Partitioning the genetic architecture of amyotrophic lateral sclerosis

Iris J. Broce,[1*] Chun C. Fan,[2] Nicholas T. Olney,[3] Catherine Lomen-Hoerth,[3] Steve Finkbeiner,[3] Nazem Atassi,[4] Merit E. Cudkowicz,[4] Sabrina Paganoni,[5] Jennifer S. Yokoyama,[3] Aimee Kao,[3] William P. Dillon,[1] Christine M. Glastonbury,[1] Christopher P. Hess,[1] Wouter van Rheenen,[6] Jan H. Veldink,[6] Ammar Al-Chalabi,[7] Ole A. Andreassen,[8] Anders M. Dale,[2,9,10] William W. Seeley,[3] Leo P. Sugrue,[1] Aaron Ofori-Kuragu,[11] Celeste M. Karch,[11] Bruce L. Miller,[3*] and Rahul S. Desikan[1*]

[1]Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA.
[2]Department of Cognitive Sciences, University of California, San Diego, La Jolla, CA, USA.
[3]Department of Neurology, University of California, San Francisco, CA, USA.
[4]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA.
[5]Neurological Clinical Research Institute (NCRI) Massachusetts General Hospital (MGH) Boston, MA, USA.
[6]Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands.
[7]King's College London, Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience and Department of Neurology, King's College Hospital, London, UK.
[8]Norwegian Centre for Mental Disorders Research (NORMENT), Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
[9]Department of Radiology, University of California, San Diego, La Jolla, CA, USA
[10]Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA
[11]Department of Psychiatry, Washington University in St Louis, St Louis, MO, USA.

***** rahul.desikan@ucsf.edu (RSD), iris.broce@ucsf.edu (IJB), bruce.Miller@ucsf.edu (BLM)

Abstract

The genetic basis of sporadic amyotrophic lateral sclerosis (ALS) is not well understood. Using large genome-wide association studies and validated tools to quantify genetic overlap, we systematically identified single nucleotide polymorphisms (SNPs) associated with ALS conditional on genetic data from 65 different traits and diseases from >3 million people. We found strong genetic enrichment between ALS and a number of disparate traits including frontotemporal dementia, coronary artery disease, C-reactive protein, celiac disease and memory function. Beyond *C9ORF72,* we detected novel genetic signal within numerous loci including *GIPC1*, *ELMO1* and *COL16A* and confirmed previously reported variants, such as *ATXN2*, *KIF5A*, *UNC13A* and *MOBP.* We found that ALS variants form a small-world co-expression network characterized by highly inter-connected 'hub' genes. This network clustered into smaller sub-networks, each associated with a unique function. Altered gene expression of several sub-networks and hubs was over-represented in neuropathological samples from ALS patients and SOD1 G93A mice. Our collective findings indicate that the genetic architecture of ALS can be partitioned into distinct components where some genes are highly important for developing disease. These findings have implications for stratification and enrichment strategies for ALS clinical trials.

Introduction

Sporadic amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by progressive muscle paralysis from selective loss of upper and lower motor neurons. Spreading rapidly, ALS can lead to respiratory failure and death in 3-5 years[1]. Given the paucity of disease modifying treatments, elucidating the genetic basis of ALS can delineate putative pharmacological targets and highlight molecular mechanisms underlying disease. Importantly, refining the genetic landscape of ALS can inform cohort stratification and enrichment strategies for clinical trials.

ALS is increasingly recognized as a complex disorder with an incompletely understood genetic architecture. Prior work suggests that sporadic ALS is genetically characterized by a few rare variants, each explaining a substantial portion of the inherited risk (oligogenic)[2-4]. However, more recent evidence indicates that low-risk, common variants underlie ALS (polygenic) [5]. Importantly, several ALS-associated variants have been implicated in other diseases suggesting genetic pleiotropy[6-8]. Furthermore, it is not known whether certain ALS genes are more important than others for influencing disease etiology.

Here, our goal was to elucidate the genetic architecture of ALS by leveraging statistical power from large GWAS from 65 distinct traits and diseases. Using these methods, we have discovered novel genetic risk loci and shown abundant genetic pleiotropy between several neurodegenerative diseases including ALS, FTD, progressive supranuclear palsy (PSP), corticobasal degeneration (CBD), Parkinson's disease (PD), and Alzheimer's disease (AD) [8-12].

Results

Selective shared genetic risk between ALS and 65 distinct traits

Using previously published stratified FDR methods (see Methods), we assessed genetic overlap between ALS and 65 distinct traits and diseases. We identified genetic enrichment in ALS SNPs across different levels of significance with 65 distinct traits and diseases (Fig. 1). Consistent with prior work[8], we found that the highest level of pleiotropic enrichment was between ALS and FTD (700-fold enrichment). Surprisingly, we also found robust genetic enrichment in ALS SNPs as a function of coronary artery disease (CAD; 300-fold enrichment), memory (225-fold enrichment), C-reactive protein (CRP; 50-fold enrichment), and PSP (50-fold enrichment). We found weaker genetic enrichment with celiac disease (CeD), CBD, body mass index (BMI), rheumatoid arthritis (RA), schizophrenia (SCZ), verbal numeric reasoning (VNR), and putamen volume (PUT). We found no enrichment between ALS and the other phenotypes. We note that these analyses reflect genetic enrichment after removing all SNPs within chromosome 9.

To identify novel ALS risk loci, we used a stratified approach. First, we computed conditional FDR, a statistical framework that is well suited for gene detection[9,13]. Conditional FDR analysis at a FDR p-value < 0.05 revealed 180 SNPs across 21 chromosomes (Fig. 2, Supplementary Table 1). Next, we performed extensive LD analyses to identify the variants underlying the genetic signals (see Supplemental Information). After accounting for LD, we identified 89 risk loci and annotated each ALS risk SNP with the closest gene(s), resulting in a total of 92 closest genes (Fig. 2, Table 1). Of these, 30 SNPs were either previously reported or were in LD with a previously reported SNP (Table 1, Supplementary Fig. 1). An additional 59 SNPs were novel or were not in LD with SNPs within previously reported loci (Table 1,

Supplementary Fig. 1). We found independent hits from SNPs previously reported within *CRIM1, NAF1, TNIPI, PARKIN, ELP3, MRSA, C9ORF72, PCDH9, A2BP1*, and *CPNE7* (Table 1).

To determine whether the ALS risk genes were associated with a single trait or multiple traits, we plotted the minimum conditional FDR associated with all traits and closest genes. As shown in Supplementary Fig. 2, across all 65 traits and diseases, we found that the ALS risk variants are associated with multiple traits and diseases. Genetic variants within *C9ORF72* were identified on 46 traits. Genetic variants within *KIAA0524* (also known as *SARM1*) were identified on 64 traits, *UNC13A* on 62 traits, *MOB3B* on 58 traits, *TBK1* on 34 traits, and *CAT* and *C21ORF2* on 27 traits.

These findings suggest that ALS has a polygenic component where several genes potentially contribute to disease risk. Genetic pleiotropy with traits like FTD and CAD can be leveraged for ALS gene detection. Importantly, there are several ALS susceptibility loci that are also associated with numerous other traits and diseases.

cis-eQTL expression

To determine the functional effects of the ALS pleiotropic risk SNPs, we evaluated *cis*-expression quantitative loci (*cis*-eQTL) in human brains free of neuropathology (Supplementary Table 2). In total, the ALS risk SNPs produced significant *cis*-eQTLs (below 1.5×10^{-3}) within 41 genes. Of these, SNPs within *SMARCA2, GGNBP2*, *NUP50*, and *TNIP1* showed overlapping annotation between the eQTL and the closest genes. Thirteen SNPs showed significant *cis*-eQTLs with multiple genes.

Biological networks associated with ALS genetic risk genes

Using GeneMANIA (www.genemania.org), an online web-portal for bioinformatic assessment of gene networks[14], we conducted a network analysis to explore the interaction and co-expression patterns associated with the ALS risk genes defined as the combination of the 1) closest genes to the SNP and 2) functional genes (i.e., SNPs with significant *cis*-eQTLs). We found that a large number of these genes showed physical protein-protein interactions (42.93%), were co-expressed (29.33%), and showed genetic interactions (13.28%) (Fig. 3). Few ALS genetic risk genes shared pathways (9.36%), were co-localized (2.60%), or predicted functional interactions between genes based on orthology (2.50%) (Supplementary Table 3).

Properties of the ALS biological networks

We assessed the network structure of the physical protein-protein interaction network, co-expression network, and genetic interactions network. Specifically, we asked whether some genes play a more influential role than others. Most complex networks have a small-world property characterized by relatively short paths between any pair of nodes (genes) [15]. In small-world networks, perturbing any given node is thought to also perturb neighboring nodes and the entire network in general. Quantitatively, a network is considered small-world if its "small-worldness" index is higher than one (a stricter rule is small-worldness >=3)[16]. Further, the clustering coefficient for the target small-world network should be higher than the clustering coefficient of a comparable random network. Also, the average shortest path length of the target network should be similar or higher (but not substantially higher) than a comparable random network.

First, we evaluated the degree to which each network assumed a small-work network structure. The co-expression interaction network consisted of 95 nodes and 132 edges, had a small-world index of 6.06, a diameter of 13, and average shortest path length of 5.12. The clustering coefficient was 0.102, which is higher than the clustering coefficient of a random graph with the same number of indices (0.031). The physical protein-protein interaction network consisted of 85 nodes and 41 edges, had a small-world index of 4.43, a diameter of 5, and average shortest path length of 82.03. The clustering coefficient was 0.326, which is also higher than the clustering coefficient of a random network with the same number of indices (0.06). Lastly, the genetic interaction network consists of 98 nodes and 472 edges, had a low small-world network index of 1.11, a diameter of 5, and average shortest path length of 2.33. The clustering coefficient for this network (0.197) was similar to the clustering coefficient of a random network with the same number of indices (0.192). Of the three networks, the co-expression network showed robust small-world network properties - the physical protein-protein interactions network had a substantially large shortest path length and the genetic-interactions network clustering coefficient did not differ from random. Therefore, in subsequent network analysis we focused on the co-expression network.

To further assess the structure of the co-expression network, we evaluated various network centrality measures, including degree centrality, eigenvector centrality, and edge-betweenness centrality. Centrality network measures define how important each node is within a given network. The degree centrality is the number of edges connected to a node. Eigenvector centrality is the extent to which a node is connected to other highly influential “hub” nodes. Edge-betweenness centrality is the extent to which a node lies on the shortest path between other nodes (see Methods). Fig. 4a shows the co-expression network. The size of each node is defined

by its eigenvector centrality (EC) value. Genes with a high EC and high degree centrality can be characterized as hubs[15,16]. Within the co-expression network, genes with high degree centrality (Supplementary Fig. 3) and high EC (Supplementary Fig. 4) were *GIPC1* (n=8, EC=1.00), *ELMO1* (n=6, EC = 0.994), *UGCG* (n=7, EC = 0.978), *SMARCA2* (n=6, EC = 0.924), *ATXN2* (n=6, EC = 0.765), and *SETD2* (n=6, EC = 0.729). In addition to these, *DOCK2* (n= 5, EC = 0.809), *KIAA0917*/*SCFD1* (n=5, EC = 0.752), *HTR2A* (n=4, EC = 0.695), and *COL16A1* (n=5, EC = 0.645) were also highly influential based on EC values.

Next, we explored whether the co-expression network could be partitioned into potentially meaningful subnetworks (Fig. 4a,b). The network was partitioned into subnetworks by removing edges with high edge-betweenness centrality (see Methods). The result was a hierarchical map, called a dendrogram (Fig. 4b). We identified a total of 8 clusters. Five clusters contained the most influential hub genes. The yellow cluster contained 5 hub genes (*HTR2A*, *GIPC1*, *SMARCA2*, *ELMO1*, and *DOCK2*). The purple cluster contained two hub genes (*UGCG* and *SETD2*), and the teal, orange, and red clusters contained a single hub gene (*COL16A1*, *ATXN2*, *KIAA0917*, respectively).

Collectively, these findings suggest that ALS genes that are co-expressed form networks with small-world properties and can be further partitioned into different clusters. Importantly, specific genes, such as *GIPC1* and *ATXN2*, act like hubs, characterized by a high degree of connectivity with other genes; perturbing these hub genes may disrupt the entire network.

Biological pathways associated with ALS risk genes within co-expression network

We used bioinformatics approaches to identify distinct and common biological pathways associated with the 8 subnetworks within the ALS co-expression network (see Methods). The top

5 functional annotations for each subnetwork are shown in Fig. 5a. The blue subnetwork was enriched for developmental processes, including neuron projection development and regulation of cell growth. The gray subnetwork was enriched for apoptotic signaling pathway, cell death, and protein phosphorylation. The green subnetwork was enriched for cell-cell signaling, response to external stimuli, and metabolic processes. The purple cluster was enriched for vascular morphology and angiogenesis. The orange, red, teal, and yellow subnetworks were enriched for vesicle-mediated processes. Lastly, the orange and red networks were also enriched for organelle membrane fusion and disassembly processes, and the teal subnetwork was also enriched for oxidative reduction and androgen metabolic processes.

Further, we explored whether genes within the distinct subnetworks were selectively enriched for neuron-specific biological, chemical, or molecular processes. As shown in Fig. 5b, the blue, green, teal, and yellow subnetworks were selectively enriched for neuronal processes, particularly involved in neuron differentiation, projection, guidance, and development. We also found that several hub genes within the yellow subnetwork (*GIPC1*, *SMARC2*, *HTR2A*) were over-represented in these processes (Supplementary Fig. 5). Taken together, these findings highlight that the ALS co-expression subnetworks are involved in distinct and overlapping biological pathways.

Differential expression of ALS risk genes in tissue from ALS patients and SOD1 G93A transgenic mice

We investigated whether particular clusters or hub genes were enriched in pathological samples from ALS patients. To do this, we assessed the differential expression of genes within the ALS co-expression network in the gray matter of motor neurons isolated from spinal cords of

patients with familial and sporadic ALS and controls (Fig. 6a, see Methods). We found that 2 genes within the teal subnetwork (*COL16A1* and *GPX3)* and one gene within the purple subnetwork *(UGCG)* were differentially expressed in tissue from controls compared to familial ALS patients. *COL16A1* was also differentially expressed in tissue from controls compared to sporadic ALS patients. Importantly, *COL16A1* and *UGCG* are hub genes.

To validate the genes identified in ALS human tissue, we evaluated expression data from a well-characterized mouse model[17]. The RNA expression data were analyzed in non-transgenic SOD1 WT and SOD1 G93A mice at 75 and 110 days (GEO accession number GSE4390). Differential expression of *UGCG* was independently replicated in the SOD1 G93A mice (Fig. 6b). Additionally, we found that *TBK1* within the purple cluster, *NDSTI* and *WIPI2* within the orange cluster, and *KIF5A* within the green cluster were enriched in SOD1 G93A mice (Fig. 6b).

Discussion

We sought to elucidate the genetic basis of sporadic ALS. By exploiting statistical power from several large GWAS of >3 million people, we identified novel susceptibility loci, each associated with a small increase in ALS risk. We found that ALS variants form a small-world co-expression network characterized by highly inter-connected 'hub' genes. This network clustered into smaller sub-networks, each with a unique function. Altered gene expression of several sub-networks and hubs was over-represented in neuropathological samples from ALS patients and SOD1 G93A mice. Our collective findings indicate that the genetic architecture of ALS can be partitioned into distinct components where some genes are highly important for developing disease.

We found that ALS has a robust polygenic component. By leveraging genetic studies from 65 different traits and diseases, we identified 89 ALS risk loci across 21 chromosomes of

which 59 are novel. Beyond *C9ORF72,* our pleiotropy analyses detected novel genetic signal within numerous loci including *GIPC1, ELMO1* and *COL16A* and confirmed previously reported variants, such as *ATXN2, KIF5A, UNC13A* and *MOBP* [4-6]. Neither as polygenic as schizophrenia[18] or Alzheimer's disease[19] nor purely oligogenic[5], it is likely that the genetic architecture of ALS is a continuum of common low-risk variants and rare high-risk variants. Although each of the ALS susceptibility loci we detected was associated with a small effect, when aggregated together into a polygenic score they may explain a substantial portion of the inherited risk underlying ALS[20].

We found genetic pleiotropy between ALS and a number of diseases and traits. In line with previous reports, we show strong genetic enrichment in ALS conditional on FTD and PSP[8,21,22]. Building on prior work showing a relationship between inflammation/immune dysfunction and motor neuron disease[23-25], we found enrichment in ALS SNPs as a function of CRP and celiac disease. Surprisingly, we also identified strong genetic overlap between ALS and CAD, and memory function. Clinically, these findings suggest that a subset of ALS patients are at elevated genetic risk for FTD whereas another (potentially overlapping) group of ALS individuals may be at high risk for CAD or immune dysfunction. Therefore, development of multiple pathway specific polygenic scores may identify individuals at risk for developing ALS who are 'enriched' for FTD, cardiovascular or immune mediated processes.

Our findings inform cohort stratification and enrichment strategies for ALS clinical trials. We found that the ALS pleiotropic and functional risk genes form a small-world co-expression network. This network can be partitioned into 8 subnetworks, each enriched for distinct biological pathways. Similar to previous research, we found functional enrichment of ALS risk genes for oxidative-mediated (teal subnetwork), neuronal (teal, blue, yellow, and green

subnetworks), and endoplasmic reticulum processes (orange subnetwork)[26,27]. Additionally, genes within the blue subnetwork were enriched for developmental and growth pathways while genes within the gray subnetwork were enriched for cell-death and apoptotic processes. Altered gene expression within the teal subnetwork was over-represented in postmortem spinal cord samples from familial and sporadic ALS patients, whereas abnormal expression of genes in the orange and purple clusters was present in SOD1 G93A mice. Clinically, these results suggest that partitioning genetic susceptibility may help identify individuals who have a higher likelihood of responding to therapies with a specific mechanism of action and support including DNA collection and sequencing in ALS clinical trials. For example, ALS patients who are enriched for genetic abnormalities within the teal subnetwork may respond to therapies targeting oxidation reduction or steroid catabolism (Fig. 5). On the other hand, vascular treatments may be effective in ALS individuals with an overabundance of altered purple cluster genes (Fig. 5).

Not all ALS genes are created equal. We show that particular genes within the ALS co-expression network are characteristic of hubs. Playing a central role within a biological system, perturbation of a hub can cause rapid degeneration of the whole network[28]. We found that these hub genes were key drivers of biological enrichment. Specifically, *HTR2A*, *SMARCA2*, and *GIPC1* were enriched for neuronal processes (Supplementary Fig. 5), and *ELMO1* and *DOCK2* were enriched for vesicle-mediated transport (Supplementary Fig. 6). Therapeutically targeting hub genes may be most effective for altering an entire biological pathway. However, given abundant pleiotropy with other traits (Supplementary Fig. 2), comprehensive biological and experimental evaluation of the entire network of a hub gene will be necessary prior to therapeutic evaluation.

This study should be interpreted within the context of its limitations. First, the ALS GWAS used contained people predominantly of European descent, while the other GWAS included people of both European and non-European descent. Therefore, these results may not be generalizable to ALS patients from other populations. Second, like most GWAS, a major limitation of our study is that we could not determine with certainty the causal genes underlying our genetic signal. Although we performed extensive LD and *cis*-eQTL analyses and included the combination of closest and functional genes in our network analyses, it is likely that genetic fine mapping and experimental approaches, such as CRISPR/Cas9 gene editing, will be needed to isolate the causal variants. Finally, given evidence that a substantial proportion of coronary disease is associated with inflammation[29], future work should evaluate whether CAD influences ALS risk through inflammation or other mediator variables.

In summary, we show that the genetic architecture of ALS has a robust polygenic component that can be partitioned into distinct subnetworks, each enriched for divergent biological pathways. We also identify several hub genes that may be key drivers of ALS pathobiology. Our findings are compatible with the hypothesis that ALS is a multi-step, non-uniform disease process. These results have implications for cohort stratification and enrichment strategies for ALS clinical trials.

Methods

Participant samples

We conducted a meta-analysis of summary data obtained from published data. We evaluated complete GWAS results in the form of summary statistics (p-values and odds ratios) for ALS and 65 distinct traits and diseases (see Supplementary Table 4). We obtained ALS

GWAS summary statistic data from 12,577 ALS cases and 23,475 controls at 18,741,501 SNPs (see Supplementary Table 4 for additional details). The ALS GWAS summary statistics and sequenced variants are publicly available through the Project MinE data browser: http://databrowser.projectmine.com. We also obtained GWAS summary statistic data for the 65 distinct traits and diseases (for additional details, please see Supplementary Table 4). The relevant institutional review boards or ethics committees approved the research protocol of the individual GWASs used in the current analysis, and all participants gave written informed consent.

Genetic Enrichment Statistical Analyses

The pleiotropic enrichment strategies implemented here were derived from previously published stratified FDR methods[13,30]. For given phenotypes A and B, pleiotropic 'enrichment' of phenotype A with phenotype B exists if the proportion of SNPs or genes associated with phenotype A increases as a function of increased association with phenotype B. To assess for enrichment, we constructed fold-enrichment plots of nominal $-\log_{10}(p)$ values for all ALS SNPs and for subsets of SNPs determined by the significance of their association with the 65 distinct traits and diseases. In fold-enrichment plots, the presence of enrichment is reflected as an upward deflection of the curve for phenotype A with increasing strength of association with phenotype B. To assess for polygenic effects below the standard GWAS significance threshold, we focused the fold-enrichment plots on SNPs with nominal $-\log_{10}(p) < 7.3$ (corresponding to $p > 5\times10^{-8}$). The enrichment seen can be directly interpreted in terms of true discovery rate (TDR = 1 – False Discovery Rate (FDR)). Given prior evidence that several genetic variants within chromosome 9 are associated with increased ALS risk, one concern is that random pruning may not sufficiently

account for these large LD blocks, resulting in artificially inflated genetic enrichment[12]. To better account for these large LD blocks, in our genetic enrichment analyses, we removed all SNPs within chromosome 9.

To identify novel ALS risk loci as a function of genetic variants associated with the 65 traits and diseases, we computed conditional FDRs[13,30], a statistical framework that is well suited for gene detection. The standard FDR framework is based on Bayesian statistics and follows the assumption that SNPs are either associated with the phenotype (non-null) or are not associated with the phenotype (null SNPs). Within a Bayesian statistical framework, the FDR is then the posterior probability of the SNP being null given its p-value is as small as or smaller than the observed one. The conditional FDR is an extension of the standard FDR, which incorporates information from GWAS summary statistics of a second phenotype to adjust its significance level. The conditional FDR is defined as the probability that a SNP is null in the first phenotype given that the p-values in the first and second phenotypes are as small as or smaller than the observed ones. Ranking SNPs by the standard FDR or by p-values gives the same ordering of SNPs. In contrast, if the primary and secondary phenotypes are related genetically, the conditional FDR reorders SNPs and results in a different ranking than that based on p-values alone. We used an overall FDR threshold of $p < .05$ to indicate statistical significance, meaning 5 expected false discoveries per 100 reported. In addition, we constructed Manhattan plots based on the ranking of the conditional FDR to illustrate the genomic location. In all analyses, we controlled for the effects of genomic inflation. Detailed information on the conditional FDR can be found in prior reports[13,30].

Functional evaluation of shared risk loci

To assess whether the SNPs associated with ALS and the 65 traits and diseases modify gene expression, we identified *cis*-expression quantitative loci (eQTLs, defined as variants within 1 Mb of a gene's transcription start site) associated with the identified ALS pleiotropic SNPs and measured their regional brain expression in a publicly available dataset of normal control brains (UK Brain Expression Consortium, http://braineac.org/)[31]. To minimize multiple comparisons, we analyzed *cis*-eQTL for the mean p-value obtained from the following brain regions: the cerebellum, frontal cortex, hippocampus, medulla, occipital cortex, putamen, substantia nigra, temporal cortex, thalamus, and white matter. To minimize false positives, we applied a Bonferroni-corrected p-value of 1.5×10^{-3}.

Biological networks associated with ALS genetic risk genes

To evaluate potential protein and genetic interactions, co-expression, co-localization, and protein domain similarity for the combined pleiotropic (i.e. closest genes from the pleiotropy analyses) and functionally expressed ALS genes (i.e., with significant *cis*-eQTLs), we used GeneMANIA (www.genemania.org), an online web-portal for bioinformatic assessment of gene networks[14]. To visualize the composite gene network, we also assessed the weights of individual components within the network[32]. Further, we evaluated whether the biological networks fell into the class of a small-world network using several diagnostic criteria. First, we computed the "small-worldness" index, using the R package 'qgraph'. The function computes the global transitivity of the target network and its average shortest path length[16,33]. It then computes the same indices on 1000 random networks. The small-worldness index is then equal to the transitivity of the target network (normalized by the random transitivity) over the average

shortest path of the target network (normalized by the random average shortest path length). A network was considered small-world if the "small-worldness" index was >= 3; [16]. In addition to the small-worldness index, we inspected whether the network had a transitivity substantially higher than comparable random networks and that its average shortest path length was similar or higher (but not substantially higher) than that computed on random networks.

Further, to determine whether some genes play a more influential role than others, we evaluated various network centrality measures, including degree centrality, eigenvector centrality, and edge-betweenness centrality. We used the R package 'igraph' for all network centrality analysis and visualization[34]. The degree of a node corresponds to the sum of its adjacent edges (i.e., connections). The eigenvector centrality of a node corresponds to the values of the first eigenvector of the graph adjacency matrix. In general, nodes with high eigenvector centralities are also connected to many other nodes which are, in turn, connected to many other nodes. Consequently, eigenvector centrality corresponds to the degree to which a node is connected to other highly influential nodes. Lastly, we partitioned the co-expression network, in particular, into subnetworks based on edge-betweenness centrality. Edge-betweenness centrality is defined by the number of shortest-paths going through an edge. Here, a subnetwork is analogous to modules within a network. Nodes within a module are densely connected to themselves (e.g., cluster) but sparsely connected to other modules. To create modules, we gradually remove the edge with the highest edge-betweenness score, since all the shortest paths from one module to another typically pass through them. The result is a hierarchical map, called a dendrogram. The leaves of the tree are the individual nodes and the root of the tree represents the whole graph.

Lastly, to evaluate biological pathways of the ALS pleiotropic genes (i.e. closest genes from the pleiotropy analyses) and functionally expressed ALS genes (i.e., with significant *cis*-eQTLs), we used FUMA (http://fuma.ctglab.nl/), a web-based platform that integrates information from multiple biological resources to facilitate functional annotation of GWAS results[35].

Gene expression alterations in tissue from ALS patients and SOD1 G93A transgenic mice

To determine whether the ALS genetic risk genes were differentially expressed in tissue from patients with ALS, we analyzed the gene expression of the target genes from postmortem spinal cord gray matter from 11 individuals (2 patients with familial ALS, 5 patients with sporadic ALS, and 4 controls; Gene Expression Omnibus [GEO] accession number GDS412[36]). Details about this dataset and analysis–including the human brain samples used, RNA extraction and hybridization methods, microarray quality control, and microarray data analysis– are described in the original manuscript[36]. To validate the genes identified in ALS human tissue, we also analyzed RNA expression data in non-transgenic SOD1 WT (n=2) and SOD1 G93A (n=2) mice at 75 and 110 days (GEO accession number GSE4390). SOD1 G93A mice are pre-symptomatic at 75 days and exhibit hindlimb paralysis at 110 days[17,37].

Code availability

Code and scripts available by request from authors.

Data availability

Summary statistics from secondary GWAS of single disorders and traits are available upon request from the corresponding author. Cis-eqtl data from the UK Brain Expression Consortium are publicly available (http://braineac.org/). Findings from biological networks were obtained using GeneMANIA (www.genemania.org), an online web-portal for bioinformatic assessment of gene networks. Biological pathways were evaluated using FUMA (http://fuma.ctglab.nl/), a web-based platform that integrates information from multiple biological resources to facilitate functional annotation of GWAS results. Expression data from sporadic and familial ALS patients and controls postmortem spinal cord gray matter are available in GEO with the accession number GDS412. RNA expression data in non-transgenic SOD1 WT and SOD1 G93A mice at 75 and 110 days are also available in GEO with the accession number GSE4390.

ACKNOWLEDGEMENTS

We salute the millions of people who battle with ALS every single day - your courage is our strength. This research is in part an EU Joint Programme - Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organizations under the aegis of NL (ZONMW) and JPND - www.jpnd.eu (United Kingdom, Medical Research Council (MR/L501529/1 (STRENGTH); MR/R024804/1 (BRAIN-MEND)) and Economic and Social Research Council ((ES/L008238/1) ALS-CarE) and through the Motor Neurone Disease Association. This study represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement n° 772376 – EScORIAL).

References

1. Brown RH, Al-Chalabi A. Amyotrophic lateral sclerosis. *N Engl J Med* **377,** 162-72 (2017).

2. van Blitterswijk M, van Es MA, Hennekam EA, Dooijes D, van Rheenen W, Medic J, et al. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet*, Sep 1;21(17):3776-84.

3. Bury JJ, Highley JR, Cooper-Knock J, Goodall EF, Higginbottom A, McDermott CJ, et al. Oligogenic inheritance of optineurin (OPTN) and C9ORF72 mutations in ALS highlights localisation of OPTN in the tdp-43-negative inclusions of C9ORF72-ALS. *Neuropathology* 2016, Apr;36(2):125-34.

4. Al-Chalabi A, van den Berg LH, Veldink J. Gene discovery in amyotrophic lateral sclerosis: Implications for clinical management. *Nat Rev Neurol* 2017, Feb;13(2):96-104.

5. van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet* 2016;48(9):1043-8.

6. Elden AC, Kim HJ, Hart MP, Chen-Plotkin AS, Johnson BS, Fang X, et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 2010, Aug 26;466(7310):1069-75.

7. McLaughlin RL, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. *Nat Commun* 2017, Mar 21;8:14774.

8. Karch CM, Wen N, Fan CC, Yokoyama JS, Kouri N, Ross OA, et al. Selective genetic overlap between amyotrophic lateral sclerosis and diseases of the frontotemporal dementia spectrum. *JAMA Neurol* 2018, Apr 9.

9. Desikan RS, Schork AJ, Wang Y, Thompson WK, Dehghan A, Ridker PM, et al. Polygenic overlap between c-reactive protein, plasma lipids, and alzheimer disease. *Circulation* 2015, Jun 9;131(23):2061-9.

10. Yokoyama JS, Karch CM, Fan CC, Bonham LW, Kouri N, Ross OA, et al. Shared genetic risk between corticobasal degeneration, progressive supranuclear palsy, and frontotemporal dementia. *Acta Neuropathol* 2017, May;133(5):825-37.

11. Ferrari R, Wang Y, Vandrovcova J, Guelfi S, Witeolar A, Karch CM, et al. Genetic architecture of sporadic frontotemporal dementia and overlap with alzheimer's and parkinson's diseases. *J Neurol Neurosurg Psychiatry* 2017;88(2):152-64.

12. Broce I, Karch CM, Wen N, Fan CC, Wang Y, Tan CH, et al. Correction: Immune-related genetic enrichment in frontotemporal dementia: An analysis of genome-wide association studies. *PLoS Med* 2018, Jan;15(1):e1002504.

13. Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS, O'Donovan MC, et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am J Hum Genet* 2013, Feb 7;92(2):197-209.

14. David Warde-Farley. The genemania prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010, Jul 1;38(Web Server issue):W214.

15. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* 1998, Jun 4;393(6684):440-2.

16. Humphries MD, Gurney K. Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PLoS One* 2008, Apr 30;3(4):e0002051.

17. Karch CM, Prudencio M, Winkler DD, Hart PJ, Borchelt DR. Role of mutant SOD1 disulfide oxidation and aggregation in the pathogenesis of familial ALS. *Proc Natl Acad Sci* U S A 2009, May 12;106(19):7774-9.

18. Schork AJ, Wang Y, Thompson WK, Dale AM, Andreassen OA. New statistical approaches exploit the polygenic architecture of schizophrenia--implications for the underlying neurobiology. *Curr Opin Neurobiol* 2016, Feb;36:89-98.

19. Escott-Price V, Sims R, Bannister C, Harold D, Vronskaya M, Majounie E, et al. Common polygenic variation enhances risk prediction for alzheimer's disease. *Brain* 2015, Dec;138(Pt 12):3673-84.

20. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet* 2018, Sep;19(9):581-90.

21. Wilke C, Baets J, De Bleecker JL, Deconinck T, Biskup S, Hayer SN, et al. Beyond ALS and FTD: The phenotypic spectrum of TBK1 mutations includes psp-like and cerebellar phenotypes. *Neurobiol Aging* 2018;62:244.e9-244.e13.

22. Chen JA, Chen Z, Won H, Huang AY, Lowe JK, Wojta K, et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol Neurodegener* 2018;13(1):41.

23. Ryberg H, An J, Darko S, Lustgarten JL, Jaffa M, Gopalakrishnan V, et al. Discovery and verification of amyotrophic lateral sclerosis biomarkers by proteomics. *Muscle Nerve* 2010, Jul;42(1):104-11.

24. Turner MR, Goldacre R, Ramagopalan S, Talbot K, Goldacre MJ. Autoimmune disease preceding amyotrophic lateral sclerosis: An epidemiologic study. *Neurology* 2013, Oct 1;81(14):1222-5.

25. Hemminki K, Li X, Sundquist J, Sundquist K. Familial risks for amyotrophic lateral sclerosis and autoimmune diseases. *Neurogenetics* 2009, Apr;10(2):111-6.

26. Zou ZY, Liu CY, Che CH, Huang HP. Toward precision medicine in amyotrophic lateral sclerosis. *Ann Transl Med* 2016, Jan;4(2):27.

27. Ferraiuolo L, Kirby J, Grierson AJ, Sendtner M, Shaw PJ. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat Rev Neurol* 2011, Nov;7(11):616.

28. Hopkins AL. Network pharmacology: The next paradigm in drug discovery. *Nat Chem Biol* 2008, Nov;4(11):682-90.

29. Harrington RA. Targeting inflammation in coronary artery disease. *N Engl J Med* 2017;377(12):1197-8.

30. Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, Kelsoe JR, et al. Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genet* 2013, Apr;9(4):e1003455.

31. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* 2014, Oct;17(10):1418-28.

32. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: A real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 2008;9 Suppl 1:S4.

33. Newman E J. The structure and function of complex networks. *SIAM Review* 2003;45(3):167-256.

34. Kolaczyk ED, Csárdi G. Statistical analysis of network data with R. New York, N.Y: Springer; 2014.

35. Kyoko Watanabe. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017;8.

36. Dangond F, Hwang D, Camelo S, Pasinelli P, Frosch MP, Stephanopoulos G, et al. Molecular signature of late-stage human ALS revealed by expression profiling of postmortem spinal cord gray matter. *Physiol Genomics* 2004, Jan 15;16(2):229-39.

37. Lukas TJ, Luo WW, Mao H, Cole N, Siddique T. Informatics-assisted protein profiling in a transgenic mouse model of amyotrophic lateral sclerosis. *Mol Cell Proteomics* 2006, Jul;5(7):1233-44.

38. Daoud H, Valdmanis PN, Gros-Louis F, Belzil V, Spiegelman D, Henrion E, et al. Resequencing of 29 candidate genes in patients with familial and sporadic amyotrophic lateral sclerosis. *Arch Neurol* 2011, May;68(5):587-93.

39. Kwee LC, Liu Y, Haynes C, Gibson JR, Stone A, Schichman SA, et al. A high-density genome-wide association screen of sporadic ALS in US veterans. *PLoS One* 2012;7(3):e32768.

40. Benyamin B, He J, Zhao Q, Gratten J, Garton F, Leo PJ, et al. Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nat Commun* 2017;8(1):611.

41. Haider SA, Faisal M. Human aging in the post-gwas era: Further insights reveal potential regulatory variants. *Biogerontology* 2015, Aug;16(4):529-41.

42. Kwee LC, Liu Y, Haynes C, Gibson JR, Stone A, Schichman SA, et al. A high-density genome-wide association screen of sporadic ALS in US veterans. *PLoS One* 2012;7(3):e32768.

43. Chen Y, Zhou Q, Gu X, Wei Q, Cao B, Liu H, et al. An association study between SCFD1 rs10139154 variant and amyotrophic lateral sclerosis in a chinese cohort. Amyotroph Lateral Scler Frontotemporal Degener 2018, Aug;19(5-6):413-8.

44. Simpson CL, Lemmens R, Miskiewicz K, Broom WJ, Hansen VK, van Vught PW, et al. Variants of the elongator protein 3 (ELP3) gene are associated with motor neuron degeneration. *Hum Mol Gene*t 2009, Feb 1;18(3):472-81.

45. Morahan JM, Yu B, Trent RJ, Pamphlett R. A genome-wide analysis of brain DNA methylation identifies new candidate genes for sporadic amyotrophic lateral sclerosis. Amyotroph Lateral Scler 2009;10(5-6):418-29.

46. Gijselinck I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G, et al. A c9orf72 promoter repeat expansion in a flanders-belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: A gene identification study. *Lancet Neurol* 2012, Jan;11(1):54-65.

47. Schymick JC, Scholz SW, Fung HC, Britton A, Arepalli S, Gibbs JR, et al. Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: First stage analysis and public release of data. *Lancet Neurol* 2007, Apr;6(4):322-8.

48. Fogh I, Ratti A, Gellera C, Lin K, Tiloca C, Moskvina V, et al. A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. *Hum Mol Genet* 2014, Apr;23(8):2220-31.

49. Lahut S, Ömür Ö, Uyan Ö, Ağım ZS, Özoğuz A, Parman Y, et al. ATXN2 and its neighbouring gene SH2B3 are associated with increased ALS risk in the turkish population. *PLoS One* 2012;7(8):e42956.

50. Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* 2015, Mar 27;347(6229):1436-41.

51. van Doormaal PTC, Ticozzi N, Weishaupt JH, Kenna K, Diekstra FP, Verde F, et al. The role of de novo mutations in the development of amyotrophic lateral sclerosis. *Hum Mutat* 2017;38(11):1534-41.

52. Freischmidt A, Wieland T, Richter B, Ruf W, Schaeffer V, Müller K, et al. Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. *Nat Neurosci* 2015, May;18(5):631-6.

53. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, et al. Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron* 2018, Mar 21;97(6):1268-1283.e6.

54. Saba R, Medina SJ, Booth SA. A functional SNP catalog of overlapping mirna-binding sites in genes implicated in prion disease and other neurodegenerative disorders. *Hum Mutat* 2014, Oct;35(10):1233-48.

55. Cronin S, Berger S, Ding J, Schymick JC, Washecka N, Hernandez DG, et al. A genome-wide association study of sporadic ALS in a homogenous irish population. *Hum Mol Genet* 2008, Mar;17(5):768-74.

56. Orozco D, Edbauer D. FUS-mediated alternative splicing in the nervous system: Consequences for ALS and FTLD. *J Mol Med* 2013, Dec;91(12):1343-54.

57. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013, Dec;45(12):1452-8.

58. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for ADHD. *BioRxiv* 2017, Jun:145581.

59. Lo M-T, Hinds DA, Tung JY, Franz C, Fan C-C, Wang Y, et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat Genet* 2016, Dec;49(1):152.

60. Clarke TK, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK biobank (N=112 117). *Mol Psychiatry* 2017, Oct;22(10):1376-84.

61. Tielbeek JJ, Johansson A, Polderman TJC, Rautiainen MR, Jansen P, Taylor M, et al. Genome-Wide association studies of a broad spectrum of antisocial behavior. *JAMA Psychiatry* 2017, Dec 1;74(12):1242-50.

62. Consortium TASDWGOTPG. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism* 2017, May;8(1):21.

63. Stahl E, Forstner A, McQuillin A, Ripke S, PGC BDWGOT, Ophoff R, et al. Genomewide association study identifies 30 loci associated with bipolar disorder. *BioRxiv* 2017, Aug:173062.

64. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015, Feb 12;518(7538):197-206.

65. Witt SH, Streit F, Jungkunz M, Frank J, Awasthi S, Reinbold CS, et al. Genome-wide association study of borderline personality disorder reveals genetic overlap with bipolar disorder, major depression and schizophrenia. *Transl Psychiatry* 2017;7(6):e1155.

66. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013, Mar;45(4):353.

67. Consortium TC, Nikpay M, Goel A, Won H-H, Hall LM, Willenborg C, et al. A comprehensive 1000 genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015, Sep;47(10):1121.

68. Kouri N, Ross OA, Dombroski B, Younkin CS, Serie DJ, Soto-Ortolaza A, et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nat Commun* 2015, Jun 16;6:7247.

69. de Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* 2017, Feb;49(2):256-61.

70. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010, Apr;42(4):295-302.

71. Jones SE, Tyrrell J, Wood AR, Beaumont RN, Ruth KS, Tuke MA, et al. Genome-Wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. *PLoS Genet* 2016, Aug;12(8):e1006125.

72. Tobacco T, Consortium G, Furberg H, Kim Y, Dackor J, Boerwinkle E, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet* 2010, Apr;42(5):441.

73. Trampush JW, Yang MLZ, Yu J, Knowles E, Davies G, Liewald DC, et al. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: A report from the COGENT consortium. *Mol Psychiatry* 2017, Jan;22(3):336.

74. Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE, et al. Genome-wide association study of cognitive functions and educational attainment in UK biobank (N=112 151). *Mol Psychiatry* 2016;21(6):758-67.

75. Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B, et al. Corrigendum: Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun* 2015, Oct;6:8739.

76. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for c-reactive protein levels. *Circulation* 2011, Feb 22;123(7):731-8.

77. Studies TICFBPG-WA, Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 2011, Sep;478(7367):103.

78. Okbay A, Baselmans BML, Neve J-ED, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* 2016, Apr;48(6):624.

79. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016, May;533(7604):539.

80. Ferrari R, Hernandez DG, Nalls MA, Rohrer JD, Ramasamy A, Kwok JB, et al. Frontotemporal dementia and its subtypes: A genome-wide association study. *Lancet Neurol* 2014, Jul;13(7):686-99.

81. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* 2018, Jun;50(7):912.

82. Consortium GLG, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet 2013*, Oct;45(11):1274.

83. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat Genet* 2014, Oct;46(11):1173.

84. Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivières S, Jahanshad N, et al. Common genetic variants influence human subcortical brain structures. *Nature* 2015, Jan;520(7546):224.

85. Hammerschlag AR, Stringer S, de Leeuw CA, Sniekers S, Taskesen E, Watanabe K, et al. Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. *Nat Genet* 2017, Jun;49(11):1584.

86. Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E, et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet* 2017, May;49(7):1107.

87. Timofeeva MN, Hung RJ, Rafnar T, Christiani DC, Field JK, Bickeböller H, et al. Influence of common genetic variation on lung cancer risk: Meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet* 2012, Nov 15;21(22):4980-95.

88. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet* 2018, May;50(5):668-81.

89. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet* 2017, Apr;49(6):834.

90. Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and brca1-mediated DNA repair. *Nat Genet* 2015, Sep;47(11):1294.

91. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011, Aug 10;476(7359):214-9.

92. Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin UM, Saad M, et al. Imputation of sequence variants for identification of genetic risks for parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet* 2011, Feb 19;377(9766):641-9.

93. Eeles RA, Olama AAA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the icogs custom genotyping array. *Nat Genet* 2013, Mar;45(4):385.

94. Ellinghaus D, Ellinghaus E, Nair RP, Stuart PE, Esko T, Metspalu A, et al. Combined analysis of genome-wide association studies for crohn disease and psoriasis identifies seven shared susceptibility loci. *Am J Hum Genet* 2012, Apr 6;90(4):636-47.

95. Höglinger GU, Melhem NM, Dickson DW, Sleiman PM, Wang LS, Klei L, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet* 2011, Jun 19;43(7):699-705.

96. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014, Feb 20;506(7488):376-81.

97. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014, Jul 24;511(7510):421-7.

98. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. *Nature* 2016, Jul;536(7614):41.

99. Jin Y, Andersen G, Yorgov D, Ferrara TM, Ben S, Brownson KM, et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat Genet* 2016;48(11):1418-24.

100. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Mägi R, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015, Feb 12;518(7538):187-96.

Figures

Fig. 1: Fold enrichment plots of enrichment versus nominal -log$_{10}$ p-values (corrected for inflation) in Amyotrophic lateral sclerosis (ALS). Fold enrichment plots of enrichment versus nominal -log$_{10}$ p-values (corrected for inflation) in ALS below the standard GWAS threshold of p-value < 5x10^{-8} as a function of significance of association with 65 distinct traits and diseases and at the level of p-value ≤ 1, p-value ≤ 0.1, p-value ≤ 0.01, respectively. Blue line indicates all SNPs.

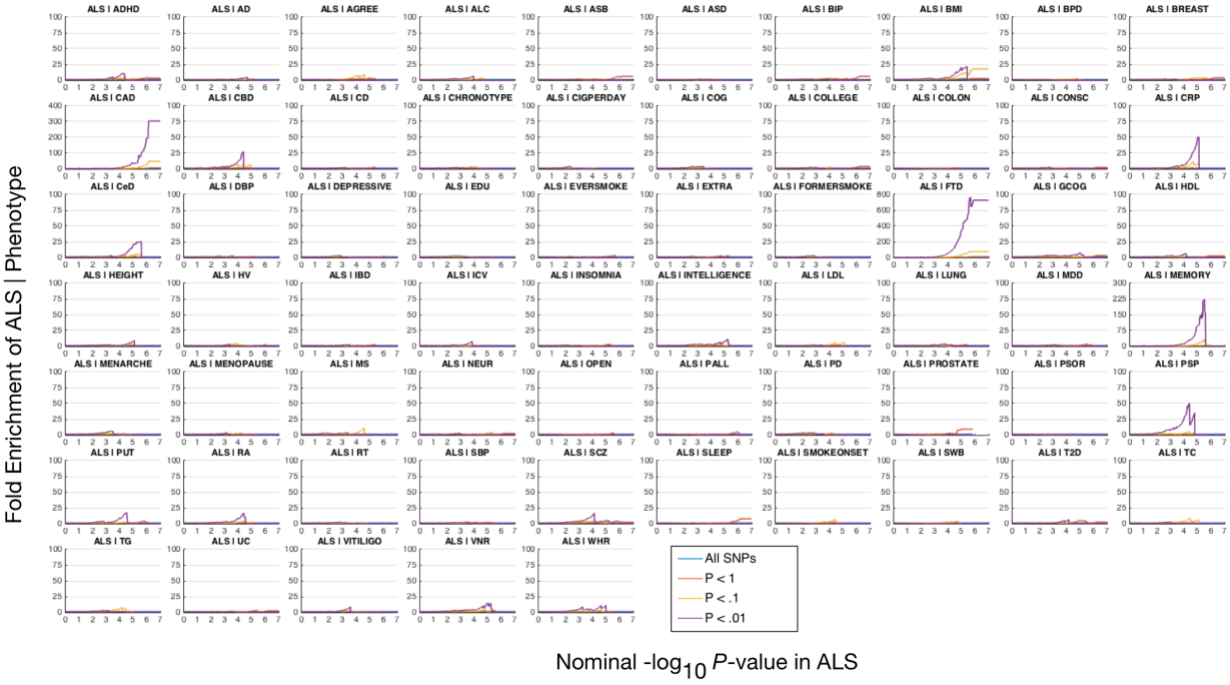


Fig. 2: 'Conditional' Manhattan plot of conditional –log$_{10}$ (FDR) values for Amyotrophic lateral sclerosis (ALS) as a function of 65 distinct traits and diseases. SNPs with conditional –log$_{10}$ FDR > 1.3 (i.e. FDR < 0.05) are shown with large points. A black line around the large points indicates the most significant SNP in each LD block and this SNP was annotated with the closest gene, which is listed above the symbols in each locus.

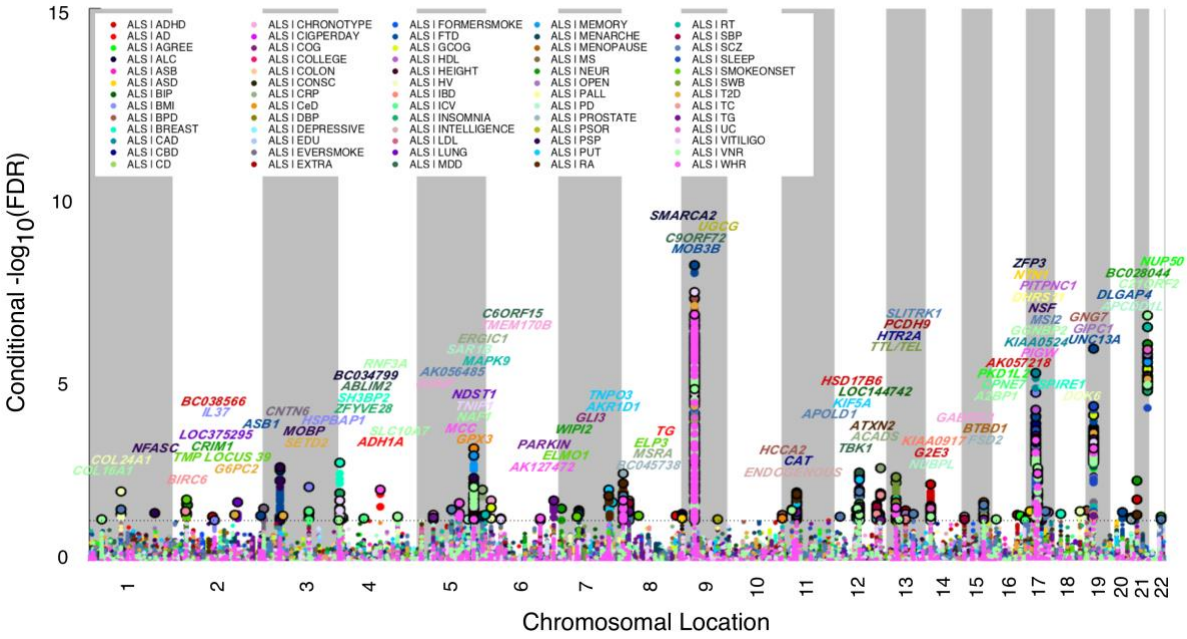


Fig. 3: Network interaction graph illustrating genetic interactions, physical interactions, and co-expression patterns associated with the ALS risk genes.

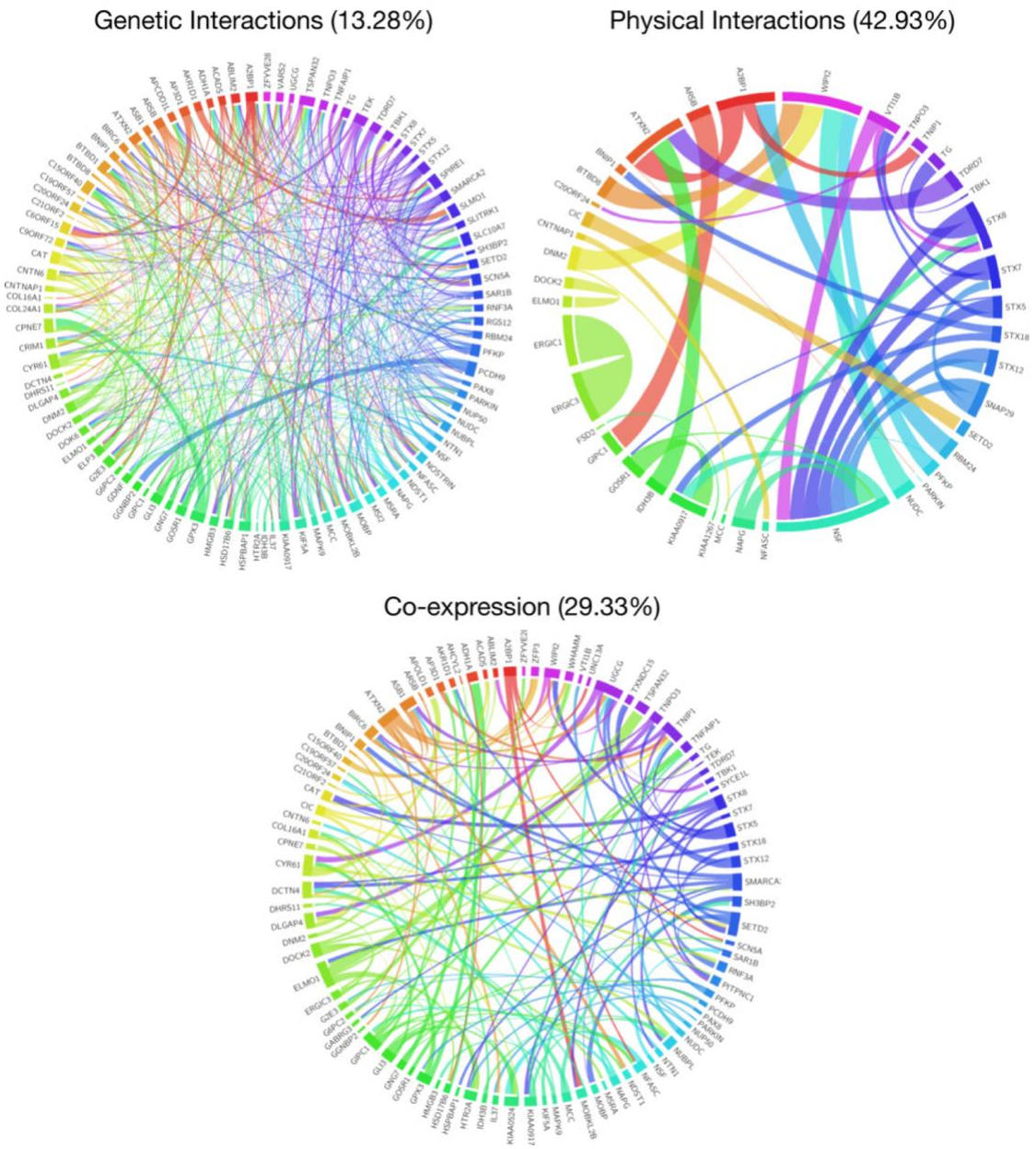
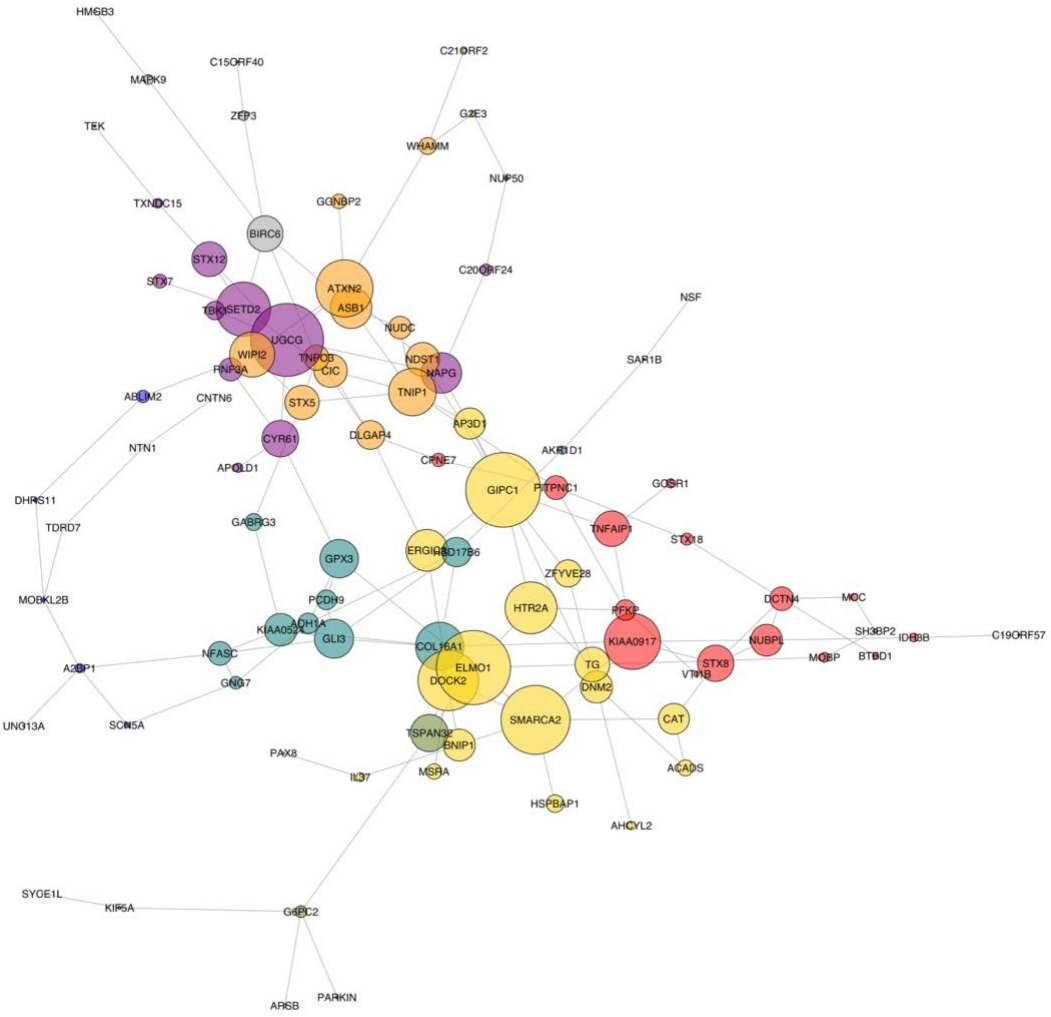


Fig. 4: Co-expression interaction network plot and dendrogram. a, Co-expression interaction network plot. Each node represents a single gene and the edges (lines between genes) represent co-expression interactions between genes. The size of each node corresponds to the eigenvector centrality score for that gene. The color of each node represents membership to a distinct subnetwork. **b,** Co-expression interaction dendrogram plot. The co-expression network was partitioned into 8 subnetworks by removing edges with high edge-betweenness centrality. Hub genes are annotated in brown and marked by a large point.



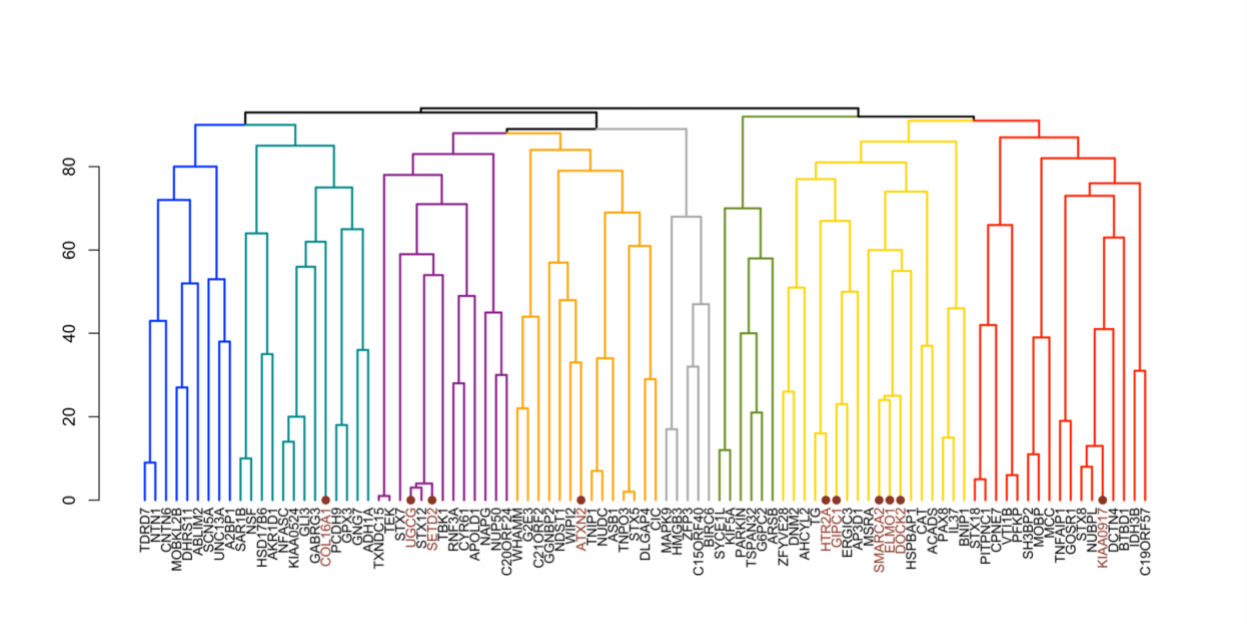
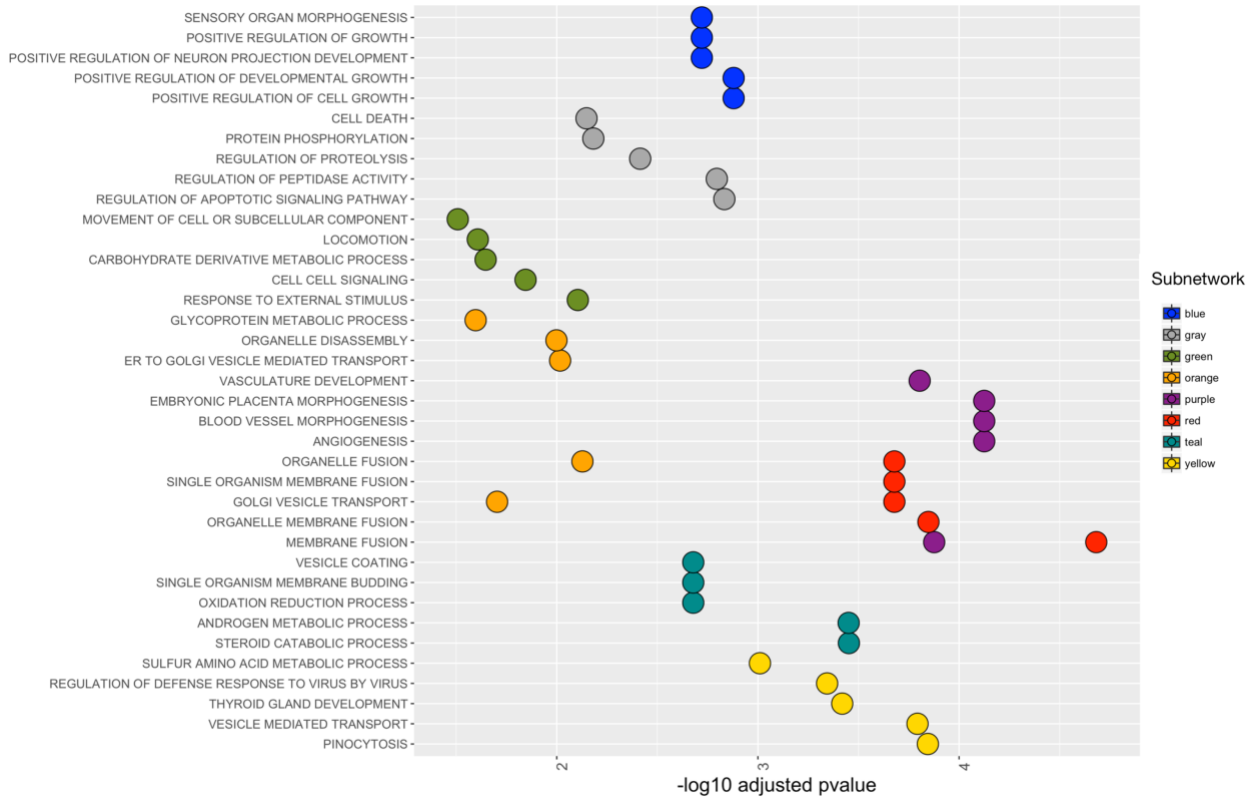

0
20
40
60
80
TDRD7
NTN1
CNTN6
MOBKL2B
DHRS11
ABLIM2
SCN5A
UNC13A
AZBP1
SAR1B
NSF
HSD17B6
AKR1D1
NFASC
KIAA0524
GLI3
GABRG3
COL16A1
PCDH9
GPX3
GNG7
ADH1A
TXNDC15
TEK
STX7
UGCG
STX12
SETD2
TBK1
RNF3A
CYR61
APOLD1
NAPG
NUP50
C20ORF24
WHAMM
G2E3
C21ORF2
GGNBP2
NDST1
WIPI2
ATXN2
TNIP1
NUDC
ASB1
TNPO3
STX5
DLGAP4
CIC
MAPK9
HMGB3
ZF3
C15ORF40
BIRC6
SYCE1L
KIF5A
PARKIN
TSPAN32
GPC2
ARSB
ZFYVE28
DNM2
AHCYL2
TG
HTR2A
GIPC1
ERGIC3
AP3D1
MSRA
SMARCA2
ELMO1
DOCK2
HSPBAP1
CAT
ACADS
PAX8
IL37
BNIP1
STX18
PITPNC1
CPNE7
VTI1B
PFKP
SH3BP2
MOBP
MCC
TNFAIP1
GOSR1
STX8
NUBPL
KIAA0917
DCTN4
BTBD1
IDH3B
C19ORF57

Fig. 5: Biological pathways associated with the ALS co-expression subnetworks. a, Biological pathways associated with each subnetwork of the ALS co-expression network classified using FUMA (http://fuma.ctglab.nl/). **b,** Neuron-specific biological pathways associated with each subnetwork also classified using FUMA.



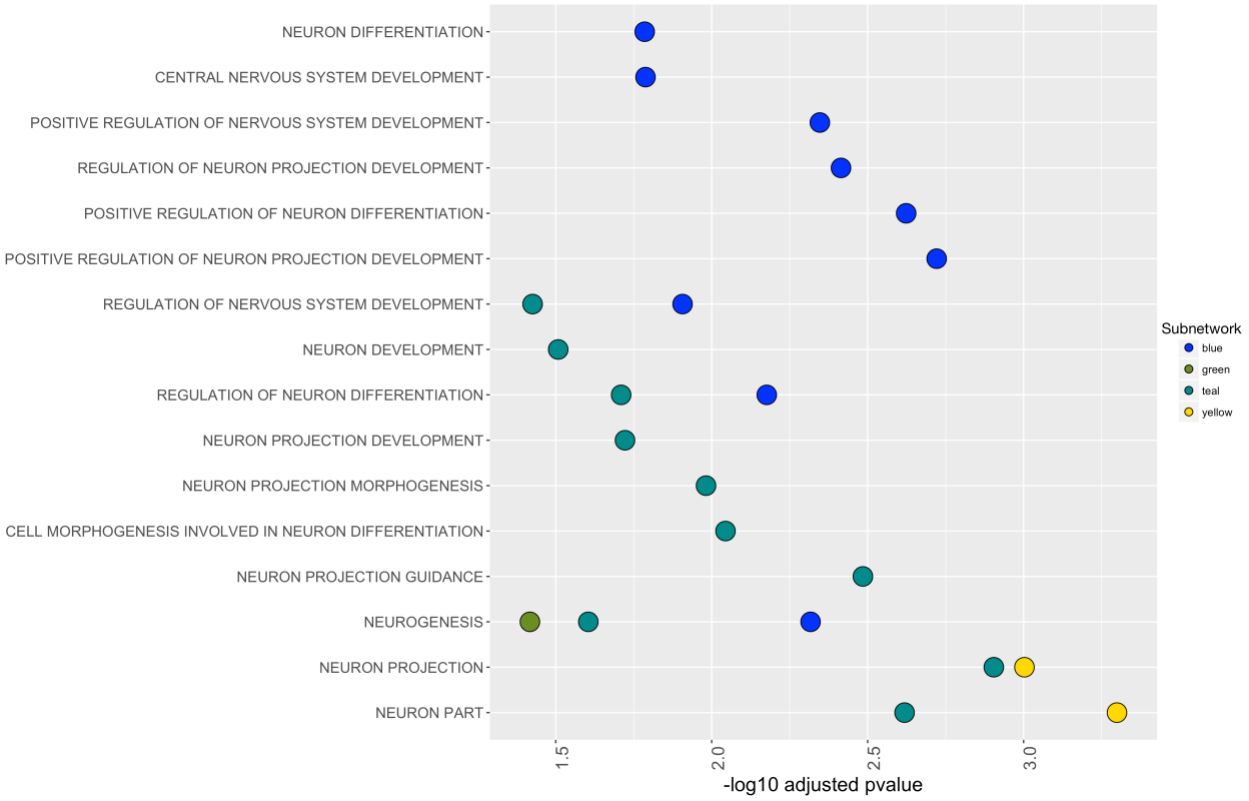
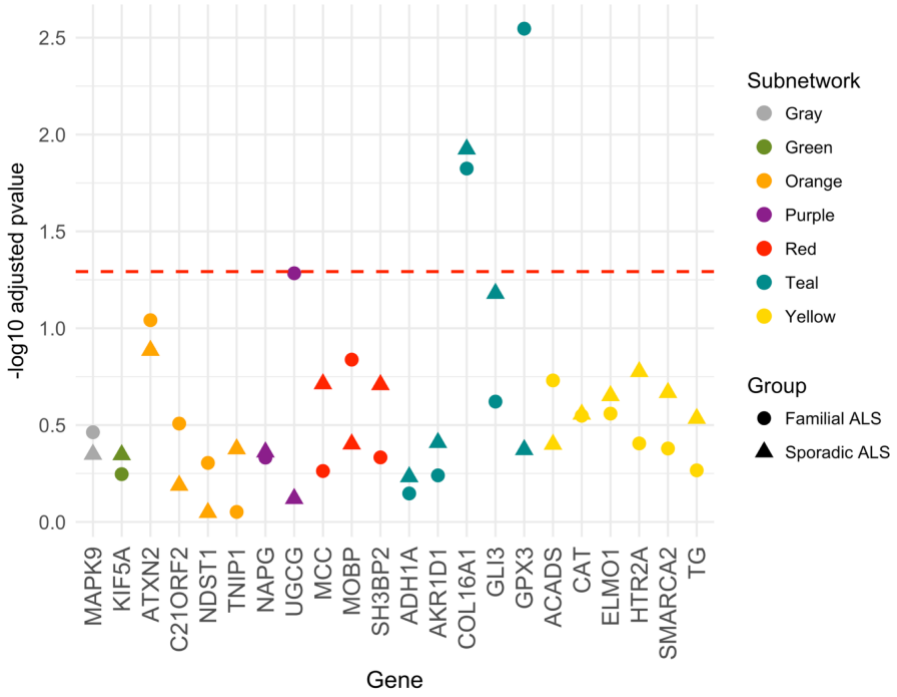

NEURON DIFFERENTIATION
CENTRAL NERVOUS SYSTEM DEVELOPMENT
POSITIVE REGULATION OF NERVOUS SYSTEM DEVELOPMENT
REGULATION OF NEURON PROJECTION DEVELOPMENT
POSITIVE REGULATION OF NEURON DIFFERENTIATION
POSITIVE REGULATION OF NEURON PROJECTION DEVELOPMENT
REGULATION OF NERVOUS SYSTEM DEVELOPMENT
NEURON DEVELOPMENT
REGULATION OF NEURON DIFFERENTIATION
NEURON PROJECTION DEVELOPMENT
NEURON PROJECTION MORPHOGENESIS
CELL MORPHOGENESIS INVOLVED IN NEURON DIFFERENTIATION
NEURON PROJECTION GUIDANCE
NEUROGENESIS
NEURON PROJECTION
NEURON PART
1.5
2.0
2.5
3.0
-log10 adjusted pvalue
Subnetwork
blue
green
teal
yellow

Fig. 6: Differential expression of ALS risk genes in diseased tissue. a, Differential expression of ALS risk genes in tissues of patients with ALS. **b,** Differential expression of ALS risk genes in SOD1 G93A transgenic mouse.



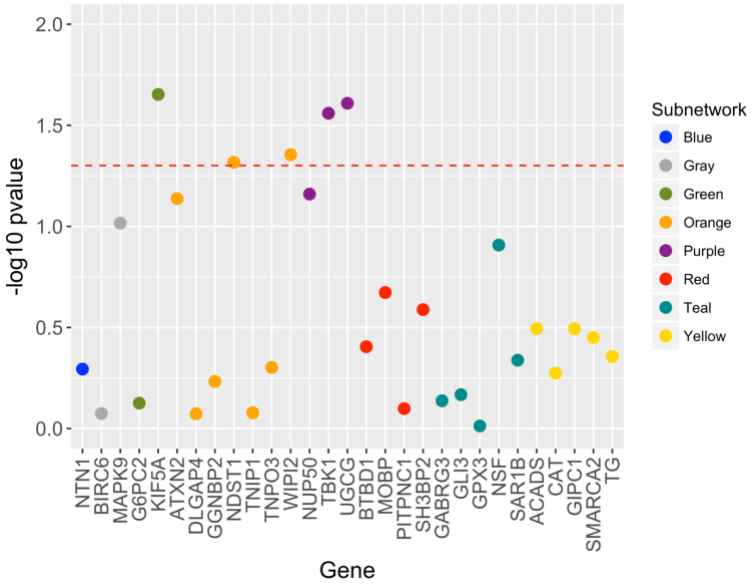


Table 1

	Disease/Trait	Pleiotropic SNP	Closest Gene	Chr	Position	ALS p-value	Min CondFDR	Previously Reported SNPs in LD	Other SNPs within Gene	Citation
1	VNR	rs10914464	COL16A1	1	32132319	1.26E-04	4.54E-02	Novel		
2	HV	rs10443173	COL24A1	1	86533754	8.20E-06	8.11E-03	Novel		
3	PSP	rs2068667	NFASC	1	204917680	3.39E-04	3.16E-02	rs2068667		[8]
4	FTD	rs515342	ASB1	2	239367296	6.22E-04	3.71E-02	rs515342		[8]
5	ADHD	rs896444	BC038566 (NCK2)	2	106241808	1.32E-04	3.89E-02	Novel		
6	TC	rs6737916	BIRC6	2	32597985	2.82E-05	2.81E-02	Novel		
7	NEUR	rs181357866	CRIM1	2	35863487	3.10E-05	1.92E-02	Independent	rs139895660	[38]
8	T2D	rs13387347	G6PC2	2	169754846	2.95E-05	3.52E-02	Novel		
9	BMI	rs7602576	IL37	2	113699617	6.79E-05	4.80E-02	Novel		
10	LUNG	rs11695294	LOC375295	2	177473618	5.12E-05	1.62E-02	Novel		

11	SMOKEONSET	rs13417671	TMP LOCUS 39 (RASGRP3)	2	34043426	1.04E-05	1.36E-02	Novel		
12	EVERSMOKE	rs17042645	CNTN6	3	1731848	1.26E-04	2.45E-02	rs149853584		[7]
13	BMI	rs11718653	HSPBAP1	3	122477858	9.41E-06	6.15E-03	Novel		
14	PSP	rs9820623	MOBP	3	39493858	1.64E-05	1.89E-03	rs13079368	rs1768208 rs616147	[8] [5] [7]
15	T2D	rs111970477	SETD2	3	47105581	6.55E-05	3.76E-02	Novel		
16	MDD	rs10938692	ABLIM2	4	8118561	2.57E-05	4.21E-02	Novel		
17	AD	rs1159918	ADH1A	4	100243009	1.98E-05	7.39E-03	Novel		
18	ALC	rs78710307	BC034799 (LOC105377671)	4	59983437	1.07E-04	4.32E-02	Novel		
19	VNR	rs12646225	RNF3A	4	696848	1.51E-04	4.91E-02	Novel		
20	BREAST	rs430979	SH3BP2	4	2814698	2.41E-06	1.40E-03	Novel		
21	VNR	rs10011222	SLC10A7	4	147441065	1.00E-04	3.84E-02	Novel		

22	INSOMNIA	rs61789437	ZFYVE28	4	2315630	2.08E-06	9.40E-03	Novel		
23	SCZ	rs34384833	AK056485 (GCNT4)	5	91238155	1.65E-05	2.50E-02	rs6453104		[39]
24	CRP	rs7735726	ERGIC1	5	172347823	1.91E-05	7.56E-03	rs538622		[8]
25	CHRONOTYPE	rs10050775	GDNF	5	38008042	1.00E-05	3.61E-02	Novel		
26	CeD	rs3828599	GPX3 (TNIPI)	5	150401796	2.45E-06	5.85E-04	rs10463311 rs3828599		[40] [7] [8] [12]
27	RT	rs35318094	MAPK9	5	179672984	1.07E-05	3.53E-02	Novel		
28	WHR	rs17326496	MCC	5	112676579	5.79E-05	1.70E-02	Novel		
29	VNR	rs10463311	NAF1	5	150410835	7.85E-07	6.47E-03	Independent	rs17111695	[8] [40]
30	LUNG	rs150949995	NDST1	5	149898499	4.13E-06	1.01E-02	Novel		
31	PD	rs7728741	SAR1B	5	133942492	2.55E-04	4.58E-02	Novel		
32	VITILIGO	rs4958888	TNIP1	5	150472842	1.45E-05	9.80E-03	Independent	rs10463311	[40] [7]
33	WHR	rs1543705	AK127472	6	127197108	1.92E-04	4.43E-02	Novel		

34	MDD	rs145705679	C6ORF15	6	31069749	1.34E-04	4.30E-02	Novel		
35	TG	rs4333390	PARKIN	6	162330919	4.22E-05	1.45E-02	Independent	rs16892673 rs777468774 rs956103217 rs7740421 rs7757630 rs7764218 rs6904956 rs564053 rs6931162	[41] [42]
36	CHRONOTYPE	rs651001	TMEM170B	6	11569402	3.57E-05	1.48E-02	Novel		
37	PUT	rs141730255	AKR1D1	7	137938550	5.24E-05	1.63E-02	Novel		
38	SMOKEONSET	rs17171046	ELMO1	7	37477863	3.90E-05	3.88E-02	Novel		
39	COG	rs141347161	GLI3	7	42417313	3.03E-05	2.69E-02	Novel		
40	PUT	rs10488631	TNPO3	7	128594183	2.60E-05	7.66E-03	rs10488631		[12] [8]
41	NEUR	rs8180839	WIPI2	7	5239970	6.14E-05	2.41E-02	Novel		
42	PROSTATE	rs6996532	BC045738	8	2417678	2.52E-06	2.82E-03	rs17070492	rs7813314	[7] [8] [43] [5]

43	SMOKEONSET	rs11786174	ELP3	8	28038789	1.84E-04	3.75E-02	Independent	rs12682496 rs60024250 rs2614046 rs13268953 rs6985069	[44] [42]
44	CRP	rs11250002	MSRA	8	10257041	5.60E-05	1.42E-02	Independent	rs968806172 (not in database)	[45]
45	AD	rs79496463	TG (TGN)	8	133917088	7.02E-05	3.69E-02	Novel		
46	MDD	rs117204439	C9ORF72	9	27607973	2.67E-10	2.78E-07	Independent	rs2225389 rs774352 rs774351 rs903603 rs2814707 rs3849942 rs774359 rs1752784 rs2782931	[8] [5] [46] [47]

47	FTD	rs895021	MOB3B (MOBKL2B)	9	27484911	6.02E-17	7.38E-09	rs3849943 rs3849941 rs2477523 rs2453555 rs774352 rs774351 rs3849942 rs2814707 rs774359 rs7019847		[8] [48] [5] [46] [47]
48	ALC	rs7024326	SMARCA2	9	1903397	1.86E-05	3.28E-02	Novel		
49	PSOR	rs7041171	UGCG	9	114701630	2.78E-05	4.72E-02	Novel		
50	CBD	rs7944397	CAT	11	34455309	1.14E-05	8.81E-03	rs12803540	rs7118388	[8]
51	INTELLIGENCE	rs71472777	ENDOGENOUS	11	24142935	1.34E-05	2.54E-02	Novel	rs12361953	LUPZP2 in AD
52	BPD	rs140988250	HCCA2	11	1489420	1.27E-05	3.51E-02	Novel		
53	CRP	rs12369156	ACADS	12	121167675	7.70E-06	1.98E-03	Novel		
54	SCZ	rs117704471	APOLD1	12	12904565	7.20E-05	3.91E-02	Novel		

55	DBP	rs593226	ATXN2	12	111993886	7.43E-05	1.94E-02	rs739496 rs10849949 rs2073950 rs2301621 rs6490162 rs628825 rs63051 rs616513 rs12369009 rs695872 rs695871	rs2239194 rs3184504 rs2239194 rs10774625 rs10849952 rs17805591 rs16941541 rs7969300	[49]
56	MDD	rs74654358	TBK1	12	64881967	7.07E-07	2.68E-03	rs74654358 rs75209514	rs76805704 rs41292021 rs55824172 rs149881816 rs56196591 rs34774243 rs145905497 rs138369490 rs17857028 rs138839127 rs142030898 rs35635889 rs144424516 rs151225287 rs139195702 rs186475789 rs141727722 rs144370662 rs187122554 rs185524052	[5] [50] [51] [41] [52] [7]
57	ADHD	rs11171998	HSD17B6	12	57240444	1.04E-04	3.63E-02	Novel		

58	PUT	rs113247976	KIF5A	12	57975700	1.17E-05	2.45E-02	rs113247976	rs117027576 rs118082508 rs116900480 rs142321490	[53] [7]
59	BIP	rs144387708	LOC144742 (HSPB8, HSP22)	12	119702200	4.58E-06	3.27E-02	Novel	rs104894345 rs104894351	
60	CBD	rs1578303	HTR2A	13	47963146	9.55E-05	3.63E-02	rs1578303		[8]
61	EXTRA	rs10492593	PCDH9	13	67494117	2.73E-05	2.64E-02	Independent	rs35892541 rs10492593	[54] [8]
62	SCZ	rs6420358	SLITRK1	13	85289468	5.84E-05	3.35E-02	Novel		
63	SWB	rs17446243	TTL/TEL	13	40748931	1.17E-05	3.41E-03	rs17446243		[8]
64	ADHD	rs447614	G2E3 KIAA0917	14	31080799	1.97E-05	5.47E-03	rs10139154	rs179552	[5] [43] [7]
65	PD	rs12886280	NUBPL	14	32298659	3.31E-06	2.18E-02	rs12886280		[8]
66	MENOPAUSE	rs2381030	BTBD1	15	83741876	2.35E-05	1.64E-02	rs6603044		[8]
67	PROSTATE	rs1877240	FSD2	15	83438878	1.04E-05	4.37E-02	Novel		

68	CHRONOTYPE	rs3098553	GABRG3	15	27876256	2.85E-05	3.82E-02	rs3097439		[7]
69	VNR	rs7193729	A2BP1	16	7145869	1.19E-04	4.76E-02	Independent	rs1551960	[55]
70	ADHD	rs56024498	AK057218	16	76927135	8.24E-05	3.46E-02	Novel		
71	ICV	rs62068675	CPNE7	16	89648291	1.51E-04	4.81E-02	Independent	rs74213330	[45]
72	AGREE	rs192688752	PKD1L2	16	81152628	2.43E-05	3.63E-02	Novel		
73	VNR	rs9903355	GGNBP2 DHRS11 PIGW	17	34937221	4.93E-06	1.99E-03	rs2285642		[8]
74	CAD	rs35714695	KIAA0524 (SARM1)	17	26719788	1.16E-08	5.75E-06	rs739439 rs35714695		[8] [5] [7]
75	SCZ	rs2240601	MSI2	17	55751112	4.82E-05	2.89E-02	rs2240601		[8]
76	PSP	rs7224296	NSF	17	44800046	5.78E-04	4.94E-02	rs7224296		[8]
77	ASD	rs9894834	NTN1	17	9129028	8.16E-05	2.83E-02	Novel		
78	HDL	rs11652752	PITPNC1	17	65375892	7.01E-05	2.93E-02	rs11652752		[56] (chr17:65373923-65689644)

79	ALC	rs7209200	ZFP3	17	4969940	6.44E-05	3.75E-02	Novel	
80	HV	rs17187386	DOK6	18	67196957	6.19E-05	2.76E-02	Novel	
81	BREAST	rs12967284	SPIRE1	18	12532098	1.62E-05	3.31E-02	Novel	
82	OPEN	rs112183647	GIPC1	19	14612725	2.25E-05	4.26E-02	Novel	
83	BPD	rs7258235	GNG7	19	2612118	1.38E-05	2.70E-02	Novel	
84	FTD	rs12608932	UNC13A	19	17752689	1.91E-08	1.28E-06	rs78549703 rs12608932	[8] [5]
85	PD	rs6015322	APCDD1L	20	57206540	6.30E-05	4.09E-02	rs6015322	[7]
86	FTD	rs2425220	DLGAP4	20	34999440	1.30E-04	3.06E-02	rs2425220	[8]
87	NEUR	rs9653747	BC028044	21	20041418	4.34E-06	4.33E-03	Novel	
88	VNR	rs75087725	C21ORF2	21	45753117	7.52E-11	1.65E-07	rs75087725	[5]
89	AGREE	rs2176039	NUP50	22	45585032	8.98E-05	3.82E-02	rs2176039	[7]