

Integrated Systems Approach Identifies Genetic Nodes and Networks in Late-Onset Alzheimer's Disease

Bin Zhang,^{1,2,3,4,14,*} Chris Gaiteri,^{4,14} Liviu-Gabriel Bodea,^{5,14} Zhi Wang,⁴ Joshua McElwee,⁶ Alexei A. Podtelevnikov,⁷ Chunsheng Zhang,⁶ Tao Xie,⁶ Linh Tran,⁴ Radu Dobrin,⁶ Eugene Fluder,⁶ Bruce Clurman,⁸ Stacey Melquist,⁶ Manikandan Narayanan,⁶ Christine Suver,⁴ Hardik Shah,^{1,2} Milind Mahajan,^{1,2,3} Tammy Gillis,⁹ Jayalakshmi Mysore,⁹ Marcy E. MacDonald,⁹ John R. Lamb,¹⁰ David A. Bennett,¹¹ Cliona Molony,⁶ David J. Stone,⁷ Vilmundur Gudnason,¹² Amanda J. Myers,¹³ Eric E. Schadt,^{1,2,3} Harald Neumann,⁵ Jun Zhu,^{1,2,3} and Valur Emilsson^{12,*}

¹Department of Genetics and Genomic Sciences

²Icahn Institute of Genomics and Multi-scale Biology

Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, New York, NY 10029, USA

³Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

⁴Sage Bionetworks, 1100 Fairview Avenue North, Seattle, WA 98109, USA

⁵Neural Regeneration Group, Institute of Reconstructive Neurobiology, University of Bonn, Sigmund-Freud-Str. 25, 53127 Bonn, Germany

⁶Merck Research Laboratories, Merck & Co. Inc., 33 Avenue Louis Pasteur, Boston, MA 02115, USA

⁷Merck Research Laboratories, Merck & Co. Inc., 770 Sumneytown Pike, WP53B-120 West Point, PA 19486, USA

⁸Fred Hutch Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109, USA

⁹Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA

¹⁰GNF Novartis, 10675 John Jay Hopkins Drive, San Diego, CA 92121, USA

¹¹Rush Alzheimer Disease Center, Rush University Medical Center, Chicago, IL 60612, USA

¹²Icelandic Heart Association and University of Iceland, Holtasari 1, IS-201 Kopavogur, Iceland

¹³Department of Psychiatry and Behavioral Sciences, Division of Neuroscience, Miller School of Medicine, University of Miami, Miami, FL 33136, USA

¹⁴These authors contributed equally to this work

*Correspondence: bin.zhang@mssm.edu (B.Z.), valur@hjarta.is (V.E.)
<http://dx.doi.org/10.1016/j.cell.2013.03.030>

SUMMARY

The genetics of complex disease produce alterations in the molecular interactions of cellular pathways whose collective effect may become clear through the organized structure of molecular networks. To characterize molecular systems associated with late-onset Alzheimer's disease (LOAD), we constructed gene-regulatory networks in 1,647 post-mortem brain tissues from LOAD patients and nondemented subjects, and we demonstrate that LOAD reconfigures specific portions of the molecular interaction structure. Through an integrative network-based approach, we rank-ordered these network structures for relevance to LOAD pathology, highlighting an immune- and microglia-specific module that is dominated by genes involved in pathogen phagocytosis, contains TYROBP as a key regulator, and is upregulated in LOAD. Mouse microglia cells overexpressing intact or truncated TYROBP revealed expression changes that significantly overlapped the human brain TYROBP network. Thus the causal network structure is a useful predictor of response to gene perturbations and presents a

framework to test models of disease mechanisms underlying LOAD.

INTRODUCTION

Complex diseases such as late-onset Alzheimer's disease (LOAD) arise from the downstream interplay of DNA-sequence variants and nongenetic factors that act through molecular networks to confer disease risk (Schadt, 2009). Despite decades of intensive research, the causal chain of mechanisms behind LOAD remains elusive. In fact, there are no effective disease-modifying or preventive therapies, and the only available treatment remains symptomatic; meanwhile, the incidence of LOAD is expected to double by 2050 (Brookmeyer et al., 2007). Progress in LOAD research is fundamentally limited by our reliance on mouse models of severe familial/early-onset Alzheimer's disease; therefore, our primary knowledge of LOAD is in actuality based on the downstream effects of three rare mutations in *APP*, *PSEN1*, and *PSEN2* (Bertram et al., 2010). Although such mouse models are necessary and helpful, the cognitive deficits in these transgenic mice are less severe than those in humans, and they do not exhibit equivalent neurodegeneration, which is the most accurate clinical marker of cognitive disease progression in humans. Correspondingly, attrition rates from early discovery to late

drug development have been very high (Schäfer and Kolkhof, 2008).

In contrast to the plethora of potential disease mechanisms detected in humans with LOAD, the search for LOAD-modifying interventions has focused primarily on compounds targeting the amyloid- β pathway. Both biological risk factors, often related to vascular health and psychosocial factors (Cechetto et al., 2008; Qiu et al., 2010), as well as genetic susceptibility play a critical role in the underlying pathophysiology of LOAD (Bertram et al., 2010). *APOE* is still the best validated susceptibility gene accounting for at least 30% of the genetic variance in LOAD (Corder et al., 1993). Genome-wide association studies (GWAS) have identified several additional genetic risk loci for LOAD that seem to cluster in patterns that suggest immunity (*CLU*, *CR1*, *CD33*, *EPHA1*, *MS4A4A/MS4A6A*), lipid processing (*APOE*, *ABCA7*), and endocytosis (*PICALM*, *BIN1*, *CD2AP*) as important causal biological processes (Bettens et al., 2013). More recently, low-frequency missense variants in *APP* and *TREM2* were found to confer strong protection or elevated risk of LOAD (Guerreiro et al., 2013; Jonsson et al., 2012, 2013). However, the overall contribution of these new common and low-frequency variants to the heritability of LOAD is very small, suggesting that a large fraction of the genetic variance beyond the *APOE* risk still remains hidden. Can we clarify the pathology of LOAD by zooming out to the pathway level to search for emergent risk of many genomic contributions? If so, how can we identify the key causal genes in these pathways?

In light of the complexity and elusiveness of LOAD pathogenesis, new approaches are needed to boost the probability of identifying causal genes and pathways. Recently, we have leveraged the molecular network structure that is reflected in genotypic and gene-expression data to uncover biologically meaningful gene modules involved in the development of complex disease (Chen et al., 2008; Emilsson et al., 2008). Targeting such causal networks in ways that restore them to a normal state has been proposed as a path to treat disease (Schadt et al., 2009), but this potential has never been realized for LOAD. However, the complexity of these networks makes it difficult to distinguish the causal from correlated disease effects or how the causal regulators propagate their effects. To better address this, we constructed molecular networks based on whole-genome gene-expression profiling and genotyping data in 1,647 autopsied brain tissues from hundreds of LOAD patients and nondemented subjects. We identified numerous modules of distinct functional categories and cellular specificity, many showing a massive remodeling effect in the LOAD brain. Next, we applied an integrative network-based approach to rank-order these modules for relevance to LOAD pathology and used a Bayesian inference to identify the key causal regulators of these remodeled networks. For instance, we identified eight causal regulators of the top-ranked immune/microglia module, including *TYROBP* (a.k.a. *DAP12*) as the highest ranking in terms of regulatory strength and differential expression in LOAD brains. We demonstrate through mouse microglia cells overexpressing intact or truncated dominant-negative *TYROBP* that downstream expression changes significantly overlapped the human *TYROBP* brain network. This study presents many of the network advantages useful in identifying

and prioritizing pathways and gene targets involved in the pathophysiology of LOAD.

RESULTS

Leveraging a Systems Approach to LOAD

We developed and applied an integrative network-based approach to identify modules of genes associated with neurodegenerative disease (Figures 1A–1C). We processed 1,647 autopsied tissues from dorsolateral prefrontal cortex (PFC), visual cortex (VC), and cerebellum (CB) in 549 brains of 376 LOAD patients and 173 nondemented healthy controls (Figure 1A). All subjects were diagnosed at intake, and each brain underwent extensive LOAD-related pathology examination. We note that the known *APOE* genotype exposure was confirmed in the Harvard Brain Tissue Resource Center (HBTRC) sample, showing an odds ratio of 3.74 per copy $\epsilon 4$ allele ($p = 4.1 \times 10^{-13}$). Each tissue sample was profiled for 39,579 transcripts representing 25,242 known and 14,337 predicted gene-expression traits, and each subject genotyped for 838,958 unique SNPs (Figure 1A). Unless otherwise noted, gene-expression analyses were adjusted for age and sex, postmortem interval (PMI) in hours, and sample pH and RNA integrity number (RIN). In the overall cohort of LOAD and nondemented brains, the mean \pm standard deviation (SD) for sample PMI, pH, and RIN were 17.8 ± 8.3 , 6.4 ± 0.3 , and 6.8 ± 0.8 , respectively. Extensive analysis of the effect of covariates on gene-expression variation in LOAD and nondemented brains was carried out, as shown in Figure S1 (available online) and described in the Extended Experimental Procedures. Here, we used a robust linear regression model for covariate corrections in all our gene-expression analyses (Experimental Procedures). Results of traditional differential expression analysis demonstrate that subsets of genes were up- or downregulated in LOAD (Figure 2A). Consistent with the known progression and regional severity of LOAD pathology (Braak and Braak, 1991), we observed that the PFC region contained the greatest number of differentially expressed genes (Figure 2B). Figure 2C summarizes the clustering or colinearity of the various LOAD pathology traits and age within the HBTRC cohort, resulting in distinct groups of clinical pathology and age as separate clusters. For instance, the number of significant correlations of expression traits to neuropathology like Braak stage within the LOAD patient group was highest in the PFC region (Figure 2D). Given these observations and the fact that PFC is more commonly affected in LOAD than CB and VC (Braak and Braak, 1991), a particular attention was paid to this region in our strategy to rank-order modules for relevance to LOAD. These massive data sets were the basis of further method development with the aim to identify and rank-order network modules and gene targets associated with LOAD pathology (Figures 1A–1C). Results of these various analysis steps are discussed in the sections that follow, and a more detailed description of methods and statistical procedures is found in the Extended Experimental Procedures.

Remodeling of the Molecular Interaction Structure in LOAD Brains

For simultaneously capturing the intra- and interregional gene-gene interactions in the LOAD or nondemented state, we

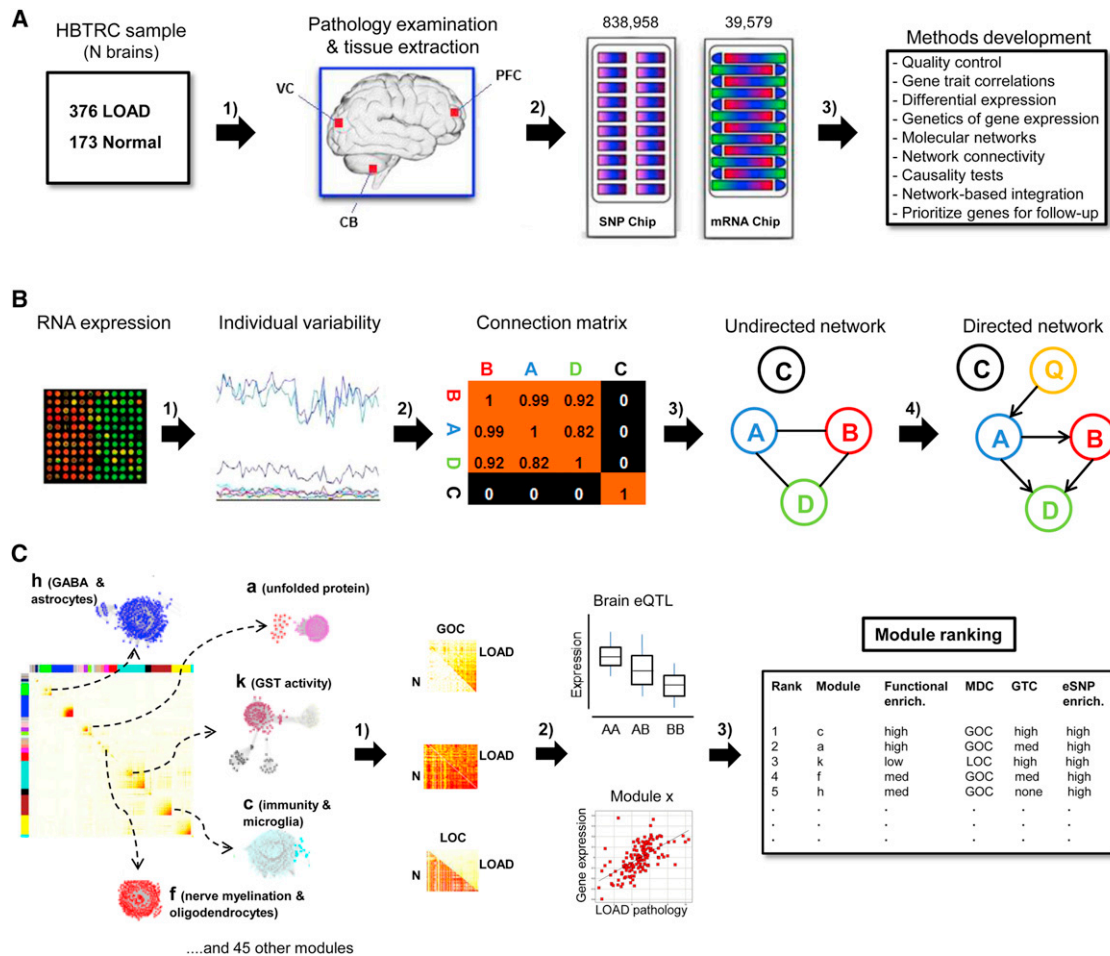


Figure 1. Sample Processing and the Integrative Network-Based Approach

(A) Five hundred and forty-nine brains were collected through the Harvard Brain Tissue Resource Center (HBTRC) from 376 LOAD patients and 173 nondemented subjects, and tissues extracted from three brain regions, the commonly affected PFC in LOAD and the less affected VC and CB (1). Each brain went through extensive neuropathology examination, and all tissues were profiled for 39,579 transcripts, and every subject genotyped for 838,958 SNPs (2). These data sets were the basis of the method development in the present study (3).

(B) From the microarray RNA expression data, we identified gene-expression traits showing individual variability in gene-expression traits as per brain region (1). Next we computed the coregulation (connectivity) strength between genes, defined the appropriate connectivity threshold (2), and performed hierarchical clustering analysis to construct the undirected coexpression network (3). Finally, we used brain eSNPs (Q) as causal anchors in the construction of directed Bayesian networks to infer a causal relationship between nodes in the network (4). A variant of the underlying causality-scoring process here can be applied to relationships among thousands of nodes to infer genome-scale networks.

(C) Comparison of LOAD and nondemented networks was performed to explore any effect on the molecular interaction structure associated with the disease. Differentially connected modules in LOAD were investigated for their functional organization (1), module relevance to clinical outcome, as well as the enrichment of brain eSNPs (2). Modules were rank-ordered (this figure does not show the true rank-order) for their strength of the functional enrichment, module correlation to neuropathology, and eSNP enrichment (3).

See also Figure S1 and Table S1.

constructed multitissue coexpression networks consisting of the top one-third ($n = 13,193$) of the most variable gene-expression traits per brain region in individuals donating tissues from all three regions (Extended Experimental Procedures). The multitissue coexpression network in LOAD brains indicated strong structurally segregated regions of the human brain molecular interactome (Figure 3A), consisting of 111 modules and each containing between 30 and 1,446 gene members (Figure 3A), whereas the network generated from nondemented samples

has 89 modules ranging in size from 30 to 2,278 genes. Figure 3B highlights a direct comparison of the two topological overlap matrices corresponding to the LOAD or nondemented associated network for a subset of 16 modules, demonstrating that LOAD reconfigures specific portions of the molecular interaction structure. To analytically detect and quantify this network reorganization across the demented and nondemented states, we developed a metric that we refer to as modular differential connectivity (MDC) (Extended Experimental Procedures). MDC is

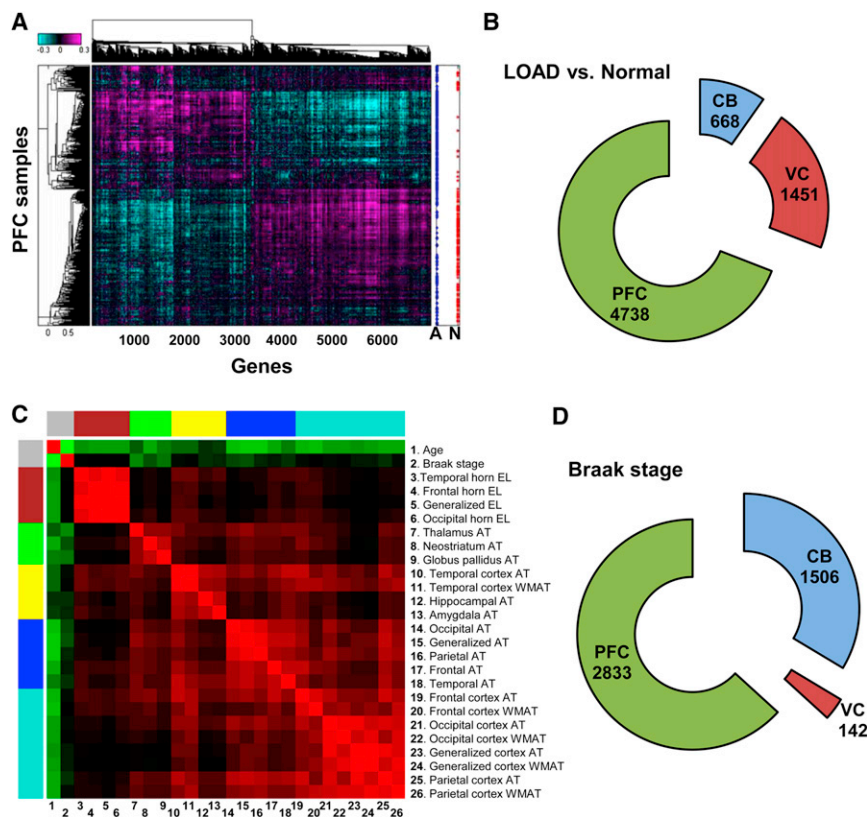


Figure 2. Differential Gene Expression in LOAD Brains and Expression Correlation to Braak Stage

(A) The heatmap shows the genes ($n = 6457$), absolute mean-log ratio > 1.5 for each profile, which most significantly differentiate disease status in PFC. The legend to the right shows the arrangement of samples with blue points denoting LOAD (A), and red points denoting nondemented state (N).

(B) The number of differentially expressed genes in LOAD compared with controls per brain region using Bonferroni adjusted $p < 0.05$ by correcting for the number of probes tested ($p \leq 2.46 \times 10^{-7}$).

(C) Clustering analysis where the rows and columns represent age, and 25 LOAD pathology traits are arranged in a symmetric fashion and sorted by the hierarchical clustering tree of the correlation matrix. The color intensity signifies the correlation strength between two traits (red positive and green negative). AT, atrophy; WMAT, white matter atrophy; EL, enlargement.

(D) Number of genes showing significant expression correlation to Braak stage as measured per brain region using Bonferroni adjusted $p < 0.05$ by correcting for the number of probes tested ($p \leq 2.46 \times 10^{-7}$).

See also Table S1.

the ratio of the average connectivity for any pair of module-sharing genes in LOAD compared to that of the same genes in the nondemented state and is a continuous measure ranging from 0 to infinity. This module-centric measure of differential connectivity between the two states is therefore fundamentally different from the gene-centric analysis of previous studies that applied hard cutoffs (Mani et al., 2008). Given the nature of the coexpression network analysis, $MDC > 1$ indicates gain of connectivity (GOC) or enhanced coregulation between genes, whereas $MDC < 1$ indicates loss of connectivity (LOC) or reduced coregulation between genes. In extreme cases where $MDC \gg 1$, e.g., the glutathione transferase (GST) module (Figure 3B), or $MDC \ll 1$, e.g., the nerve myelination module (Figure 3B), the corresponding genes do not form a coherent cluster in the nondemented state or LOAD, respectively. Thus, new modules are created in LOAD, whereas in other cases, a portion of the network is completely disrupted. The statistical significance of the MDC metrics was computed through the false discovery rate (FDR) procedure as described in the [Extended Experimental Procedures](#). Based on 10% FDR, 54% of all modules showed GOC, whereas 4.5% of modules exhibited LOC. The structures of the remaining 41.5% of the modules were found to be conserved across the LOAD and nondemented states by this MDC measure. We note a negligible overlap of only 6% between signatures of differential connectivity and standard differential gene expression in LOAD brains, implying that the observed disruption in coregulation of genes reflects a previously untapped marker of LOAD neuropathology.

Functional Organization of the Network and Its Relevance to LOAD Pathology

As observed in previous network-based studies (Chen et al., 2008; Emilsson et al., 2008; Zhang and Horvath, 2005), we find that brain gene expression is organized into modules of distinct functional categories (Figure 3C). Overrepresentation of canonical pathways and biological processes in modules was measured through Fisher's exact test (FET) and corrected for number of modules and functional categories tested. Figure 3C highlights significant overrepresentation of functional categories in modules showing GOC, LOC, or conserved connectivity and containing at least 50 genes. The multifactorial basis of LOAD neuropathology involves biological processes active in both the central nervous system (CNS) and the metabolic and vascular peripheral system that have often progressed silently for many years (Huang and Mucke, 2012; Murray et al., 2011). In fact, we find that multiple functional categories, including the immune response, unfolded protein, vascular system, extracellular matrix, neurogenesis (brain development), glucose homeostasis, synaptic transmission, and olfactory sensory perception categories in the GOC modules, are highly enriched in the LOAD-associated modules (Figure 3C), whereas the LOC modules are enriched for genes involved in nerve myelination, cell cycle, γ -aminobutyric acid (GABA) metabolism, and neurotrophin signaling (Figure 3C). Many of these functional categories have previously been implicated in LOAD and/or CNS-related function (Ansari and Scheff, 2010; Cechetto et al., 2008; Dodel et al., 2003; Luchsinger, 2008; Morawski et al.,

2012; Schiffman et al., 2002), again reinforcing the complex multifactorial basis of the underlying pathophysiology. The functional categories enriched in the conserved modules included “muscle contraction” (actin-related system), coated vesicle, cadherin, and zinc ion metabolism (Figure 3C).

CNS cell-type-specific gene signatures, from the Allen Brain Atlas (<http://www.brain-map.org/>), were enriched in distinct network modules as previously observed (Oldham et al., 2008): neurons in the synaptic transmission modules (11-fold, $p = 3.7 \times 10^{-24}$), astrocytes in the GABA biosynthesis module (22 fold, $p = 1.5 \times 10^{-15}$), oligodendrocytes in the nerve myelination module (30 fold, $p = 2.5 \times 10^{-30}$), choroid plexus cell types in the extracellular matrix module (35 fold, $p = 3.9 \times 10^{-15}$), and microglia signatures responding to amyloid- β treatment (Walker et al., 2006) in the immune module (10-fold, $p = 4.5 \times 10^{-20}$) (Figure 3C). In contrast to the GOC and LOC modules, conserved modules were not enriched for any CNS-specific cell types (Figure 3C). Pathways enriched in the brain modules and not previously implicated in LOAD may therefore represent novel disease mechanisms including, for instance, the glucuronosyl transferase activity and the dynein complex (Figure 3C). Moreover, the comprehensive representation of gene-gene interactions in the LOAD-associated networks can uncover novel gene members in pathways already implicated in LOAD, thus allowing us to work out a known pathologic mechanism in more detail than ever before. In summary, the immune module shows the statistically most significant functional enrichment of all modules (Figure 3C) and as such may have a more comprehensive representation of its respective pathway members.

Table S1 contains extensive information regarding the functional enrichment and gene membership of modules containing at least 50 unique gene symbols. We highlight some specific findings of interest from Figure 3C: (1) The enrichment of pathways related to olfactory sensory perception in a LOAD-associated module is of interest given that the processing of olfactory function is affected in subjects who are genetically at risk of developing LOAD long before the symptoms of dementia are manifested (Schiffman et al., 2002). (2) The APOE transcript is located in the LOC module enriched for astrocyte signatures and GABA metabolism, consistent with the observation that astrocytes are the major source of APOE in the CNS (Boyles et al., 1985). The close connectivity of APOE and GABA metabolism in the brain network may therefore have some bearing on the observation that GABA interneuron dysfunction is particularly severe in APOE4 carriers (Li et al., 2009). (3) The previously identified macrophage-enriched metabolic network (MEMN) in peripheral tissues and strongly supported as causal for a number of metabolic and vascular traits related to obesity, diabetes, and heart disease (Chen et al., 2008; Emilsson et al., 2008) is remarkably enriched within the brain immune/microglia module (3.9-fold, $p = 2.4 \times 10^{-46}$). This is of interest given the strong epidemiological evidence for metabolic- and vascular-based exposure on LOAD (Huang and Mucke, 2012; Murray et al., 2011). (4) The postsynaptic density proteome in the human neocortex of 748 proteins overrepresented with risk loci known to underlie cognitive, affective, and motor phenotypes (Bayés et al., 2011) is significantly enriched in the synaptic transmission module (3-fold, $p = 1.6 \times 10^{-32}$). It is still unclear how and which

of these different biological processes mentioned above interact to affect LOAD; however, it is likely that only a few downstream mechanisms on which many diverse effects converge are causally related to LOAD (Huang and Mucke, 2012; Murray et al., 2011). The accumulated data show a strikingly coherent organization of molecular processes in the LOAD-associated network.

The coexpression network structure, its changes between nondemented and LOAD brains, and the genetic loci responsible for the expression covariation behind these networks collectively reflect molecular processes associated with LOAD. By linking the network modules to clinical outcome or LOAD neuropathology via a multiple regression analysis (Extended Experimental Procedures), we can infer key molecular processes associated with LOAD. A covariance matrix of the average expression correlation ($|r|$) between 49 modules, comprised of at least 100 probes, and 25 LOAD-related traits is shown in Figure 4A. We performed principal component analysis (PCA) to estimate the module-trait correlation and used the FDR method to assess the significance (see Extended Experimental Procedures). Of all modules, the immune/microglia showed correlation to the greatest number of LOAD-related neuropathology traits (Figure 4B). Expression of the PFC immune/microglia module correlated to atrophy levels in multiple brain regions, including frontal cortex ($r = 0.27$, FDR = 0.018) and parietal ($r = 0.20$, FDR = 0.016), temporal ($r = 0.19$, FDR = 0.022), and neostriatum regions ($r = 0.28$, FDR = 3.3×10^{-9}), as well as ventricular enlargement ($r = 0.17$, FDR = 0.031). Several modules, however, showed correlation to a more restricted type of neuropathology, including the modules characteristic for the glucuronosyl transferase correlated to Braak stage ($r = 0.18$, FDR = 9.8×10^{-5}), NAD(P) homeostasis to Braak stage ($r = 0.25$, FDR = 1.4×10^{-7}), neurogenesis to ventricular enlargement ($r = 0.19$, FDR = 5.1×10^{-5}), and GST to ventricular enlargement ($r = 0.22$, FDR = 4×10^{-6}). The significance of functional enrichment in modules and the number of neuropathology traits correlated with modules were considered important criteria in rank-ordering modules for their potential to affect LOAD.

Bayesian Networks and the Immune Module as an Effector in LOAD

Causal probabilistic Bayesian networks were constructed and used as an alternative approach to delineate potential regulatory mechanisms. In order to establish a causal relationship or dependency between nodes in the network, we constructed a directed probabilistic Bayesian network through the application of brain *cis* expression (e)SNPs as causal anchors. Because *cis*-eSNPs are in linkage disequilibrium (LD) with causal variants that affect the expression levels of a neighboring gene or they are the causal variant themselves, they serve as an excellent source of natural perturbation to infer causal relationships among genes and between genes and higher-order phenotypes like disease (Chen et al., 2008; Emilsson et al., 2008). We detected a total of 11,318 unique *cis*-eSNP transcripts in the three brain regions, at FDR of 10% (Figure S2A), which is the largest number of brain eSNP transcripts detected to date in a single study (Webster et al., 2009). The methodology to identify *cis*- and *trans*-eSNPs is detailed in Extended Experimental Procedures, whereas Table S1 lists all *cis*- and *trans*-acting eSNPs detected in the present

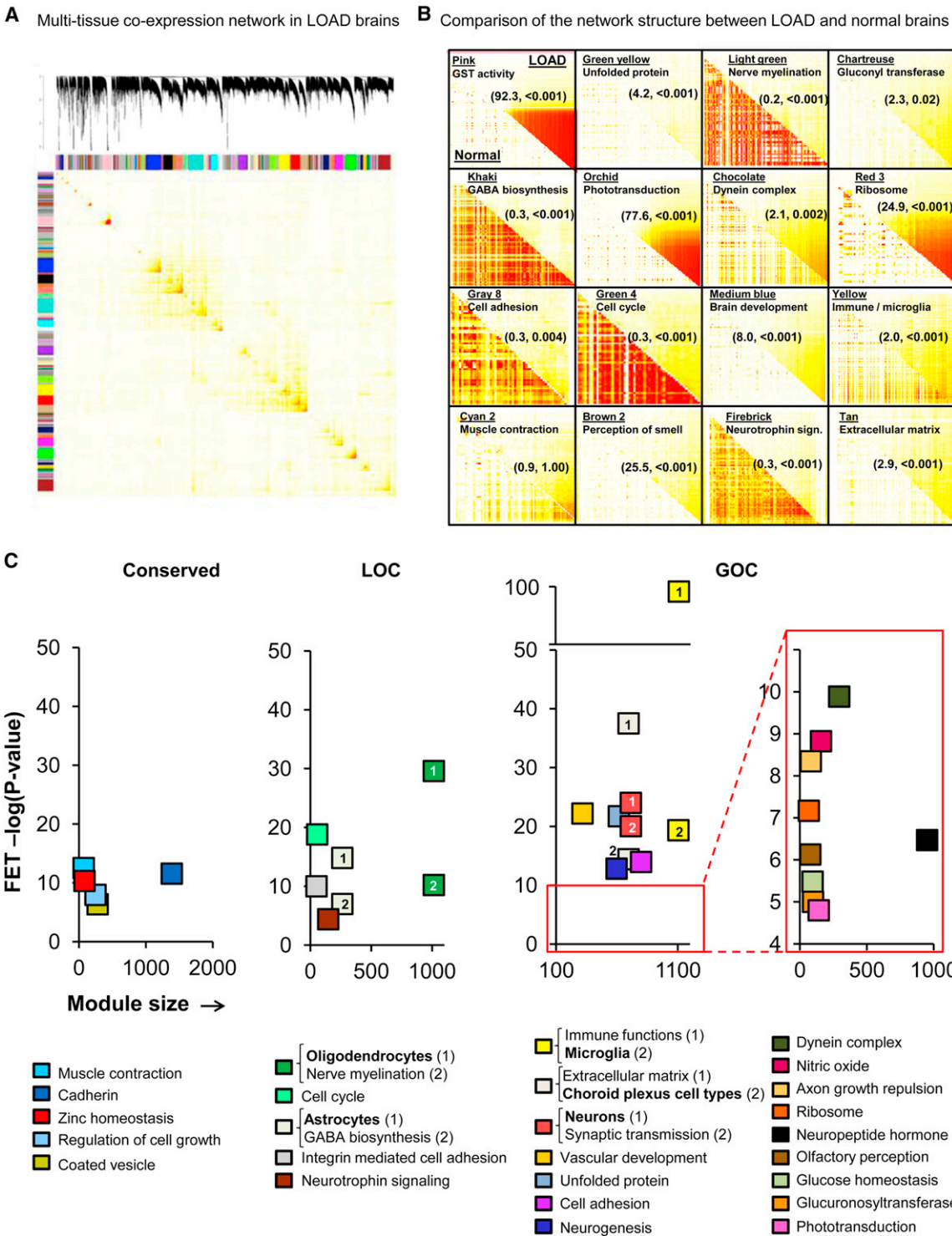


Figure 3. Multitissue Gene Coexpression Network in LOAD Brains

(A) The topological overlap matrix (TOM) plot corresponds to the LOAD multitissue coexpression network. The rows and columns represent the same set of the top one-third (13,193) of the most variably expressed genes in each of the three brain tissues and states, expressed in a symmetric fashion and sorted by the hierarchical clustering tree of the LOAD network.

(B) Individual TOM covariance matrices of 15 differentially connected and one conserved modules in LOAD (the upper right triangle of each module) versus that in the nondemented state (the lower left triangle of each module). Differential connectivity (MDC) and FDR estimate is specified in each panel in parenthesis (MDC, FDR).

(legend continued on next page)

study at FDR of 10%. There was between 70% and 80% sharing of *cis*-eSNP transcripts across different brain regions, and 37% overlapped all brain regions (Figure S2A). Importantly, we find a variable and often strong enrichment of brain eSNPs in many of the LOAD-associated modules compared to all probes on the array, suggesting the possibility that these variants determine the differential connectivity observed in LOAD. For instance, in the PFC region (Figure 4C), there were five modules showing significant enrichment for *cis*-eSNPs, including the unfolded protein (3.8-fold, $p = 3.8 \times 10^{-81}$), nerve myelination (2.5-fold, $p = 2.9 \times 10^{-40}$), immune function (2.2-fold, $p = 4.3 \times 10^{-30}$), GABA metabolism (2.7-fold, $p = 2.3 \times 10^{-13}$), and extracellular matrix (1.6-fold, $p = 2.3 \times 10^{-7}$) modules (Figure 4C). The enrichment of *cis*-eSNPs in the differentially connected LOAD modules in the VC and CB regions is shown in Figure S2B. For the present study, however, a particular attention was paid to the *cis*-eSNPs for their applicability as priors in the construction of Bayesian networks (Extended Experimental Procedures and schematic Figure S3).

We constructed Bayesian networks for each coexpression module. Although many of the LOAD-associated network modules are of potential interest, the reconstruction of the Bayesian network for the immune/microglia module is highlighted given that it has the strongest disease association based on clinical covariates and network-associated properties: (1) significant differential connectivity of the cortex-specific immune modules in LOAD (MDC between 49% and 100% GOC at FDR < 0.001); (2) the immune/microglia module showed the most significant enrichment of functional categories; (3) the highest degree of gene-expression correlation to several measures of LOAD neuropathology; (4) the PFC version of the module was highly enriched for brain eSNPs. To increase the predictive power of inflammation-related regulatory networks, we further built up the directed Bayesian network for the inflammation modules derived from the individual brain regions. Figure 5 highlights the interactions within and between the five predominant immunologic families in the PFC-based putative microglia module. To generate this roadmap to the complex structure of the immune/microglia module, genes that were not direct members of one of these five core pathways were assigned to the family with which they have the greatest number of causal interactions. The immune module was dissected into five families representing functional immune pathways that were labeled according to their main function as “complement,” “Fc” for Fc-receptors, “MHC” for major histocompatibility complex, “cytokines” for cytokines/chemokines, and “toll-like” for toll-like receptors (Figure 5).

Highlighting the Microglia Pathway with TYROBP as Causal Regulator

The Bayesian inference enabled us to compute the causal regulators of the differential connectivity in individual modules, defined as the genes controlling many downstream nodes in

the respective network (see Extended Experimental Procedures). The causal regulators of the highest scoring immune/microglia module were rank-ordered based on the number of downstream nodes, i.e., the power of regulating other genes, as well as differential expression in LOAD brains. Here, we used a combined score as $G_j = \prod_i g_{ji}$, where g_{ji} is the discriminant value of a j in the case i and is defined as $(\max_i(r_{ji}) + 1 - r_{ji}) / \sum_j r_{ji}$ (Duda et al., 2000). In comparison to the average gene/node in a given network, the causal regulators are expected to have a stronger effect on the clinical outcome as they direct the expression of a significant portion of the network module they reside in. The size of the gene membership for the different regional-specific immune modules ranges from 386 in CB to 1,108 in the PFC, with 247 of the genes in the CB detected in all regions ($p = 1 \times 10^{-19}$). The identity of the key causal regulators is somewhat variable across each brain-regional version of the microglia module of which *CTSC*, *HCK*, *TYROBP*, *SERPINA1*, *S100A11*, *LY86*, *DOCK2*, and *FCER1G* were common to all immune modules, regardless of brain region. Through the combined ranking score based on regulatory strength and differential expression in PFC of LOAD brains, *TYROBP* scored the highest (Figure S4A). Table 1 lists the 20 top-ranking PFC modules and their respective key causal regulators. Expression of *TYROBP* is restricted to cells involved in the innate immunity, including the microglial cells in the brain (Schleinitz et al., 2009). Here, *TYROBP* was significantly upregulated in LOAD brains in the HBTRC sample (1.18-fold, $p = 0.028$), and the direction of this effect was replicated (1.17-fold, $p = 5.1 \times 10^{-5}$) in an independent multicenter cohorts study (see Extended Experimental Procedures and Figure S4B). Additionally, we observed a progression of *TYROBP* expression changes across mild cognitive impairment (MCI) in the replication study (Figure S4B). Estimating what constitutes a “large” or “small” change in gene-expression levels is challenging in microarray analyses. We note, however, that *TYROBP* was the 124th most differentially expressed probe out of 48,803 probes assayed in the replication study cohort. Moreover, *TYROBP* was more differentially expressed in LOAD brains than the classical markers of microglia, *AIF1* and *CD68*, indicating that there was not a relative downregulation of *TYROBP* despite elevated microgliosis in LOAD brains (Perry et al., 2010).

The majority of the common causal regulators were located either in the “Fc” pathway and associated/clustering genes (*HCK*, *SERPINA1*, *S100A11*, *DOCK2*, and *FCER1G*) or the “complement” pathway (*TYROBP*) in the immune/microglia network (Figure 5). Recent reports (we note that our submission predates these reports) show a striking association of a low-frequency DNA variant in *TREM2* to LOAD (Guerreiro et al., 2013; Jonsson et al., 2013). More specifically, *TREM2* is known to associate and signal via *TYROBP*, the key regulator of the immune/microglia network activated in LOAD. Thus, our data-driven, network-based approach places both *TREM2* and *TYROBP* in a gene network that literally unifies them with previous top GWAS risk

(C) Significant (FET p value after correcting for number of modules and functional categories/pathways tested) enrichment of functional categories in conserved modules (left most panel), LOC modules (center panel), or GOC modules (right most panel). The y axis represents the $-\log(p \text{ value})$ of enrichment, whereas the x axis denotes the number of genes per module. Each module contains at least 50 unique gene symbols.

See also Table S1.

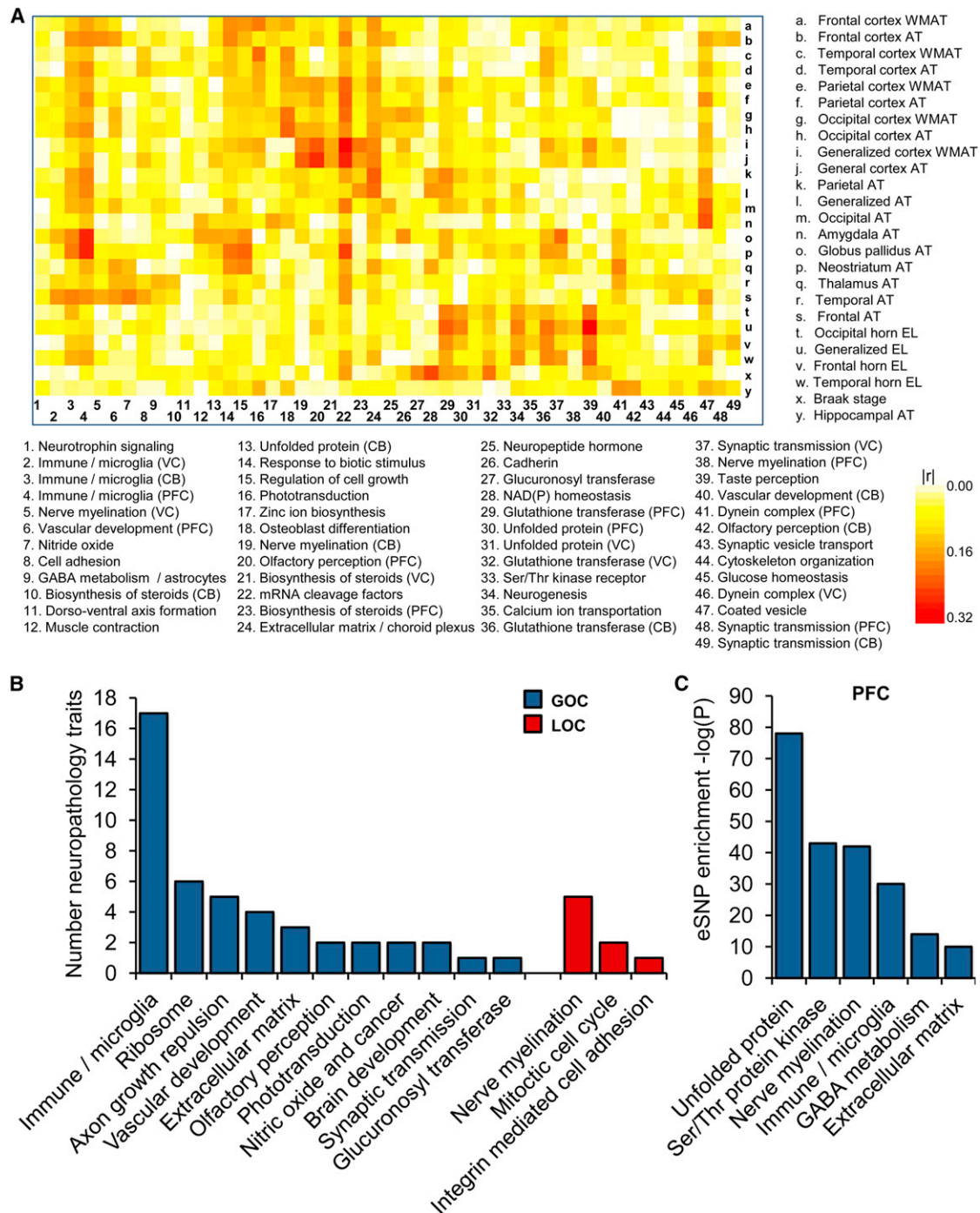


Figure 4. Module Relevance to LOAD Pathology and Enrichment of Brain eSNPs

(A) A heatmap of the correlations ($|r|$) between 49 module principal components (PCs) and 25 LOAD-related neuropathology traits. These modules contain at least 100 probes. AT, atrophy; WMAT, white matter atrophy; EL, enlargement.

(B) Number of significant module-dependent correlations to LOAD-related neuropathology of all differentially connected modules with at least 100 members and showing significant correlation to at least a single neuropathology trait (see [Extended Experimental Procedures](#)). The total number of traits associated with a module was used to rank-order modules for relevance to LOAD pathology.

(C) We tested the enrichment of brain eSNPs in the differentially connected modules of the multitissue coexpression network in LOAD as per brain region. Here we present a significant enrichment of brain eSNPs in many of the PFC modules. We used the FET analysis to assess the significance of the overlap between each module and *cis*-eSNPs, correcting for the number of modules tested.

See also [Figure S2](#) and [Table S1](#).

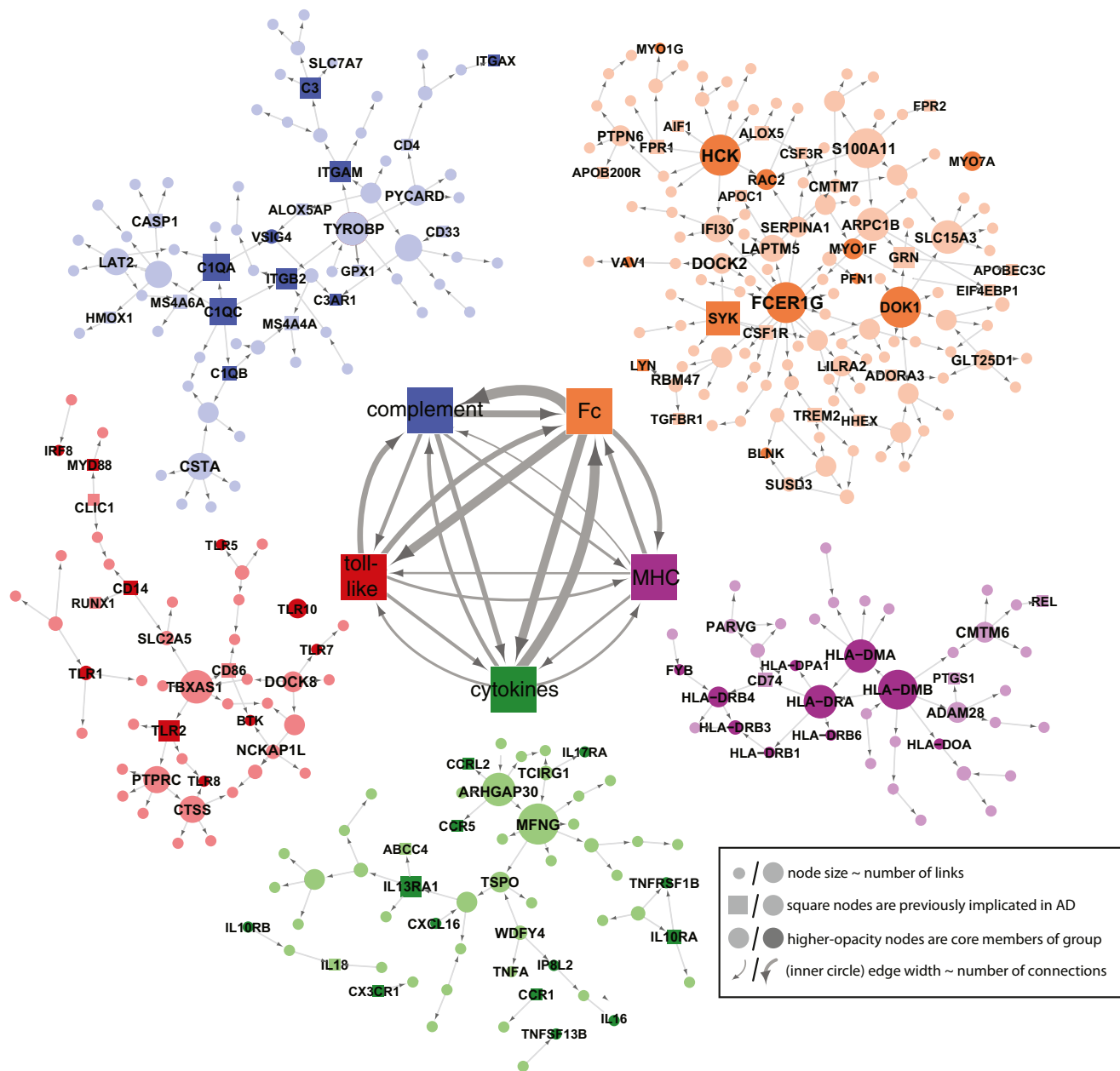


Figure 5. The Bayesian Brain Immune and Microglia Module

A module that correlates with multiple LOAD clinical covariates and is enriched for immune functions and pathways related to microglia activity (PFC module shown). (Inner networks) The PFC module is enriched in genes that can be classified as members of the complement cascade (“complement”), toll-like receptor signaling (“toll-like”), chemokines/cytokines (“chemokine”), the major histocompatibility complex (“MHC”), or Fc-receptor system (“Fc”). The direction and strength of interactions between these pathways are collected across all gene-gene causal relationships that span different pathways. The minimum line width corresponds to a single interaction (MHC to toll-like) and scales linearly to a maximum of 17 interactions (Fc to complement). (Outer networks) Each color-coded group of genes consists of the core members of the different families and genes that are causally related to a given family. Core family members of each pathway are shaded darkly, whereas square nodes in any family denote literature-supported nodes (at least two PubMed abstracts implicating the gene or final protein complex in LOAD or a model of LOAD). Labeled nodes are either highly connected in the original network, literature-implicated LOAD genes, or core members of one of the five immune families. Node size is proportional to connectivity in the module. See also Figure S5.

loci including *MS4A4A*, *MS4A6A*, and *CD33* (Figure 5). These new results provide exciting convergent evidence for the specific microglia network that we had directly implicated as activated in LOAD and reinforce the potential causality of this pathway in

LOAD pathology. In fact, the dissection of the immune/microglia module into distinct families and key causal regulators points toward an important function of the microglia pathways involving genes of the “complement” and/or “Fc” network clusters.

Table 1. Top 20 Modules in PFC Ranked for Relevance to LOAD Pathology

Module	Rank	Top Functional and Cellular Category	N PFC Genes	MDC	Highlighted Causal Regulators ^a
Yellow	1	immune and microglia	1,102	1.49	<i>TYROBP</i> , <i>DOCK2</i> , <i>FCER1G</i>
Pink	2	glutathione transferase	113	92.67	<i>GSTA4</i> , <i>ABCC2</i> , <i>TIMELESS</i>
Gray 1	3	cell junction	51	0.82 ^b	<i>ACBD5</i> , <i>LMAN1</i> , <i>MLL3S</i>
Seashell	4	coated vesicle	278	1.29 ^b	<i>KIFAP3</i> , <i>PCTK2</i> , <i>SNCA</i>
Red 3	5	ribosome	50	24.93	<i>RPS27</i> , <i>RPS18</i> , <i>PCBP2</i>
Green yellow	6	unfolded protein	721	4.50	<i>STIP1</i> , <i>HSPA1A</i> , <i>DOPEY1</i>
Red	7	nerve myelination and oligodendrocytes	987	0.68	<i>ENPP2</i> , <i>PSEN1</i> , <i>GAB2</i>
Gold 2	8	axon growth repulsion	80	3.27	<i>TUBB4</i> , <i>ACTL9</i> , <i>ACTG1</i>
Tan	9	extracellular matrix and choroid plexus cells	700	2.88	<i>SLC22A2</i> , <i>AGTR1</i> , <i>ZIC2</i>
Gold 3	10	dynein complex	67	12.12	<i>TEKT1</i> , <i>FANK1</i> , <i>HYDIN</i>
Light yellow	11	mRNA cleavage	96	6.01	<i>MED6</i> , <i>STAT1P1</i> , <i>SFRS3</i>
Brown 2	12	olfactory perception	77	25.51	<i>PPP2R5A</i> , <i>C1ORF143</i> , <i>RNASE11</i>
Dark cyan	13	steroid biosynthesis	110	1.39 ^b	<i>LAMP2</i> , <i>P2RX7</i> , <i>MID1IP1</i>
Khaki	14	GABA biosynthesis and astrocytes	267	0.29	<i>GJA1</i> , <i>STON2</i> , <i>CST3</i>
Grey 60	15	Ser/Thr kinase receptor	495	4.64	<i>CREBBP</i> , <i>ABCC11</i> , <i>MDGA1</i>
Purple	16	synaptic transmission and neurons	805	1.22	<i>SNAP91</i> , <i>BSN</i> , <i>GLS</i>
Green 4	17	cell cycle	50	0.33	<i>DTL</i> , <i>UBE2C</i> , <i>BUB1</i>
Honey dew	18	muscle contraction	128	1.10 ^b	<i>RFX4</i> , <i>DGCR6</i> , <i>AQP4</i>
Red 2	19	zinc homeostasis	83	1.17 ^b	<i>MT1M</i> , <i>MT1JP</i> , <i>MT1P3</i>
Beige	20	glucose homeostasis	95	12.64	<i>AMPD1</i> , <i>EGR2</i> , <i>PDGFB</i>

This table lists the top 20 rank-ordered modules consisting of at least 50 genes from PFC, or if majority of genes are from PFC in mixed modules with a total of 50 genes or more. See also [Table S1](#) and [Figure S4](#).

^aSelected set of maximum three causal regulators per module.

^bMDC FDR > 10% and therefore not considered significant.

Figure S5 (genes marked in red) highlights many of the key genes in the pathogen phagocytosis pathways found in the immune/microglia module. It is notable how comprehensive representation of specific signal transduction pathways is observed within the two immune families of this module. The strategic network position of *TYROBP* as a causal regulator of many genes mirrors its bottleneck position in several microglia activation-signaling cascades. Extrapolating from this data-driven interaction, it is possible that *TYROBP* may be associated with neuronal pruning activity of the complement system that may be reawakened in LOAD via amyloid- β and tau aggregates ([Perry et al., 2010](#)). In this manner, the network structure can become a data-driven hypothesis generator for disease-relevant interactions.

Structure of Causal Networks Guides Differential Expression in a Distance-Dependent Manner

To test our prediction that *TYROBP* can direct LOAD-associated gene networks, we contrast both the molecular function and genome-wide effects of *TYROBP* with those predicted by the structure of causal networks inferred from human LOAD brains. For this, microglia cells derived from mouse embryonic stem cells were genetically modified by lentiviral vectors to overexpress either full-length or a truncated version of *Tyrbp* that lacks both intracellular immunoreceptor tyrosine-based activation motif (ITAM) motifs ([Extended Experimental Procedures](#) and [Figure S6](#)). To assess the genome-wide gene-expression changes in response to the perturbation of *Tyrbp*, we derived

gene-expression data from the RNA sequencing of mouse microglia cell lines overexpressing (1) vehicle, (2) the full-length *Tyrbp*, or (3) dominant-negative truncated *Tyrbp*. We identified 2,638 and 3,415 differentially expressed genes for the overexpression of full-length *Tyrbp* and truncated *Tyrbp*, respectively ([Table S1](#)), at FDR < 2.5%. Roughly one-third (858 to 1,092) of these genes are found in the most variable gene set in the brain data set used for the network reconstruction. The PFC variant of the human immune/microglia module was highly enriched for genes that are differentially expressed in the full-length or truncated *Tyrbp* experiments ($p < 1 \times 10^{-15}$) ([Figure 6A](#)). We projected results of RNA-sequencing experiments onto a large Bayesian brain network of ~8,000 nodes that contains the microglia module as well as many other modules. In this large network, we could track differential expression of genes that are predicted to be downstream of *TYROBP* at various network path distances ([Figure 6B](#)). The highest predictive power for differential expression is in the primary neighborhood of the perturbed gene, and this power decreases for genes that are farther away in the network. The enrichment for differentially expressed genes in the network neighborhood of *TYROBP* and strong negative correlation between the fraction of confirmed targets and path distance ($r = -0.82$, $p = 4 \times 10^{-7}$) ([Figure 6B](#)) show that our causal network structure is a significant and useful predictor of response to gene perturbations, even in a challenging cross-species setting. Thus, both the structure and direction of links

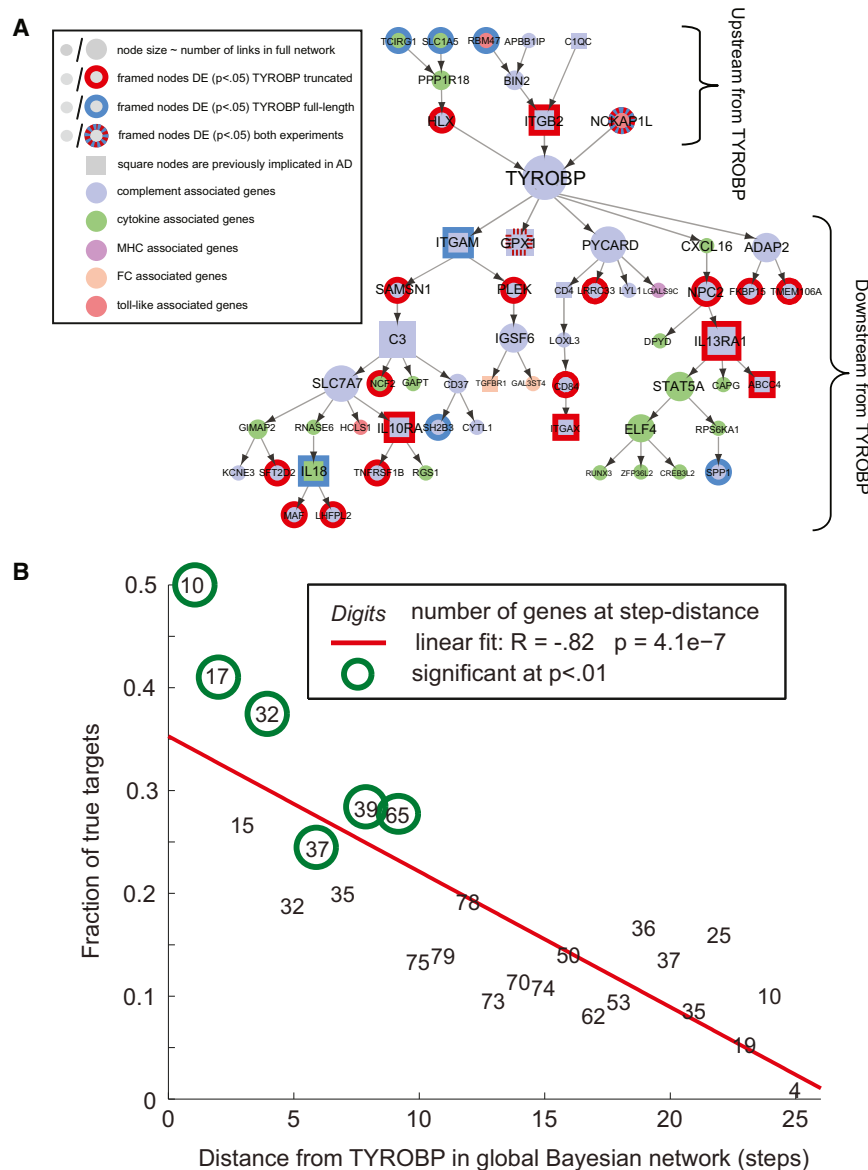


Figure 6. Structure of Causal Networks Guides Differential Expression in a Distance-Dependent Manner

(A) Within the microglia module, we show all genes that receive direct or indirect causal inputs to/from *TYROBP*. Genes that were differentially expressed in either full-length or truncated *Tyrobp* experiments are circled (p value < 0.05 , $n = 4/4/4$ for control/truncated/full-length RNA-sequenced samples). Possible reasons for differentially expressed (DE) predicted upstream genes are mouse-human network differences, network inaccuracy, or presence of feedback loops, which are not represented in a Bayesian framework.

(B) We mapped results of RNA-sequencing experiments onto a large Bayesian network of $\sim 8,000$ nodes that contains the microglia module as well as many other modules. In this large network, we could track differential expression of genes that are predicted to be downstream of *TYROBP* at various network distances (link distances). There was a strong negative correlation ($r = -0.82$, $p = 4 \times 10^{-7}$) between the differentially expressed genes in the microglia and the path distance from *TYROBP* in the brain immune network.

See also Figure S6 and Table S1.

in these causal networks provide significant information on the effects of complex signal transduction mechanisms.

The inferred network structure has significant predictive power for nodes that are several links away from *TYROBP*. We studied the enrichment of functional categories in the gene sets responding to the *Tyrobp* perturbation experiments and applied Bonferroni-corrected p values for statistical significance (Extended Experimental Procedures). Approximately 99% of the differentially expressed genes from the microglia overexpressing intact *Tyrobp* were downregulated compared to the control vehicle. This set was enriched for genes involved in RNA metabolism ($p = 6.2 \times 10^{-5}$) and cell-cycle mitosis ($p = 2.7 \times 10^{-3}$). In the microglia cells overexpressing the dominant-negative truncated *Tyrobp*, 2,856 upregulated genes were enriched for the vacuole/autophagy ($p = 1.7 \times 10^{-8}$) and mitochondrion ($p = 4.6 \times 10^{-4}$), whereas 559 genes involved in

that *TYROBP* may be a therapeutic target in prevention of neuronal damage in LOAD.

DISCUSSION

The construction of gene-regulatory networks in a large sampling of human brain specimens has revealed many facets of the molecular-interaction structure in LOAD, when compared to that in nondemented brains. A comprehensive characterization of gene-network connectivity and its regulation and association to disease can provide critical insights into the underlying mechanisms and identify genes that may serve as effective targets for therapeutic intervention. For instance, targeting genes that are the most central (highly connected) may be more effective in disrupting disease-related networks for the purpose of therapy, but that could be at the cost of

histone assembly ($p = 1.6 \times 10^{-31}$) were downregulated. Moreover, the *Tyrobp*-regulatory effect reflects a degree of symmetry as 658 genes, related to the vacuole/autophagy ($p = 5 \times 10^{-3}$), were downregulated by active *Tyrobp* but upregulated in cells expressing dominant-negative truncated *Tyrobp*. These findings are of interest because they link the far downstream effects of *TYROBP* to known molecular pathology in LOAD, such as abnormalities in the cell cycle, mitochondrion, and autophagy (Coskun et al., 2004; Webber et al., 2005). The accumulated data suggest

more adverse effects. In summary, the utility of network-based approaches to complex disease includes the following: (1) elucidating the biological function and molecular context of a particular set of causal genes, (2) establishing a framework to map interaction between genes and network modules, (3) providing an objective filter for rank-ordering genes based on connectivity or other network features, (4) defining dynamic changes and corresponding causal regulators of the altered network structure associated with disease condition, (5) identifying modules and pathways causally related to disease, and (6) revealing tissue-to-tissue interactions that can aid in the identification of key target tissues for disease (Dobrin et al., 2009). The present study utilizes many of these network advantages to highlight and prioritize pathways and gene targets causally related to LOAD.

Our network-based integrative analysis not only highlighted the immune/microglia module as the molecular system most strongly associated with the pathophysiology of LOAD but also identified the key network regulators, including *TYROBP*. In a separate in vitro study, we have found that the microglia-expressed *TYROBP* is directly involved in amyloid- β turnover and neuronal damage (our unpublished data). Of interest, mutations in *TYROBP* or *TREM2* cause Nasu-Hakola disease (Bianchin et al., 2010), a rare Mendelian disease characterized by bone reabsorption dysfunction and chronic inflammatory neurodegeneration, leading to death in the fourth or fifth decade of life. The exact pathomechanism underlying Nasu-Hakola disease is still unclear, but it was hypothesized that failure of proper microglial clearance is causal for the lethal effect of neurodegeneration. Thus, dysfunctional immune/microglia pathways might not be unique to LOAD. To test the generalization of this concept, we explored the connection of the immune/microglia module to Huntington disease (HD), another neurodegenerative disease. HD pathology, caused by expanded alleles of a variable stretch of trinucleotide (CAG) repeat length in *HTT* (The Huntington's Disease Collaborative Research Group, 1993), features astrogliosis and neurodegeneration of the striatum, prefrontal cortex, and hippocampus. We constructed molecular networks in the PFC from 194 HD patients genotyped for CAG allele size (see Extended Experimental Procedures) and found that the PFC version of the immune/microglia module was well conserved between LOAD and HD in terms of gene annotation (75% overlap, p value $< 1 \times 10^{-300}$). This module, however, did not show any alteration in connectivity in HD brains compared to the disease-free controls used in our LOAD study. Moreover, through a PCA, we did not detect any gene-expression correlation of the HD brain immune/microglia module to expanded CAG repeat length ($r = -0.05$, FDR = 56%), a key biomarker for predicting HD severity (Gusella and MacDonald, 2006). Thus, based on the comparison to HD, the disease-related effect of the immune/microglia module appears to be specific to LOAD (and possibly Nasu-Hakola disease).

Immune activation in LOAD may have multifaceted activity: long-term use of nonsteroid anti-inflammatory drugs (NSAIDs) before onset of the disease decreases risk (Etminan et al., 2003), and microglia effector function via interfering with reactive oxygen production, cytokines, and complement cascade

members has been postulated to damage healthy neurons and synapses (Cameron and Landreth, 2010). Close association and positive feedback between amyloid- β and microglia (Meyer-Luehmann et al., 2008) further cloud the cause and effect relationships of inflammation to disease progression. Without a causal framework for these observations, it is difficult to find optimal molecular targets that direct LOAD inflammation. Therefore, we integrated clinical factors with whole-genome genotype and molecular trait data to identify a network module containing several microglia-signaling cascades functionally related to the reactive oxygen burst during pathogen phagocytosis. We highlight the causal regulator *TYROBP* that exerts control over multiple genes within this module and pathways involved in LOAD, thus validating our network structure and its relevance to LOAD pathology. This approach appears to offer insights for drug-discovery programs that can affect neurodegenerative diseases, such as LOAD.

EXPERIMENTAL PROCEDURES

Raw gene-expression data together with information related to demographics, disease state, and technical covariates are available via the GEO database (GEO accession number GSE44772; GSE44768, GSE44770, and GSE44771). A brief description of key methods and sample description are provided below, whereas complete details are found in the Extended Experimental Procedures.

Data Sets and Sample Processing

We compiled six disease- and tissue-specific gene expression data sets consisting of 1,647 postmortem specimens from three brain regions (PCF [BA9], VC [BA17], and CB) in LOAD and nondemented subjects recruited through the HBTRC. Each subject was diagnosed at intake and via extensive neuropathology examination. Tissues were profiled on a custom-made Agilent 44K array of 39,579 gene-specific DNA probes, and each subject genotyped for 838,958 SNPs.

Molecular Networks and Causal Regulators

We constructed both multitissue and single-tissue coexpression networks from the top one-third ($n = 13,193$) of the most variably expressed genes in each tissue and condition. We computed the MDC in LOAD brains as:

$$\delta_{\Omega}(X, Y) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N k_{ij}^X}{\sum_{i=1}^{N-1} \sum_{j=i+1}^N k_{ij}^Y},$$

where k_{ij} is the connectivity between two genes i and j in a given network, and assessed the statistical significance through the FDR method. We constructed causal probabilistic Bayesian networks from individual coexpression modules and used brain *cis*-eSNPs as priors to infer directionality between nodes (see Figure S3). For this, we identified 11,318 unique *cis*-eSNPs transcripts at FDR of 10% (Extended Experimental Procedures), all listed in Table S1. The Bayesian inference allowed us to compute the causal regulators of the differential connectivity in individual modules by examining the number of N-hub downstream nodes.

Mouse Microglia Cultivation, Cell Transduction, and RNA Sequencing

Genome-wide gene expression of messenger RNA (mRNA) from cultivated microglia cells overexpressing intact or genetically modified *TYROBP* was sequenced with a TruSeq Kit for RNA capture and HiSeq 2000 for the sequencing. Read mapping was done with the TopHat (Trapnell et al., 2009) RNA-seq aligner.

ACCESSION NUMBERS

Raw gene-expression data together with information related to demographics, disease state, and technical covariates are available via the GEO database (GEO accession number GSE44772; see also accession numbers GSE44768, GSE44770, and GSE44771).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2013.03.030>.

ACKNOWLEDGMENTS

H.N. and L.-G.B. were supported by the Hertie-Foundation and the Deutsche Forschungsgemeinschaft (FOR1336, SFB704, KFO177). H.N. is a member of the DFG-funded excellence cluster ImmunoSensation. We thank Jessica Schumacher and Rita Hass for technical assistance. We are grateful to the HBTRC for the generous gift of human postmortem brain samples. The authors are also grateful to the participants in the Religious Orders Study and the Memory and Aging Project. This work is supported by the National Institutes of Health (R01 AG034504, R01 AG030146, P30 AG10161, R01 AG17917, R01 AG15819, K08 AG034290, P30 AG10161, R01 AG11101, and NS032765), the Illinois Department of Public Health, and start-up funding from the University of Miami, Miller School of Medicine. J.M., A.A.P., C.Z., T.X., R.D., E.F., S.M., M.N., C.M., and D.J.S. are employees and shareholders of Merck & Co., Inc. J.R.L. is an employee and shareholder of Novartis.

Received: April 11, 2012

Revised: October 17, 2012

Accepted: March 22, 2013

Published: April 25, 2013

REFERENCES

- Ansari, M.A., and Scheff, S.W. (2010). Oxidative stress in the progression of Alzheimer disease in the frontal cortex. *J. Neuropathol. Exp. Neurol.* 69, 155–167.
- Bayés, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D., Whittle, I.R., Choudhary, J.S., and Grant, S.G. (2011). Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* 14, 19–21.
- Bertram, L., Lill, C.M., and Tanzi, R.E. (2010). The genetics of Alzheimer disease: back to the future. *Neuron* 68, 270–281.
- Bettens, K., Sleegers, K., and Van Broeckhoven, C. (2013). Genetic insights in Alzheimer's disease. *Lancet Neurol.* 12, 92–104.
- Bianchin, M.M., Martin, K.C., de Souza, A.C., de Oliveira, M.A., and Rieder, C.R. (2010). Nasu-Hakola disease and primary microglial dysfunction. *Nat. Rev. Neurol.* 6, 193–201.
- Boyles, J.K., Pitas, R.E., Wilson, E., Mahley, R.W., and Taylor, J.M. (1985). Apolipoprotein E associated with astrocytic glia of the central nervous system and with nonmyelinating glia of the peripheral nervous system. *J. Clin. Invest.* 76, 1501–1513.
- Braak, H., and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 82, 239–259.
- Brookmeyer, R., Johnson, E., Ziegler-Graham, K., and Arrighi, H.M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement.* 3, 186–191.
- Cameron, B., and Landreth, G.E. (2010). Inflammation, microglia, and Alzheimer's disease. *Neurobiol. Dis.* 37, 503–509.
- Cechetto, D.F., Hachinski, V., and Whitehead, S.N. (2008). Vascular risk factors and Alzheimer's disease. *Expert Rev. Neurother.* 8, 743–750.
- Chen, Y., Zhu, J., Lum, P.Y., Yang, X., Pinto, S., MacNeil, D.J., Zhang, C., Lamb, J., Edwards, S., Sieberts, S.K., et al. (2008). Variations in DNA elucidate molecular networks that cause disease. *Nature* 452, 429–435.
- Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., and Pericak-Vance, M.A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261, 921–923.
- Coskun, P.E., Beal, M.F., and Wallace, D.C. (2004). Alzheimer's brains harbor somatic mtDNA control-region mutations that suppress mitochondrial transcription and replication. *Proc. Natl. Acad. Sci. USA* 101, 10726–10731.
- Dobrin, R., Zhu, J., Molony, C., Argman, C., Parrish, M.L., Carlson, S., Allan, M.F., Pomp, D., and Schadt, E.E. (2009). Multi-tissue coexpression networks reveal unexpected subnetworks associated with disease. *Genome Biol.* 10, R55.
- Dodel, R.C., Hampel, H., and Du, Y. (2003). Immunotherapy for Alzheimer's disease. *Lancet Neurol.* 2, 215–220.
- Duda, R.O., Hart, P.E., and Stork, D.G. (2000). *Pattern Classification*, Second Edition (New York: John Wiley & Sons, Inc.).
- Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. *Nature* 452, 423–428.
- Etmann, M., Gill, S., and Samii, A. (2003). Effect of non-steroidal anti-inflammatory drugs on risk of Alzheimer's disease: systematic review and meta-analysis of observational studies. *BMJ* 327, 128.
- Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S., Younkin, S., et al.; Alzheimer Genetic Analysis Group. (2013). TREM2 variants in Alzheimer's disease. *N. Engl. J. Med.* 368, 117–127.
- Gusella, J.F., and MacDonald, M.E. (2006). Huntington's disease: seeing the pathogenic process through a genetic lens. *Trends Biochem. Sci.* 31, 533–540.
- Huang, Y., and Mucke, L. (2012). Alzheimer mechanisms and therapeutic strategies. *Cell* 148, 1204–1222.
- Jonsson, T., Atwal, J.K., Steinberg, S., Snaedal, J., Jonsson, P.V., Bjornsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J., et al. (2012). A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* 488, 96–99.
- Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P.V., Snaedal, J., Bjornsson, S., Huttenlocher, J., Levey, A.I., Lah, J.J., et al. (2013). Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* 368, 107–116.
- Li, G.F., Bien-Ly, N., Andrews-Zwilling, Y., Xu, Q., Bernardo, A., Ring, K., Halabisky, B., Deng, C., Mahley, R.W., and Huang, Y. (2009). GABAergic interneuron dysfunction impairs hippocampal neurogenesis in adult apolipoprotein E4 knockin mice. *Cell Stem Cell* 5, 634–645.
- Luchsinger, J.A. (2008). Adiposity, hyperinsulinemia, diabetes and Alzheimer's disease: an epidemiological perspective. *Eur. J. Pharmacol.* 585, 119–129.
- Mani, K.M., Lefebvre, C., Wang, K., Lim, W.K., Basso, K., Dalla-Favera, R., and Califano, A. (2008). A systems biology approach to prediction of oncogenes and molecular perturbation targets in B-cell lymphomas. *Mol. Syst. Biol.* 4, 169.
- Meyer-Luehmann, M., Spires-Jones, T.L., Prada, C., Garcia-Alloza, M., de Calignon, A., Rozkalne, A., Koenigsnecht-Talboo, J., Holtzman, D.M., Bacskai, B.J., and Hyman, B.T. (2008). Rapid appearance and local toxicity of amyloid-beta plaques in a mouse model of Alzheimer's disease. *Nature* 451, 720–724.
- Morawski, M., Brückner, G., Jäger, C., Seeger, G., Matthews, R.T., and Arendt, T. (2012). Involvement of perineuronal and perisynaptic extracellular matrix in Alzheimer's disease neuropathology. *Brain Pathol.* 22, 547–561.
- Murray, I.V., Proza, J.F., Sohrabji, F., and Lawler, J.M. (2011). Vascular and metabolic dysfunction in Alzheimer's disease: a review. *Exp. Biol. Med.* (Maywood) 236, 772–782.

- Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D.H. (2008). Functional organization of the transcriptome in human brain. *Nat. Neurosci.* *11*, 1271–1282.
- Perry, V.H., Nicoll, J.A., and Holmes, C. (2010). Microglia in neurodegenerative disease. *Nat Rev Neurol* *6*, 193–201.
- Qiu, C., Xu, W., and Fratiglioni, L. (2010). Vascular and psychosocial factors in Alzheimer's disease: epidemiological evidence toward intervention. *J. Alzheimers Dis.* *20*, 689–697.
- Schadt, E.E. (2009). Molecular networks as sensors and drivers of common human diseases. *Nature* *461*, 218–223.
- Schadt, E.E., Friend, S.H., and Shaywitz, D.A. (2009). A network view of disease and compound screening. *Nat. Rev. Drug Discov.* *8*, 286–295.
- Schäfer, S., and Kolkhof, P. (2008). Failure is an option: learning from unsuccessful proof-of-concept trials. *Drug Discov. Today* *13*, 913–916.
- Schiffman, S.S., Graham, B.G., Sattely-Miller, E.A., Zervakis, J., and Welsh-Bohmer, K. (2002). Taste, smell and neuropsychological performance of individuals at familial risk for Alzheimer's disease. *Neurobiol. Aging* *23*, 397–404.
- Schleinitz, N., Chiche, L., Guia, S., Bouvier, G., Vernier, J., Morice, A., Hous-saint, E., Harlé, J.R., Kaplanski, G., Montero-Julian, F.A., and Vély, F. (2009). Pattern of DAP12 expression in leukocytes from both healthy and systemic lupus erythematosus patients. *PLoS ONE* *4*, e6264.
- The Huntington's Disease Collaborative Research Group. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* *72*, 971–983.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* *25*, 1105–1111.
- Walker, D.G., Link, J., Lue, L.F., Dalsing-Hernandez, J.E., and Boyes, B.E. (2006). Gene expression changes by amyloid beta peptide-stimulated human postmortem brain microglia identify activation of multiple inflammatory processes. *J. Leukoc. Biol.* *79*, 596–610.
- Webber, K.M., Raina, A.K., Marlatt, M.W., Zhu, X., Prat, M.I., Morelli, L., Casadesus, G., Perry, G., and Smith, M.A. (2005). The cell cycle in Alzheimer disease: a unique target for neuropharmacology. *Mech. Ageing Dev.* *126*, 1019–1025.
- Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., et al.; NACC-Neuropathology Group. (2009). Genetic control of human brain transcript expression in Alzheimer disease. *Am. J. Hum. Genet.* *84*, 445–458.
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* *4*, e17.

EXTENDED EXPERIMENTAL PROCEDURES

Study Populations and Clinical Data

One thousand six hundred and forty-seven frozen tissue samples from cerebellum (CB), visual cortex (VC) BA17, and dorsolateral prefrontal cortex (PFC) BA9 were provided by the Harvard Brain Tissue Resource Center (HBTRC) at McLean Hospital (Belmont, MA, USA). These regions were selected because PFC is commonly affected in LOAD, whereas VC and CB remain less affected throughout most of the disease progression (Braak and Braak, 1991). All autopsied brains were collected from subjects with LOAD diagnosis or from normal nondemented subjects, for whom both the donor and the next of kin had completed the HBTRC informed consent (<http://www.brainbank.mclean.org/>). Tissue collection and the research were conducted according to the HBTRC guidelines (<http://www.brainbank.mclean.org/>). The HBTRC samples were primarily of Caucasian ancestry, as only eight non-Caucasian outliers were identified and therefore excluded for further analysis. Postmortem interval (PMI) was 17.8 ± 8.3 hr, sample pH was 6.4 ± 0.3 and RNA integrity number (RIN) was 6.8 ± 0.8 for the average sample in the overall cohort. These were composed of 376 LOAD patients and 173 nondemented subjects. Braak stage, general and regional atrophy, gray and white matter atrophy and ventricular enlargement were assessed and cataloged by pathologists at McLean Hospital (Belmont, MA, USA). The clustering of LOAD-related pathology traits and age is shown in Figure 2. It is of note that age formed a separate cluster from most of the LOAD neuropathology, however hippocampal atrophy was positively correlated with age ($r = 0.28$) while Braak staging of LOAD pathology was negatively correlated with age ($r = -0.15$). The negative correlation between age and Braak stage in the HBTRC sample most likely reflects the higher degree of Tau-related neurofibrillary tangle load associated with the earlier onset and more severe LOAD cases that skew the correlation structure, as has been previously noted in combined familial and sporadic LOAD cases (Nochlin et al., 1993). Finally, for the immune/microglia comparative study we used a brain tissue sampling (PFC, CB, and VC) of 194 Huntington disease (HD) patients from the HBTRC, where each of the 582 HD patient sample went through identical procedure as described for the LOAD and nondemented brain samples. In addition, the severity of pathology in the HD autopsy brains was determined from the Vonsattel grading system (Vonsattel et al., 1985), and the *HTT* CAG repeat allele size was determined using a modification of previously described polymerase chain reaction amplification assay (Warner et al., 1993).

Brain samples for testing the replication of the *TYROBP* differential expression were obtained from 21 National Alzheimer's Coordinating Center (NACC) brain banks and from the Miami Brain Bank as previously described (Myers et al., 2007; Webster et al., 2009). The Rush Alzheimer's Disease Center series consists of two cohorts: the Religious Orders Study (ROS) and the Memory and Aging Project (MAP) (Bennett et al., 2005, 2006a, 2006b). cRNA was hybridized as before to Illumina HT-12 Expression Bead Chip (48,803 transcripts) via standard protocols using an Illumina Bead Station 500GX (Webster et al., 2009). Disease status included 377 with LOAD, 119 classified with mild cognitive impairment (MCI), and 359 healthy nondemented controls. Brain regions sampled in these patients were 726 prefrontal and 129 temporal cortex samples.

Array Design, Gene-Expression Normalization, Covariate Analysis, and Genotyping

RNA preparation and array hybridizations applied custom microarrays manufactured by Agilent Technologies consisting of 4,720 control probes and 39,579 probes targeting transcripts representing 25,242 known and 14,337 predicted genes. One microgram of total RNA from each of the 2,229 (LOAD, nondemented, HD) brain tissues was reverse transcribed and labeled with either Cy3 or Cy5 fluorochrome. Purified Cy3 or Cy5 complementary RNA was hybridized to at least two single microarrays with fluor reversal for 24 hr in a hybridization chamber, washed, and scanned using a laser confocal scanner. Arrays were quantified on the basis of spot intensity relative to background, adjusted for experimental variation between arrays using average intensity over multiple channels, and fitted to an error model to determine significance (type I error) as previously described (Emilsson et al., 2008). This microarray dataset is MIAME compliant and the raw gene expression data for all hybridizations together with information related to demographics, disease state (LOAD or non-demented) and technical covariates (pH, RIN, PMI, batch, preservation method) are available via the GEO database (GEO accession number GSE44772 including the regional-specific subseries GSE44768, GSE44770, GSE44771).

Gene expression was reported as the mean-log ratio of individual microarray intensities relative to average intensities of all samples. Gene expression data were generated using Rosetta Resolver gene-expression analysis software (version 7.0, Rosetta Biosoftware) and MATLAB (The MathWorks). To remove bias in expression profiles related to potential latent variables unrelated to underlying biological processes, we implemented a normalization method based on control probes present on the microarrays. We separated the control probes into two classes: (1) specialty probes, such as spike-in probes or other probes designed to monitor the quality of the microarrays; and (2) border probes used to describe the geometry of the microarray. We then identified the principal components (PC) explaining the variability of each control probe class and then identified the same components from a randomly permuted data set. We performed 10,000 permutations for each set of control probes and selected Principal Components with p values defined as $\{\text{number of } [\text{var}(\text{random-PC}_i) > \text{var}(\text{PC}_i)] / (\text{number of permutations}) < 10^{-4}\}$. The expression data in each tissue for each probe are thus the residuals from a linear model fitting incorporating the significant PCs.

Next we analyzed the contribution of each of the experimental covariate's age, gender, sex, and technical covariates. The Kolmogorov-Smirnov (KS) test was used to capture the difference of the correlation p value distributions for each covariate per brain region and condition. Figure S1 shows the correlation p values between a specific covariate and all the probes based on

the expression data before (red curve, raw) and after (black curve, corrected) the adjustment in LOAD or nondemented brains. Particular attention was paid to established covariates of mRNA quality that have been demonstrated in brain tissues, which included: (1) Extreme confounding of demographic variables in disease versus controls; it was not possible to match samples based on demographic variables including age, sex, and ethnicity as many of these factors differ in their distributions substantially across the disease groups. Ethnicity was determined using STRUCTURE resulting in the exclusion of eight outliers (non-Caucasians), which were removed from all studies. (2) Pre-/postmortem quantitative and quality factors including PMI, RIN, tissue pH status, manner of death, and/or agonal state. Seven samples with RIN < 5 were removed from the analysis. (3) Batch effects; day-to-day variability in the amplification process can influence the observed gene expression, even when profiled against a common-pool. For this particular data set, the batch effects are further exacerbated by the two additional facts: (1) the demographic compositions of the two batches differ, and (2) for technical reasons, as it is preferable to have your pool sample be a representation of the greater population (mean-log ratio of individual microarray intensities was relative to average intensities of all samples in the final set). In summary, the distribution of p values obtained from modeling the covariates listed above strongly suggest that adjusting for them would improve the downstream analysis steps. Therefore we chose a robust linear regression model for covariate corrections as $\text{rlm}(\text{expression} \sim \text{RIN} + \text{pH} + \text{PMI} + \text{age} + \text{batch} + \text{preservation} + \text{gender})$, and then the residuals were used for further testing including the genetics of gene expression, network construction, differential gene expression, and gene-expression clinical trait correlations carried out in the present study.

DNA isolation was performed as previously described (Emilsson et al., 2008). Each sample was genotyped on two different platforms, the IlluminaHumanHap650Y array and a custom Perlegen 300K array (a focused panel for detection of singleton SNPs). Counting only the union of markers from both genotyping platforms (114,925 SNPs were in the intersection), a total of 838,958 unique SNPs were used for the analysis. Finally, APOE genotypes were obtained through restriction fragment length polymorphism (RFLP) analysis (Lai et al., 1998).

Constructing the Multitissue Coexpression Networks

We constructed a multitissue coexpression network that simultaneously captured the intra- and intertissue gene-gene interactions between the LOAD and normal states. These networks can be characterized only when multiple tissues are monitored in a population of individuals donating all tissues (PFC, CB, and VC). For construction of the multitissue brain network, we used the top one-third (T1) or 13,193 of the most variable genes (interindividual variability) in a given tissue and physiological state (LOAD or nondemented). The most variable gene expression traits in the cohort were defined based on a previously described and validated error model testing the intensity between the experimental and reference channel, to obtain p values for each individual expression trait and then compute the standard deviation (SD) of $-\log_{10}(\text{p value})$ for each trait over all samples profiled for a given tissue. Then we rank-ordered all of the genes profiled in each tissue based on this SD value (rank-ordered in descending order). Genes that fall at the top of this rank ordered list can be considered as the most differentially expressed or variable genes in the study. We have previously shown that this type of ranking approach well captures the most active genes in a given set of tissue samples (Emilsson et al., 2008; He et al., 2003).

Specifically, for each of the three T1 data sets we first assigned each gene a unique identifier by combining probe ID and tissue name, then aligned the samples, and finally merged the three sets into a multitissue expression data set, which included 39,579 genes and 129 AD samples. The combined 39,579 genes were then mapped to the three tissues in normal nondemented data set in a supervised manner, which led to another data set of 39,579 genes and 99 normal samples. The LOAD and normal multitissue data sets were independently processed through the weighted gene coexpression network analysis (Zhang and Horvath, 2005). The weighted network analysis begins with a matrix of the Pearson correlations between all gene pairs, then converts the correlation matrix into an adjacency matrix using a power function $f(x) = x^\beta$. The parameter β of the power function is determined in such a way that the resulting adjacency matrix, i.e., the weighted coexpression network, is approximately scale-free. To measure how well a network satisfies a scale-free topology, we use the fitting index proposed by Zhang et al. (Zhang and Horvath, 2005), i.e., the model fitting index R^2 of the linear model that regresses $\log(p(k))$ on $\log(k)$ where k is connectivity and $p(k)$ is the frequency distribution of connectivity. The fitting index of a perfect scale-free network is 1. The connectivity between genes or k_{ij} is a transformed correlation between the expression profiles of two genes, $|r(i,j)|^\beta$, with r as the Pearson correlation coefficient. The parameter β (>0) of the power function is determined in such a way that the global probability distribution of the resulted connectivity values for all the gene pairs is scale free. Thus, k_{ij} is a continuous value ranging from 0 to 1.

To explore the modular structures of the co-expression network, the adjacency matrix is further transformed into a topological overlap matrix (TOM) (Ravasz et al., 2002). As the topological overlap between two genes reflects not only their direct interaction but also their indirect interactions through all the other genes in the network, previous studies (Ravasz et al., 2002; Zhang and Horvath, 2005) have shown that topological overlap leads to more cohesive and biologically more meaningful modules. To identify modules of highly coregulated genes, we used average linkage hierarchical clustering to group genes based on the topological overlap of their connectivity, followed by a dynamic cut-tree algorithm to dynamically cut clustering dendrogram branches into gene modules (Langfelder et al., 2008). To distinguish between modules, each module was assigned a unique color identifier, with the remaining, less well connected genes colored gray. To compare and contrast two multitissue networks, we combined their TOM heatmaps into a single large network. In the combined heatmap, the upper panel shows the hierarchical clustering on the TOM of the LOAD network, whereas the color bar below represents the gene modules. Similarly, the lower panel represents the TOM from normal multitissue network. The color intensity in the map represents the interaction strength between genes. This connectivity map

highlights how genes in the multitissue transcriptional networks fall into distinct network modules, where genes within a given module are more highly interconnected with each other (blocks along the diagonal of the matrix) than with genes in other modules.

Defining the Modular Differential Connectivity

We developed two measures to quantify the difference between the connectivity among a set of genes (or module, denoted Ω) in LOAD versus normal nondemented networks. Given a set of N genes and two networks, x and y , the first MDC measurement is the ratio of the average connectivity among the N genes in the network x to that among the same gene set in network y , specified by the formula below:

$$\delta_{\Omega}(x, y) = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N k_{ij}^x}{\sum_{i=1}^{N-1} \sum_{j=i+1}^N k_{ij}^y}, \quad (1)$$

where, k_{ij} is the connectivity between two genes i and j in a given network.

Obviously, $\delta_{\Omega}(x, y)$ and $\delta_{\Omega}(y, x)$ are reciprocal, i.e., $\delta_{\Omega}(x, y) = 1/\delta_{\Omega}(y, x)$.

The significance or false discovery rate (FDR) of the statistic MDC can be accessed by permuting the data underlying the two networks. We differentiate two scenarios, gain of connectivity ($\delta_{\Omega}(x, y) > 1$) and loss of connectivity ($\delta_{\Omega}(x, y) < 1$). Given M permutations, FDR of MDC is computed as follows:

$$FDR(\delta_{\Omega}(x, y) > 1) = \frac{1}{M} \sum_{p=1}^M \delta_{\Omega}(x, y) > \delta_{\Omega}(x^p, y^p) \sum_p \delta_{\Omega}(x, y) > \delta_{\Omega}(x^p, y^p)$$

or

$$FDR(\delta_{\Omega}(x, y) < 1) = \frac{1}{M} \sum_{p=1}^M \delta_{\Omega}(x, y) > \delta_{\Omega}(x^p, y^p) \sum_p \delta_{\Omega}(x, y) < \delta_{\Omega}(x^p, y^p),$$

where x^p and y^p are the networks derived from the permuted data. To rigorously assess significance of MDC, we estimate two types of FDR estimates, one based on shuffled samples, i.e., networks with nonrandom nodes but random connections, and the other based on shuffled gene labels, i.e., networks with random nodes but nonrandom connections, and then we select the larger value as the final FDR estimate.

Differential Gene Expression and Module Relevance to LOAD Pathology

To identify differential expression of individual genes between LOAD and nondemented controls, Agilent gene-expression data generated for each brain region were first adjusted using robust linear regressions for age, gender, PMI, tissue preservation method, tissue pH, RIN, and sample preparation batch. The residuals following regression were analyzed by ANOVA to identify reporters that showed differential expression between LOAD and control samples in each tissue, and a Monte Carlo analysis was performed by permuting sample order to estimate the FDR. Differentially expressed reporters were defined as those with a Bonferroni corrected p value < 0.05 ; this corresponds to a nominal $p < 2.46 \times 10^{-7}$ and is well below the 0% FDR threshold for all three brain regions tested.

To examine how each gene module was related to LOAD neuropathology traits, we first performed principal component analysis (PCA) for each module and then computed module-trait relevance using two complementary approaches: (1) the correlation between the first principal component (Module Eigengene) and each trait and (2) the correlation (the square root of R square) between the top principal components and each trait through multivariate regression model. The significance (p value) and FDR of each correlation was also calculated. FDR was estimated through random permutation of sample names of the trait data. A module is associated with a trait if both correlation p value and FDR are below 0.05. The total number of traits associated with a module is used to quantify the association of a module with LOAD.

Detection of eSNPs in the Different Brain Regions

Each of the 838,958 SNPs was tested for association to all 39,579 expression traits using Kruskal Wallis test based on the residuals, left after correction for covariates using robust linear regression as described in the section “*Array Design, Gene Expression Normalization, Covariate Analysis, and Genotyping*.” Expression quantitative trait loci (e)QTLs for gene expression traits were determined by identifying the SNP most strongly associated with each expression trait profiled on the array over all the 838,958 genotyped SNPs. *Cis* analysis was limited to SNPs located within 1 Mb of either side of the transcription start or end within the gene body, whereas *trans* effects were defined as the associated SNPs located ≥ 1 Mb distance from the physical location of the corresponding probe. The association p value was adjusted to control for testing of multiple SNPs and expression traits using two different methods: (1) a highly conservative Bonferroni correction method to constrain the study-wise significance level, and (2) an empirical FDR method (Storey and Tibshirani, 2003) that constrains the overall rate of false positive events. For *cis*-eQTL, to achieve a study-wise significance level

of 0.05, the Bonferroni adjusted p value threshold was computed as $0.05/(39,579 \times N_i)$, where N_i denotes the number of SNPs tested for trait i within the two Mb window, over all 39,579 expression traits tested. The nominal p value to meet this significance threshold is 3.0×10^{-8} . The Bonferroni adjustment method can be conservative when there is dependence among the expression traits and among the SNP genotypes. Given that strong correlation structures exist among expression traits and among SNP genotypes in a given linkage disequilibrium (LD) block, the Bonferroni adjustment may be overly conservative. Therefore, we used an empirical FDR method based on permutations that accounts for the correlation structures among the expression traits and among the SNP genotypes. We constrained the empirically determined FDR to be less than 10%. Here, the FDR was estimated as the ratio of the average number of eQTLs found in data sets with randomized sample labels to the number of eQTLs identified in the original data set. Since the number of tests was large, we found the empirical null distribution was very stable and ten permutation runs were sufficient for convergence to estimate FDR. FDR computation was performed separately for *cis* and *trans* associations resulting in nominal p value cutoffs of 5.0×10^{-5} and 1.0×10^{-8} for *cis*- and *trans*-eQTLs, respectively. The significance of the *trans*-eQTL was also assessed by the Bonferroni method and by constraining the empirically determined FDR to be less than 10%. In the case of *trans*-eQTL, all 838,958 SNPs were tested for association to each of the 39,280 expression traits. Therefore, the Bonferroni adjusted p value threshold was computed as $0.05/(838,958 \times 39,579) = 1.5 \times 10^{-12}$. Table S1 lists all *cis*- and *trans*-acting eSNPs detected in each brain region and condition at FDR of 10% and provides nominal p values as well.

For the present work, we pay particular attention to *cis* acting eSNPs for the following reasons: (1) for construction of causal probabilistic Bayesian networks as described in the section “Reconstruction of the Bayesian Networks and Identification of Key Causal Regulators” and (2) for studying the enrichment of eSNPs in LOAD associated modules. Brain eSNPs were compared to the 111 gene modules identified from the multitissue gene coexpression network analysis and the enrichment of *cis*-eSNPs tested using the Fisher’s exact test (FET) to assess the significance of the overlap between each module and *cis* eSNPs, correcting for number of modules as eSNPs are regarded as a single category.

Reconstruction of the Bayesian Networks and Identification of Key Causal Regulators

Reconstruction of the Bayesian Networks

Bayesian networks are directed acyclic graphs in which the edges of the graph are defined by conditional probabilities that characterize the distribution of states of each node given the state of its parents. The network topology defines a partitioned joint probability distribution over all nodes in a network, such that the probability distribution of states of a node depends only on the states of its parent nodes: formally, a joint probability distribution $p(X)$ on a set of nodes X can be decomposed as $p(X) = \prod_i p(X^i | \text{Pa}(X^i))$, where $\text{Pa}(X^i)$ represents the parent set of X^i . In our networks, each node represents transcription expression of a gene. These conditional probabilities reflect not only relationships between genes, but also the stochastic nature of these relationships, as well as noise in the data used to reconstruct the network.

Bayes formula allows us to determine the likelihood of a network model M given observed data D as a function of our prior belief that the model is correct and the probability of the observed data given the model: $P(M | D \sim P(D | M) * P(M))$. The number of possible network structures grows super-exponentially with the number of nodes, so an exhaustive search of all possible structures to find the one best supported by the data is not feasible, even for a relatively small number of nodes. We employed Monte Carlo Markov Chain (MCMC) (Madigan and York, 1995) simulation to identify potentially thousands of different plausible networks, which are then combined to obtain a consensus network (see below). Each reconstruction begins with a null network. Small random changes are then made to the network by flipping, adding, or deleting individual edges, ultimately accepting those changes that lead to an overall improvement in the fit of the network to the data. We assess whether a change improves the network model using the Bayesian Information Criterion (BIC) (Schwarz, 1978), which avoids overfitting by imposing a cost on the addition of new parameters. This is equivalent to imposing a lower prior probability $P(M)$ on models with larger numbers of parameters.

Although edges in Bayesian networks are directed, we can’t infer causal relationships from the structure directly in general. For example, in a network with two nodes, X^1 and X^2 , the two models $X^1 \rightarrow X^2$ and $X^2 \rightarrow X^1$ have equal probability distributions as $p(X^1, X^2) = p(X^2 | X^1) p(X^1) = p(X^1 | X^2) p(X^2)$. Thus, by the data itself, we can’t infer whether X^1 is causal to X^2 , or vice versa. In a more general case, a network with three nodes, X^1 , X^2 , and X^3 , there are multiple groups of structures that are mathematically equivalent. For example, the following three different models, M1 : $X^1 \rightarrow X^2, X^2 \rightarrow X^3$, M2 : $X^2 \rightarrow X^1, X^2 \rightarrow X^3$, and M3 : $X^2 \rightarrow X^1, X^3 \rightarrow X^2$, are Markov equivalent (which means that they all encode for the same conditional independent relationships). In the above case, all three structures encode the same conditional independent relationship, $X^1 \perp\!\!\!\perp X^3 | X^2$, X^1 and X^3 are independent conditioning on X^2 , and they are mathematically equal

$$\begin{aligned} p(X) &= p(M1 | D) = p(X^2 | X^1) p(X^1) p(X^3 | X^2) \\ &= p(M2 | D) = p(X^1 | X^2) p(X^2) p(X^3 | X^2) . \\ &= p(M3 | D) = p(X^2 | X^3) p(X^3) p(X^1 | X^2) \end{aligned}$$

Thus, we cannot infer whether X^1 is causal to X^2 or vice versa from these types of structures. However, there is a class of structures, V-shape structure (e.g., Mv : $X^1 \rightarrow X^2, X^3 \rightarrow X^2$), which has no Markov equivalent structure. In this case, we can infer causal relationships. There are more parameters to estimate in the Mv model than M1, M2, or M3, which means a large penalty in BIC score for the Mv model. In practice, a large sample size is needed to differentiate the Mv model from the M1, M2, or M3 models.

Incorporating Genetic Data as a Structure Prior in the Bayesian Network Reconstruction Process

In general, Bayesian networks can only be solved to Markov equivalent structures, so that it is often not possible to determine the causal direction of a link between two nodes even though Bayesian networks are directed graphs. However, the Bayesian network reconstruction algorithm can take advantage of the experimental design by incorporating genetic data to break the symmetry among nodes in the network that lead to Markov equivalent structures, thereby providing a way to infer causal directions in the network in an unambiguous fashion (Zhu et al., 2004). We modified the reconstruction algorithm to incorporate eSNP data as priors (see schematic in Figure S3), in the following way: genes with *cis*-eSNP (Schadt et al., 2008) are allowed to be parent nodes of genes without *cis*-eSNPs, but genes without *cis*-eSNPs are not allowed to be parents of genes with *cis*-eSNPs, $p(\text{trans} \rightarrow \text{cis}) = 0$. We have shown that integrating genetic data such as *cis*-acting eSNPs or eQTLs (excluding edges into certain nodes) improves the quality of the network reconstruction by simulations (Zhu et al., 2007) and by experimental validations (Zhu et al., 2004, 2008). We note that in applying this particular version of the Bayesian network reconstruction algorithm (i.e., incorporating genetic information as a prior) and if genetic information is not available or is ignored, the population is simply treated as a population with random genetic perturbations.

Averaging Network Models

Searching optimal Bayesian network structures in a given data set is a NP-hard problem. We employed an MCMC method to do local search of optimal structures as described above. As the method is stochastic, the resulting structure will be different for each run. In our process, 1,000 Bayesian networks were reconstructed using different random seeds to start the stochastic reconstruction process. From the resulting set of 1,000 networks generated by this process, edges that appeared in greater than 30% of the networks were used to define a consensus network. A 30% cutoff threshold for edge inclusion was based on our simulation study (Zhu et al., 2007), where a 30% cutoff yields the best tradeoff between recall rate and precision. The consensus network resulting from the averaging process may not be a Bayesian network (a directed acyclic graph). To ensure the consensus network structure is a directed acyclic graph, edges in this consensus network were removed if and only if (1) the edge was involved in a loop, and (2) the edge was the most weakly supported of all edges making up the loop.

Bayesian Network for Individual Coexpression Module

The computational complexity of our MCMC method described above is expressed as $O(N^4)$, where the number of nodes included in the network reconstruction process is N . It is practically impossible to construct a global Bayesian network including all 39,000 genes from three different brain regions. Thus, we constructed a Bayesian network for each individual coexpression module. Following the procedure described above, 1,000 Bayesian networks were reconstructed using different random seeds to start the reconstruction process. From the resulting set of 1,000 networks generated by this process, edges that appeared in greater than 30% of the networks were used to define a consensus network. Our previous simulation study shows that the 30% inclusion threshold results in a stable structure and achieves the best tradeoff between precision and recall (Zhu et al., 2007).

Identification of Key Causal Regulators

For each Bayesian network of individual modules, we further identified the regulators by examining the number of N-hop downstream nodes (NHDN) for each gene in the directed network (Wang et al., 2012; Yang et al., 2010; Zhu et al., 2007). For a given network, let μ be the numbers of N-hop downstream nodes and d be the out degrees for all the genes. Genes with a number of NHDN greater than $\bar{\mu} + \sigma(\mu)$ were nominated as causal regulators. The regulators with degree above $\bar{d} + 2\sigma(d)$, where d denotes the number of downstream genes, become key causal regulators of a corresponding network module associated with LOAD differential connectivity. These criteria identified genes with number of downstream nodes and number of out links significantly above the corresponding average value.

Mouse Microglia Cultivation, Cell Transduction, and Flow Cytometry Analysis

Embryonic stem cell-derived microglia cells (ESdM) were generated from C57BL/6 ES cells (ATCC number SCRC-1002) and were shown to be very similar to primary cultured microglia both phenotypically and functionally as previously described (Beutner et al., 2010). ESdM were cultured under serum-free conditions in medium consisting of DMEM/F12 (GIBCO) supplemented with 1% N2 supplement (Invitrogen), 0.5 mM L-Glutamine (GIBCO), 15 $\mu\text{g}/\text{ml}$ D-glucose (Sigma), and 100 $\mu\text{g}/\text{ml}$ penicillin/streptomycin (GIBCO). After reaching a confluency of more than 80%, cells were split using cell scrapers (Sarstedt), suspended in medium, centrifuged, and plated in fresh medium.

Lentiviral vectors of the third generation (pLenti6/V5, Invitrogen) were used for transduction of microglia (ESdM). The CMV promoter in the pLenti6/V5 vector was replaced by the elongation factor-1 α (EF1 α) promoter. Full-length mouse *Tyrbp* (cds from nt 115 to 393, accession NM_011662) and truncated mouse *Tyrbp* (with the ITAM motifs removed, cds from nt 115 to 324, accession NM_011662) were tagged with the FLAG sequence at the 5', HA sequence at the 3' and the IgK leader sequence and cloned independently downstream of the EF1 α promoter. An IRES-enhanced green fluorescent protein (eGFP) gene was inserted directly downstream of the *Tyrbp* sequences. The correct nature of all cloned sequences was confirmed by automated sequencing (Seqlab, Germany). For overexpression studies, microglia (ESdM) were transduced using the corresponding IRES-eGFP constructs containing either full length mouse *Tyrbp* or truncated *Tyrbp* (ITAM motifs removed). For a control, microglia (ESdM) were transduced with the modified pLenti vector expressing only IRES-eGFP. Lentiviral transduction was performed using the ViraPower transduction kit (Invitrogen). Supernatant was removed 24 hr after infection and replaced with medium. Cells expressing eGFP were isolated by flow cytometry sorting (BD, DiVa cell sorter) to obtain a pure population. We confirmed successful overexpression of the transduced

genes through flow cytometry analysis. Further, we detected comparable protein expression levels in the flow cytometry between the full-length TYROBP and truncated TYROBP transduced microglial cell lines as shown in [Figure S6](#).

RNA Sequencing, Sample Preparation, and Data Analysis

To access the genome-wide gene expression changes in response to the perturbation of *Tyrobp*, we prepared twelve samples from mouse microglia cell lines, four from each of the three previously described groups: full-length *Tyrobp*-overexpressing cells, truncated *Tyrobp*-overexpressing cells, and GFP-expressing control cells. For each sample, about one microgram of total RNA was used for the preparation of the sequence library using RNA TruSeq Kit supplied by Illumina (Cat # 1004814). Briefly, rRNA was depleted from total RNA using ribozero kit (Invitrogen) to enrich polyadenylated coding RNA and noncoding RNA. The ribominus RNA was then fragmented in the presence of divalent cations at 94°C. The fragmented RNA was converted into double-stranded cDNA. After polishing the ends of the cDNA, adenine base was added at the 3' ends following which Illumina supplied specific adaptors were ligated. The adaptor ligated DNA was size selected to get an average of 200 bp insert size using AmpPure beads and amplified by 15 cycle PCR. The PCR DNA was then purified using AmpPure beads to get the final sequence library ready for sequencing. The insert size and DNA concentration of the sequence library was determined on Agilent Bioanalyzer. Each RNA sequence library was layered on one of the eight lanes of the Illumina flow cell at appropriate concentration and bridge amplified to get around 35–40 million raw reads. The DNA reads on the flow cell were then sequenced on HiSeq 2000 using 100 bp single end recipe. The rate of sequencing was around 1.2 hr per base sequenced. Read mapping was done using the TopHat ([Trapnell et al., 2009](#)) RNA-seq aligner. We allow up to two mismatches in each aligned segment and keep only reads mapping less than five times in the genome using the default quality cutoffs. The resulting alignments to genomic intervals and predicted splice sites are stored as a binary BAM file ([Li et al., 2009](#)).

Differential gene expression with a statistical significance $p < 0.05$ (12 tests), based on a t test, was used to identify the differentially expressed gene sets. This yielded 2,638 and 3,415 differentially expressed genes for the full-length *Tyrobp* and the truncated *Tyrobp* expressions, respectively. The mean FDR for the two signatures was 2.4% and 1.8%, respectively. To compute the FDR, we first constructed a pair of random case and control groups, each of which was comprised of two samples from a true case group and two from a true control group, then selected the most differentially expressed genes based on the same criteria. The FDR was the percentage of the original differentially expressed genes, which were found in the gene pool based on the permutation. Such a procedure was applied to all 36 possible permutations.

Acknowledgments for Additional Brain Samples

Additional tissues include samples from the following sites: The Kathleen Price Bryan Brain Bank, Duke University Medical Center (NIA grant #AG05128, NINDS grant # NS39764, NIMH MH60451 also funded by Glaxo Smith Kline); Christine Hulette, MD, Director, John F. Ervin University of Michigan (NIH grant P50-AG08671); Dr. Roger Albin, Lisa Bain, Eszter Gombosi, Sun Health Research Institute Brain Donation Program of Sun City, Arizona (NIA #P30 AG19610; Arizona Alzheimer's Disease Core Center, Arizona Department of Health Services, contract 211002, Arizona Alzheimer's Research Center; Arizona Biomedical Research Commission, contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium; Michael J. Fox Foundation for Parkinson's Research); Joseph Rogers, PhD, Thomas G. Beach, MD, PhD, Lucia I. Sue University of Miami/NPF Brain Endowment Bank; Deborah C. Mash, MD, Margaret J. Basile, Mitsuko Tanaka Oregon Health & Science University; Randy Wotljer, PhD Newcastle Brain Tissue Resource (funding via the Medical Research Council, local NHS trusts and Newcastle University); C.M. Morris, MD, Ian G McKeith, Robert H. Perry MRC London Brain Bank for Neurodegenerative Diseases (funding via the Medical Research Council); Simon Lovestone, MD PhD, Safa Al-Sarraj, MD, Claire Troakes, South West Dementia Brain Bank (funding via numerous sources including the Higher Education Funding Council for England (HEFCE), Alzheimer's Research Trust (ART), BRACE as well as North Bristol NHS Trust Research and Innovation Department and DeNDROn); Seth Love, MD, Patrick Kehoe, PhD, Laura Palmer, The Netherlands Brain Bank (funding via numerous sources including Stichting MS Research, Brain Net Europe, Hersenstichting Nederland Breinbrekend Werk, International Parkinson Fonds, Internationale Stichting Alzheimer Onderzoek); Inge Huitinga, MD, Marleen Rademaker, Michiel Kooreman, Institut de Neuropatologia, Servei Anatomia Patologica, Universitat de Barcelona; Isidre Ferrer Abizanda, MD, PhD, Susana Casas Boluda.

SUPPLEMENTAL REFERENCES

- Bennett, D.A., Schneider, J.A., Buchman, A.S., Mendes de Leon, C., Bienias, J.L., and Wilson, R.S. (2005). The Rush Memory and Aging Project: study design and baseline characteristics of the study cohort. *Neuroepidemiology* 25, 163–175.
- Bennett, D.A., Schneider, J.A., Aggarwal, N.T., Arvanitakis, Z., Shah, R.C., Kelly, J.F., Fox, J.H., Cochran, E.J., Arends, D., Treinkman, A.D., and Wilson, R.S. (2006a). Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology* 27, 169–176.
- Bennett, D.A., Schneider, J.A., Arvanitakis, Z., Kelly, J.F., Aggarwal, N.T., Shah, R.C., and Wilson, R.S. (2006b). Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* 66, 1837–1844.
- Beutner, C., Roy, K., Linnartz, B., Napoli, I., and Neumann, H. (2010). Generation of microglial cells from mouse embryonic stem cells. *Nat. Protoc.* 5, 1481–1494.
- Braak, H., and Braak, E. (1991). Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 82, 239–259.

- Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S., et al. (2008). Genetics of gene expression and its effect on disease. *Nature* 452, 423–428.
- He, Y.D., Dai, H., Schadt, E.E., Cavet, G., Edwards, S.W., Stepaniants, S.B., Duenwald, S., Kleinhanz, R., Jones, A.R., Shoemaker, D.D., and Stoughton, R.B. (2003). Microarray standard data set and figures of merit for comparing data processing methods and experiment designs. *Bioinformatics* 19, 956–965.
- Lai, E., Riley, J., Purvis, I., and Roses, A. (1998). A 4-Mb high-density single nucleotide polymorphism-based map around human APOE. *Genomics* 54, 31–38.
- Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 24, 719–720.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Madigan, D., and York, J. (1995). Bayesian graphical models for discrete data. *Int. Stat. Rev.* 63, 215–232.
- Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P., et al. (2007). A survey of genetic human cortical gene expression. *Nat. Genet.* 39, 1494–1499.
- Nochlin, D., van Belle, G., Bird, T.D., and Sumi, S.M. (1993). Comparison of the severity of neuropathologic changes in familial and sporadic Alzheimer's disease. *Alzheimer Dis. Assoc. Disord.* 7, 212–222.
- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555.
- Schadt, E.E., Molony, C., Chudin, E., Hao, K., Yang, X., Lum, P.Y., Kasarskis, A., Zhang, B., Wang, S., Suver, C., et al. (2008). Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 6, e107.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 461–464, 461–464.
- Storey, J.D., and Tibshirani, R. (2003). Statistical methods for identifying differentially expressed genes in DNA microarrays. *Methods Mol. Biol.* 224, 149–157.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–1111.
- Vonsattel, J.P., Myers, R.H., Stevens, T.J., Ferrante, R.J., Bird, E.D., and Richardson, E.P., Jr. (1985). Neuropathological classification of Huntington's disease. *J. Neuropathol. Exp. Neurol.* 44, 559–577.
- Wang, I.M., Zhang, B., Yang, X., Zhu, J., Stepaniants, S.B., Zhang, C., Meng, Q., Peters, M., He, Y., Ni, C., et al. (2012). Systems analysis of eleven rodent disease models reveals an inflammatoric signature and key drivers. *Mol. Syst. Biol.* 8, 594.
- Warner, J.P., Barron, L.H., and Brock, D.J. (1993). A new polymerase chain reaction (PCR) assay for the trinucleotide repeat that is unstable and expanded on Huntington's disease chromosomes. *Mol. Cell. Probes* 7, 235–239.
- Webster, J.A., Gibbs, J.R., Clarke, J., Ray, M., Zhang, W., Holmans, P., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., et al.; NACC-Neuropathology Group. (2009). Genetic control of human brain transcript expression in Alzheimer disease. *Am. J. Hum. Genet.* 84, 445–458.
- Yang, X., Zhang, B., Molony, C., Chudin, E., Hao, K., Zhu, J., Gaedigk, A., Suver, C., Zhong, H., Leeder, J.S., et al. (2010). Systematic genetic and genomic analysis of cytochrome P450 enzyme activities in human liver. *Genome Res.* 20, 1020–1036.
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, e17.
- Zhu, J., Lum, P.Y., Lamb, J., GuhaThakurta, D., Edwards, S.W., Thieringer, R., Berger, J.P., Wu, M.S., Thompson, J., Sachs, A.B., and Schadt, E.E. (2004). An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet. Genome Res.* 105, 363–374.
- Zhu, J., Wiener, M.C., Zhang, C., Fridman, A., Minch, E., Lum, P.Y., Sachs, J.R., and Schadt, E.E. (2007). Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput. Biol.* 3, e69.
- Zhu, J., Zhang, B., Smith, E.N., Drees, B., Brem, R.B., Kruglyak, L., Bumgarner, R.E., and Schadt, E.E. (2008). Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat. Genet.* 40, 854–861.

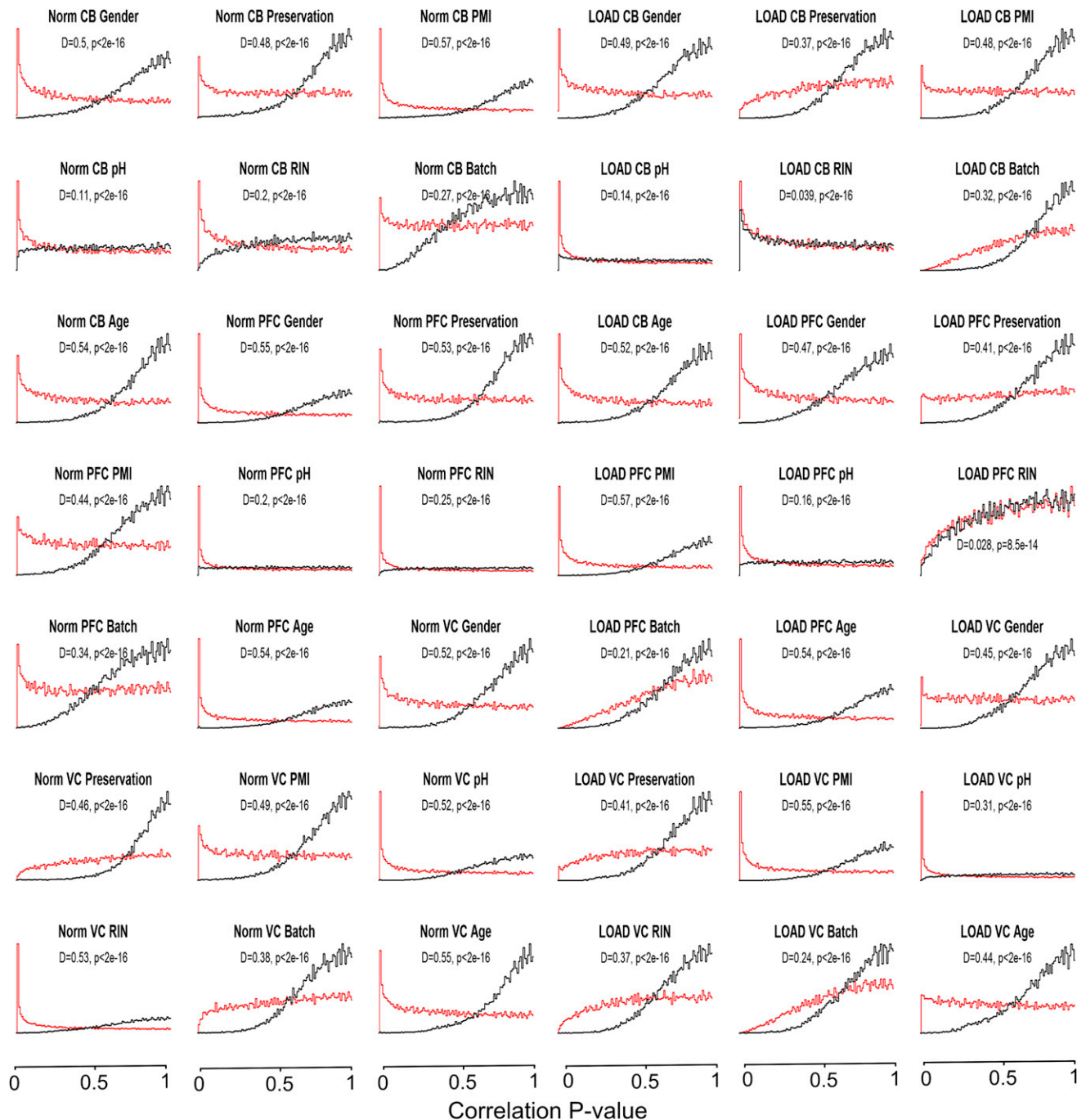


Figure S1. Correcting for Covariate Effects on the Expression Data in the Nondemented and LOAD Brains, Related to Figure 1

Each plot shows the distribution of the correlation p values (x axis) between a specific covariate and all the probes based on the expression data before (red curve, raw) and after the adjustment (black curve, corrected). The Kolmogorov-Smirnov (KS) test was used to capture the difference of the correlation p value distributions. The D statistics and p values of the KS test are shown here. y axis denotes the frequency, or the fraction of genes showing correlation to a given covariate.

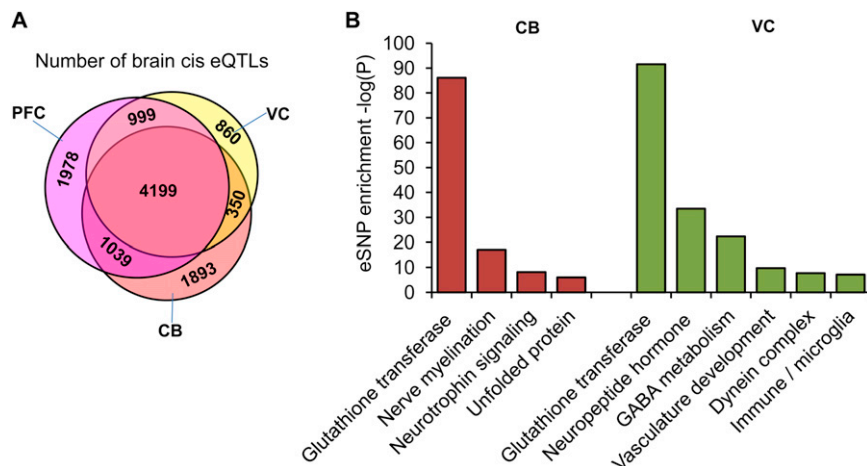


Figure S2. Brain *cis*-eSNPs and Their Enrichment in Modules, Related to Figure 4

(A) The number and overlap of all *cis*-eSNPs detected in the different brain regions in the present study at FDR of 10%. Table S1 lists all *cis*- and *trans*-eSNPs detected in the present study. Detailed statistical procedure related to the eSNP detection is provided in the [Extended Experimental Procedures](#).

(B) We tested the enrichment of brain eSNPs in the differentially connected modules of the multitissue coexpression network in LOAD as per brain region. Here we show the VC and CB modules with at least 100 gene members and showing significant enrichment of brain eSNPs (Figure 4C highlights eSNP enrichment in PFC).

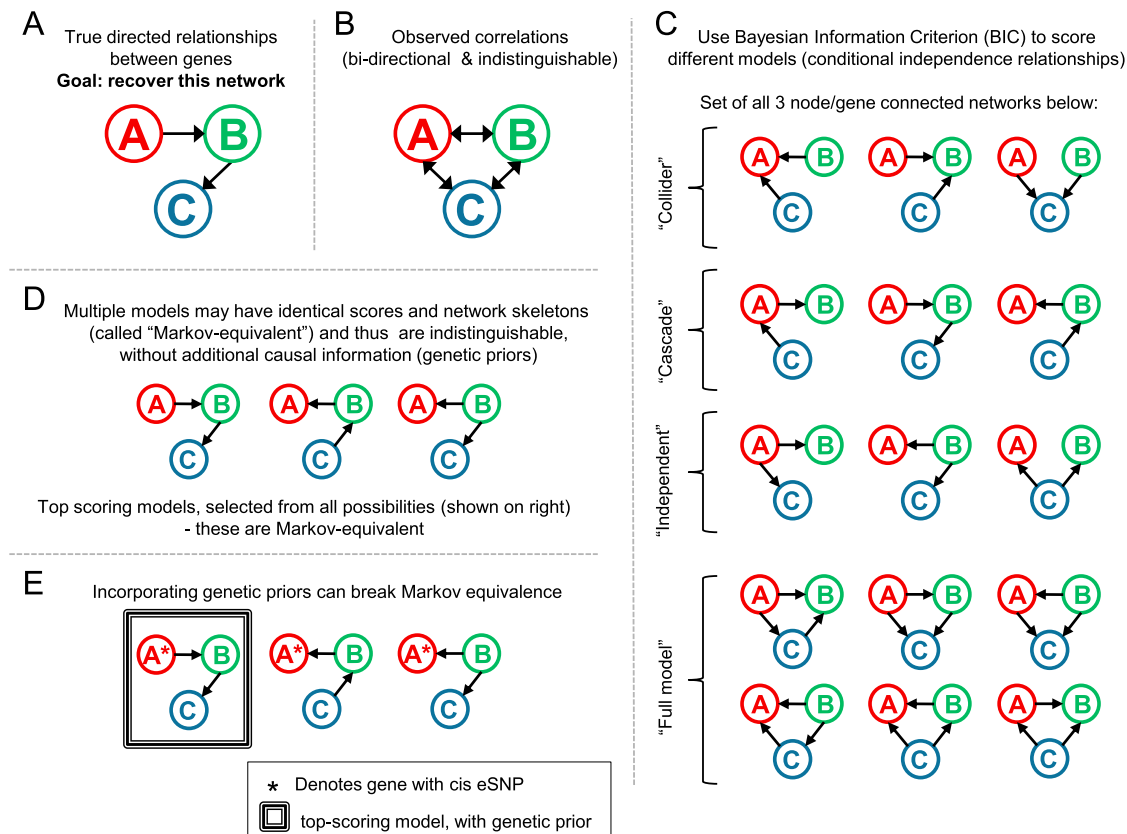


Figure S3. Schematic Drawing for Deriving Causal Gene-Gene Relationships from Gene-Expression and eSNP Data, Related to Experimental Procedures

Correlations between gene expression values alone do not indicate the true causal structure of gene networks. Our goal is to identify the true causal structure between transcripts (A) given observed correlation between gene expression (B), through a combination of model evaluation and incorporation of genetic priors as "causal anchors" (C–E). There are many possible causal relationships between even small groups of genes (C), and we use the Bayesian Information Criterion (BIC) to evaluate the posterior probability of each structure (shown in C), given the expression data. Several of these model structures may be statistically indistinguishable (D). Therefore, additional information is required to determine which of these models most accurately represents the data. We use a hard prior to put genes with strong *cis*-eSNPs as head nodes (causal anchors) to distinguish these models (E). This is justified because genetic perturbations (*cis*-eSNPs in the case of this data set) always precede changes in expression, and conversely expression changes do not lead to DNA alterations.

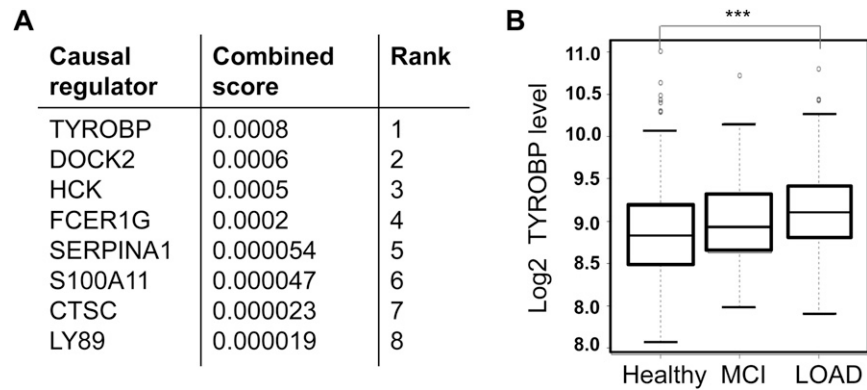


Figure S4. Ranking of Causal Regulators, Differential Expression, and Overexpression of *TYROBP*, Related to Table 1

(A) There were eight common causal regulators for all immune networks. We used a combined ranking score based on degree regulated nodes in the network related to each of the common causal regulators and the significance of the differential expression in LOAD brains, where *TYROBP* ranked the highest (see Results).

(B) *TYROBP* levels (median and interquartile ranges) in PFC were compared between 377 LOAD patients, 119 with the diagnosis of mild cognitive impairment (MCI), and 359 nondemented controls. The raw t test p value was 5.1×10^{-5} , whereas the Benjamini-Hochberg adjusted p value = 5×10^{-4} . See Extended Experimental Procedures for details on study groups.

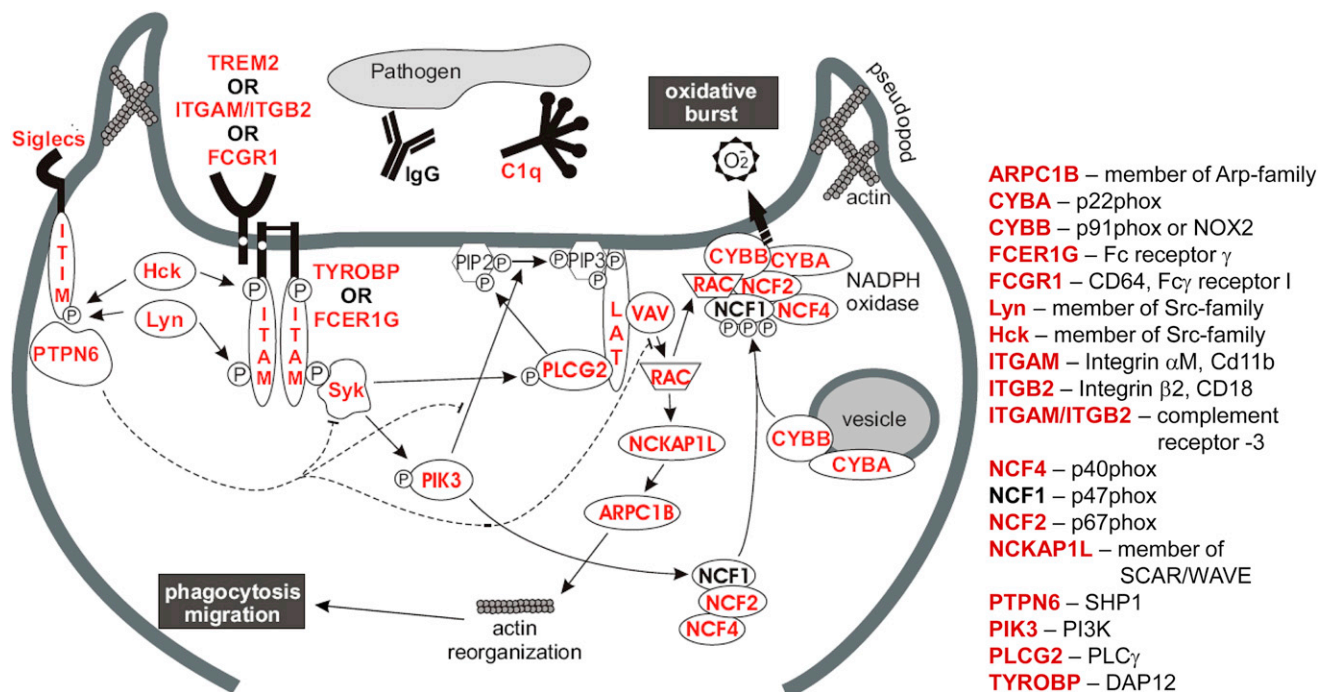


Figure S5. The Microglia Pathogen Phagocytosis Pathway, Related to Figure 5

Genes derived from the LOAD immune- and microglia-enriched module are marked in red. Pathogens are recognized by complement C1q or immunoglobulin (IgG) that bind to microglia complement receptors (e.g., ITGAM/ITGB2) or Fc-receptors (e.g., FCGR1) that signal via the immunoreceptor tyrosine-based activation motif (ITAM)-containing adaptor molecules TYROBP or FCER1G, respectively. Alternatively, pathogens are directly recognized by classical innate immune receptors (e.g., TREM2) that require the interaction with TYROBP for further signaling.

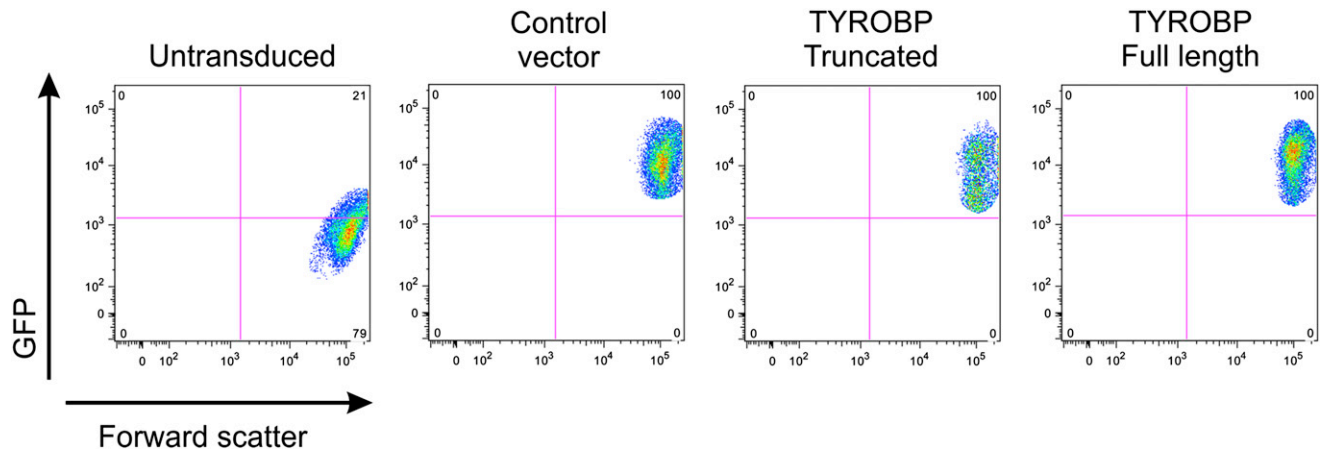


Figure S6. Overexpression of Intact or Truncated TYROBP in Mouse Microglia Cells, Related to Figure 6

Flow cytometry analysis of microglia transduced with full-length mouse TYROBP-IRES-eGFP (TYROBP full-length), truncated mouse TYROBP-IRES-eGFP (TYROBP truncated), control vector, or untransduced microglia. Transduced eGFP proteins were detected at similar levels in the microglial cells overexpressing either full-length TYROBP, truncated TYROBP, or the control vector. Untransduced microglia cells were used as control.