nature neuroscience

Systems genetics identifies a convergent gene network for cognition and neurodevelopmental disease

Michael R Johnson[1], Kirill Shkura[1,2,23], Sarah R Langley[1,3,23], Andree Delahaye-Duriez[1,2,4,23], Prashant Srivastava[1,23], W David Hill[5,6,23], Owen J L Rackham[3,23], Gail Davies[5,6], Sarah E Harris[5,7], Aida Moreno-Moral[2], Maxime Rotival[2], Doug Speed[8], Slavé Petrovski[9], Anaïs Katz[1,2], Caroline Hayward[10,11], David J Porteous[5,7,11], Blair H Smith[12], Sandosh Padmanabhan[13], Lynne J Hocking[14], John M Starr[5,15], David C Liewald[5], Alessia Visconti[16], Mario Falchi[16], Leonardo Bottolo[17,18], Tiziana Rossetti[2], Bénédicte Danis[19], Manuela Mazzuferi[19], Patrik Foerch[19], Alexander Grote[20], Christoph Helmstaedter[21], Albert J Becker[22], Rafal M Kaminski[19], Ian J Deary[5,6] & Enrico Petretto[2,3]

© 2015 Nature America, Inc. All rights reserved.

Genetic determinants of cognition are poorly characterized, and their relationship to genes that confer risk for neurodevelopmental disease is unclear. Here we performed a systems-level analysis of genome-wide gene expression data to infer gene-regulatory networks conserved across species and brain regions. Two of these networks, M1 and M3, showed replicable enrichment for common genetic variants underlying healthy human cognitive abilities, including memory. Using exome sequence data from 6,871 trios, we found that M3 genes were also enriched for mutations ascertained from patients with neurodevelopmental disease generally, and intellectual disability and epileptic encephalopathy in particular. M3 consists of 150 genes whose expression is tightly developmentally regulated, but which are collectively poorly annotated for known functional pathways. These results illustrate how systems-level analyses can reveal previously unappreciated relationships between neurodevelopmental disease–associated genes in the developed human brain, and provide empirical support for a convergent gene-regulatory network influencing cognition and neurodevelopmental disease.

Cognition refers to human mental abilities such as memory, attention, processing speed, reasoning and executive function. Performance on cognitive tasks varies between individuals, and is highly heritable[1] and polygenic[2,3]. To date, however, progress in identifying molecular genetic contributions to healthy human cognitive abilities has been limited[4,5].

A distinction can be made between cognitive domains such as the ability to apply acquired knowledge and learned skills (so-called crystallized abilities), and fluid cognitive abilities such as the capacity to establish new memories, reason in novel situations or perform cognitive tasks accurately and quickly[6]. Within individuals, performance on different measures of cognitive ability tends to be positively correlated such that people who do well in one domain, such as memory, tend to do well in other domains[7]. Seemingly disparate domains of cognitive ability also show high levels of genetic correlation in twin studies, typically in excess of 0.6 (ref. 8), and analyses using genome-wide similarity between unrelated individuals (genome-wide complex trait analysis) has also demonstrated substantial genetic correlation between diverse cognitive and learning abilities[9,10]. These studies suggest genes that influence human cognition may exert pleiotropic effects across diverse cognitive domains, such that genes regulating one cognitive ability might influence other cognitive abilities.

As impairment of cognitive function is a core clinical feature of many neurodevelopmental diseases including schizophrenia[11], autism[12], epilepsy[13] and intellectual disability (by definition), we sought to investigate gene-regulatory networks for human cognition and to determine their relationship to neurodevelopmental disease. An overview of our experimental design is provided in **Supplementary Figure 1**.

[1]Division of Brain Sciences, Imperial College Faculty of Medicine, London, UK. [2]Medical Research Council (MRC) Clinical Sciences Centre, Imperial College London, London, UK. [3]Duke-NUS Medical School, Singapore, Republic of Singapore. [4]Université Paris 13, Sorbonne Paris Cité, UFR de Santé, Médecine et Biologie Humaine, Paris, France. [5]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [6]Department of Psychology, University of Edinburgh, Edinburgh, UK. [7]Medical Genetics Section, Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [8]University College London Genetics Institute, London, UK. [9]Department of Medicine, Austin Hospital and Royal Melbourne Hospital, University of Melbourne, Melbourne, Victoria, Australia. [10]Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [11]Generation Scotland, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK. [12]Division of Population Health Sciences, University of Dundee, Dundee, UK. [13]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [14]Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK. [15]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK. [16]Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK. [17]Department of Mathematics, Imperial College, London, UK. [18]Department of Medical Genetics, University of Cambridge, Cambridge, UK. [19]Neuroscience TA, UCB Pharma, Braine-l'Alleud, Belgium. [20]Department of Neurosurgery, University of Bonn, Bonn, Germany. [21]Department of Neuropsychology, University of Bonn, Bonn, Germany. [22]Department of Neuropathology, University of Bonn, Bonn, Germany. [23]These authors contributed equally to this work. Correspondence should be addressed to M.R.J. (m.johnson@imperial.ac.uk) or E.P. (enrico.petretto@duke-nus.edu.sg).

Received 8 September; accepted 13 November; published online 21 December 2015; doi:10.1038/nn.4205

RESULTS

Gene coexpression network analysis

We hypothesized that unsupervised genome-wide coexpression network analysis starting from the human hippocampus may be informative for genes and pathways that influence cognition. Specifically, gene coexpression network analysis could prioritize sets of genes preferentially enriched for common variants (that is, single-nucleotide polymorphisms; SNPs) associated with cognitive abilities and thus reveal genetic pathways that influence variable cognitive performance.

As starting material, we used 122 fresh-frozen whole-hippocampus samples surgically resected *en bloc* from patients with temporal lobe epilepsy (TLE) (**Supplementary Table 1**). We used surgical hippocampus samples from living patients to avoid potential unwanted effects on gene expression related to the variable agonal state or time to autopsy associated with post-mortem samples. In addition, we used several gene expression data sets (detailed below) to assess the reproducibility of the identified gene networks in non-TLE hippocampi both across species and across brain regions.

We first determined gene coexpression networks in the human hippocampus by weighted gene coexpression network analysis (WGCNA), which groups sets of covarying genes across the sample set into coexpression 'modules'[14]. Applied to the full set of 122 samples, WGCNA grouped the human hippocampus transcriptome into 24 distinct coexpression modules (M1–M24), which varied in size from 29 to 1,148 genes (**Fig. 1a** and **Supplementary Table 2**).

To identify which of the 24 hippocampus modules from patients with TLE had coexpression patterns unrelated to epilepsy, for each module we compared its coexpression topology in patients with TLE with that from hippocampus samples ascertained from persons with no history of psychiatric or neurological illness[15]. We undertook this comparative network analysis using the default network dissimilarity measure in WGCNA based on the topological overlap matrix (TOM)[14]. We calculated empirical *P* values for the validity (that is, reproducibility) of modules by comparing the average topological overlap for module genes to the average connectivity of 10,000 randomly sampled networks (Online Methods). After Bonferroni adjustment for the number of modules tested, we found that 16 of the 24 modules were significantly preserved in 63 nondiseased human postmortem hippocampus samples (empirical $P \leq 0.002$; **Fig. 1a** and **Supplementary Table 3**), suggesting the coexpression of genes in these 16 modules is unrelated to epilepsy. Additionally, preservation of these 16 coexpression modules in a distinct human hippocampus gene expression data set provides an independent line of evidence to support the validity of these modules.

As molecular pathways underlying cognitive processes might be evolutionarily conserved[16–18], and indeed the rodent hippocampus has long been the primary model for studying molecular processes related to learning and memory[19], we next aimed to identify which of the human hippocampus coexpression modules are preserved in the healthy mouse hippocampus. To this aim, we carried out high-throughput sequencing of mRNA (RNA-seq) on snap-frozen hippocampus samples from 100 healthy adult mice and assessed the coexpression patterns between the mouse orthologs of human hippocampus module genes (Online Methods). Of the 16 human hippocampus modules preserved between nondiseased postmortem hippocampus and surgical hippocampus samples from patients with TLE, four modules (M1, M3, M11 and M19) were also significantly preserved in the healthy mouse hippocampus (empirical $P \leq 0.002$; **Fig. 1a** and **Supplementary Table 3**).

To assess whether the four cross-species conserved hippocampus modules (M1, M3, M11 and M19) are specific to the hippocampus or more widely expressed and coexpressed across the human cortex, we then analyzed genome-wide gene expression data from 102 postmortem human brains from the UK Brain Expression Consortium (UKBEC)[20] across the following brain regions: cerebellum, temporal cortex, occipital cortex and frontal cortex. We treated each brain region as an independent data set and adjusted gene expression levels in UKBEC for age, gender, postmortem interval, cause of death and brain bank identifier. Comparative network analysis, performed as described above, showed preservation of all four hippocampus coexpression modules in multiple other brain regions (**Supplementary Table 4**). Therefore, despite the modules being originally reconstructed from hippocampus gene expression data, these results suggest the modules are not specific to the hippocampus and thus might be capturing functions that are more widely distributed in the human cortex.

We analyzed biological terms and canonical pathways enriched among the genes in all 24 hippocampus modules from TLE patients (**Supplementary Table 5**). The different hippocampal coexpression modules demonstrated notable functional specificity. Of the four modules conserved in healthy hippocampi across species (M1, M3, M11 and M19), only M1 (n = 1,148 genes) and M3 (n = 150 genes) were enriched for functional categories explicitly related to synaptic processes (**Fig. 1b**). Module M1 was highly enriched for Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways 'calcium signaling' (Benjamini-Hochberg (BH)-corrected $P = 7.3 \times 10^{-7}$, ratio of enrichment (r) = 3.0), 'axon guidance' (BH $P = 9.0 \times 10^{-5}$, r = 2.5) and 'long-term potentiation' (BH $P = 4.0 \times 10^{-3}$, r = 5.0), and for the gene ontology (GO) terms 'synapse' (BH $P = 6.9 \times 10^{-15}$, r = 2.5), 'neuron projection' (BH $P = 1.4 \times 10^{-14}$, r = 2.2) and 'synaptic vesicle' (BH $P = 2.9 \times 10^{-8}$, r = 3.5). Module M3 was enriched for genes belonging to 'postsynaptic density' (PSD) (BH $P = 9.0 \times 10^{-4}$, r = 6.6) and 'Reelin signaling pathway' (BH P = 0.049, r = 12.5). We therefore investigated whether M1 and M3 were enriched for genes encoding postsynaptic complexes using data on 671 proteins in human neocortical PSD and 79 proteins related to NMDA receptor–activity-regulated cytoskeleton complexes (NMDAR-ARC) previously implicated in neurodevelopmental disease, memory and intelligence[5,17,21,22]. We found that genes comprising the PSD and NMDAR-ARC complexes were significantly overrepresented in M1 (Fisher's exact test (FET) $P = 5.4 \times 10^{-13}$, odds ratio (OR) = 2.10, 95% confidence interval (CI) 1.73–2.55 and $P = 2.6 \times 10^{-8}$, OR = 4.25, 95% CI 2.57–6.90, respectively) but not in M3 (**Fig. 1c**). However, manual annotation of gene function for M3 genes revealed that 58 of the 121 genes with a reported putative function had a biological activity potentially related to neural processes (**Supplementary Table 6**), suggesting M3 is also capturing previously unknown connectivity between genes that share related functions. Analysis of physical interactions between the protein products of genes in M1 and M3 using the InWeb database[23] revealed significant enrichment for direct protein-protein interactions for M1 (551 of 1,148 genes, P = 0.001) and M3 (17 of 150 genes, P = 0.02), providing further evidence to support the validity of these two coexpression modules.

In summary, these comparative genome-wide network analyses starting from human surgical hippocampus samples identify four modules (M1, M3, M11 and M19) that are cross-species–conserved and whose constituent genes are widely coexpressed across the human brain. Two of these modules (M1 and M3) are highlighted as having potential function related to neural activity.

Integrated cognitive GWAS data and gene network analysis

To determine the relationship between the four cross-species preserved coexpression modules (M1, M3, M11 and M19) and human

© 2015 Nature America, Inc. All rights reserved.

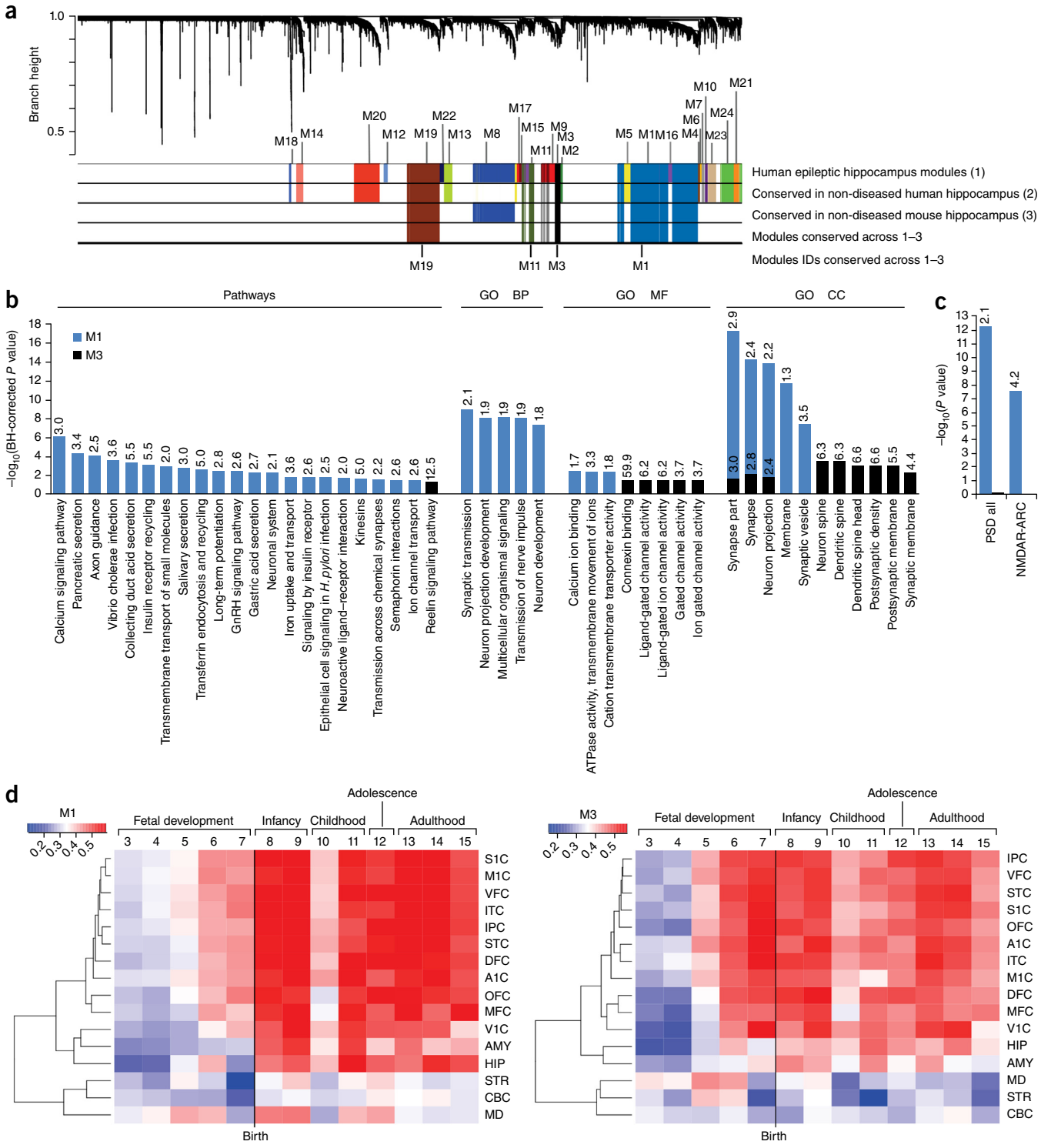

Figure 1 Gene coexpression network analysis. (**a**) Dendrogram showing clustering of coexpressed genes (modules) based on human surgical hippocampus samples. Top color bar: 24 modules (M1–M24) generated by unsupervised hierarchical clustering of the surgical hippocampal transcriptome; second color bar: 16 (of 24) modules whose gene coexpression relationships are significantly preserved in nondiseased postmortem human hippocampus; third color bar: 5 (of 24) human surgical hippocampus modules whose gene coexpression relationships are preserved in the healthy mouse hippocampus; bottom color bar: the 4 coexpression modules conserved across all three expression data sets (1–3). (**b**) KEGG and Pathway Commons (Pathways) and gene ontology (GO) enrichments for M1 (blue) and M3 (black). BP, biological process; MF, molecular function; CC, cellular component. For each functional category the ratio of enrichment is reported on top of each bar. (**c**) Enrichment of proteins comprising the postsynaptic density (PSD) and NMDAR-ARC complexes in M1 (blue) and M3 (black). ORs of enrichment are reported on top of each bar. (**d**) Heatmap of gradient of expression of modules M1 and M3 spanning fetal development to late adulthood and in topographically distinct cortical regions. A1C, auditory cortex; AMY, amygdala; CBC, cerebellar cortex; DFC, dorsolateral prefrontal cortex; HIP, hippocampus; IPC, posterior inferior parietal cortex; ITC, inferior temporal cortex; M1C, primary motor cortex; MD, mediodorsal nucleus of thalamus; MFC, medial prefrontal cortex; OFC, orbital prefrontal cortex; S1C, primary somatosensory cortex; STC, superior temporal cortex; STR, striatum; V1C, primary visual cortex; VFC, ventrolateral prefrontal cortex.

© 2015 Nature America, Inc. All rights reserved.

cognitive function we tested each module for enrichment of genetic association with four cognitive phenotypes (general fluid cognitive ability, processing speed, crystalized cognitive ability and verbal delayed recall) in two independent cohorts of cognitively healthy subjects. Our 'discovery' cohort consisted of genome-wide association study (GWAS) data relating to 6,732 (after quality control) cognitively healthy subjects participating in the 'Generation Scotland: Scottish Family Health Study' (GS:SFHS)[24]. The 'replication' cohort consisted of independent GWAS data relating to 1,003 (after quality control) cognitively healthy subjects participating in the Lothian Birth Cohort 1936 (LBC1936)[25]. Mean age at assessment was 55 years (s.d. = 11.35) in GS:SFHS and 69.6 years (s.d. = 0.8) in LBC1936. Descriptions of how we derived the cognitive phenotypes and GWAS analysis are available in Online Methods.

Table 1 Module enrichment for genetic association with cognitive abilities

		Discovery cohort GS:SFHS n = 6,732 subjects[a]			Replication cohort LBC1936 n = 1,003 subjects[a]		
Module	Phenotype	Genes[b]	Z score	P value[c] (FDR)	Genes[b]	Z score	P value[c]
M1	General fluid cognitive ability	983	**2.33**	**0.010 (5.3%)**	1,051	0.73	0.230
	Processing speed	983	**1.79**	**0.040 (8.9%)**	1,051	0.51	0.300
	Crystalized cognitive ability	983	**1.96**	**0.020 (6.4%)**	1,051	**1.70**	**0.045**
	Delayed recall	1,051	**2.14**	**0.016 (6.4%)**	1,046	**2.51**	**0.006**
M3	General fluid cognitive ability	135	**2.95**	**0.002 (2.4%)**	142	**2.66**	**0.004**
	Processing speed	135	**2.80**	**0.003 (2.4%)**	142	1.02	0.150
	Crystalized cognitive ability	135	**1.60**	**0.050 (8.9%)**	142	–0.10	0.540
	Delayed recall	142	**1.77**	**0.038 (8.9%)**	139	**2.56**	**0.005**
M11	General fluid cognitive ability	121	0.27	0.390 (52%)			
	Processing speed	121	–0.63	0.740 (78%)			
	Crystalized cognitive ability	121	1.62	0.050 (8.9%)	133	1.09	0.140
	Delayed recall	133	0.04	0.480 (59%)			
M19	General fluid cognitive ability	466	1.28	0.100 (16%)			
	Processing speed	466	–1.29	0.900 (90%)			
	Crystalized cognitive ability	466	–0.27	0.610 (69%)			
	Delayed recall	504	0.42	0.340 (49%)			

[a]Total number of participants after genotype quality control. [b]Genes in the module with ≥1 genotyped SNP within the transcription start and end positions of the gene (NCBI36, hg18). [c]P value for enrichment of association determined by 100,000 bootstrap samples.
Bold, enrichment of association $P < 0.05$; FDR was calculated to account for the number of modules and cognitive domains tested (16 tests); modules with FDR < 10% in the discovery cohort were taken forward for replication in LBC1936.

To test each module's association to the four cognitive phenotypes we first used versatile gene-based association study[26] (VEGAS) to account for the number of SNPs in each gene and the linkage disequilibrium (LD) between those SNPs followed by GWAS enrichment analysis using the Z-score enrichment method[27] (Online Methods). As a negative control, and to assess specificity of the GWAS-enrichments, each module was also tested against five large GWAS of clinical phenotypes with no known relationship to healthy cognitive performance (waist:hip ratio, fasting glucose homeostasis, glucose challenge homeostasis, systolic blood pressure and diastolic blood pressure; (**Supplementary Table 7**).

In the larger discovery cohort (GS:SFHS), we found nominal enrichment of association ($P < 0.05$) for M1 with general fluid cognitive ability, processing speed, crystalized cognitive ability and verbal delayed recall, and for M3 with general fluid cognitive ability, processing speed and verbal delayed recall (**Table 1**). Neither M1 nor M3 was enriched for association to any of the five noncognitive control phenotypes despite the substantial sample size and power of these GWAS studies (**Supplementary Table 7**). M11 and M19 were not significantly ($P < 0.05$) enriched for association with any cognitive phenotype. We adopted a false discovery rate (FDR) adjustment based on the number of modules and phenotypes tested in the discovery cohort GS:SFHS, and modules significantly enriched for association at FDR <10% were taken forward for replication in LBC1936. The strongest replicable enrichment of association was between M3 and general fluid cognitive ability (GS:SFHS $P = 0.002$, Z score = 2.95; LBC1936 $P = 0.004$, Z score = 2.66) (**Table 1**). In addition, we observed replicable enrichment of association between M3 and delayed recall (GS:SFHS $P = 0.038$, Z score = 1.77; LBC1936 $P = 0.005$, Z score = 2.56). For M1, we observed replicable enrichment of association with delayed recall (GS:SFHS $P = 0.016$, Z score = 2.14; LBC1936 $P = 0.006$, Z score = 2.51) and crystalized cognitive ability (GS:SFHS $P = 0.020$, Z score = 1.96; LBC1936 $P = 0.045$, Z score = 1.70).

These results suggest modules M1 and M3 are enriched for genes related to general cognitive ability including memory. We therefore further explored M1 and M3 by investigating their expression in different stages of human brain development following the method of Pletikos[28] and by undertaking a detailed analysis of brain region expression of M1 and M3 genes. Using data from Kang and colleagues[29] consisting of gene expression measurements from 11 topographically defined cortical areas from 53 human brains spanning 10 weeks post-conception (PCW) to 82 years of age (Online Methods), we observed a clear developmental gradient of expression of both M1 and M3 beginning in early mid-fetal development ($16 \leq PCW \leq 19$), maximal by birth and then persisting through all post-natal periods (**Fig. 1d**). Consistent with the coexpression analyses using UKBEC data (**Supplementary Table 4**), we observed that following birth M1 and M3 genes are highly expressed across the human cortex with the exception of striatum, mediodorsal nucleus of thalamus and cerebellar cortex. The developmentally regulated expression of M1 and M3 genes across diverse brain regions is consistent with the genetic evidence (**Table 1**) suggesting these modules play a broader role in human cognitive abilities beyond hippocampal memory.

The tightly regulated developmental trajectory of expression of M1 and M3 led us to explore their transcriptional control. Using the WebGestalt toolkit[30] to test for enrichment of transcription factor binding sites (TFBS) among M1 and M3 genes, we found M1 was highly enriched for NRSF/REST (repressor element 1-silencing transcription factor) targets (BH $P = 0.0006$), and this was confirmed using a set of previously published and experimentally derived targets of REST[31] (enrichment $P = 0.007$). For M3, the maximum TFBS enrichment was for SRY (sex determining region Y) transcription factor (BH $P = 0.01$). However, using publicly available data on sex-biased gene expression in the brain[29] we found no evidence of enrichment for male-specific genes in M3 (data not shown). In addition, we found no significant enrichment for experimentally derived REST targets in M3 ($P = 0.67$), suggesting different processes underlie the transcriptional regulation of M1 and M3 in the brain.

Burden of neurodevelopmental *de novo* mutations in gene networks

Extensive epidemiological and genetic evidence suggest that clinically distinct neurodevelopmental disorders could be thought of as reflecting different patterns of symptoms (or impairments) of a shared neurodevelopmental continuum[32]. The co-occurrence of clinical symptoms and diagnostic overlap between neuropsychiatric disorders has also meant that diseases such as epilepsy are increasingly considered within the neurodevelopmental spectrum[33]. Since cognitive impairment is a core component of many neurodevelopmental

© 2015 Nature America, Inc. All rights reserved.

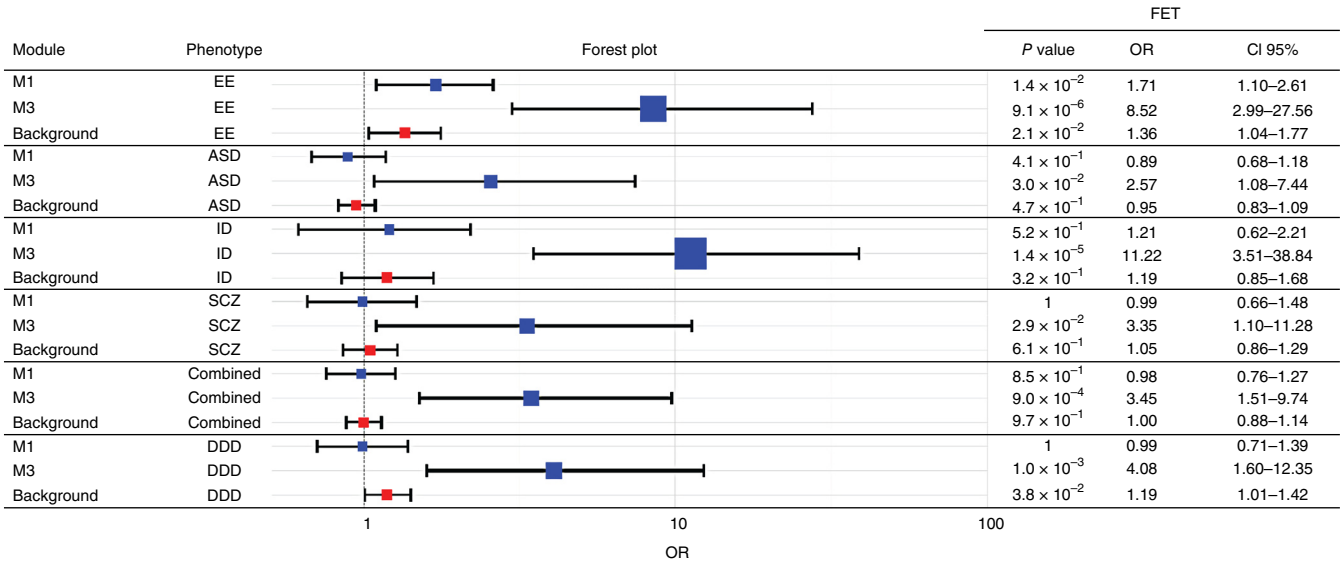

Figure 2 Enrichment of nsDNM from patients with neurodevelopmental disease. Statistical significance of overrepresentation of nsDNM in cases compared to controls is reported using Fisher's exact test for epileptic encephalopathy (EE, 356 trios), autism spectrum disorders (ASD, 4,186 trios), intellectual disability (ID, 192 trios), schizophrenia (SCZ, 1,004 trios) and across all four neurodevelopmental disorders consisting of EE, ID, ASD and SCZ (combined, 5,738 trios). The nsDNM of the DDD study (1,133 trios) were not combined with the other neurodevelopmental disorders as some of the patients of the DDD study had congenital abnormalities without neuropsychiatric features. *P* value, OR and 95% CI are reported for M1, M3 and all genes expressed in the human surgical hippocampus samples (background). In the forest plot, the magnitude of the ORs are represented by the area of the squares and the 95% CI by horizontal lines. Blue, modules; red, background.

disorders including schizophrenia[11], autism[12] and epilepsy[13], we set out to explore the relationship between the four cross-species conserved gene coexpression modules (and in particular M1 and M3) and susceptibility to neurodevelopmental disease.

To this aim, we first assessed if any of the modules were enriched for genes intolerant to functional mutation using the residual variation intolerance score (RVIS)[34]; genes considered to be intolerant to mutation according to their RVIS are more likely to be associated with developmental disease when mutated[34,35]. Using the individual RVIS for each gene in a module we calculated a module-level RVIS and compared the distribution of RVIS scores for each module to the distribution of intolerance scores from all hippocampus-expressed protein-coding genes outside of that module (Online Methods). Of the four cross-species conserved modules, three (M1, M3 and M11) were significantly enriched for intolerant genes (**Supplementary Table 8**), meaning that these modules contain an excess of genes intolerant to functional genetic variation relative to the genome-wide expectation. Given their cross-species preservation of coexpression, this finding suggests selective constraints on these modules in terms of both their coding sequence and transcriptional regulation.

We then investigated the relationship between the four cross-species conserved modules and neurodevelopmental disease by testing each module for enrichment of validated non-polymorphic *de novo* single nucleotide variant mutations (DNMs) identified in neurodevelopmental whole-exome sequencing studies that used similar sequencing technologies, coverage criteria and variant-calling methodology (Online Methods). The neurodevelopmental disease cohort consisted of 5,738 non-overlapping published parent-offspring trios across four disease phenotypes; autism spectrum disorder (n = 4,186 trios), schizophrenia (n = 1,004 trios), intellectual disability (n = 192 trios) and epileptic encephalopathy (n = 356 trios) (see Online Methods for cohort references). Additionally, we considered DNMs from an independent cohort of 1,133 trios with severe, previously undiagnosed developmental disease from the deciphering developmental disorders study[36,37]. As controls, we used 1,891 non-neurological control samples from seven published studies[38–44].

Then we tested each module's genetic relationship to disease using two statistical approaches. First, we compared rates of DNMs in each module relative to random expectation based on the collective consensus coding sequence (CCDS) of module genes. We calculated the expected number of DNMs for each gene set (that is, module) based on the length of CCDS sequence of genes in the set and the overall frequency of DNM in all CCDS genes. Then to estimate the enrichment we used the ratio between the observed number of DNMs in the gene set and the expected number based on this length model using binomial exact test (BET, two-tailed). Second, to accommodate for sequence context factors such as the inherent mutability of genes in a module, we adopted a FET (two-tailed) to empirically compare the rates of DNMs overlapping the CCDS real estate of a module in case cohorts and control cohorts. This approach also can identify modules comprising genes that are preferentially depleted of DNMs in healthy controls. For each module, we report DNM enrichments by both approaches and by considering three main classes of mutation: (i) predicted deleterious DNM (pdDNM) consisting of loss-of-function (nonsense and splice-site mutations) and predicted functional missense mutations, (ii) nonsynonymous DNM (nsDNM) consisting of all missense, nonsense and splice-site mutations and (iii) synonymous DNM (as a negative control). For completeness, we also report enrichments considering only loss-of-function (that is, nonsense and splice-site) mutations, although we expect limited power to detect significant enrichments given that single nucleotide DNMs in this class were relatively uncommon in the neurodevelopmental disease cohorts used here. Finally, to assess specificity of the module-level enrichment results, for each class of DNM detailed above, we calculated an enrichment of DNM among all genes significantly expressed in the human hippocampus (termed 'background' genes), taking the conservative route of including in this set of genes all the genes contributing to the individual coexpression modules.

© 2015 Nature America, Inc. All rights reserved.

Table 2 Genes in M3 impacted by neurodevelopmental-ascertained nonsynonymous *de novo* mutation

Gene symbol	Total nsDNM	Single nucleotide variant and predicted effect			Sift score	Polyphen score	Neurodevelopmental disease cohort			
SCN2A	20	2:166,245,137	A>T	SV			ASD			
		2:166,201,379	C>A	SG			ASD			
		2:166,210,819	G>T	SG			ASD			
		2:166,152,367	G>A	MS	0.11	0.025	ASD			
		2:166,152,578	A>G	MS	0	0.999	ASD			
		2:166,170,231	G>A	MS	0	0.999	ASD			
		2:166,201,312	G>A	MS	0	0.999	ASD			
		2:166,231,378	*T*>C	MS	0	1	ASD			
		2:166,201,311	C>T	MS	0	0.999	ASD			
		2:166,234,111	C>T	MS	0	0.996	ASD			
		2:166,234,116	A>G	MS	0	0.999		EE		
		2:166,198,975	G>A	MS	0	0.838		EE		
		2:166,201,311	C>T	MS	0	0.999			ID	
		2:166,231,415	G>A	SG					ID	
		2:166,187,838	A>G	SV						SCZ
		2:166,153,563	C>T	SG						DDD
		2:166,165,305	G>A	SV						DDD
		2:166,245,954	G>A	MS	0	0.997				DDD
		2:166,243,484	T>A	MS	0	0.972				DDD
		2:166,210,714	T>C	MS	0	0.719				DDD
GABRB3	7	15:27,017,557	C>T	MS	0.04	0.444	ASD			
		15:26,828,534	C>T	MS	0	0.584	ASD			
		15:26,866,594	T>C	MS	0.15	0.999		EE		
		15:26,806,254	T>C	MS	0	1		EE		
		15:26,866,564	C>T	MS	0	0.994		EE		
		15:26,828,484	T>C	MS	0	0.967		EE		
		15:26,806,242	A>G	MS	0	0.999				DDD
RYR2	7	1:237,870,440	C>A	MS	0.23	0.034	ASD			
		1:237,666,734	C>T	MS	0.02	0.947	ASD			
		1:237,868,631	C>T	SG				EE		
		1:237,995,907	G>A	MS	0	0.998			ID	
		1:237,982,492	G>T	MS	0	0.998				DDD
		1:237,982,471	A>G	MS	0	0.658				DDD
		1:237,693,752	G>A	MS	0.08	0.36				DDD
GNAO1	6	16:56,388,838	G>A	MS	0	0.316	ASD			
		16:56,385,380	A>C	MS	0	0.999		EE		
		16:56,385,396	T>C	MS	0	0.996		EE		
		16:56,370,728	G>A	MS	0.02	0.964				SCZ
		16:56,370,674	C>T	MS	0	1				DDD
		16:56,309,901	T>G	MS	0	0.799				DDD
TCF4	5	18:52,921,925	G>A	SG					ID	
		18:52,896,230	C>T	MS	0	1			ID	
		18:53,070,725	G>A	MS	0	0.942			ID	
		18:52,899,819	G>A	SG						DDD
		18:52,895,593	C>T	SV						DDD
GRIN2A	3	16:9,928,084	G>C	MS	0	0.921			ID	
		16:9,923,342	G>C	MS	0.01	0.999			ID	
		16:9,857,517	A>G	MS	0.01	0.816				SCZ
TCF20	2	22:42,564,699	G>A	MS	1	0			ID	
		22:42,575,645	G>A	SG						DDD
PPP6R2	2	22:50,857,408	C>T	MS	0.01	0.862	ASD			
		22:50,857,843	T>C	MS	0.01	0.898		EE		
NUAK1	2	12:106,461,269	G>A	SG			ASD			
		12:106,460,608	G>A	MS	0.02	0.997	ASD			
MYCBP2	2	13:77,700,568	A>G	MS	0.54	0.039	ASD			
		13:77,657,240	G>A	MS	0.14	0				DDD
KCNB1	2	20:47,990,976	G>A	MS	0	1		EE		
		20:47,990,924	T>G	MS	0	1				DDD
GNB5	2	15:52,427,874	T>C	MS	0	1	ASD			
		15:52,416,801	T>C	MS	0.38	0.68				SCZ
DLG2	2	11:83,497,765	G>C	MS	0	0.786	ASD			
		11:83,194,295	C>T	SV						SCZ
BRSK2	1	11:1,471,005	G>C	SV			ASD			
CAMK1D	1	10:12,595,343	C>A	MS	0.06	0.003	ASD			
CERS6	1	2:169,417,831	A>G	MS	0.11	0.229	ASD			
CNST	1	1:246,754,937	G>A	MS	0.07	0.09	ASD			
DENND5B	1	12:31,613,279	G>C	MS	0.08	0.305	ASD			
DUSP3	1	17:41,847,180	G>A	MS	0	0.921	ASD			
GLTSCR1L	1	6:42,796,946	C>G	MS	0	1	ASD			

(continued)

© 2015 Nature America, Inc. All rights reserved.

Table 2 (Continued)

Gene symbol	Total nsDNM	Single nucleotide variant and predicted effect			Sift score	Polyphen score	Neurodevelopmental disease cohort				
GRIA2	1	4:158,254,055	C>T	SG			ASD				
GSK3B	1	3:119,582,433	G>T	MS	0.01	0.521	ASD				
HNRNPR	1	1:23,637,156	G>A	MS	0	0	ASD				
KLHL28	1	14:45,400,640	A>G	MS	0.99	0.324	ASD				
MAP1B	1	5:71,491,094	G>T	MS	0.33	0	ASD				
MCM4	1	8:48,883,381	G>C	MS	0.04	0.363	ASD				
NT5C3A	1	7:33,055,445	A>G	MS	0.14	0.546	ASD				
PAPD5	1	16:50,263,085	G>A	MS	0.09	0.027	ASD				
PIAS1	1	15:68,378,807	G>A	MS	0.16	1	ASD				
PUM1	1	1:31,437,728	G>A	MS	0	0.999	ASD				
UPF3A	1	13:115,057,116	G>A	MS	0	1	ASD				
GABRB1	1	4:47,405,630	T>C	MS	0	0.998		EE			
SGK223	1	8:8,234,597	C>A	MS	0.01	0.36		EE			
HIVEP3	1	1:42,047,669	G>A	SG					SCZ		
PCDHAC2	1	5:140,346,499	G>T	SG					SCZ		
SSBP3	1	1:54,870,560	G>A	SG					SCZ		
TAF13	1	1:109,607,282	G>A	SG					SCZ		
TNRC6C	1	17:76,083,048	C>G	MS	0.01	0.808			SCZ		
PHACTR1	1	6:12,933,928	G>A	MS	0.02	0					DDD
PLEKHB2	1	2:131,884,360	G>A	SV							DDD
ROBO2	1	3:77,637,907	C>T	MS	0.18	0.784					DDD
SPIN1	1	9:91,083,440	A>G	MS	0	1					DDD
USP14	1	18:203,143	C>T	SG							DDD

M3 genes reported with nsDNM identified in heterogeneous neurodevelopmental phenotypes. We detail the number and kind of nsDNM and for each single nucleotide variant, Sift and Polyphen2 scores were calculated using the Ensembl SNP Effect Predictor tool[49]. ASD, autism spectrum disorder; ID, intellectual disability; EE, epilepsy; SCZ, schizophrenia; SV, splice variant; SG, stop gain; MS, missense.

We observed that module M3 was strongly and specifically enriched for genes that, when mutated, are associated with intellectual disability and epileptic encephalopathy, and that this enrichment holds true for both pdDNM (intellectual disability BET $P = 6.6 \times 10^{-5}$, FET $P = 3.1 \times 10^{-4}$, OR = 10.29, 95% CI 2.56–48.91; epileptic encephalopathy BET $P = 1.9 \times 10^{-6}$, FET $P = 7.1 \times 10^{-5}$, OR = 9.1, 95% CI 2.64–39.47) and all nsDNM (intellectual disability BET $P = 3.3 \times 10^{-5}$, FET $P = 1.4 \times 10^{-5}$, OR = 11.22, 95% CI 3.51–38.84; epileptic encephalopathy BET $P = 1.3 \times 10^{-5}$, FET $P = 9.1 \times 10^{-6}$, OR = 8.52, 95% CI 2.99–27.56) (**Fig. 2** and **Supplementary Table 9**). These enrichments remained significant after adjustment for the number of modules and phenotypes tested. M1 was not significantly enriched for any neurodevelopmental disease above the background (**Fig. 2**). There was no enrichment in M3 of disease-ascertained synonymous DNM for neither intellectual disability (BET $P = 0.251$, FET $P = 0.239$) nor epileptic encephalopathy (BET $P = 0.576$, FET $P = 0.522$), or any other neurodevelopmental phenotype (**Supplementary Table 9**).

For autism spectrum disorder and schizophrenia, there was a trend toward enrichment of disease-ascertained DNM in M3, but estimates of the 95% CI of the OR overlapped with those from background genes (**Fig. 2**). However, when combining all 5,738 trios with neurodevelopmental disease (that is, intellectual disability + epileptic encephalopathy + autism spectrum disorder + schizophrenia) we observed significant enrichment of nsDNM in M3 above background (BET $P = 3.54 \times 10^{-6}$, FET $P = 9.0 \times 10^{-4}$, OR = 3.54, 95% CI 1.51–9.74) (**Fig. 2**), suggesting M3 is enriched for genes impacted by DNM associated with neurodevelopmental disease broadly, and with intellectual disability and epileptic encephalopathy in particular. Consistent with this interpretation, M3 was also significantly enriched for nsDNM ascertained from unselected developmental phenotypes from the independent DDD study[36,37] (BET $P = 2.2 \times 10^{-3}$, FET $P = 1.0 \times 10^{-3}$, OR = 4.08, 95% CI 1.60–12.35) (**Fig. 2** and **Supplementary Table 9**).

In total, almost a third of genes in M3 (43 of 150) were impacted by one or more nsDNM across the five disease cohorts considered here (intellectual disability, epileptic encephalopathy, autism spectrum disorder, schizophrenia and DDD). These 43 genes and their corresponding mutation (with functional consequence) and disease phenotype are shown in **Table 2** and **Figure 3**. Among the 43 genes in M3 impacted by nsDNM several genes including *SCN2A*, *GABRB3*, *GNAO1, TCF4*, *GRIN2A* and *UPF3A* are known to be implicated in neurodevelopmental disease. Thus, starting from an unsupervised gene network perspective, M3 revealed previously unappreciated coexpression between genes for heterogeneous neurodevelopmental disorders in the developed human brain.

The finding that M3 is highly enriched for genes that confer risk for neurodevelopmental disease when mutated led us to explore the relationship between M3 and neuropsychiatric disease using GWAS data relating to the Psychiatric Genomics Consortium traits attention deficit–hyperactivity disorder, bipolar disorder, major depressive disorder and schizophrenia[45] as well as GWAS data relating to common forms of epilepsy from the International League Against Epilepsy (ILAE) Consortium on Complex Epilepsies[46] and those from a risk and age of onset of Alzheimer's disease[47]. We tested the enrichment in M3 of association to each phenotype as previously described (Online Methods). After Bonferroni correction for multiple testing, the only significant association was between M3 and schizophrenia (enrichment $P = 0.003$, Z score = 2.76) (**Supplementary Table 10**). The corresponding enrichment statistics for schizophrenia trio–ascertained DNM were as follows: pdDNM BET $P = 2.14 \times 10^{-3}$, FET $P = 0.013$, OR = 4.52, 95% CI 1.25–20.27 and nsDNM BET $P = 0.08$, FET $P = 0.029$, OR = 3.35, 95% CI 1.1–11.28. This suggested that M3 may be enriched for genes in which both common and rare variants contribute risk for schizophrenia.

DISCUSSION

Using a stepwise procedure we prioritized gene networks whose gene coexpression relationships were significantly reproducible across brain regions and species to facilitate the identification of functionally conserved and replicable networks. We demonstrated replicable

© 2015 Nature America, Inc. All rights reserved.

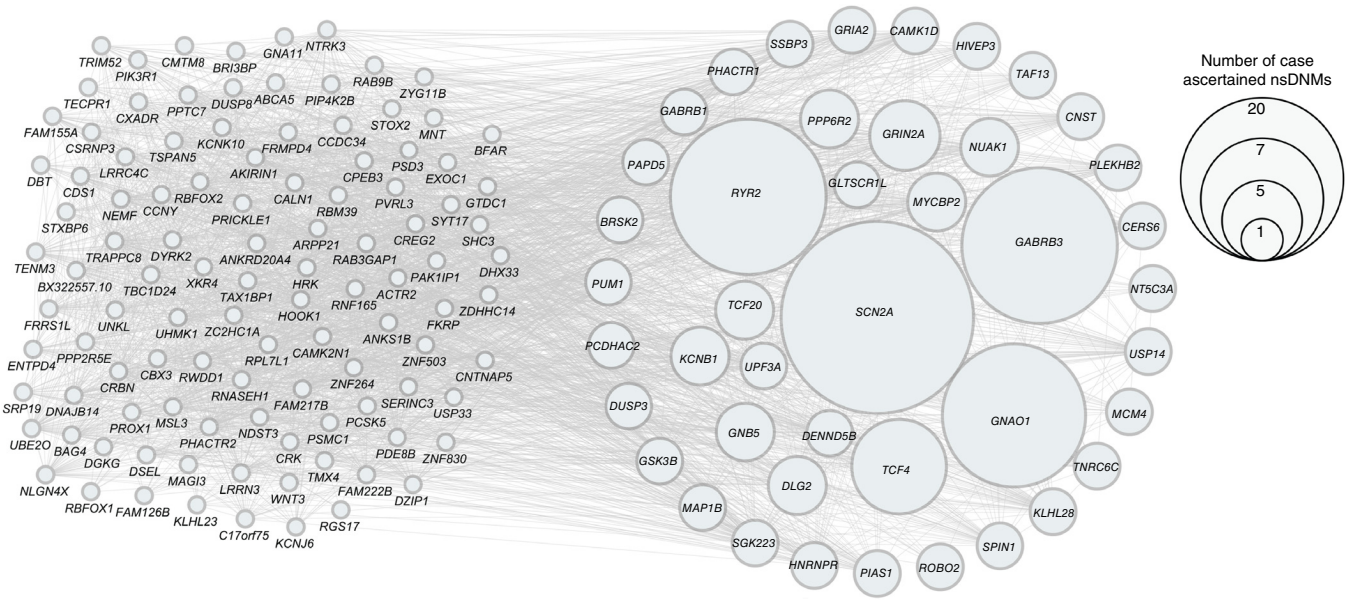

Figure 3 Graphical representation of the M3 coexpression network and its relationship to neurodevelopmental disease. Genes in M3 impacted by single nucleotide variant nsDNM from neurodevelopmental disease cases are drawn separately in a circle (right). The area of each node is proportional to the number of nsDNM for that gene across the full cohort of 6,871 parent-offspring trios (Online Methods). Individual nsDNM, their predicted affect and corresponding neurodevelopmental disease phenotypes are detailed in **Table 2**.

association between two of these coexpression networks (M1 and M3) and healthy human cognitive abilities. As M1 is functionally enriched for genes involved in synaptic processes, these findings provide systems-level evidence for a relationship between long-term potentiation and postsynaptic processes and human cognition, as previously suggested by an analysis of known postsynaptic signaling complexes[5]. In contrast to the functional specialization of M1, M3 is relatively poorly annotated for known functional categories or canonical pathways and our study revealed previously unappreciated coexpression relationships between genes influencing cognitive abilities. The finding that M1 and M3 influence cognitive abilities generally (as opposed to influencing specific cognitive domains such as memory) is in agreement with the evidence from twin and genome-wide complex trait analysis demonstrating high genetic correlation between diverse cognitive and learning abilities[9,10,48]. The widespread expression and coexpression of M1 and M3 genes across the human cortex, and their tight developmental regulation, is also consistent with these modules playing a role across cognitive domains.

By analyzing *de novo* mutations reported in whole-exome sequencing studies of neurodevelopmental disease parent-offspring trio cohorts, we found that rare genetic risk variants for neurodevelopmental disease also converged on module M3. Almost one-third of genes in M3 were impacted by one or more nonsynonymous DNM ascertained from neurodevelopmental disease cases. Among the individual genes in M3 mutated in two or more cases, most were associated with more than one neurodevelopmental phenotype (**Table 2**). These results reveal a convergence of genetic risk variants contributing to healthy human cognitive abilities and neurodevelopmental disease on a common set of genes under tight developmental regulation and widely coexpressed in the human cortex. Nonspecific (or pleotropic) effects of pathogenic mutations have recently emerged as a key theme among neurodevelopmental disease genes[35]. Here we provided empirical evidence to suggest this pleiotropy also extends to healthy cognitive function, although the underlying mechanisms for mutational nonspecificity remain unknown.

One observation from our study is the extent to which the expression of M1 and M3 genes is temporally specified. After birth, expression of M1 and M3 genes appeared remarkably stable over time, consistent with an enduring role for these genes in cognitive function throughout life. This is in keeping with the finding of the modules' association with cognition in two independent cohorts that differ in their age at assessment (**Table 1**). Whereas studies have suggested that sequence variation in genes that are developmentally regulated can be related to susceptibility to neurodevelopmental disease[42,43], here we showed that genes under tight developmental regulation and later coexpressed in the developed human brain are also related to this class of disorder as well as healthy cognitive processes. These observations provide a starting point for the identification of gene-regulatory factors that influence cognition and neurodevelopmental disease.

Our analyses integrating DNMs with gene-regulatory networks revealed that M3 was associated most strongly with intellectual disability and epileptic encephalopathy and to a lesser extent with neurodevelopmental disease in general. This is consistent with the hypothesis that genetic variation affecting quantitative variation in cognitive abilities overlaps with that underlying related monogenic phenotypes. However, when considering common risk variants (that is, SNPs) for neuropsychiatric disease, we observed an association between M3 and schizophrenia but not with common forms of epilepsy. Potential explanations for the lack of GWAS enrichment of association between M3 and common epilepsy include different gene contributions to severe childhood epileptic encephalopathy arising from rare *de novo* mutations compared to the (mostly) adult epilepsies considered in the ILAE study[46], and/or insufficient power to detect common variant associations using the ILAE GWAS (which, despite consisting of only 8,696 epilepsy cases and 26,157 controls is the largest epilepsy GWAS yet undertaken). Further studies will be required to clarify the specific contribution of M3 genes to disease risk across the allelic spectrum, and to elucidate the role of both rare and common sequence variants in the complex inheritance of childhood and adult epilepsy.

© 2015 Nature America, Inc. All rights reserved.

In conclusion, starting from an unsupervised analysis of gene expression variation in the hippocampus and across the brain, we identified two cross-species conserved gene coexpression networks (M1 and M3) associated with healthy human cognitive abilities, and we identified one of these (M3) as a convergent gene network for both cognition and neurodevelopmental disease. Our experimental framework, which integrates gene-network analysis with genetic susceptibility data, can be applied generally to any human behavioral or cognitive phenotype for which relevant genetic data (GWAS, whole-exome sequencing, etc.) are available. We have therefore made our human hippocampal gene network and data accessible via an integrated web tool (Neurodevelopmental disease Brain Integrated Gene Networks, http://www.nbign.co.uk). This framework and underlying data may help to tackle the fundamental challenge of understanding how genetic risk variants for neurodevelopmental disease and related cognitive phenotypes exert their effects in the developed human brain.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We acknowledge funding from UK Imperial National Institute for Health Research Biomedical Research Centre (M.R.J.), the UK Medical Research Council (M.R.J., E.P. and D.S.), The National Genome Research Network (NGFNplus: EMINet, grant 01GS08122; A.J.B.), EuroEpinomics (A.J.B.), UCB Pharma (M.R.J. and E.P.) and the Singapore Ministry of Health (E.P.). We thank the Lothian Birth Cohort 1936 research team for data collection and collation. The Lothian Birth Cohort 1936 is supported by Age UK (Disconnected Mind project). The work at The University of Edinburgh was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). We acknowledge funding from the Biotechnology and Biological Sciences Research Council and MRC. Generation Scotland received core funding from the Chief Scientist Office of the Scottish Government Health Directorate CZD/16/6 and the Scottish Funding Council HR03006. Genotyping of the GS:SFHS samples was carried out by staff at the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the UK's Medical Research Council. We thank all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. A.D.-D. was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme.

AUTHOR CONTRIBUTIONS

M.R.J. and E.P. conceived, designed and coordinated the study. K.S., P.S. and A.D.-D. carried out network and comparative genomics analyses. S.P. and O.J.L.R. carried out RVIS analysis. S.R.L. carried out GWAS-enrichment analysis with support from A.K., D.S. and W.D.H., A.D.-D. carried out enrichment analysis of neuropsychiatric *de novo* mutations with support from S.P. and K.S. M.R., A.V., M.F., L.B., T.R. and A.M.-M. provided technical support and contributed to methodology. M.M., P.F., B.D. and R.M.K. contributed mouse RNA-seq data. C.H., A.G. and A.J.B. contributed human hippocampus expression data and clinical information. I.J.D., W.D.H., G.D., S.E.H., C.H., D.J.P., B.H.S., S.P., L.J.H., J.M.S. and D.C.L. contributed GWAS data for the cognitive phenotypes. O.J.L.R. designed and implemented the web server. M.R.J. and E.P. wrote and revised the manuscript with input from K.S., S.R.L., A.D.-D., P.S., W.D.H. and I.J.D.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Deary, I.J., Johnson, W. & Houlihan, L.M. Genetic foundations of human intelligence. *Hum. Genet.* **126**, 215–232 (2009).
2. Davies, G. *et al.* Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Mol. Psychiatry* **16**, 996–1005 (2011).
3. Plomin, R., Haworth, C.M., Meaburn, E.L., Price, T.S. & Davis, O.S. & Wellcome Trust Case Control Consortium 2. Common DNA markers can account for more than half of the genetic influence on cognitive abilities. *Psychol. Sci.* **24**, 562–568 (2013).
4. Davies, G. *et al.* Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949). *Mol. Psychiatry* **20**, 183–192 (2015).
5. Hill, W.D. *et al.* Human cognitive ability is influenced by genetic variation in components of postsynaptic signalling complexes assembled by NMDA receptors and MAGUK proteins. *Transl. Psychiatry* **4**, e341 (2014).
6. Christoforou, A. *et al.* GWAS-based pathway analysis differentiates between fluid and crystallized intelligence. *Genes Brain Behav.* **13**, 663–674 (2014).
7. Carroll, J. *Human cognitive abilities: A survey of factor-analytic studies* (Cambridge University Press, 1993).
8. Plomin, R. & Deary, I.J. Genetics and intelligence differences: five special findings. *Mol. Psychiatry* **20**, 98–108 (2015).
9. Trzaskowski, M. *et al.* DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. *Behav. Genet.* **43**, 267–273 (2013).
10. Trzaskowski, M., Shakeshaft, N.G. & Plomin, R. Intelligence indexes generalist genes for cognitive abilities. *Intelligence* **41**, 560–565 (2013).
11. Kahn, R.S. & Keefe, R.S.E. Schizophrenia is a cognitive illness: time for a change in focus. *JAMA Psychiatry* **70**, 1107–1112 (2013).
12. Doherty, J.L. & Owen, M.J. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome Med.* **6**, 29 (2014).
13. Helmstaedter, C. & Witt, J.-A. Clinical neuropsychology in epilepsy: theoretical and practical issues. *Handb. Clin. Neurol.* **107**, 437–459 (2012).
14. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
15. Li, J.Z. *et al.* Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. *Proc. Natl. Acad. Sci. USA* **110**, 9950–9955 (2013).
16. Nithianantharajah, J. *et al.* Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nat. Neurosci.* **16**, 16–24 (2013).
17. Bayés, A. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* **14**, 19–21 (2011).
18. Bayés, A. *et al.* Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. *PLoS One* **7**, e46683 (2012).
19. Bliss, T.V.P. & Collingridge, G.L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
20. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418–1428 (2014).
21. Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179–184 (2014).
22. Kirov, G. *et al. De novo* CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* **17**, 142–153 (2012).
23. Rossin, E.J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* **7**, e1001273 (2011).
24. Smith, B.H. *et al.* Cohort profile: generation Scotland: Scottish family health study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700 (2013).
25. Deary, I.J., Gow, A.J., Pattie, A. & Starr, J.M. Cohort profile: the Lothian birth cohorts of 1921 and 1936. *Int. J. Epidemiol.* **41**, 1576–1584 (2012).
26. Liu, J.Z. *et al.*; AMFS Investigators. A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.* **87**, 139–145 (2010).
27. Nam, D., Kim, J., Kim, S.-Y. & Kim, S. GSA-SNP: a general approach for gene set analysis of polymorphisms. *Nucleic Acids Res.* **38**, W749–54 (2010).
28. Pletikos, M. *et al.* Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* **81**, 321–332 (2014).
29. Kang, H.J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
30. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
31. Satoh, J., Kawana, N. & Yamamoto, Y. ChIP-Seq data mining: remarkable differences in NRSF/REST target genes between human ESC and ESC-derived neurons. *Bioinform. Biol. Insights* **7**, 357–368 (2013).
32. Moreno-De-Luca, A. *et al.* Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. *Lancet Neurol.* **12**, 406–414 (2013).
33. Johnson, M.R. & Shorvon, S.D. Heredity in epilepsy: neurodevelopment, comorbidity, and the neurological trait. *Epilepsy Behav.* **22**, 421–427 (2011).
34. Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S. & Goldstein, D.B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* **9**, e1003709 (2013).
35. Zhu, X., Need, A.C., Petrovski, S. & Goldstein, D.B. One gene, many neuropsychiatric disorders: lessons from Mendelian diseases. *Nat. Neurosci.* **17**, 773–781 (2014).

© 2015 Nature America, Inc. All rights reserved.

36. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. *Nature* **519**, 223–228 (2014).
37. Wright, C.F. *et al.* Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet* **385**, 1305–1314 (2015).
38. Sanders, S.J. *et al. De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
39. Iossifov, I. *et al.* De novo gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
40. O'Roak, B.J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
41. Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
42. Gulsuner, S. *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
43. Xu, B. *et al. De novo* gene mutations highlight patterns of genetic and neural complexity in schizophrenia. *Nat. Genet.* **44**, 1365–1369 (2012).
44. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–221 (2014).
45. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
46. International League Against Epilepsy Consortium on Complex Epilepsies. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **13**, 893–903 (2014).
47. Li, H. *et al.* Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease. *Arch. Neurol.* **65**, 45–53 (2008).
48. Rimfeld, K., Kovas, Y., Dale, P.S. & Plomin, R. Pleiotropy across academic subjects at the end of compulsory education. *Sci. Rep.* **5**, 11713 (2015).
49. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).

© 2015 Nature America, Inc. All rights reserved.

ONLINE METHODS

Human surgical hippocampus gene expression data generation. Ethics approval for the study was given by the NHS Tayside committee on research ethics (reference 05/S1401/89). Genome-wide gene expression data were generated from 122 snap-frozen whole hippocampus samples surgically removed from patients who had undergone *en bloc* amygdalahippocampectomy for mesial TLE as previously described[50]. Informed consent was obtained from all patients and the study was approved by statutory Ethics Committees and Institutional Review Boards. Clinical data recorded for each patient included date of birth, gender, handedness, age at epilepsy onset, laterality of TLE, operation date, age at operation, preoperative seizure frequency, antiepileptic drug therapy at the time of surgery and neuropathology. Genome-wide gene expression was assayed as previously described[50]. Expression data were normalized by quantile normalization with background subtraction. Prior to network analysis, the data were filtered as follows: first, non-expressed probes were removed using the internal *P* values of detection provided by Illumina BeadArray Reader. Probes were retained if they passed 95% confidence threshold in at least 30% of the samples. Second, probes were removed if their sequences did not map uniquely to the reference genome or if the target regions contained at least one known SNP, as accessed by ReMOAT[51]. Third, the coefficient of variation (s.d./mean) in gene expression was used to remove the 5% of probes showing the lowest variation in gene expression in the TLE cohort. These filtering steps defined a final data set of 11,837 probes, representing 9,616 protein-coding unique genes (Ensembl version 72), which were then used for network analysis and as the 'background' gene set for enrichment analyses.

Gene coexpression network analysis of human surgical hippocampus samples. Before inferring gene coexpression networks, we used principal component (PC) analysis to calculate summary variables describing the variation in the microarray expression of the 11,837 probes and estimate the potential effects of clinical covariates on global gene expression variability. The first three PCs explained the following fraction of variation in gene expression: PC1, 25%; PC2, 15%; and PC3, 8%, with other components explaining <5% of the variability in gene expression. We assessed the impact of clinical covariates age, gender, epilepsy severity, anti-epileptic drug (AED) load and hippocampal 'pathology type' (that is, Ammons Horn Sclerosis alone or in association with reactive astrogliosis and/or neuronal loss) on global gene expression by calculating univariate correlations between PC1–PC3 and each clinical covariate. After Bonferroni correction for multiple testing, 'pathology type' was the only covariate to show a significant effect on gene expression in epileptic hippocampus ($P = 1.1 \times 10^{-4}$, $R^2 = 0.24$ on PC1 of global gene expression). PC1 summarized 25% of the global variation in gene expression, and since 'pathology type' explained only a limited fraction of this variability ($R^2 = 0.24$), this was considered the only relevant covariate. This is in keeping with our previous analyses where we observed no significant effects from clinical covariates (apart from epilepsy pathology as shown here)[50]. Gene expression levels were therefore adjusted to remove the effect of 'pathology type' by fitting linear models on gene expression and accounting for pathology using the *lm* function in R. The residuals from the linear model were then used in the coexpression network analysis.

Genes were then grouped into modules using weighted gene coexpression network analysis (WGCNA)[14] on the set of 11,837 probes in 122 human hippocampus samples. WGCNA builds undirected coexpression networks where the nodes of the network correspond to genes and edges between genes are determined by the pairwise correlations between the genes' expression levels. To avoid outlier bias, Tukey's biweight method[52] was used to compute robust pairwise correlations of gene expression. The strength of relationships between probes is defined as the adjacency matrix, which is calculated by applying a power function (connection strength = $|\text{correlation}|^{\beta}$) on the biweight correlation matrix. The power function reduces the strength of weak correlations while preserving connection strength of highly correlated probes. Higher values of β increase this effect and increase specificity of gene interactions, whereas a lower β increases sensitivity. For the network analysis in the surgical hippocampus and for the comparative networks analyses in different data sets (see below), the beta was chosen to optimize the scale free property and the sparsity of connections between genes in each data set. Then, the adjacency matrix was used to calculate the topological overlap matrix (TOM), which measures the number of neighbors that a pair of probes have in common, relative to the rest of the probes. Average hierarchical clustering was used to group genes based on the dissimilarity of gene connectivity, defined as 1 – TOM. The dynamic cut-tree method[53] was used to cut the dendrogram on a branch-by-branch basis to produce coexpression clusters.

Reproducibility of TLE hippocampal modules in control (nondiseased) human and mouse hippocampus samples. Several independent hippocampal gene-expression data sets were used to establish module reproducibility. To establish reproducibility of modules in nondiseased human hippocampus we used human postmortem hippocampus microarray expression data from 63 healthy postmortem human brains publicly available from Pritzker Neuropsychiatric Disorders Research Consortium (http://www.pritzkerneuropsych.org/?page_id=1196). To investigate module conservation across species, we generated mRNA-sequencing (RNA-seq) expression data from 100 healthy mouse hippocampi as follows: total RNA was isolated from snap frozen hippocampi from 100 healthy (Crl: NMRI(Han)-FR) mice. Mouse hippocampus samples were ascertained strictly in accordance with statutory ethical guidelines/regulations. cDNA and sample preparation for RNA sequencing followed manufacturer protocol (TruSeq RNA kit, Illumina). Samples were sequenced on an Illumina HiSeq 2000 sequencer as paired-end 75-nucleotide reads. Raw reads were mapped to the reference mouse genome (mm10) using TopHat[54] version 2.0.8. Read counts per gene were calculated for each sample using HTseq version 0.5.3 (http://www-huber.embl.de/users/anders/HTSeq) and subsequently normalized across all the samples using trimmed mean of M value (TMM) approach[55]. For each replication gene expression data set we checked whether human surgical modules had higher connectivity in the replication data sets than expected by chance. For each replication gene expression data set, the adjacency matrix was calculated using biweight correlations and the β value was chosen to optimize scale free property of the networks. The adjacency matrix was used to calculate topological overlap matrix (TOM) using WGCNA. For each of the 24 networks (M1–M24) detected in the 122 TLE subjects, empirical *P* values for the significance of the coexpression relationships were calculated by comparing the average topological overlap for network genes in the replication data sets (human or mouse) to the average connectivity of 10,000 randomly sampled networks[56]. The randomly sampled networks had the same size of the networks detected in the TLE patients (M1–M24).

Module coexpression across brain regions. To determine whether coexpression of genes in modules M1 and M3 are preserved across topographically distinct cortical regions, we analyzed genome-wide gene expression data from four brain regions (cerebellum, temporal cortex, occipital cortex and frontal cortex) using 102 postmortem human brains from the UK Brain Expression Consortium (UKBEC) (GSE60862)[20]. Each brain region was treated as an independent data set. Raw expression profiles from the Affymetrix Human Exon 1.0 ST Array were processed to transcript-level expression with Affymetrix Power Tools (APT) (http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx) using probe logarithmic intensity error (plier) normalisation[57] with probe G+C content correction. Only the most reliable 'core' set of probes was used to generate transcript-level expression profiles as defined by Affymetrix. Exons were considered as 'expressed' if more than 50% of the samples had detection above background *P* values below 0.01, as calculated using APT. Gene-level expression was obtained by taking the median of the expression values of multiple exons mapping to the same gene. Expression profiles from each brain region were analyzed as independent data sets and were processed separately. This means that some genes were considered as 'expressed' in some brain regions and not in others (number of unique Ensembl genes expressed per brain region as follows: frontal cortex, 14,800; temporal cortex, 14,777; cerebellum, 15,162; and occipital cortex, 14,815).

Gene expression profiles were corrected for measured clinical covariates: age, gender, postmortem interval, cause of death and the source of samples (that is, brain-bank identifier). The data were also adjusted for any potential batch effects using probabilistic estimation of expression residuals (PEER)[58]. PEER uses factor analysis to infer hidden determinants that explain large proportions of variability in the data. This approach allows expression data to be corrected for the effects of measured covariates such as age and sex as well as other potential sources of bias such as batch effects, environmental influences, sample history and other unknown factors[58]. Comparative network analysis was undertaken as previously (above) using the default network dissimilarity measure in WGCNA based on the TOM[14], and empirical *P* values for the reproducibility of networks calculated by comparing the average topological overlap for module genes to the average connectivity of 10,000 randomly sampled networks.

© 2015 Nature America, Inc. All rights reserved.

Spatiotemporal analysis of module expression. To determine the spatiotemporal expression dynamics of modules, we used quantile-normalized gene-level expression values (log_2 transformed) from GSE60862 (ref. 29). These transcriptome data were generated using Affymetrix Human Exon 1.0 ST array analysis of 16 brain regions comprising the cerebellar cortex, mediodorsal nucleus of the thalamus, striatum, amygdala, hippocampus and 11 areas of the neocortex. The data were generated from 1,263 samples collected from 53 clinically unremarkable postmortem human brains, spanning embryonic development to late adulthood (from 10 weeks after conception to 82 years of age, which corresponded to periods 3–15, as previously designated)[29]. The log_2-transformed gene expression data follows a bimodal distribution contributed by low (likely nonfunctional) and high expressed genes[59]. We used the expectation maximization (EM) algorithm to model gene expression levels as mixture of normal distributions and identify the underlying distributions of low and high expressed genes. Only the genes, with mean of log_2-transformed expression values over the 95% percentile of distribution of low-expressed genes (here > 5.61) were considered for further analysis (n = 8,704). The EM algorithm was implemented using normalMixEM function from the mixtools R package. Spatiotemporal dynamics of coexpression modules M1 and M3 across 16 brain regions and 13 developmental time points were illustrated as a heatmap (**Fig. 1d**), as previously described[28]. Module expression for each region and developmental time point was calculated by averaging the scaled expression across all genes in a module. The resultant heatmap graphs illustrate the changes in expression of genes of a coexpression module across brain development and cortical regions.

Functional enrichment analysis of networks. Co-expression modules were functionally annotated using WebGestalt[30] with terms of Kyoto Encyclopedia of Genes and Genomes (KEGG)[60], 'Pathway Commons' and GO[61] terms. For each data set, we conservatively used all hippocampus-expressed genes (including those that contributed to the individual coexpression modules) as the background in the functional enrichment analyses. For each gene set (module), the ratio of enrichment (r), $r = k/k_e$ is calculated as the number of genes in the module (k) over the expected value (k_e) of genes in the reference as determined by WebGestalt[30].

Assessment of overrepresentation of synaptic genes in modules. Enrichment of postsynaptic genes in the modules was assessed by hypergeometric test (two-tailed). The list of genes encoding ARCs and NMDAR was sourced from a published study (80 genes; see supplementary table 9 in ref. 17). The postsynaptic density (PSD) gene list used was the consensus human PSD genes (supplementary table 2 in ref. 22) that had an Ensembl gene identifier (745 out of 748 genes). PSD and ARC-NMDAR–encoding genes were tested for overrepresentation in the modules using the list of brain expressed genes as the background gene set (n = 9,616 genes).

Genome-wide association study of cognitive phenotypes. We analyzed four cognitive phenotypes in two independent community-based cohorts, discovery cohort GS:SFHS[24] and replication LBC1936 (ref. 25). The same four cognitive phenotypes were analyzed in both LBC1936 and GS:SFHS; these were general fluid cognitive ability, crystallized ability, memory (delayed recall) and information processing speed. For LBC1936, the general fluid factor was derived using the six nonverbal tests from the Wechsler Adult Intelligence scale III^{uk} (ref. 62): matrix reasoning, digit span backward, symbol search, digit symbol coding, block design and letter-number sequencing. The raw scores from each of these tests were used in a PC analysis where the first unrotated PC was extracted using regression analysis. Next, each participant's score on this PC was linearly regressed against age, sex and the first four multidimensional scaling components (to control for population stratification) used as predictor variables. The residuals from this model were then used in subsequent analyses. For crystallized ability, the National Adult Reading Test (NART)[63] was used. For memory and information processing speed, the delayed memory section from the logical memory section and the digit symbol section of the WAIS-III^{UK} (ref. 62) were used, respectively. For each of these single tests, the effect of age, sex and population stratification was controlled for using regression approaches (as described above), and the standardized residuals from the regression model were used in the downstream analyses.

In GS:SFHS, for general fluid cognitive ability, the raw scores from the digit symbol substitution task[62], the delayed and immediate sections of the logical memory test[64], verbal fluency[65], and the Mill Hill vocabulary scale[66] were subjected to a PC analysis where the first unrotated PC was extracted using regression. This PC was then used as the dependent variable in a linear regression model with age, sex and the first six principal components (to control for population stratification) used as predictor variables. The residuals from this model were then extracted and carried forward for subsequent analyses. Whereas different tests were used in the construction of the general factor in GS:SFHS and in LBC1936, correlations between general factors constructed from different test batteries is high[67,68]. As with LBC1936, for crystallized ability, memory and information-processing speed only a single test was used. For crystallized ability this was the Mill Hill vocabulary scale[66], for memory the delayed section of the logical memory test[64], and for information processing speed the digit symbol substitution task[62] was used. As for general cognitive ability, the effects of age, sex and population stratification were controlled for using regression approaches. Using these cognitive phenotypes we then undertook a standard GWAS of cognitive phenotypes in GS:SFHS and LBC1936 separately, as follows.

GWAS in GS:SFHS. GS:SFHS was composed of families recruited from the population of Scotland between 2006 and 2011. 7,953 unrelated individuals aged between 35 and 65 years were recruited from Glasgow, Tayside, Ayrshire, Arran and the northeast of Scotland. 95% of subjects were contacted through their general practitioner, with the remaining 5% contacted through word of mouth. These individuals' family members were also recruited, yielding a sample size of 24,084 with an age range of 18–100 years of age. A description of GS:SFHS is available in refs. 24 and 69. DNA from blood (or saliva from clinical and postal participants) was extracted following informed consent from 10,000 Caucasian participants who were born in the UK. DNA was processed and stored using the standard operating procedures at the Wellcome Trust Clinical Research Facility Genetics Core in Edinburgh[70]. Genotyping was undertaken on Illumina HumanOmniExpressExome-8 v1.0 DNA Analysis BeadChip. In order to ensure comparability between the LBC1936 cohort and GS:SFHS, the UCSC Batch Coordinate Conversion (liftOver) (https://genome.ucsc.edu/cgi-bin/hgLiftOver) tool was used to convert the hg 19 build of GS to hg18. To control for the effect of shared environment subjects who were related to another participant were removed (estimated kinship > 0.025) leaving a total of 6,816 unrelated participants. After quality control 594,756 SNPs with a minor allele frequency > 0.01 were included in the analysis. Cognitive phenotypes were derived as described above and the effects of age, sex and population stratification controlled for as described previously. The standardized residuals were used for subsequent single-SNP GWAS which was performed using PLINK[71]. Single SNP P values of association to individual cognitive scores were then used in the GWAS enrichment analysis (see below).

GWAS in LBC1936. The LBC1936 cohort consisted of 1,091 cognitively healthy individuals (548 men and 543 women) assessed on cognitive and medical traits at a mean age 69.6 years (s.d. = 0.8). Informed consent was obtained from all subjects. All subjects were of Caucasian descent and almost all lived independently in the Lothian region (Edinburgh city and surrounding area) of Scotland. Genotyping using the Illumina 610-Quadv1 array was performed at the Wellcome Trust Clinical Research Facility, Edinburgh. Quality control measures were as follows: individuals were excluded from the study based on unresolved gender discrepancy, relatedness (so that no pair remained with estimated kinship >0.025), SNP call rate (≤0.95) and evidence of non-Caucasian descent. A total of 542,050 SNPs meeting the following conditions were included in the analysis: call rate ≥ 0.98, minor allele frequency ≥ 0.01 and Hardy-Weinberg equilibrium test with $P \geq 0.001$. After quality control, we included 1,003 participants in the association analysis. Derivation of the cognitive phenotypes is described above, followed by correction for age, sex and population stratification. The standardized residuals were used for genotype-phenotype analyses by PLINK[71]. Single SNP P values of association to individual cognitive scores were then used in the GWAS enrichment analysis (see below).

GWAS enrichment analysis. To test for enrichment of genetic association in a gene set (that is, coexpression module) we used VEGAS[26] to generate a gene-based association statistic (P value) controlled for the number of SNPs in each gene and the LD between those SNPs. In all analyses gene-based P values were calculated using VEGAS and the top 10% option with 100,000 iterations and a gene window consisting of the transcriptional start and stop position of each gene. For both GS:SFHS and LBC1936, the genotype data from the GWAS participants was used to control for LD (rather than the default HapMap population) as this is expected to provide a more accurate estimate of the LD structure, which can be specific of the population cohort analyzed. For the other GWAS for which raw genotype data were not available (the Psychiatric Genomics Consortium traits, International League Against Epilepsy Consortium on Complex Epilepsies (**Supplementary**

© 2015 Nature America, Inc. All rights reserved.

doi:10.1038/nn.4205

Table 9), and the noncognitive control GWAS data sets of waist:hip ratio, fasting glucose homeostasis, glucose challenge homeostasis, systolic blood pressure and diastolic blood pressure (**Supplementary Table 6**)) the default HapMap population was used to control for LD in the VEGAS analysis. The GWAS-enrichment statistic was calculated for a given module from the gene-based association *P* values (from VEGAS) using the *Z* test–based bootstrapping method[27] (one-sided) where, for each network, 100,000 random gene sets of same size as the network were sampled from the list of all hippocampus expressed genes (n = 9,616). *P* values of enrichment for the discovery cohort were considered significant if they passed FDR correction for the number of modules tested, as indicated in each case.

Using RVIS to assess the genic intolerance properties of specific modules. The extent of human-specific genic constraint was estimated for each of the 24 coexpression modules by using the genic protein-coding intolerance scores (RVIS)[34]. RVIS was only calculated for protein-coding genes that had at least one protein-coding transcript that was publically approved among the CCDS release 9 database[72], and that had ≥70% of their CCDS real estate adequately covered among the population database adopted in their original manuscript (Exome Sequencing Project (ESP)-6500)[34]. This resulted in scores for 16,956 assessable CCDS release 9 genes, thus all RVIS comparisons are restricted to these 16,956 'assessable' genes. We found that 89.4% of the genes across all modules had an assessable RVIS score. To determine whether a module was enriched for genes that are relatively more intolerant to functional variation than the rest of the genes expressed in the human hippocampus (n = 8,414 with CCDS), a two-tailed Mann-Whitney *U* test was used to compare the distribution of genic RVIS scores for each module to the distribution from the rest of the hippocampus-expressed protein-coding genes outside of the module (module-level RVIS results are reported in **Supplementary Table 7**).

Assessing the relationship between coexpression modules and neurodevelopmental disorder ascertained *de novo* mutations. We collated published DNM data sets to determine whether any relationships exists between coexpression modules and the DNMs reported in neurodevelopmental trio whole-exome sequencing studies. Collectively, the neurodevelopmental disease cohort consisted of 5,738 non-overlapping published parent-offspring trios across four disease phenotypes; autism spectrum disorder (ASD, n = 4,186)[44,73], schizophrenia (SCZ, n = 1,004)[21,42,43,74,75] intellectual disability (ID, n = 192)[41,76,77] and epileptic encephalopathy (EE, n = 356)[78,79]. Additionally, we considered DNMs from an independent cohort of 1,133 trios with severe, previously undiagnosed developmental disease from the DDD study[36,37]. For controls, we used 1,891 nonneurological control samples from seven published studies[38–44].

Each module's genetic relationship to disease was tested using two approaches. First, we compared rates of DNMs in each module compared to random expectation based on the CCDS of module genes. In the absence of individual trio data across the different studies, we cannot determine the effectively sequenced real estate for each gene so we took the conservative route by assuming each gene has 100% of its CDDS sequence covered across all trios, appreciating that some genes will not have been adequately covered due to reasons such as capture kit specifications or low coverage. Thus, the expected numbers of DNM for each gene set is calculated based on the length of CCDS sequence of genes in the set and the overall frequency of DNM in all CCDS genes. Then to estimate the enrichment we used the ratio between the observed number of DNM in the gene set and the expected number based on this length model using BET (two-tailed). Second, to accommodate for sequence context factors such as the inherent mutability of genes in a module, we adopted a FET (two-tailed) to empirically compare the rates of DNMs overlapping the CCDS real estate of a module in case- and control cohorts. This approach is also able to capture modules comprised of genes that are preferentially depleted of DNMs in healthy control cohorts. For each module, we report single nucleotide variant DNM enrichments by both approaches and by considering three main classes of DNM: (i) pdDNM consisting of loss-of-function (that is, nonsense and splice-site mutations) plus with missense mutations with SIFT[80] score ≤0.05 and Polyphen2 (ref. 81) score ≥ 0.5, (ii) nsDNM consisting of all missense, nonsense and splice-site single nucleotide variant mutations and (iii) synonymous DNM (as a negative control). Polyphen2 and SIFT scores were obtained using the Variant Effect Predictor Ensembl tool[49]. For completeness, we also calculated enrichments considering only loss-of-function (nonsense and splice-site) mutations but because DNMs in this class were relatively infrequent, when considered alone, we expect limited power to detect significant enrichments. Finally, to establish specificity of the module-level results, we calculated enrichment of DNM for each class of DNM among all genes significantly expressed in the human hippocampus (background genes, n = 9,616) taking the conservative route of including among this set of genes all genes contributing to the individual modules. **Supplementary Code** is provided for the major functions used in the analytical workflow.

A **Supplementary Methods Checklist** is available.

50. Johnson, M.R. *et al.* Systems genetics identifies Sestrin 3 as a regulator of a proconvulsant gene network in human epileptic hippocampus. *Nat. Commun.* **6**, 6031 (2015).
51. Barbosa-Morais, N.L. *et al.* A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res.* **38**, e17 (2010).
52. Hardin, J., Mitani, A., Hicks, L. & VanKoten, B. A robust measure of correlation between two genes on a microarray. *BMC Bioinformatics* **8**, 220 (2007).
53. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008).
54. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
55. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
56. North, B.V., Curtis, D. & Sham, P.C. A note on the calculation of empirical P values from Monte Carlo procedures. *Am. J. Hum. Genet.* **71**, 439–441 (2002).
57. Therneau, T.M. & Ballman, K.V. What does PLIER really do? *Cancer Inform.* **6**, 423–431 (2008).
58. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507 (2012).
59. Hebenstreit, D. *et al.* RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol. Syst. Biol.* **7**, 497 (2011).
60. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, D277–D280 (2004).
61. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
62. Wechsler, D. *Wechsler Adult Intelligence Scale - third edition.* (London: The Psychological Corporation, 1998).
63. Nelson, H.E. & Willison, J. *National Adult Reading Test (NART) Test Manual.* (Windsor, NFER-Nelson, 1991).
64. Wechsler, D. *Wechsler Memory Scale III UK.* (London: The Psychological Corporation, 1998).
65. Lezak, M.D., Howieson, D.B., Bigler, E.D. & Tranel, D. *Neuropsychological Assessment* (Oxford University Press, 2004).
66. Raven, J.C., Court, J.H. & Raven, J. *Manual for Raven's Progressive Matrices and Vocabulary Scales* (H.K. Lewis, 1977).
67. Johnson, W., Bouchard, T.J., Krueger, R.F., McGue, M. & Gottesman, I.I. Just one g: consistent results from three test batteries. *Intelligence* **32**, 95–107 (2004).
68. Johnson, W., Nijenhuis, J.T. & Bouchard, T.J. Still just 1 g: consistent results from five test batteries. *Intelligence* **36**, 81–95 (2008).
69. Smith, B.H. *et al.* Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med. Genet.* **7**, 74 (2006).
70. Kerr, S.M. *et al.* Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. *BMC Med. Genet.* **14**, 38 (2013).
71. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
72. Pruitt, K.D. *et al.* The consensus coding sequence (CCDS) project: identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* **19**, 1316–1323 (2009).
73. De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
74. Girard, S.L. *et al.* Increased exonic *de novo* mutation rate in individuals with schizophrenia. *Nat. Genet.* **43**, 860–863 (2011).
75. Girard, S.L. *et al.* Mutation burden of rare variants in schizophrenia candidate genes. *PLoS One* **10**, e0128988 (2015).
76. de Ligt, J. *et al.* Diagnostic exome sequencing in persons with severe intellectual disability. *N. Engl. J. Med.* **367**, 1921–1929 (2012).
77. Hamdan, F.F. *et al. De novo* mutations in moderate or severe intellectual disability. *PLoS Genet.* **10**, e1004772 (2014).
78. Allen, A.S. *et al.* De novo mutations in epileptic encephalopathies. *Nature* **501**, 217–221 (2013).
79. EuroEPINOMICS-RES. Consortium, Epilepsy Phenome/Genome Project & Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am. J. Hum. Genet.* **95**, 360–370 (2014).
80. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
81. Adzhubei, I.A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).

© 2015 Nature America, Inc. All rights reserved.