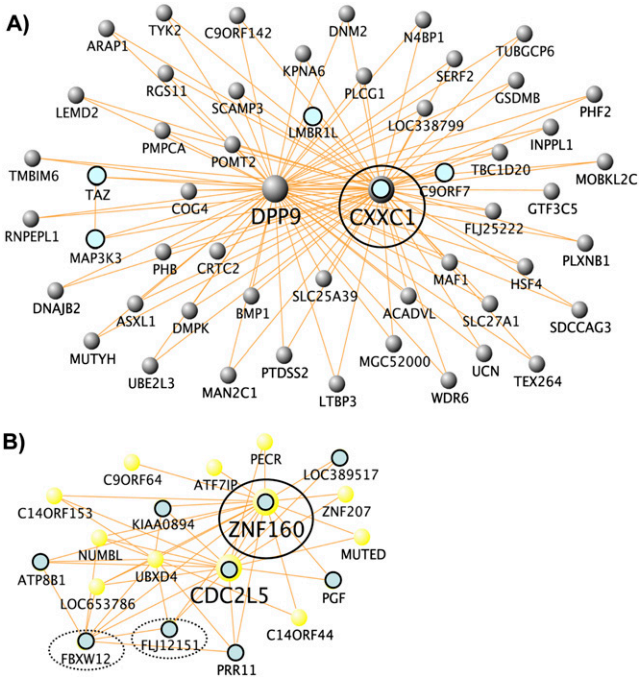


**Fig. S3.** M9h is preserved across studies. (*A*) Network depiction of a subset of the black module from ref. 1 shows that this module shares a common hub (CXXC1) with M9h. Labels as in Fig. 3*A*, except only approximately 100 connections are shown. (*B*) Network depiction of a subset of the yellow module from ref. 2 shows that this module contains multiple hub genes of M9h (i.e., ZNF160). Labels as in Fig. 3*A*, except only approximately 60 connections are shown.

1. Webster JA, et al.; NACC-Neuropathology Group (2009) Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet* 84:445–458.
2. Berchtold NC, et al. (2008) Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc Natl Acad Sci USA* 105:15605–15610.

**Table S1. Data sets used in the human and mouse networks**

| Species | GEO no. | Index | Arrays | Used* | Chip | Cel | Description | Subset/Comment | Refs. |
|---|---|---|---|---|---|---|---|---|---|
| Human | GSE1133 | 1 | 158 | 22 | HG-U133A | Y | Samples from multiple brain/CNS regions of controls | Cortex, GP, caudate, medulla, thals, cingulate, pons | 42 |
| Human | GSE1297 | 2 | 28 | 26 | HG-U133A | Y | AD at various stages of severity in CA1 | Used all data | 41 |
| Human | GSE1572 | 3 | 30 | 24 (24) | HG-U95A | Y | FC of controls aged 26–106 y | Used all data | 43 |
| Human | GSE3526A | 4 | 353 | 33 | HG-U133 Plus 2 | Y | Samples from multiple brain/CNS regions of controls | VTA, medulla, SN, corpus callosum, thals, midbrain, VN | — |
| Human | GSE3526B | 5 | 353 | 25 (21) | HG-U133 Plus 2 | Y | Samples from multiple brain/CNS regions of controls | Mostly cortex, (2 × HP, 2 × amygdala) | — |
| Human | GSE4036 | 6 | 28 | 24 | HG-U133 Plus 2 | Y | CB of schizophrenia patients and controls | Used all data | — |
| Human | GSE4757 | 7 | 20 | 20 | HG-U133 Plus 2 | Y | LCM of NFT-laden and NFT-free pyramidal cells in entorhinal cortex of mid-AD patients | Used all data | — |
| Human | GSE5281A | 8 | 161 | 34 | HG-U133 Plus 2 | Y | LCM of pyramidal cells from HP, EC, MTG, PC, SFG, and PVC | Controls (only 2 × EC, 2 × MTG) | 44,45 |
| Human | GSE5281B | 9 | 161 | 34 | HG-U133 Plus 2 | Y | LCM of pyramidal cells from HP, EC, MTG, PC, SFG, and PVC | AD subjects (only 3 × MTG, 0 × EC) | 44,45 |
| Human | GSE5392A | 10 | 82 | 34 (34) | HG-U133A | Y | OFC and DLPFC prefrontal cortex from controls | Controls | 46 |
| Human | GSE5392B | 11 | 82 | 31 | HG-U133A | Y | OFC and DLPFC from adults with bipolar disorder | Bipolar subjects | 46 |
| Human | GSE9770 | 12 | 34 | 29 | HG-U133 Plus 2 | Y | LCM of pyramidal cells from 6 brain regions of patients with MCI | Used all data | — |
| Human | GSE2164B | 13 | 87 | 40 (40) | HG-U95A | Y | Male and female DLPFC, ACC, and CB | DLPFC and ACC | 47 |
| Human | GSE3790A | 14 | 202 | 27 | HG-U133A | N | CB, FC [BA4, BA9] and CN from HD patients and controls | CN | 48 |
| Human | GSE3790B | 15 | 202 | 38 (18) | HG-U133A | N | CB, FC [BA4, BA9] and CN from HD patients and controls | FC | 48 |
| Human | GSE3790C | 16 | 202 | 38 | HG-U133A | N | CB, FC [BA4, BA9] and CN from HD patients and controls | CB | 48 |
| Human | GSE7621 | 17 | 25 | 23 | HG-U133 Plus 2 | Y | Substantia nigra from PD patients | Used all data | 49 |
| Human | GSE8397 | 18 | 47 | 27 | HG-U133A | Y | SFG and SN of PD patients and controls | SN only (mostly med/lat SN of PD patients) | 50 |
| Mouse | GSE1482 | 1 | 30 | 29 | MG-U74A | Y | HP of control and $NF1^{+/-}$ mice between 10 and 32 d postnatal | Used all data | — |
| Mouse | GSE1782A | 2 | 64 | 18 | MG-U74A | N | CA1 and DG of mice in control and fear conditioned situations | CA1 | 51 |
| Mouse | GSE1782B | 3 | 64 | 21 | MG-U74A | N | CA1 and DG of mice in control and fear conditioned situations | DG | 51 |
| Mouse | GSE2392 | 4 | 22 | 22 | MG-U74A | Y | Whole brain 0, 4, 8, 24, 72 h post traumatic brain injury | Used all data | 52 |
| Mouse | GSE3248 | 5 | 48 | 44 | MG-U74A | Y | CB of control and HD mutant mice | Used all data | 53 |
| Mouse | GSE3327A | 6 | 87 | 24 | MG-U74A | Y | Samples from 7 brain regions of 6 inbred strains | Amygdala and cingulate cortex | 54 |
| Mouse | GSE3327B | 7 | 87 | 34 | MG-U74A | Y | Samples from 7 brain regions of 6 inbred strains | BN, HP, hypothalamus | 54 |
| Mouse | GSE3594C | 8 | 150 | 21 (15) | MG-U74A | N | Samples from 24 neural and 10 nonneural tissues | Amygdala and cortex | 55 |
| Mouse | GSE3963A | 9 | 48 | 22 | MG-U74A | N | HP and amygdala of fear conditioned and control mice | HP | 56 |
| Mouse | GSE3963B | 10 | 48 | 20 | MG-U74A | N | HP and amygdala of fear conditioned and control mice | Amygdala | 56 |
| Mouse | GSE4269 | 11 | 36 | 26 | MG-U74A | N | Time course: PVC of light deprived/nondeprived mice | Used all data | 57 |
| Mouse | GSE4734 | 12 | 61 | 34 | MG-U430A | N | Mouse strain and brain region comparison | BN, hypothalamus, and PG | 58 |
| Mouse | GSE5429 | 13 | 32 | 29 | MG-U74A | Y | HP from 8 inbred strains | Used all data | 59 |
| Mouse | GSE6285 | 14 | 24 | 22 (22) | MG-U430A | Y | Whole brain of mice fed 4 different diets | Used all data | 60 |
| Mouse | GSE6514A | 15 | 90 | 37 (20) | MG-U430A | Y | Cerebral cortex/hypothalamus during spontaneous sleep and prolonged wakefulness | Cortex | 61 |

**Table S1. Cont.**

| Species | GEO no. | Index | Arrays | Used* | Chip | Cel | Description | Subset/Comment | Refs. |
|---|---|---|---|---|---|---|---|---|---|
| Mouse | GSE6514B | 16 | 90 | 42 | MG-U430A | Y | Cerebral cortex/hypothalamus during spontaneous sleep and prolonged wakefulness | Hypothalamus | 61 |
| Mouse | GSE9444A | 17 | 113 | 21 | MG-U430A | Y | Sleep deprivation and the brain (omit 850–867 -> liver) | AKR/J mice | 62 |
| Mouse | GSE9444B | 18 | 113 | 23 | MG-U430A | Y | Sleep deprivation and the brain (omit 850–867 -> liver) | C57BL/6J mice | 62 |
| Mouse | GSE9444C | 19 | 113 | 23 | MG-U430A | Y | Sleep deprivation and the brain | DBA/2J mice | 62 |
| Mouse | GSE10263 | 20 | 26 | 25 | MG-U430A | Y | Mutant HD on striatal gene expression in mice | Used all data | 63 |

Data sets used in the human and mouse networks. A list and extended summary of all of the data sets, as well as the number of samples listed on GEO vs. the number used in each data set in the meta-analysis, and whether or not cel files (raw data) was available. For subsetted data sets and data sets where a large portion of the samples were omitted, the "Subset/Comment" column summarizes which phenotypes are included in the final expression file. References for each data set, when listed on the GEO website, are presented in the *SI Text*. ACC, anterior cingulate cortex; BN, bed nucleus of stria terminalus; CB, cerebellum; CN, caudate nucleus; CNS, central nervous system; DG, dentate gyrus; DLPFC, dorsolateral prefrontal cortex; EC, entorhinal cortex; FC, frontal cortex; GP, globus pallidus; HP, hippocampus; LCM, laser capture microdissection; MCI, mild cognitive impairment; MTG, medial temporal gyrus; NFT, neurofibrillary tangles; OFC, oribitofrontal cortex; PC, posterior cingulate; PD, Parkinson disease; PG, periaqueductal gray; PVC, primary visual cortex; SFG, superior frontal gyrus; thals, thalamus, subthalamus, and hypothalamus; VN, vestibular nucleus; VTA, ventral tegmental area.
*In the "Used" column, the number in parentheses represents the number of samples included in the "cortex/control only" analysis.

**Table S2. Selected GO and IPA annotations for each module**

| Module | Network | Gene category | List hits | List total | Population hits | Population total | EASE score | Bonferroni correction |
|---|---|---|---|---|---|---|---|---|
| M1h | Human | Mitochondrion | 60 | 363 | 445 | 5,952 | 4.12E-09 | 1.14E-05 |
| M1h | Human | NADH dehydrogenase (ubiquinone) activity | 12 | 363 | 29 | 6,071 | 3.81E-07 | 1.05E-03 |
| M1h | Human | CNS-specific functions | 24 | 90 | 168 | 1,671 | 8.81E-06 | 2.44E-02 |
| M1h | Human | Hs_glycolysis and gluconeogenesis | 11 | 79 | 27 | 849 | 5.66E-05 | 1.57E-01 |
| M2h | Human | Short-chain dehydrogenases/ reductases family signature | 6 | 93 | 12 | 1818 | 1.78E-04 | 5.13E-01 |
| M2h | Human | Myelin | 5 | 112 | 7 | 1987 | 2.84E-04 | 8.17E-01 |
| M2h | Human | Cytoskeleton organization and biogenesis | 23 | 360 | 166 | 6,014 | 3.15E-04 | 9.07E-01 |
| M3h | Human | Homeostasis | 17 | 422 | 55 | 6,014 | 5.87E-07 | 2.05E-03 |
| M3h | Human | Amino acid metabolism - *Homo sapiens* | 29 | 82 | 170 | 1,168 | 2.13E-06 | 7.42E-03 |
| M3h | Human | Copper ion homeostasis | 7 | 422 | 9 | 6,014 | 7.96E-06 | 2.78E-02 |
| **M4h** | **Human** | **Mitochondrion** | **125** | **474** | **445** | **5,952** | **4.70E-40** | **1.53E-36** |
| **M4h** | **Human** | **NADH dehydrogenase (ubiquinone) activity** | **27** | **482** | **29** | **6,071** | **3.63E-26** | **1.18E-22** |
| **M4h** | **Human** | **Proteasome** | **16** | **397** | **34** | **4,833** | **1.69E-08** | **5.52E-05** |
| M4h | Human | *Homo sapiens* 19 | 65 | 507 | 408 | 6,518 | 2.12E-08 | 6.91E-05 |
| **M4h** | **Human** | **Acetylation** | **28** | **397** | **120** | **4,833** | **8.10E-07** | **2.64E-03** |
| *M4h* | *Human* | *Mitochondrial ribosome* | *8* | *474* | *20* | *5,952* | *5.95E-04* | *1.00E+00* |
| **M4m** | **Mouse** | **Mitochondrion** | **77** | **259** | **334** | **3,963** | **6.71E-25** | **1.29E-21** |
| **M4m** | **Mouse** | **NADH dehydrogenase (ubiquinone) activity** | **18** | **261** | **28** | **4,043** | **3.75E-14** | **7.20E-11** |
| *M4m* | *Mouse* | *Mitochondrial ribosome* | *14* | *259* | *23* | *3,963* | *1.79E-10* | *3.43E-07* |
| M4m | Mouse | Ribosome - Homo sapiens | 17 | 95 | 26 | 862 | 2.54E-10 | 4.87E-07 |
| **M4m** | **Mouse** | **Acetylation** | **23** | **200** | **96** | **3,350** | **1.85E-08** | **3.55E-05** |
| M4m | Mouse | RNA binding | 34 | 261 | 219 | 4,043 | 2.28E-06 | 4.38E-03 |
| **M4m** | **Mouse** | **Proteasome** | **11** | **200** | **32** | **3,350** | **8.92E-06** | **1.71E-02** |
| **M5h** | **Human** | **Mitochondrion** | **109** | **670** | **445** | **5,952** | **4.26E-16** | **1.70E-12** |
| M5h | Human | Coated vesicle | 30 | 670 | 57 | 5,952 | 1.17E-13 | 4.67E-10 |
| M5h | Human | NADH dehydrogenase (ubiquinone) activity | 19 | 686 | 29 | 6,071 | 7.41E-11 | 2.95E-07 |
| M5h | Human | ATP synthesis | 16 | 557 | 21 | 4,833 | 1.93E-10 | 7.66E-07 |
| M5h | Human | Clathrin coat | 12 | 670 | 17 | 5,952 | 2.19E-07 | 8.72E-04 |
| M5h | Human | Synapse | 14 | 670 | 24 | 5,952 | 3.13E-07 | 1.25E-03 |
| *M5h* | *Human* | *Homo sapiens* 19 | *78* | *723* | *408* | *6,518* | *1.05E-06* | *4.18E-03* |
| *M5h* | *Human* | *Proteasome subunit* | *15* | *334* | *35* | *2,897* | *1.30E-05* | *5.17E-02* |
| *M5m* | *Mouse* | *Homo sapiens* 19 | *47* | *265* | *280* | *4,290* | *2.25E-10* | *5.00E-07* |
| **M5m** | **Mouse** | **Mitochondrion** | **51** | **238** | **334** | **3,963** | **2.86E-10** | **6.35E-07** |
| M5m | Mouse | Primary active transporter activity | 22 | 251 | 112 | 4,043 | 3.17E-06 | 7.04E-03 |
| *M5m* | *Mouse* | *Proteasome* | *11* | *199* | *32* | *3,350* | *8.52E-06* | *1.89E-02* |
| *M6h* | *Human* | *Coated vesicle* | *19* | *349* | *57* | *5,952* | *1.57E-09* | *4.36E-06* |
| M6h | Human | Neuronal transmission | 27 | 107 | 144 | 1,671 | 2.66E-07 | 7.38E-04 |
| *M6h* | *Human* | *Neurogenesis* | *39* | *360* | *275* | *6,014* | *7.15E-07* | *1.98E-03* |
| M6m | Mouse | Homo sapiens 19 | 49 | 241 | 280 | 4,290 | 4.39E-13 | 8.78E-10 |
| M6m | Mouse | Synapse | 8 | 221 | 17 | 3,963 | 1.78E-05 | 3.57E-02 |
| *M6m* | *Mouse* | *Coated vesicle* | *12* | *221* | *47* | *3,963* | *3.55E-05* | *7.10E-02* |
| *M6m* | *Mouse* | *neurogenesis* | *26* | *221* | *199* | *4,002* | *6.62E-05* | *1.32E-01* |
| M7h | Human | mRNA binding | 7 | 77 | 73 | 6,071 | 2.76E-04 | 2.79E-01 |
| M7h | Human | Nucleus | 36 | 78 | 1,622 | 5,952 | 4.34E-04 | 4.38E-01 |
| M7h | Human | Neuronal cell recognition | 3 | 79 | 4 | 6,014 | 9.80E-04 | 9.91E-01 |
| M8h | Human | Defense response | 67 | 299 | 373 | 6,014 | 4.26E-21 | 1.13E-17 |
| M8h | Human | Glycoprotein | 100 | 278 | 1,045 | 4,833 | 1.98E-08 | 5.27E-05 |
| M8h | Human | Regulation of cell proliferation | 22 | 299 | 150 | 6,014 | 1.30E-05 | 3.46E-02 |
| M8m | Mouse | Heat shock protein activity | 8 | 107 | 20 | 4,043 | 4.09E-07 | 5.45E-04 |

**Table S2. Cont.**

| Module | Network | Gene category | List hits | List total | Population hits | Population total | EASE score | Bonferroni correction |
|---|---|---|---|---|---|---|---|---|
| M8m | Mouse | Regulation of transcription | 37 | 107 | 624 | 4,002 | 1.68E-06 | 2.24E-03 |
| M9h | Human | Homo sapiens 19 | 20 | 134 | 408 | 6,518 | 5.65E-04 | 7.93E-01 |
| M9h | Human | Nuclear division | 8 | 120 | 93 | 6,014 | 2.31E-03 | 1.00E+00 |
| M9h | Human | Alternative splicing | 35 | 88 | 1,214 | 4,833 | 2.62E-03 | 1.00E+00 |
| M9h | Human | M phase | 8 | 120 | 99 | 6,014 | 3.30E-03 | 1.00E+00 |
| M9h | Human | Regulation of cell cycle | 14 | 120 | 283 | 6,014 | 3.71E-03 | 1.00E+00 |
| M10h | Human | Defense response | 80 | 244 | 373 | 6,014 | 4.34E-38 | 9.96E-35 |
| M10h | Human | Antigen presentation | 14 | 244 | 21 | 6,014 | 8.52E-14 | 1.96E-10 |
| M10h | Human | Glycoprotein | 96 | 222 | 1,045 | 4,833 | 2.33E-13 | 5.35E-10 |
| M10h | Human | Polymorphism | 79 | 222 | 894 | 4,833 | 9.74E-10 | 2.24E-06 |
| M10h | Human | Ribosome - Homo sapiens | 17 | 52 | 71 | 1,168 | 8.72E-09 | 2.00E-05 |
| M10m | Mouse | Molecular function unknown | 4 | 17 | 176 | 4,043 | 2.99E-02 | 1.00E+00 |
| M10m | Mouse | Development | 6 | 17 | 706 | 4,002 | 1.36E-01 | 1.00E+00 |
| M10m | Mouse | Polymorphism | 5 | 13 | 642 | 3,350 | 1.83E-01 | 1.00E+00 |
| M11h | Human | Mitochondrion | 115 | 682 | 445 | 5,952 | 2.63E-18 | 1.08E-14 |
| M11h | Human | Proteasome subunit | 21 | 357 | 35 | 2,897 | 2.12E-10 | 8.67E-07 |
| M11h | Human | Main pathways of carbohydrate metabolism | 27 | 685 | 61 | 6,014 | 4.51E-10 | 1.84E-06 |
| M11h | Human | ATP synthesis - Homo sapiens | 19 | 214 | 29 | 1,168 | 1.36E-07 | 5.58E-04 |
| **M12h** | **Human** | **Ribosome - Homo sapiens** | **64** | **83** | **71** | **1,168** | **1.12E-77** | **1.51E-74** |
| M12h | Human | Antigen presentation | 10 | 148 | 21 | 6,014 | 5.58E-10 | 7.51E-07 |
| M12m | Mouse | Mitochondrion | 74 | 298 | 334 | 3,963 | 9.58E-19 | 2.09E-15 |
| M12m | Mouse | RNA binding | 46 | 300 | 219 | 4,043 | 7.68E-11 | 1.68E-07 |
| **M12m** | **Mouse** | **Ribosome - Homo sapiens** | **18** | **105** | **26** | **862** | **8.87E-11** | **1.94E-07** |
| M12m | Mouse | Mitochondrial ribosome | 14 | 298 | 23 | 3,963 | 1.05E-09 | 2.30E-06 |
| M12m | Mouse | NADH dehydrogenase (ubiquinone) activity | 14 | 300 | 28 | 4,043 | 2.09E-08 | 4.56E-05 |
| M12m | Mouse | Acetylation | 22 | 237 | 96 | 3,350 | 1.76E-06 | 3.84E-03 |
| M12m | Mouse | Proteasome | 12 | 237 | 32 | 3,350 | 5.77E-06 | 1.26E-02 |
| M13h | Human | Transport | 142 | 561 | 1,026 | 6,014 | 1.99E-07 | 7.55E-04 |
| M13h | Human | Secretory pathway | 23 | 561 | 86 | 6,014 | 7.80E-06 | 2.96E-02 |
| M13h | Human | *Homo sapiens* Xq | 30 | 611 | 146 | 6,518 | 6.62E-05 | 2.51E-01 |
| M13h | Human | Mitochondrion | 67 | 557 | 445 | 5,952 | 6.90E-05 | 2.61E-01 |
| M13m | Mouse | Neuronal transmission | 22 | 71 | 96 | 1,175 | 2.56E-08 | 6.03E-05 |
| M13m | Mouse | Signal transduction | 86 | 226 | 924 | 4,002 | 2.37E-07 | 5.58E-04 |
| M13m | Mouse | Acetylcholine receptor | 7 | 95 | 9 | 1,660 | 2.06E-06 | 4.85E-03 |
| M13m | Mouse | GTPase activity | 17 | 227 | 97 | 4,043 | 7.29E-05 | 1.71E-01 |
| M14h | Human | Intracellular | 346 | 458 | 3,968 | 5,952 | 1.29E-05 | 4.08E-02 |
| M14h | Human | *Homo sapiens* 4q | 29 | 501 | 165 | 6,518 | 4.74E-05 | 1.50E-01 |
| M14h | Human | Heterogeneous nuclear ribonucleoprotein complex | 5 | 458 | 7 | 5,952 | 9.96E-04 | 1.00E+00 |
| *M14h* | *Human* | *RNA-binding* | *22* | *347* | *151* | *4,833* | *2.27E-03* | *1.00E+00* |
| *M14h* | *Human* | *Nucleus* | *148* | *458* | *1,622* | *5,952* | *8.87E-03* | *1.00E+00* |
| M14m | Mouse | 6.3.2.19 [ubiquitin] | 8 | 68 | 12 | 807 | 1.14E-05 | 3.45E-02 |
| M14m | Mouse | Protein transport | 42 | 437 | 196 | 4,002 | 1.94E-05 | 5.87E-02 |
| *M14m* | *Mouse* | *RNA-binding* | *45* | *450* | *219* | *4,043* | *4.49E-05* | *1.36E-01* |
| *M14m* | *Mouse* | *Nucleus* | *156* | *427* | *1,131* | *3,963* | *1.22E-04* | *3.70E-01* |
| *M15h* | *Human* | *Glycoprotein* | *83* | *188* | *1,045* | *4,833* | *3.19E-12* | *6.93E-09* |
| M15h | Human | Signal transducer activity | 71 | 241 | 1,117 | 6,071 | 2.25E-05 | 4.90E-02 |
| *M15h* | *Human* | *G protein coupled receptor* | *17* | *188* | *128* | *4,833* | *2.74E-05* | *5.96E-02* |
| M15h | Human | Carcinoembryonic antigen | 6 | 99 | 10 | 2,228 | 3.15E-05 | 6.85E-02 |
| M15h | Human | Homo sapiens 19 | 33 | 252 | 408 | 6,518 | 8.02E-05 | 1.74E-01 |
| *M15m* | *Mouse* | *G protein coupled receptor* | *21* | *242* | *84* | *3,350* | *1.05E-06* | *2.69E-03* |
| *M15m* | *Mouse* | *Glycoprotein* | *72* | *242* | *729* | *3,350* | *2.24E-03* | *1.00E+00* |

Selected GO categorizations for each module in the mouse and human networks. At least one GO category corresponding to each significant cellular process or component is included for each relevant module. Categories with significant overlap between matched modules in the mouse and human modules are highlighted in bold (corrected $P < 0.05$) or italics (uncorrected $P < 0.01$). GO categories were found using EASE (ref. 10), with the EASE score used as an approximation for $P$ value. Categories confirmed using Ingenuity pathways analysis are underlined. EASE, Expression Analysis Systematic Explorer.

**Table S3. Interspecies marker genes for brain cell types**

| Marker | References |
|---|---|
| Oligodendrocyte (M2h) | |
| MOG | 2, 27, 64 |
| EVI2A | 2, 27 |
| MAG | 2, 27, 64 |
| GPR37 | 2, 27, 64 |
| MBP | 2, 27, 64 |
| RNF20 | 2, 27, 64 |
| SCD | 2 |
| PLP1 | 2, 27 |
| CLDN11 | 27 |
| CSRP1 | 64 |
| FNBP1 | 2 |
| TMEM 39 | 2 |
| Neuron (M13h) | |
| SNX10 | 2, 27 |
| HMP20 | 27, 64 |
| PGRMC1 | 2, 27 |
| GUCY1B3 | 2, 27 |
| MAPK1 | 2, 27 |
| PPP3CB | 27 |
| ACTR3 | 27 |
| DLD | – |
| ARMCX2 | 27 |
| SEH1L | 27 |
| SUMO3 | 2 |
| MAP2K1 | 2, 64 |
| Astrocyte (M3h) | |
| ZFP36L1 | 2, 27 |
| SOX2 | 27 |
| F3 | 2, 27 |
| TOB1 | 2 |
| HIF1A | 2 |
| TSPAN6 | 2 |
| GJA1 | 2, 27, 64 |
| ITGA6 | 2, 27 |
| HRSP 3 | 2, 27 |
| TJP1 | 27 |
| MCL1 | – |
| ZFP36L2 | 2, 27 |
| Microglia (M10h) | |
| TYROBP | 2, 65 |
| ARHGDIB | 2 |
| LY86 | 2 |
| S100A11 | – |
| CD21 | 2, 66 |
| FCGR1A | 2, 66 |
| CYBA | 2 |
| CSF1R | 2, 66 |
| TSPO | – |
| LAPTM3 | 2 |
| HMOX1 | 66 |
| GPX1 | 65 |

Genes were ranked based on interspecies hub status for the indicated cell-type related modules, omitting genes that showed high coexpression with multiple such modules. The "References" column indicates which studies have previously found these genes to be markers for their respective cell types, demonstrating that these genes show a high overlap with known markers ($P < 10^{-10}$ for all cell types). In ref. 2, genes in modules M9A, M15A, M16A, and M4A with $R > 0.5$ were considered markers of oligodendrocytes, astrocytes, neurons, and microglia, respectively. Neither refs. 27 nor 64 had microglia as a cell type, whereas refs. 65 and 66 only tested microglia in mouse and human, respectively.

**Table S4. Top 20 interspecies marker genes for each module**

| Module | Gene | *P* value rank | |
|---|---|---|---|
| | | Mouse | Human |
| M1 | *SNAP91* | 11 | 5 |
| M1 | *GUCY1B3* | 10 | 15 |
| M1 | *TRIM37* | 15 | 2 |
| M1 | *SCN1A* | 18 | 4 |
| M1 | *NPTN* | 29 | 10 |
| M1 | *ITPR1* | 35 | 23 |
| M1 | *GABRG2* | 41 | 17 |
| M1 | *SNAP25* | 5 | 44 |
| M1 | *SERPINI1* | 20 | 72 |
| M1 | *SH3GL2* | 74 | 35 |
| M1 | *GABRA1* | 6 | 76 |
| M1 | *PLCB1* | 37 | 89 |
| M1 | *FBXW7* | 58 | 90 |
| M1 | *MAT2B* | 1 | 92 |
| M1 | *RNF11* | 28 | 93 |
| M1 | *CCNI* | 22 | 94 |
| M1 | *NARS* | 99 | 22 |
| M1 | *USP33* | 8 | 115 |
| M1 | *OLFM1* | 116 | 48 |
| M1 | *CAB39* | 25 | 144 |
| M2 | *MOG* | 6 | 18 |
| M2 | *EVI2A* | 5 | 25 |
| M2 | *MAG* | 13 | 34 |
| M2 | *GPR37* | 35 | 5 |
| M2 | *RDX* | 41 | 40 |
| M2 | *SYPL1* | 4 | 44 |
| M2 | *MBP* | 25 | 49 |
| M2 | *RNF13* | 19 | 50 |
| M2 | *SCD* | 30 | 51 |
| M2 | *PLP1* | 57 | 11 |
| M2 | *SEPT4* | 21 | 58 |
| M2 | *STXBP3* | 17 | 61 |
| M2 | *CLDN11* | 28 | 62 |
| M2 | *CSRP1* | 11 | 89 |
| M2 | *S100B* | 66 | 93 |
| M2 | *LITAF* | 108 | 38 |
| M2 | *GSN* | 111 | 48 |
| M2 | *FNBP1* | 116 | 31 |
| M2 | *LAMP2* | 123 | 7 |
| M2 | *SPG20* | 92 | 129 |
| M3 | *ZFP36L1* | 10 | 26 |
| M3 | *TMEM123* | 28 | 35 |
| M3 | *SOX2* | 9 | 36 |
| M3 | *PCAF* | 67 | 42 |
| M3 | *F3* | 75 | 57 |
| M3 | *TOB1* | 13 | 78 |
| M3 | *HIF1A* | 51 | 92 |
| M3 | *TSPAN6* | 109 | 99 |
| M3 | *SEPT2* | 110 | 61 |
| M3 | *GJA1* | 142 | 1 |
| M3 | *ITGA6* | 146 | 82 |
| M3 | *HRSP12* | 147 | 60 |
| M3 | *MAPRE1* | 156 | 76 |
| M3 | *TJP1* | 160 | 126 |
| M3 | *ZFP36L2* | 171 | 113 |
| M3 | *PTTG1IP* | 177 | 24 |
| M3 | *SERP1* | 151 | 188 |
| M3 | *TRAM1* | 143 | 195 |
| M3 | *TMED10* | 55 | 208 |

**Table S4. Cont.**

| Module | Gene | *P* value rank | |
|---|---|---|---|
| | | Mouse | Human |
| M3 | *ABCA1* | 27 | 219 |
| M4 | *TCEB2* | 3 | 13 |
| M4 | *SNRPD2* | 2 | 20 |
| M4 | *UQCRQ* | 27 | 12 |
| M4 | *NDUFA7* | 33 | 10 |
| M4 | *CCDC56* | 5 | 37 |
| M4 | *NDUFA8* | 15 | 39 |
| M4 | *DDT* | 6 | 40 |
| M4 | *NDUFB2* | 42 | 17 |
| M4 | *C14orf156* | 45 | 29 |
| M4 | *NDUFV2* | 47 | 27 |
| M4 | *VPS28* | 22 | 49 |
| M4 | *NEDD8* | 34 | 51 |
| M4 | *NDUFB7* | 49 | 53 |
| M4 | *PSMD4* | 55 | 7 |
| M4 | *PSMB3* | 16 | 57 |
| M4 | *NDUFC1* | 4 | 62 |
| M4 | *EIF3K* | 70 | 4 |
| M4 | *HSPC171* | 56 | 70 |
| M4 | *MGST3* | 12 | 71 |
| M4 | *NDUFB11* | 51 | 75 |
| M5 | *CDK5* | 1 | 3 |
| M5 | *GNG3* | 12 | 5 |
| M5 | *UCHL1* | 6 | 19 |
| M5 | *ACOT7* | 20 | 1 |
| M5 | *SULT4A1* | 21 | 16 |
| M5 | *EEF1A2* | 22 | 9 |
| M5 | *ACTL6B* | 18 | 24 |
| M5 | *PKM2* | 35 | 23 |
| M5 | *SLC4A3* | 43 | 20 |
| M5 | *PLD3* | 47 | 6 |
| M5 | *ATP6V0A1* | 42 | 51 |
| M5 | *RABAC1* | 53 | 33 |
| M5 | *ATP6V0B* | 39 | 55 |
| M5 | *SNCB* | 55 | 4 |
| M5 | *ARHGDIG* | 61 | 28 |
| M5 | *TAGLN3* | 73 | 39 |
| M5 | *ASNA1* | 27 | 77 |
| M5 | *UQCRC1* | 78 | 7 |
| M5 | *PFKM* | 24 | 79 |
| M5 | *PCSK1N* | 56 | 80 |
| M6 | *CPNE6* | 7 | 5 |
| M6 | *ICAM5* | 1 | 11 |
| M6 | *HPCA* | 12 | 7 |
| M6 | *PRKCG* | 3 | 12 |
| M6 | *GRIA1* | 13 | 2 |
| M6 | *EFNB3* | 9 | 18 |
| M6 | *NELL2* | 6 | 22 |
| M6 | *CAMK2B* | 23 | 25 |
| M6 | *ARF3* | 21 | 27 |
| M6 | *ST6GALNAC5* | 26 | 28 |
| M6 | *SPTBN2* | 18 | 31 |
| M6 | *RGS14* | 4 | 34 |
| M6 | *SYN2* | 5 | 39 |
| M6 | *NCDN* | 31 | 49 |
| M6 | *CRMP1* | 52 | 13 |
| M6 | *PTPRN* | 56 | 16 |
| M6 | *SLC22A17* | 62 | 62 |
| M6 | *CACNB3* | 15 | 69 |

**Table S4. Cont.**

| Module | Gene | *P* value rank | |
|---|---|---|---|
| | | Mouse | Human |
| M6 | *DLG3* | 10 | 70 |
| M6 | *MAST3* | 11 | 83 |
| M7 | *ZNF148* | 1 | 1 |
| M7 | *UGCG* | 36 | 47 |
| M7 | *TCF4* | 53 | 14 |
| M7 | *NUCKS1* | 57 | 19 |
| M7 | *GRIA3* | 58 | 66 |
| M7 | *GRIA2* | 17 | 82 |
| M7 | *NARG1* | 83 | 49 |
| M7 | *RBBP6* | 100 | 8 |
| M7 | *TNPO2* | 112 | 45 |
| M7 | *PIK3R1* | 133 | 6 |
| M7 | *PLEKHA5* | 144 | 63 |
| M7 | *SFRS7* | 92 | 164 |
| M7 | *KCNMA1* | 46 | 171 |
| M7 | *ODZ3* | 175 | 70 |
| M7 | *MEF2C* | 187 | 153 |
| M7 | *SLC4A7* | 73 | 190 |
| M7 | *PAPSS2* | 160 | 202 |
| M7 | *KPNB1* | 206 | 44 |
| M7 | *ZNF638* | 212 | 189 |
| M7 | *PCYOX1* | 30 | 217 |
| M8 | *CSDA* | 1 | 1 |
| M8 | *SRGN* | 20 | 4 |
| M8 | *CEBPD* | 21 | 2 |
| M8 | *CDKN1A* | 27 | 6 |
| M8 | *CEBPB* | 32 | 10 |
| M8 | *MCL1* | 7 | 36 |
| M8 | *FOS* | 18 | 37 |
| M8 | *BCL6* | 43 | 28 |
| M8 | *TGM2* | 45 | 47 |
| M8 | *TIPARP* | 47 | 35 |
| M8 | *CLDN5* | 5 | 59 |
| M8 | *ANXA1* | 9 | 62 |
| M8 | *ANXA2* | 65 | 26 |
| M8 | *GADD45A* | 69 | 20 |
| M8 | *MGP* | 53 | 70 |
| M8 | *CDH5* | 42 | 86 |
| M8 | *JUNB* | 48 | 87 |
| M8 | *DUSP1* | 11 | 95 |
| M8 | *S100A10* | 95 | 83 |
| M8 | *ACTA2* | 82 | 97 |
| M9 | *CXXC1* | 1 | 2 |
| M9 | *PHF1* | 6 | 11 |
| M9 | *CTTN* | 41 | 45 |
| M9 | *ZNF444* | 57 | 26 |
| M9 | *PCBP4* | 132 | 89 |
| M9 | *AATK* | 97 | 137 |
| M9 | *SIDT2* | 165 | 76 |
| M9 | *SF3A2* | 152 | 166 |
| M9 | *SFRS14* | 167 | 101 |
| M9 | *NAGA* | 34 | 170 |
| M9 | *RALGDS* | 179 | 39 |
| M9 | *PIGT* | 192 | 138 |
| M9 | *GGA1* | 198 | 12 |
| M9 | *PGF* | 199 | 1 |
| M9 | *ARHGEF1* | 207 | 7 |
| M9 | *CSK* | 208 | 189 |
| M9 | *NAP1L4* | 33 | 223 |

**Table S4. Cont.**

| Module | Gene | *P* value rank | |
|---|---|---|---|
| | | Mouse | Human |
| M9 | *RXRA* | 151 | 227 |
| M9 | *EMID1* | 201 | 232 |
| M9 | *METTL7A* | 250 | 86 |
| M10 | *TYROBP* | 1 | 4 |
| M10 | *ARHGDIB* | 4 | 8 |
| M10 | *LY86* | 2 | 9 |
| M10 | *S100A11* | 9 | 7 |
| M10 | *CD14* | 5 | 10 |
| M10 | *FCGR1A* | 21 | 16 |
| M10 | *CYBA* | 18 | 22 |
| M10 | *CSF1R* | 36 | 18 |
| M10 | *TSPO* | 3 | 37 |
| M10 | *LAPTM5* | 40 | 1 |
| M10 | *HMOX1* | 19 | 45 |
| M10 | *GPX1* | 50 | 39 |
| M10 | *IFITM3* | 20 | 50 |
| M10 | *IRF8* | 53 | 26 |
| M10 | *ITGB2* | 56 | 5 |
| M10 | *RAC2* | 60 | 46 |
| M10 | *GFAP* | 16 | 61 |
| M10 | *C1QB* | 66 | 3 |
| M10 | *CTSS* | 8 | 66 |
| M10 | *VAMP8* | 68 | 27 |
| M11 | *HPRT1* | 1 | 5 |
| M11 | *DLD* | 4 | 26 |
| M11 | *ATP6AP2* | 10 | 31 |
| M11 | *PCMT1* | 33 | 8 |
| M11 | *ATP5C1* | 45 | 22 |
| M11 | *G3BP2* | 7 | 47 |
| M11 | *ATP6V1C1* | 30 | 50 |
| M11 | *PREPL* | 51 | 32 |
| M11 | *SERINC1* | 52 | 34 |
| M11 | *PPP2CA* | 59 | 57 |
| M11 | *SUCLA2* | 60 | 3 |
| M11 | *GLRB* | 69 | 15 |
| M11 | *TIMM17A* | 72 | 38 |
| M11 | *PSMC6* | 80 | 69 |
| M11 | *ITFG1* | 81 | 63 |
| M11 | *UBE2V2* | 3 | 82 |
| M11 | *TMEM30A* | 29 | 89 |
| M11 | *SLC30A9* | 95 | 37 |
| M11 | *ATP6V1D* | 92 | 101 |
| M11 | *UBE2N* | 91 | 105 |
| M12 | *RPS19* | 1 | 1 |
| M12 | *RPS11* | 2 | 7 |
| M12 | *RPS15A* | 5 | 10 |
| M12 | *RPL7* | 11 | 14 |
| M12 | *RPL7A* | 7 | 15 |
| M12 | *RPL8* | 17 | 5 |
| M12 | *BTF3* | 3 | 21 |
| M12 | *SNRPG* | 21 | 19 |
| M12 | *RPL13A* | 25 | 4 |
| M12 | *RPL37* | 10 | 26 |
| M12 | *RPL6* | 26 | 24 |
| M12 | *RPS3* | 29 | 2 |
| M12 | *RPL41* | 34 | 13 |
| M12 | *RPLP1* | 42 | 3 |
| M12 | *CHMP2A* | 43 | 33 |
| M12 | *CNBP* | 52 | 45 |

**Table S4. Cont.**

| Module | Gene | *P* value rank | |
|---|---|---|---|
| | | Mouse | Human |
| M12 | *UBL5* | 28 | 57 |
| M12 | *NACA* | 66 | 11 |
| M12 | *SSR2* | 72 | 64 |
| M12 | *C2orf28* | 68 | 73 |
| M13 | *SNX10* | 15 | 21 |
| M13 | *HMP19* | 24 | 2 |
| M13 | *PGRMC1* | 5 | 25 |
| M13 | *MAPK1* | 4 | 37 |
| M13 | *PPP3CB* | 54 | 38 |
| M13 | *ACTR3* | 58 | 45 |
| M13 | *ARMCX2* | 52 | 76 |
| M13 | *SEH1L* | 21 | 77 |
| M13 | *SUMO3* | 81 | 84 |
| M13 | *MAP2K1* | 44 | 85 |
| M13 | *SLC1A1* | 86 | 60 |
| M13 | *HTR2C* | 90 | 22 |
| M13 | *KRAS* | 78 | 90 |
| M13 | *GAP43* | 97 | 95 |
| M13 | *RCN2* | 80 | 98 |
| M13 | *TOMM70A* | 101 | 73 |
| M13 | *GABRB3* | 117 | 71 |
| M13 | *PTPRO* | 131 | 33 |
| M13 | *GLOD4* | 95 | 132 |
| M13 | *RTN1* | 141 | 118 |
| M14 | *SH3BGRL* | 7 | 17 |
| M14 | *USP1* | 1 | 23 |
| M14 | *HNRPH2* | 18 | 27 |
| M14 | *RAD21* | 41 | 12 |
| M14 | *EIF1AX* | 46 | 33 |
| M14 | *SMARCA5* | 50 | 13 |
| M14 | *DDX5* | 56 | 35 |
| M14 | *UBE2E1* | 60 | 10 |
| M14 | *DEK* | 67 | 73 |
| M14 | *FMR1* | 10 | 79 |
| M14 | *PLS3* | 80 | 21 |
| M14 | *SFRS11* | 83 | 85 |
| M14 | *VBP1* | 87 | 99 |
| M14 | *TCF12* | 68 | 101 |
| M14 | *MGEA5* | 96 | 108 |
| M14 | *OGT* | 110 | 105 |
| M14 | *TTRAP* | 111 | 29 |
| M14 | *TOP2B* | 113 | 26 |
| M14 | *YTHDC1* | 16 | 113 |
| M14 | *VPS4B* | 33 | 121 |
| M15 | *GPR12* | 2 | 15 |
| M15 | *DOLPP1* | 10 | 53 |
| M15 | *DNPEP* | 57 | 11 |
| M15 | *ZC3H3* | 71 | 62 |
| M15 | *MVK* | 79 | 4 |
| M15 | *FGF4* | 56 | 101 |
| M15 | *KRT2* | 121 | 21 |
| M15 | *BCAN* | 74 | 122 |
| M15 | *HGS* | 16 | 137 |
| M15 | *XPO6* | 138 | 17 |
| M15 | *SPRR1B* | 37 | 152 |
| M15 | *MYH14* | 159 | 164 |
| M15 | *PRKACA* | 170 | 72 |
| M15 | *PML* | 173 | 87 |
| M15 | *G6PC3* | 143 | 180 |

**Table S4. Cont.**

| Module | Gene | *P* value rank | |
|---|---|---|---|
| | | Mouse | Human |
| M15 | *NUDCD3* | 8 | 181 |
| M15 | *KLC2* | 189 | 147 |
| M15 | *IMPDH1* | 195 | 142 |
| M15 | *HTR4* | 200 | 166 |
| M15 | *CCKBR* | 154 | 204 |

Top 20 interspecies marker genes for each module. For each module, genes were scored based on maximum significance of MM between species (ranks close to 1 represent higher MM). Note that we expect some genes to show between-species preservation for all modules, both due to chance and because of the way mouse MEs were defined. Human-specific modules (i.e., M7 and M9) have the least significant preservation, as measured by the larger minimum ranks of the marker genes for these modules.

**Table S5. Known highly specific marker genes confirmed in our network**

| Gene | List | Cell type | Top module | | Top rank | |
|---|---|---|---|---|---|---|
| | | | Human | Mouse | Human | Mouse |
| *ALDOC* | Cahoy (M) | Astrocyte | **M3** | M9 | 544 | 648 |
| *AQP4* | Cahoy (M) | Astrocyte | **M3** | M12 | 21 | 344 |
| *ATP1A2* | Cahoy (M) | Astrocyte | **M3** | M6 | 4 | 264 |
| *DIO2* | Cahoy (M) | Astrocyte | **M3** | **M8** | 736 | 23 |
| *F3* | Cahoy (M) | Astrocyte | M8 | **M3** | 45 | 75 |
| *GFAP* | Cahoy (M) | Astrocyte | **M3** | **M10** | 43 | 16 |
| *HAPLN1* | Cahoy (M) | Astrocyte | M1 | **M8** | 664 | 412 |
| *MERTK* | Cahoy (M) | Astrocyte | **M3** | M6 | 85 | 480 |
| *PAPSS2* | Cahoy (M) | Astrocyte | M7 | M7 | 202 | 160 |
| *PLA2G7* | Cahoy (M) | Astrocyte | M7 | **M2** | 504 | 137 |
| *PPP1R3C* | Cahoy (M) | Astrocyte | **M3** | **M8** | 62 | 145 |
| *SLC15A2* | Cahoy (M) | Astrocyte | **M3** | M5 | 314 | 1005 |
| *SLC1A2* | Cahoy (M) | Astrocyte | M7 | M6 | 46 | 619 |
| *SLC4A4* | Cahoy (M) | Astrocyte | **M3** | M7 | 133 | 726 |
| *AHCYL1* | Oldham (H) | Astrocyte | **M3** | M1 | 8 | 172 |
| *EDG1* | Oldham (H) | Astrocyte | **M3** | M15 | 9 | 1048 |
| *NTRK2* | Oldham (H) | Astrocyte | **M3** | M15 | 11 | 212 |
| *PON2* | Oldham (H) | Astrocyte | **M3** | **M2** | 2 | 345 |
| *PPAP2B* | Oldham (H) | Astrocyte | **M3** | M6 | 17 | 245 |
| *SDC4* | Oldham (H) | Astrocyte | **M3** | M4 | 39 | 504 |
| *EPHA7* | Cahoy (M) | Neuron | **M13** | **M13** | 344 | 28 |
| *GABRA1* | Cahoy (M) | Neuron | **M1** | **M1** | 76 | 6 |
| *GABRG2* | Cahoy (M) | Neuron | **M1** | **M13** | 17 | 30 |
| *HTR2C* | Cahoy (M) | Neuron | **M13** | **M13** | 22 | 90 |
| *MEF2C* | Cahoy (M) | Neuron | M7 | M7 | 153 | 187 |
| *MYT1L* | Cahoy (M) | Neuron | **M1** | **M6** | 16 | 168 |
| *NEUROD6* | Cahoy (M) | Neuron | **M6** | **M1** | 116 | 13 |
| *NOV* | Cahoy (M) | Neuron | **M13** | **M6** | 330 | 101 |
| *PCSK2* | Cahoy (M) | Neuron | **M1** | **M5** | 110 | 54 |
| *SCG2* | Cahoy (M) | Neuron | **M13** | M3 | 20 | 144 |
| *SLA* | Cahoy (M) | Neuron | M10 | M15 | 17 | 389 |
| *SLC12A5* | Cahoy (M) | Neuron | **M1** | **M6** | 6 | 1244 |
| *SNAP25* | Cahoy (M) | Neuron | **M11** | **M1** | 17 | 5 |
| *SSTR2* | Cahoy (M) | Neuron | **M6** | M8 | 314 | 175 |
| *STMN2* | Cahoy (M) | Neuron | **M11** | **M5** | 24 | 218 |
| *SYT1* | Cahoy (M) | Neuron | **M11** | **M1** | 51 | 63 |
| *VIP* | Cahoy (M) | Neuron | **M13** | **M1** | 505 | 421 |
| *DNM1L* | Oldham (H) | Neuron | **M11** | **M13** | 117 | 8 |
| *FGF12* | Oldham (H) | Neuron | **M11** | M11 | 126 | 361 |
| *GABRG2* | Oldham (H) | Neuron | **M1** | **M13** | 17 | 30 |
| *MAPK1* | Oldham (H) | Neuron | **M13** | **M13** | 37 | 4 |
| *NUDT21* | Oldham (H) | Neuron | **M11** | **M4** | 121 | 391 |
| *PITPNA* | Oldham (H) | Neuron | **M5** | M15 | 56 | 608 |

**Table S5. Cont.**

| Gene | List | Cell type | Top module | | Top rank | |
|---|---|---|---|---|---|---|
| | | | Human | Mouse | Human | Mouse |
| *RAB5A* | Oldham (H) | Neuron | M14 | **M11** | 509 | 26 |
| *SYN2* | Oldham (H) | Neuron | M6 | **M6** | 39 | 5 |
| *YWHAZ* | Oldham (H) | Neuron | **M13** | M7 | 7 | 280 |
| *CLDN11* | Cahoy (M) | Oligoden. | **M2** | **M2** | 62 | 28 |
| *ERBB3* | Cahoy (M) | Oligoden. | **M2** | **M2** | 9 | 212 |
| *EVI2A* | Cahoy (M) | Oligoden. | **M2** | **M2** | 25 | 5 |
| *GSN* | Cahoy (M) | Oligoden. | **M2** | **M2** | 48 | 111 |
| *MAG* | Cahoy (M) | Oligoden. | **M2** | **M2** | 34 | 13 |
| *MAL* | Cahoy (M) | Oligoden. | **M2** | **M2** | 4 | 153 |
| *MBP* | Cahoy (M) | Oligoden. | **M2** | **M2** | 49 | 25 |
| *MOBP* | Cahoy (M) | Oligoden. | **M2** | **M2** | 71 | 162 |
| *MOG* | Cahoy (M) | Oligoden. | **M2** | **M2** | 18 | 6 |
| *PLA2G4A* | Cahoy (M) | Oligoden. | **M8** | **M3** | 307 | 768 |
| *PLP1* | Cahoy (M) | Oligoden. | **M2** | **M2** | 11 | 57 |
| *PRKCQ* | Cahoy (M) | Oligoden. | **M2** | M6 | 263 | 934 |
| *SRPK3* | Cahoy (M) | Oligoden. | **M15** | M9 | 328 | 759 |
| *UGT8* | Cahoy (M) | Oligoden. | **M2** | **M2** | 17 | 335 |
| *CRYAB* | Oldham (H) | Oligoden. | **M2** | **M10** | 15 | 153 |
| *ENPP2* | Oldham (H) | Oligoden. | **M2** | **M2** | 1.5 | 302 |
| *HSPA2* | Oldham (H) | Oligoden. | **M2** | M13 | 1.5 | 227 |
| *MAL* | Oldham (H) | Oligoden. | **M2** | **M2** | 4 | 153 |
| *NPC1* | Oldham (H) | Oligoden. | **M2** | **M2** | 8 | 200 |
| *PMP22* | Oldham (H) | Oligoden. | **M2** | **M2** | 3 | 272 |

Known highly specific marker genes confirmed in our network. The top marker genes for astrocytes, neurons, and oligodendrocytes in mouse (figure *3* from ref. 27) and human (figure *4* from ref. 2), which are also present in our networks, are listed in the "Gene" column. The "top module" column represents the module to which this gene has the most significant MM in each network. Genes in modules consistent with expectations are labeled in bold. The "Top Rank" columns list the ranked MM for each gene in its listed module, with lower ranks representing more significant MM values.

**Table S6. Human specific markers for major cell types**

| Gene | Correlation in human | *P*-value in human | Correlation in mouse | *P*-value in mouse |
|---|---|---|---|---|
| Oligodendrocytes | | | | |
| *ALCAM* | 0.51 | 6.5E-85 | −0.17 | 2.5E-01 |
| *AMPD3* | 0.38 | 3.0E-34 | −0.06 | 1.5E-03 |
| *ANGPTL2* | 0.41 | 1.3E-25 | −0.12 | 7.5E-05 |
| *CBFB* | 0.37 | 2.2E-38 | −0.12 | 8.5E-09 |
| *CDKN1B* | 0.30 | 2.9E-59 | −0.05 | 7.5E-02 |
| *CHD1L* | 0.32 | 1.8E-26 | 0.01 | 1.3E-00 |
| *CLK1* | 0.33 | 9.9E-54 | −0.07 | 4.6E-03 |
| *COL4A5* | 0.65 | 2.7E-182 | 0.00 | 6.7E-01 |
| *CSTF2T* | 0.21 | 1.5E-22 | −0.05 | 6.6E-09 |
| *DYNC1I2* | 0.48 | 6.0E-76 | −0.10 | 6.4E-17 |
| *ELOVL5* | 0.30 | 4.8E-44 | 0.04 | 1.6E-01 |
| *HSPA2* | 0.80 | 0.0E+00 | −0.07 | 2.6E-02 |
| *INPP1* | 0.44 | 5.4E-53 | −0.05 | 1.5E-06 |
| *IQGAP1* | 0.31 | 9.4E-27 | −0.08 | 5.7E-01 |
| *IVNS1ABP* | 0.31 | 2.7E-40 | −0.07 | 2.8E-22 |
| *LRP2* | 0.44 | 5.9E-47 | −0.03 | 2.4E-01 |
| *MAN2A1* | 0.58 | 6.4E-151 | −0.06 | 7.0E-02 |
| *MYLK* | 0.45 | 1.9E-50 | −0.02 | 1.0E-02 |
| *NCAM1* | 0.38 | 1.8E-33 | −0.13 | 2.7E-06 |
| *P2RX7* | 0.41 | 9.7E-26 | −0.09 | 6.9E-10 |
| *PSEN1* | 0.62 | 6.8E-177 | 0.03 | 6.8E-01 |
| *PTP4A2* | 0.45 | 1.9E-61 | 0.00 | 2.7E-10 |
| *RNF103* | 0.25 | 3.9E-19 | −0.04 | 1.8E-12 |
| *STAG2* | 0.20 | 3.9E-23 | −0.07 | 4.5E-02 |
| *THBS2* | 0.44 | 4.9E-50 | −0.04 | 1.3E-01 |
| *TXNIP* | 0.25 | 1.9E-17 | −0.06 | 1.0E-03 |
| *ZFYVE16* | 0.32 | 2.2E-37 | 0.00 | 8.7E-01 |
| *HBEGF* | 0.22 | 8.0E-14 | −0.12 | 9.2E-05 |
| *KIAA0174* | 0.35 | 6.3E-27 | −0.05 | 2.4E-08 |
| Microglia | | | | |
| *CD53* | 0.49 | 2.1E-94 | −0.05 | 1.4E-04 |
| *CYFIP1* | 0.41 | 3.1E-49 | −0.12 | 3.4E-01 |
| *ITGAM* | 0.28 | 2.3E-26 | 0.00 | 1.7E-01 |
| *KCTD12* | 0.43 | 4.5E-53 | −0.15 | 2.8E-04 |
| *SLA* | 0.50 | 6.2E-78 | 0.01 | 1.7E-01 |
| *SLC2A5* | 0.48 | 3.3E-86 | 0.03 | 8.6E-02 |
| *STAB1* | 0.43 | 7.2E-46 | 0.01 | 9.2E-01 |
| Astrocytes | | | | |
| *ABLIM1* | 0.34 | 1.1E-35 | −0.03 | 1.0E+00 |
| *AQP1* | 0.27 | 1.6E-38 | 0.00 | 3.1E-01 |
| *CD99* | 0.36 | 3.1E-70 | −0.09 | 4.2E-01 |
| *CRYL1* | 0.29 | 2.5E-30 | −0.05 | 8.1E-02 |
| *FTH1* | 0.21 | 9.8E-17 | −0.05 | 2.1E-03 |
| *FYN* | 0.39 | 1.1E-44 | 0.04 | 6.1E-02 |
| *GRAMD3* | 0.63 | 1.2E-185 | 0.04 | 8.7E-01 |
| *IGFBP7* | 0.35 | 1.1E-42 | −0.06 | 1.4E+00 |
| *LEPROT* | 0.31 | 1.1E-19 | 0.04 | 1.3E+00 |
| *PIK3C2A* | 0.53 | 2.5E-64 | −0.02 | 7.7E-02 |
| *PRKCA* | 0.34 | 3.0E-54 | −0.04 | 5.0E-02 |
| *RAB31* | 0.55 | 1.2E-107 | −0.04 | 1.0E+00 |
| *RYR3* | 0.45 | 4.5E-73 | −0.01 | 1.3E+00 |
| *SRI* | 0.42 | 7.0E-59 | −0.08 | 9.0E-01 |
| *TCF7L2* | 0.45 | 2.1E-43 | −0.01 | 1.9E-05 |
| *TGFBR3* | 0.41 | 1.1E-51 | −0.01 | 3.2E-02 |
| *UNG* | 0.38 | 4.4E-47 | −0.02 | 1.6E-12 |
| Neurons | | | | |
| *ACLY* | 0.24 | 2.0E-20 | 0.02 | 1.3E+00 |
| *ACP1* | 0.29 | 5.0E-34 | −0.03 | 2.8E-01 |

**Table S6. Cont.**

| Gene | Correlation in human | *P*-value in human | Correlation in mouse | *P*-value in mouse |
|---|---|---|---|---|
| *ADK* | 0.22 | 1.5E-15 | 0.04 | 4.6E-01 |
| *CCDC6* | 0.32 | 8.8E-36 | −0.01 | 7.3E-01 |
| *COPS8* | 0.30 | 3.5E-34 | −0.08 | 5.2E-01 |
| *GHITM* | 0.37 | 7.2E-66 | −0.07 | 5.9E-04 |
| *HSPA4L* | 0.28 | 1.4E-24 | −0.02 | 9.0E-03 |
| *HSPA8* | 0.25 | 1.1E-20 | 0.02 | 8.0E-01 |
| *MCFD2* | 0.39 | 2.1E-46 | 0.05 | 2.0E-01 |
| *P15RS* | 0.28 | 1.3E-22 | −0.03 | 2.4E-03 |
| *PEG10* | 0.38 | 3.6E-84 | 0.01 | 3.4E-01 |
| *PGK1* | 0.24 | 3.5E-24 | −0.04 | 1.0E+00 |
| *PRKCI* | 0.31 | 3.5E-22 | −0.07 | 4.0E-02 |
| *UQCRC2* | 0.30 | 5.9E-37 | 0.01 | 1.2E+00 |

Human specific marker genes for four major cell types: oligodendrocytes, microglia, astrocytes, and neurons. All genes in these lists pass four criteria: (*i*) Member of the human module for the cell type, (*ii*) not significantly correlated ($R < 0.05$, $P > 0.05$) with the corresponding mouse module, (*iii*) validated in a human cortex network (2), and (*iv*) not validated in corresponding mouse comparison studies (27, 64, 65). SLA and STAB1 were further confirmed as markers for human microglia in ref. 66.

**Table S7. Top confirmed M9h genes**

| Gene | Rank | Gene | Rank |
|---|---|---|---|
| *ZNF160* | 1 | *USP4* | 26 |
| *DCLRE1C* | 2 | *PRR14* | 27 |
| *SRRM2* | 3 | *SCAMP2* | 28 |
| *POGZ* | 4 | *ELAVL3* | 29 |
| *CLCN7* | 5 | *C21orf2* | 30 |
| *TAZ* | 6 | *TRIOBP* | 31 |
| *CXXC1* | 7 | *PIAS4* | 32 |
| *PRR11* | 8 | *ZNF688* | 33 |
| *WDR6* | 9 | *GAK* | 34 |
| *ZNF444* | 10 | *SORBS3* | 35 |
| *BRD3* | 11 | *PLCG1* | 36 |
| *CTDSP2* | 12 | *LMBR1L* | 37 |
| *CEP164* | 13 | *ZNF692* | 38 |
| *SAPS2* | 14 | *LAS1L* | 39 |
| *MAP3K11* | 15 | *RING1* | 40 |
| *FBXW4* | 16 | *PHF1* | 41 |
| *ZNF148* | 17 | *CELSR3* | 42 |
| *C9orf7* | 18 | *FLJ21865* | 43 |
| *ZBTB20* | 19 | *ECHDC2* | 44 |
| *MAP3K3* | 20 | *EFEMP2* | 45 |
| *DBT* | 21 | *GOSR1* | 46 |
| *TNS1* | 22 | *APBA3* | 47 |
| *EXOC7* | 23 | *MAML1* | 48 |
| *WHSC2* | 24 | *GLT25D1* | 49 |
| *AKAP8L* | 25 | *BCAT2* | 50 |

Top confirmed M9h genes. Genes in this list are ranked based on high correlation with the ME of M9h and its corresponding confirmation modules (the red module from ref. 12, the black module from ref. 40, and the yellow module from ref. 39), and high correlation with M9h hubs in the Celsius database (ref. 67)