LETTER

doi:10.1038/nature10110

Transcriptomic analysis of autistic brain reveals convergent molecular pathology

Irina Voineagu[1], Xinchen Wang[2], Patrick Johnston[3], Jennifer K. Lowe[1], Yuan Tian[1], Steve Horvath[4], Jonathan Mill[3], Rita M. Cantor[4], Benjamin J. Blencowe[2] & Daniel H. Geschwind[1,4]

Autism spectrum disorder (ASD) is a common, highly heritable neurodevelopmental condition characterized by marked genetic heterogeneity[1–3]. Thus, a fundamental question is whether autism represents an aetiologically heterogeneous disorder in which the myriad genetic or environmental risk factors perturb common underlying molecular pathways in the brain[4]. Here, we demonstrate consistent differences in transcriptome organization between autistic and normal brain by gene co-expression network analysis. Remarkably, regional patterns of gene expression that typically distinguish frontal and temporal cortex are significantly attenuated in the ASD brain, suggesting abnormalities in cortical patterning. We further identify discrete modules of co-expressed genes associated with autism: a neuronal module enriched for known autism susceptibility genes, including the neuronal specific splicing factor *A2BP1* (also known as *FOX1*), and a module enriched for immune genes and glial markers. Using high-throughput RNA sequencing we demonstrate dysregulated splicing of *A2BP1*-dependent alternative exons in the ASD brain. Moreover, using a published autism genome-wide association study (GWAS) data set, we show that the neuronal module is enriched for genetically associated variants, providing independent support for the causal involvement of these genes in autism. In contrast, the immune-glial module showed no enrichment for autism GWAS signals, indicating a non-genetic aetiology for this process. Collectively, our results provide strong evidence for convergent molecular abnormalities in ASD, and implicate transcriptional and splicing dysregulation as underlying mechanisms of neuronal dysfunction in this disorder.

We analysed post-mortem brain tissue samples from 19 autism cases and 17 controls from the Autism Tissue Project and the Harvard brain bank (Supplementary Table 1) using Illumina microarrays. For each individual, we profiled three regions previously implicated in autism[5]: superior temporal gyrus (STG, also known as Brodmann's area (BA) 41/42), prefrontal cortex (BA9) and cerebellar vermis. After filtering for high-quality array data (Methods), we retained 58 cortex samples (29 autism, 29 controls) and 21 cerebellum samples (11 autism, 10 controls) for further analysis (see Methods for detailed sample description). We identified 444 genes showing significant expression changes in autism cortex samples (DS1, Fig. 1b), and only 2 genes were differentially expressed between the autism and control groups in cerebellum (Methods), indicating that gene expression changes associated with autism were more pronounced in the cerebral cortex, which became the focus of further analysis (Supplementary Table 2). There was no significant difference in age, post mortem interval (PMI), or RNA integrity numbers (RIN) between autism and control cortex samples (Supplementary Fig. 1, Methods).

Supervised hierarchical clustering based on the top 200 differentially expressed genes showed distinct clustering of the majority of autism cortex samples (Fig. 1a), including one case that was simultaneously found to have a 15q duplication (Methods, Supplementary Table 1), which is known to cause 1% of ASD[6]. Cortex samples from ten of the cases coalesced in a single tight-clustering branch of the dendrogram. Clustering was independent of age, sex, RIN, PMI, co-morbidity of seizures, or medication (Fig. 1a and Supplementary Fig. 2c). It is interesting to note that the two ASD cases that cluster with controls (Fig. 1a) are the least severe cases, as assessed by global functioning (Supplementary Table 12). We observed a highly significant overlap between differentially expressed genes in frontal and temporal cortex ($P = 10^{-44}$; Fig. 1b), supporting the robustness of the data and indicating that the autism-specific expression changes are consistent across these cortical areas. We also validated a cross section of the differentially expressed genes by quantitative reverse transcription PCR (RT–PCR) and confirmed microarray-predicted changes in 83% of the genes tested (Methods, Supplementary Fig. 2b). Gene ontology enrichment analysis (Methods) showed that the 209 genes downregulated in autistic cortex were enriched for gene ontology categories related to synaptic function, whereas the upregulated genes ($N = 235$) showed enrichment for gene ontology categories implicated in immune and inflammatory response (Supplementary Table 3).

To test whether these findings were replicable, and to further validate the results in an independent data set, we obtained tissue from an additional frontal cortex region (BA44/45) from nine ASD cases and five controls (DS2; Supplementary Table 4). Three of the cases and all of the controls used for validation were independent from our initial cohort. Ninety-seven genes were differentially expressed in BA44/45 in DS2, and 81 of these were also differentially expressed in our initial cohort ($P = 1.2 \times 10^{-93}$, hypergeometric test; Fig. 1b, c). Remarkably, the direction of expression differences between autism and controls was the same as in the initial cohort for all but 2 of the 81 overlapping differentially expressed probes. Hierarchical clustering of DS2 samples based on either the top 200 genes differentially expressed in the initial cohort or the 81 overlapping genes showed distinct separation of cases from controls (Supplementary Fig. 6). In addition, comparison of these differentially expressed results with another, smaller study of the STG in ASD[7], revealed significant consistency at the level of differentially expressed genes, including downregulation of *DLX1* and *AHI1* (Supplementary Table 5). Thus, differential expression analysis produced robust and highly reproducible results, warranting further refined analysis.

We next applied weighted-gene co-expression network analysis (WGCNA)[8,9] to integrate the expression differences observed between autistic and control cerebral cortex into a higher order, systems level context. We first asked whether there are global differences in the organization of the brain transcriptome between autistic and control brain by constructing separate co-expression networks for the autism and control groups (Methods). The control brain network showed high similarity with the previously described human brain co-expression networks (Supplementary Table 7), consistent with the existence of

[1]Program in Neurogenetics and Neurobehavioral Genetics, Department of Neurology and Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, California 90095-1769, USA. [2]Banting and Best Department of Medical Research, Donnelly Centre, University of Toronto, Toronto, Ontario M5G 1L6, Canada. [3]Institute of Psychiatry, King's College London, London SE5 8AF, UK. [4]Department of Human Genetics, University of California Los Angeles, Los Angeles, California 90095, USA.

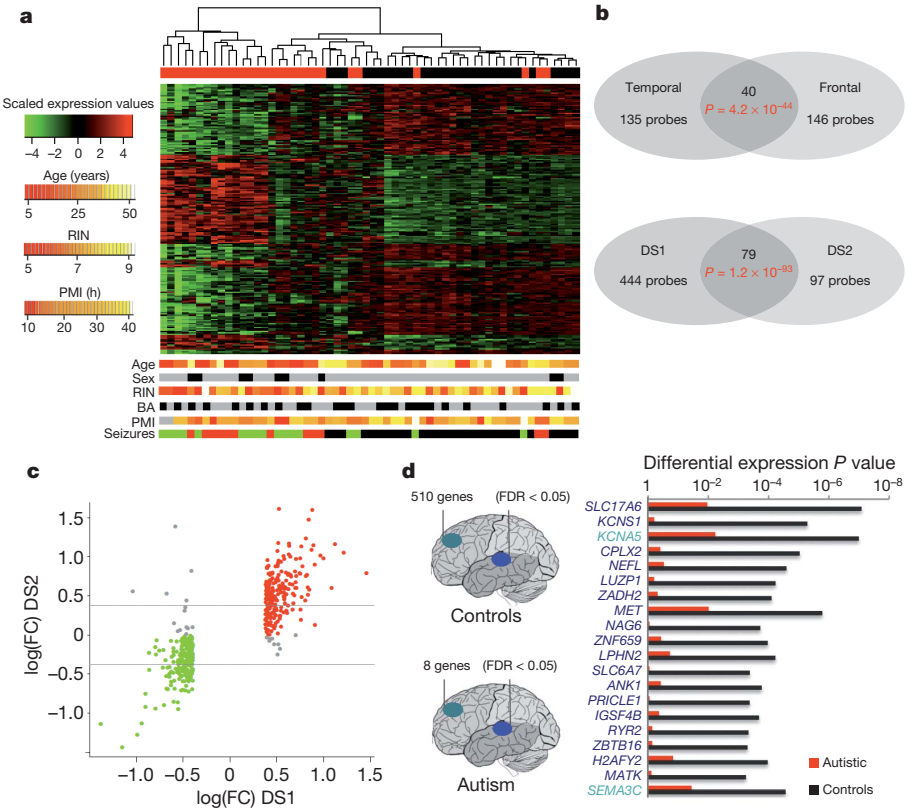

Figure 1 | Gene expression changes in autism cerebral cortex **a**, Heat map of top 200 genes differentially expressed between autism and control cortex samples. Scaled expression values are colour-coded according to the legend on the left. The dendrogram depicts hierarchical clustering based on the top 200 differentially expressed genes. The top bar (A/C) indicates the disease status: red, autism; black, control. The bottom bars show additional variables for each sample: sex (grey, male; black, female), brain area (black, temporal; grey, frontal), co-morbidity of seizures (green, autism case with seizure disorder; red, autism case without seizure disorder; black, control), age, RNA integrity number (RIN) and post mortem interval (PMI). BA, Brodmann's area. The corresponding scale for quantitative variables is shown on the left. **b**, Top, Venn diagram depicting the overlap between genes differentially expressed in frontal and temporal cortex. Bottom, Venn diagram describing the overlap between genes differentially expressed in the initial cohort (DS1) and the replication cohort (DS2). Differential expression in the initial cohort was assessed at an FDR < 0.05 and fold change >1.3. The statistical criteria were relaxed to $P < 0.05$ for the replication data set because it involved fewer samples. **c**, Expression fold changes for all genes differentially expressed in the initial cohort are plotted on the x-axis against the fold changes for the same genes in the replication cohort on the y-axis. Green, genes downregulated in the autism group in both data sets; red, genes upregulated in the autism group in both data sets; grey, genes with opposite direction of variation in the two data sets. Horizontal lines show fold change threshold for significance. **d**, Diagram depicting the number of genes showing significant expression differences between frontal and temporal cortex in control samples (top) and autism samples (bottom) at FDR < 0.05 (left). The top 20 genes differentially expressed between frontal and temporal cortex in control samples (right). All of the genes shown are also differentially expressed between frontal and temporal cortex in fetal midgestation brain[10], but show no significant expression differences between frontal and temporal cortex in autism. The horizontal bars depict P values for differential expression between frontal and temporal cortex in the autism and control groups.

robust modules of co-expressed genes related to specific cell types and biological functions[8]. Similarly, the majority (87%) of the autism modules showed significant overlap with the previously described human brain modules (Supplementary Table 6), indicating that many features reflecting the general organization of the autism brain transcriptome are consistent with that of the normal human brain.

The expression levels of each module were summarized by the first principal component (the module eigengene), and were used to assess whether modules are related to clinical phenotypes or other experimental variables, such as brain region. Two of the control module eigengenes (cM6, cM13) showed significant differences ($P < 0.05$) between the two cortical regions as expected, whereas none of the ASD modules showed any differences between frontal and temporal cortex. This led us to explore the hypothesis that the normal molecular distinctions between the two cortical regions tested were altered in ASD compared with controls. Remarkably, whereas 174 genes were differentially expressed between control BA9 and BA41 (false discovery rate (FDR) $< 1\%$), none of the genes were differentially expressed in the same regional comparison among the ASD cases. This was not simply an issue of statistical thresholds, as relaxing the statistical criteria for differential expression to an FDR of 5% identified over 500 differentially expressed genes in controls, and only 8 in ASD brains, confirming the large difference observed in regional cortical differential gene expression between ASD cases and controls (Fig. 1d, Methods). Analysis of differential expression from a data set[10] of gene expression in developing fetal human brain showed a highly significant ($P = 5.8 \times 10^{-9}$) overlap of differentially expressed genes with those found in controls in this study, independently confirming that these genes differentiate normal temporal and frontal lobes. We evaluated the homogeneity of gene expression variance across the autism and control groups using Bartlett's test (Methods) which indicated that increased variance was not the major factor responsible for the striking difference in regional gene expression between ASD and controls (Supplementary Fig. 7 and Supplementary Data).

These data suggest that typical regional differences, many of which are observed during fetal development[10], are attenuated in frontal and temporal lobe in autism brain, pointing to abnormal developmental patterning as a potential pathophysiological driver in ASD. This is especially interesting in light of a recent anatomical study of five cases with adult autism which demonstrated a reduction in typical ultrastructural differences between three frontal cortical regions in autism[11]. Together, these independent studies provide both molecular and

structural evidence suggesting a relative diminution of cortical regional identity in autism.

To identify discrete groups of co-expressed genes showing transcriptional differences between autism and controls, we constructed a co-expression network using the entire data set, composed of both autism and control samples (Methods). As previously shown for complex diseases[12,13] co-expression networks allow analysis of gene expression variation related to multiple disease-related and genetic traits. We assessed module eigengene relationship to autism disease status, age, gender, cause of death, co-morbidity of seizures, family history of psychiatric disease, and medication, providing a complementary assessment of these potential confounders to that performed in the standard differential expression analysis (Supplementary Table 9).

The comparison between autism and control groups revealed two network modules whose eigengenes were highly correlated with disease status, and not any of the potential confounding variables (Supplementary Table 9). We found that the top module (M12) showed highly significant enrichment for neuronal markers (Supplementary Table 9), and high overlap with two neuronal modules previously identified as part of the human brain transcriptional network[8]: a *PVALB*+ interneuron module and a module of genes involved in synaptic function. The M12 eigengene was under-expressed in autism cases, indicating that genes in this module were downregulated in the autistic brain (Fig. 2). Consistent with the pathways identified to be down-regulated in autism by differential expression analysis (Supplementary Table 3), the functional enrichment of M12 included the gene ontology categories involved in synaptic function, vesicular transport and neuronal projection.

Remarkably, unlike differentially expressed genes, M12 showed significant overrepresentation of known autism susceptibility genes[2] (Supplementary Table 10; $P = 6.1 \times 10^{-4}$), including *CADPS2*, *AHI1*, *CNTNAP2*, and *SLC25A12*, supporting the increased power of the network-based approach to identify disease-relevant transcriptional changes. A further advantage of network analysis over standard analysis of differential expression is that it allows one to infer the functional relevance of genes based on their network position[9]. The hubs of M12, that is, the genes with the highest rank of M12 membership[8], were *A2BP1*, *APBA2*, *SCAMP5*, *CNTNAP1*, *KLC2*, and *CHRM1* (Supplementary Data). The first three of these genes have previously been implicated in autism[14–16], whereas the fourth is a homologue of

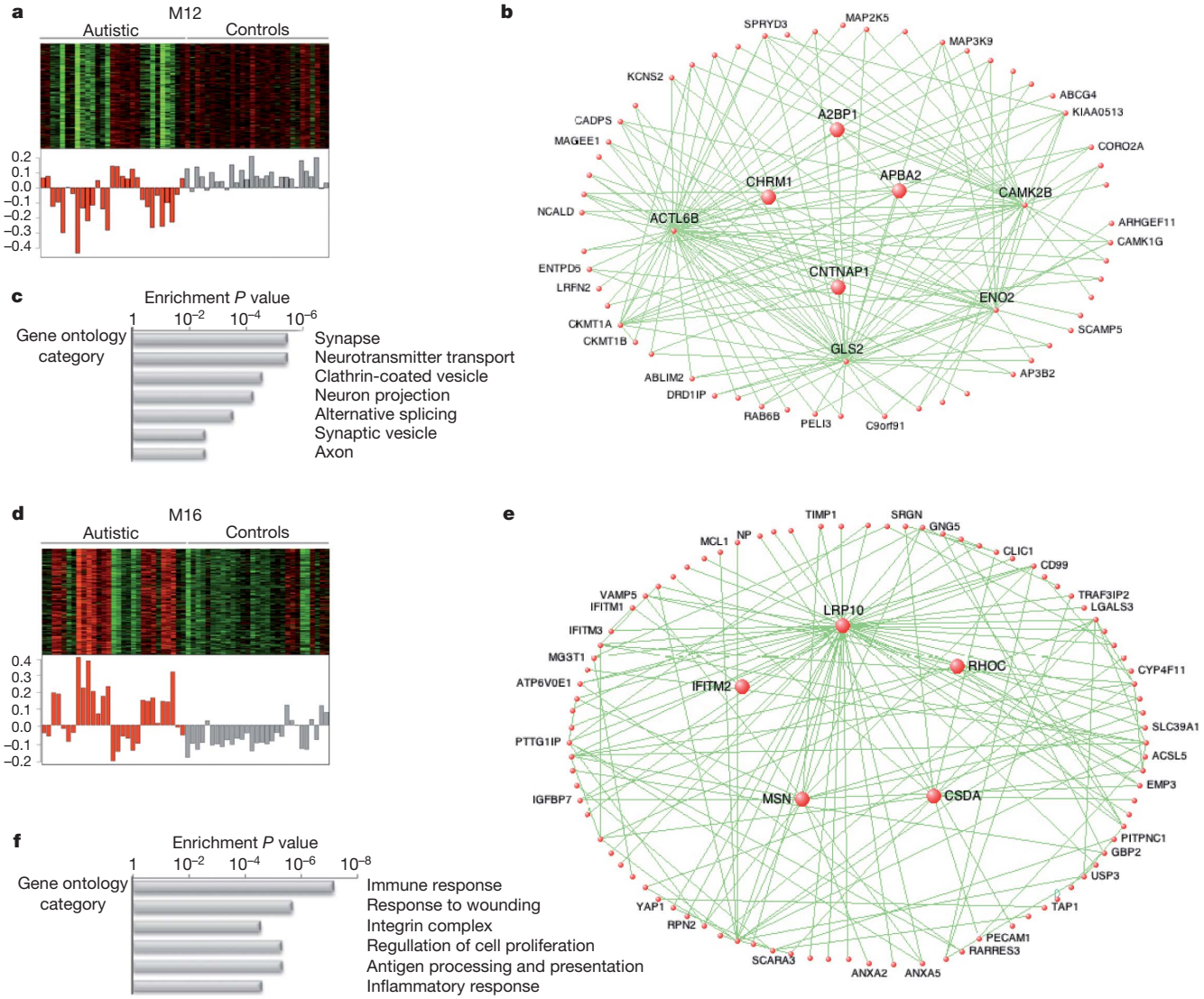

Figure 2 | Gene co-expression modules associated with autism **a**, **d**, Heat map of genes belonging to the co-expression module (top). Corresponding module eigengene values (*y*-axis) across samples (*x*-axis) (bottom). Red, autism; grey, controls. **b**, **e**, Visualization of the M12 and M16 modules, respectively. The top 150 connections are shown for each module. Genes with the highest correlation with the module eigengene value (that is, intramodular hubs) are shown in larger size. **c**, **f**, Relevant gene ontology categories enriched in the M12 and M16 modules.

the autism susceptibility gene *CNTNAP2* (ref. 17). We highlight the group of genes most strongly connected to the known ASD genes (Supplementary Fig. 5) and emphasize the downregulation of several interneuron markers, such as *DLX1* and *PVALB*, as candidates for future genetic and pathologic investigations.

The second module of co-expressed genes highly related to autism disease status, M16, was enriched for astrocyte markers and markers of activated microglia (Supplementary Table 9), as well as for genes belonging to immune and inflammatory gene ontology categories (Fig. 2). This module, which was upregulated in ASD brain, showed significant similarity to two modules identified in previous studies of normal human brain[8]: an astrocyte module and a microglial module. Consistent with this functional annotation, two of the hubs of the M16 module were known astrocyte markers (*ADFP*, also known as *PLIN2*, and *IFITM2*).

One of the hubs of the M12 module was *A2BP1*, a neural- and muscle-specific alternative splicing regulator[18] and the only splicing factor previously implicated in ASD[16]. Because *A2BP1* was downregulated in several ASD cases (Supplementary Fig. 8), this observation provided a unique opportunity to identify potential disease-relevant *A2BP1* targets. Whereas *A2BP1*-regulated alternative exons have been predicted genome-wide[19], few genes have been experimentally validated as *A2BP1* targets[20]. To identify potential *A2BP1*-dependent differential splicing events in ASD brain, we performed high-throughput RNA sequencing (RNA-Seq) on three autism samples with significant downregulation of *A2BP1* (average fold change by quantitative RT–PCR = 5.9) and three control samples with average *A2BP1* levels. We identified 212 significant alternative splicing events (Supplementary Data). Among these, 36 had been defined[19] as predicted targets of *A2BP1/2*, which represents a highly significant overlap (36/176; $P = 2.2 \times 10^{-16}$). In addition, five previously validated *A2BP1* targets showed evidence of alternative splicing, four of which (*ATP5C1*, *ATP2B1*, *GRIN1* and *MEF2C*) were confirmed as having differential splicing between ASD samples with low *A2BP1* expression and control samples, indicating that we were able to identify a high proportion of the expected *A2BP1*-dependent differential splicing events. We also observe that alternative exons with increased skipping in ASD relative to control cases are significantly enriched for *A2BP1* motifs in adjacent, downstream intronic sequences ($P = 1.09 \times 10^{-7}$, Fisher's exact test), consistent with previous data[19].

The top gene ontology categories enriched among ASD differential splicing genes highly overlapped with the gene ontology categories found to be enriched in the M12 module (Fig. 3b). In addition, *A2BP1* target genes showed enrichment for actin-binding proteins and genes involved in cytoskeleton reorganization (Fig. 3b). Among top predicted *A2BP1*-dependent differential splicing events (Fig. 3a) are *CAMK2G*, which also belongs to the M12 module, as well as *NRCAM* and *GRIN1*. The latter are proteins involved in synaptogenesis, in which allelic variants have been associated with autism and schizophrenia, respectively[21,22].

RT–PCR assays confirmed a high proportion (85%) of the tested differential splicing changes involving predicted A2BP1 targets (Supplementary Fig. 8). We further tested the differential splicing events validated by RT–PCR in three independent ASD cases with decreased *A2BP1* levels and confirmed the predicted changes in alternative splicing (Supplementary Fig. 8), indicating that the observed differential splicing events are indeed associated with reduced *A2BP1* levels, rather than due to inter-individual variability. The RNA-Seq data thus provides validation of the functional groups of genes identified by co-expression analysis, and evidence for a convergence of transcriptional and alternative-splicing abnormalities in the synaptic and signalling pathogenesis of ASD.

To test whether our findings are more generalizable, and determine whether the autism-associated transcriptional differences observed are likely to be causal, versus collateral effects or environmentally-induced changes, we tested whether our co-expression modules or

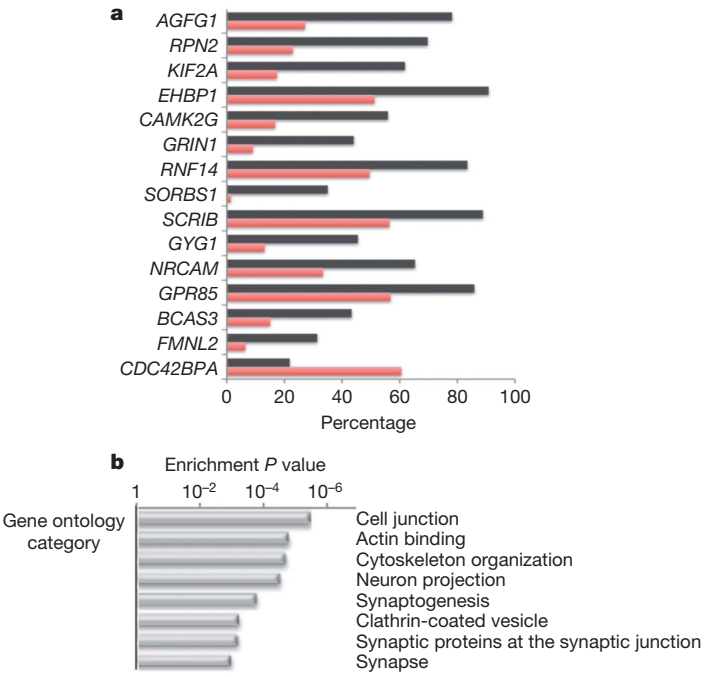

Figure 3 | A2BP1-dependent differential splicing events **a**, Top A2BP1-specific differential splicing events. Differential splicing events showing the most significant differences in alternative splicing between low-A2BP1 autism cases and controls as well as differential splicing differences consistent with the A2BP1 binding site position. The horizontal axis depicts the percentage of transcripts including the alternative exon. Red, autism samples; black, control samples. **b**, Relevant gene ontology categories enriched in the set of genes containing exons differentially spliced between low-A2BP1 autism cases and controls.

the differentially expressed genes show enrichment for autism genetic association signals. M12 showed highly significant enrichment for association signals ($P = 5 \times 10^{-4}$), but neither M16 nor the list of differentially expressed genes showed such enrichment (Fig. 4). As a negative control, we performed the same set-enrichment analysis using two GWAS studies for non-psychiatric disease performed on the same genotyping platform: a genome-wide association for hair colour[23], and a GWAS study of warfarin maintenance dose[24] finding no significant enrichment of the association signal (Fig. 4b, Supplementary Fig. 4). These results indicate that (1) M12 consists of a set of genes that are supported by independent lines of evidence to be causally involved in ASD pathophysiology, and (2) the upregulation of immune response genes in the autistic brain observed by us and others[25] has no evidence of a common genetic component.

Our system-level analysis of the ASD brain transcriptome demonstrates the existence of convergent molecular abnormalities in ASD for the first time, providing a molecular neuropathological basis for the disease, whose genetic, epigenetic, or environmental aetiologies can now be directly explored. The genome-wide analysis performed here significantly extends previous findings implicating synaptic dysfunction, as well as microglial and immune dysregulation in ASD[6] by providing an unbiased systematic assessment of transcriptional alterations and their genetic basis. We show that the transcriptome changes observed in ASD brain converge with GWAS data in supporting the genetic basis of synaptic and neuronal signalling dysfunction in ASD, whereas immune changes have a less pronounced genetic component and thus are most likely either secondary phenomena or caused by environmental factors. Because immune molecules and cells such as microglia have a role in synaptic development and function[26], we speculate that the observed immune upregulation may be related to abnormal ongoing plasticity in the ASD brain. The striking attenuation of gene expression differences observed here between frontal and temporal cortex in ASD is likely to represent a defect of developmental patterning and provides a strong rationale for further studies to assess the pervasiveness

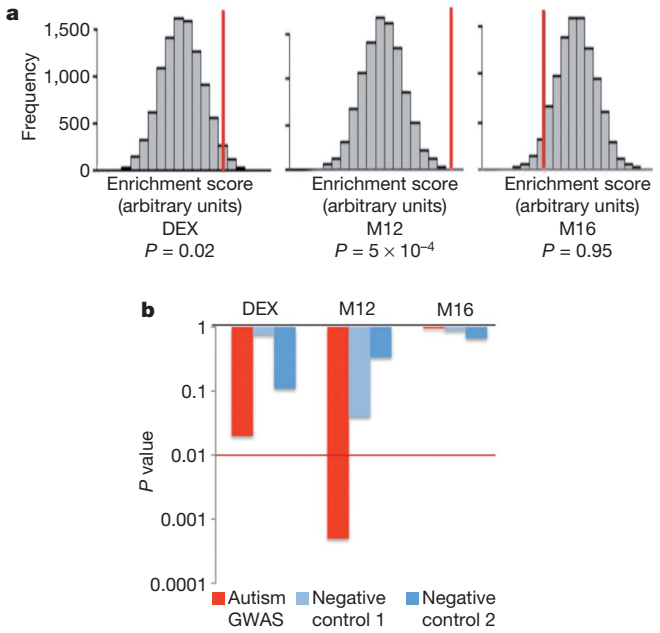

Figure 4 | GWAS set enrichment analysis **a**, GWAS set enrichment analysis using the discovery AGRE cohort from ref. 27. For each gene set (DEX, differentially expressed genes; M12 and M16) the null distribution of the enrichment score generated by 10,000 random permutations is shown (*x*-axis) and the enrichment score for the gene set is depicted by a red vertical line. A *P* value <0.01 was considered significant to correct for multiple comparisons. **b**, GWAS signal enrichment of differentially expressed genes and the autism-associated co-expression modules M12 and M16. Enrichment *P* values are shown for an autism GWAS data set (ref. 27, AGRE discovery cohort) as well as two control data sets consisting of GWAS studies of non-psychiatric traits: ref. 23 (Negative control 1) and ref. 24 (Negative control 2). The red line marks the *P* value threshold for significance.

of transcriptional patterning abnormalities across the ASD brain. We also demonstrate for the first time alterations in differential splicing associated with *A2BP1* levels in the ASD brain, and show that many of the affected exons belong to genes involved in synaptic function. Finally, given current evidence of genetic overlap between ASD and other neurodevelopmental disorders including schizophrenia and attention deficit hyperactivity disorder (ADHD), the data provide a new pathway-based framework from which to assess the enrichment of genetic association signals in other allied psychiatric disorders.

METHODS SUMMARY

Brain tissue. Post-mortem brain tissue was obtained from the Autism Tissue Project and the Harvard Brain Bank as well as the MRC London Brain bank for Neurodegenerative Disease. Detailed information on the autism cases included in this study is available in Methods.

Microarrays and RNA-seq. Total RNA was extracted from 100 mg of tissue using a Qiagen miRNA kit according to the manufacturer's protocol. Expression profiles were obtained using Illumina Ref8 v3 microarrays. RNA-seq was performed on the Illumina GAIIx, as per the manufacturer's instructions. Further detailed information on data analysis is available in Methods.

Full detailed Methods accompany this paper as Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 12 December 2010; accepted 13 April 2011.
Published online 25 May 2011.

1. Durand, C. M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genet.* **39,** 25–27 (2006).
2. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466,** 368–372 (2010).
3. Sebat, J. *et al.* Strong association of de novo copy number mutations with autism. *Science* **316,** 445–449 (2007).
4. Geschwind, D. H. Autism: many genes, common pathways? *Cell* **135,** 391–395 (2008).
5. Amaral, D. G., Schumann, C. M. & Nordahl, C. W. Neuroanatomy of autism. *Trends Neurosci.* **31,** 137–145 (2008).
6. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Rev. Genet.* **9,** 341–355 (2008).
7. Garbett, K. *et al.* Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiol. Dis.* **30,** 303–311 (2008).
8. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature Neurosci.* **11,** 1271–1282 (2008).
9. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4,** 17 (2005).
10. Johnson, M. B. *et al.* Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62,** 494–509 (2009).
11. Zikopoulos, B. & Barbas, H. Changes in prefrontal axons may disrupt the network in autism. *J. Neurosci.* **30,** 14595–14609 (2010).
12. Chen, Y. *et al.* Variations in DNA elucidate molecular networks that cause disease. *Nature* **452,** 429–435 (2008).
13. Plaisier, C. L. *et al.* A systems genetics approach implicates *USF1*, *FADS3*, and other causal candidate genes for familial combined hyperlipidemia. *PLoS Genet.* **5,** e1000642 (2009).
14. Babatz, T. D., Kumar, R. A., Sudi, J., Dobyns, W. B. & Christian, S. L. Copy number and sequence variants implicate *APBA2* as an autism candidate gene. *Autism Res.* **2,** 359–364 (2009).
15. Castermans, D. *et al. SCAMP5*, *NBEA* and *AMISYN*: three candidate genes for autism involved in secretion of large dense-core vesicles. *Hum. Mol. Genet.* **19,** 1368–1378 (2010).
16. Martin, C. L. *et al.* Cytogenetic and molecular characterization of *A2BP1/FOX1* as a candidate gene for autism. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **144B,** 869–876 (2007).
17. Alarcón, M. *et al.* Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *Am. J. Hum. Genet.* **82,** 150–159 (2008).
18. Underwood, J. G., Boutz, P. L., Dougherty, J. D., Stoilov, P. & Black, D. L. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol. Cell. Biol.* **25,** 10005–10016 (2005).
19. Zhang, C. *et al.* Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* **22,** 2550–2563 (2008).
20. Lee, J. A., Tang, Z. Z. & Black, D. L. An inducible change in Fox-1/A2BP1 splicing modulates the alternative splicing of downstream neuronal target exons. *Genes Dev.* **23,** 2284–2293 (2009).
21. Moy, S. S., Nonneman, R. J., Young, N. B., Demyanenko, G. P. & Maness, P. F. Impaired sociability and cognitive function in *Nrcam*-null mice. *Behav. Brain Res.* **205,** 123–131 (2009).
22. Zhao, X. *et al.* Significant association between the genetic variations in the 5′ end of the N-methyl-D-aspartate receptor subunit gene *GRIN1* and schizophrenia. *Biol. Psychiatry* **59,** 747–753 (2006).
23. Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4,** e1000074 (2008).
24. Cooper, G. M. *et al.* A genome-wide scan for common genetic variants with a large influence on warfarin maintenance dose. *Blood* **112,** 1022–1027 (2008).
25. Morgan, J. T. *et al.* Microglial activation and increased microglial density observed in the dorsolateral prefrontal cortex in autism. *Biol. Psychiatry* **68,** 368–376 (2010).
26. Boulanger, L. M. Immune proteins in brain development and synaptic plasticity. *Neuron* **64,** 93–109 (2009).
27. Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* **459,** 528–533 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful for the efforts of the Autism Tissue Program (ATP) of Autism Speaks and the families that have enrolled in the ATP, which made this work possible. We also thank S. Scherer and R. Wintle for sharing their SNP genotyping data on the AGP samples with us before its publication. We would also like to thank B. Abrahams for help in the initial stages of the project, B. Fogel, G. Konopka, N. Barbosa-Morais and J. Bomar for critically reading the manuscript, M. Lazaro for help with tissue dissection, and C. Vijayendran and K. Winden for useful discussions. This work was funded by an Autism Center of Excellence Network Grant from NIMH 5R01MH081754-03 and NIMH R37MH060233 to D.H.G. and by grants from the Canadian Institutes of Health Research and Genome Canada through the Ontario Genomics Institute to B.J.B. and others.

Author Contributions I.V. and D.H.G. designed the study and wrote the manuscript. I.V. performed experiments, analysed the data and conducted the GWAS set enrichment analysis. X.W. and B.J.B. analysed the RNA sequencing data. J.K.L. contributed to the GWAS set enrichment analysis. Y.T. performed some of the microarray qRT-PCR validation experiments. R.M.C. supervised the GWAS set enrichment analysis. S.H. supervised the WGCNA analysis. P.J. and J.M. provided dissected tissue for the replication experiment. All authors discussed the results and commented on the manuscript.

Author Information All microarray and RNA-seq data are deposited in GEO under accession number GSE28521. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.H.G. (dhg@mednet.ucla.edu).

METHODS

Brain tissue samples. Brain tissue samples from 19 autism cases and 17 controls were obtained from the Autism Tissue Project (ATP) and the Harvard Brain Bank. For each brain, tissue was obtained from frontal cortex (BA9), temporal cortex (BA41/42 or BA22) and cerebellum (vermis), with the exception of three controls lacking the cerebellum sample (Supplementary Table 1). For the replication experiment, frontal cortex tissue (BA44/45) from nine ASD cases and five controls were obtained from the ATP and MRC London Brain bank for Neurodegenerative Disease respectively (Supplementary Table 4).

For all of the autism cases, clinical information is available upon request from ATP (http://www.autismtissueprogram.org), including the ADI-R diagnostic scores. Supplementary Table 12 contains a summary of clinical characteristics. Although autism cases with known genetic causes were not included in this study, one case with a chromosome 15q duplication was identified for AN17138 by high density small nucleotide polymorphism (SNP) arrays[28] during the course of this study. The ATP cases were genotyped with high-density SNP arrays and with two exceptions all are Caucasians. The two Asian samples cluster with the other ASD cases in the current study, and are not distinguishable from the Caucasian cases based on clustering by gene expression.

RNA extractions and microarrays. Total RNA was extracted from approximately 100 mg of frozen tissue, using the Qiagen miRNA kit. RNA concentration was assessed by a NanoDrop spectrophotometer and RNA quality was measured using an Agilent Bioanalyzer. All RNA samples included in the expression analysis had an RNA integrity number (RIN) > 5. cDNA labelling and hybridizations on Illumina Ref8 v3 microarrays were performed according to the manufacturer's protocol.

Microarray data analysis. Microarray data analysis was performed using the R software and Bioconductor packages. Raw expression data were $\log_2$ transformed and normalized by quantile normalization. Data quality control criteria included high inter-array correlation (Pearson correlation coefficients > 0.85) and detection of outlier arrays based on mean inter-array correlation and hierarchical clustering. Probes were considered robustly expressed if the detection P value was <0.05 for at least half of the samples in the data set. Cortex samples (58: 29 autism, 29 controls) and cerebellum samples (21: 11 autism, 10 controls) fulfilled all data quality control criteria. The 29 autism cortex samples included tissue from 13 ASD cases with both frontal and temporal cortex and 3 ASD cases with frontal cortex only (in total 16 frontal cortex and 13 temporal cortex ASD samples). The 29 autism control samples also included tissue from 13 controls with both frontal and temporal cortex and 3 controls with frontal cortex only (in total 16 frontal cortex and 13 temporal cortex control samples).

Initially, all samples were normalized together to assess clustering by brain region. As expected, we observed distinct clustering of cortex and cerebellum samples (Supplementary Fig. 2A). For subsequent analyses, cortex samples and cerebellum samples were normalized and analysed separately.

Differential expression. Differential expression was assessed using the SAM package (significance analysis of microarrays, http://www-stat.stanford.edu/~tibs/SAM) and unless otherwise specified the significance threshold was FDR < 0.05 and fold changes > 1.3. Given that SAM is less sensitive in detecting differentially expressed genes for small number of samples, for the replication cohort, the differential expression was assessed by a linear regression method (Limma package, http://bioconductor.org/packages/release/bioc/html/limma.html). Our results showing high degree of overlap between genes differentially expressed in the two data sets indicate that the expression differences observed are independent of the analysis methods.

Because 444 genes were differentially expressed between autism and controls in cortex and only 2 genes were differentially expressed between the two groups in cerebellum (FDR < 0.05), we tested whether this difference was due to the smaller number of cerebellum samples, by relaxing the statistical criteria to FDR < 0.25. We found fewer than 10 differentially expressed genes in cerebellum using the relaxed statistical criteria, supporting the conclusion that genome-wide expression changes in autism were more pronounced in cerebral cortex than in cerebellum.

To account for the fact that the control group of DS1 contained samples from a single female whereas the autism DS1 group included four females, we eliminated from differential expression analysis all probes showing evidence of gender-specific gene expression ($n = 70$). We also applied linear regression of expression values against age and sex, and then assessed differential expression between the autism and control groups using the residual values. We observed a 96% overlap between differentially expressed genes using either the residual values or the raw data, indicating that neither age nor sex were major drivers of expression differences between the autism and control groups.

Differential expression between frontal and temporal cortex was assessed by a paired modified t-test (SAM) using the 13 autism and 13 control cases for which RNA samples from both cortex areas passed the quality control criteria. For each of the 510 genes that were differentially expressed in control samples between frontal and temporal cortex, we compared the variance of autism and control expression values in frontal cortex and temporal cortex. The homogeneity of variance (homoscedasticity) of gene expression was assessed using the Barlett test in R. Fifty one genes showed a significant difference in variance ($P < 0.05$, Barlett test) between autism and control groups both in frontal and temporal cortex, and the Barlett test P-values for these genes are listed in Supplementary Data.

WGCNA. Unsigned co-expression networks were built using the WGCNA package in R. Probes with evidence of robust expression (9,914; see above) were included in the network. Network construction was performed using the blockwiseModules function in the WGCNA package[29], which allows the network construction for the entire data set. For each set of genes a pair-wise correlation matrix is computed, and an adjacency matrix is calculated by raising the correlation matrix to a power. The power of 10 was chosen using the scale-free topology criterion[9] and was used for all three networks: the network built using autism samples only, controls samples only or all samples. An advantage of weighted correlation networks is the fact that the results are highly robust with respect to the choice of the power parameter. For each pair of genes, a robust measure of network interconnectedness (topological overlap measure) was calculated based on the adjacency matrix. The topological overlap based dissimilarity was then used as input for average linkage hierarchical clustering. Finally, modules were defined as branches of the resulting clustering tree. To cut the branches, we used the hybrid dynamic tree-cutting because it leads to robustly defined modules[31]. To obtain moderately large and distinct modules, we set the minimum module size to 40 genes and the minimum height for merging modules at 0.1. Each module was summarized by the first principal component of the scaled (standardized) module expression profiles. Thus, the module eigengene explains the maximum amount of variation of the module expression levels. For each module, we defined the module membership measure (also known as module eigengene based connectivity kME) as the correlation between gene expression values and the module eigengene. Genes were assigned to a module if they had a high module membership to the module (kME > 0.7). An advantage of this definition (and the kME measure) is that it allows genes to be part of more than one module. Genes that did not fulfil these criteria for any of the modules are assigned to the grey module. For the cell type marker enrichment analysis we used the markers defined experimentally in refs 32 and 33 which were previously used to annotate human brain network modules[34,35].

Module visualization: the topological overlap measure was calculated for the top 100 genes in each module ranked by kME. The resulting list of gene pairs was filtered so that both genes in a pair had the highest kME for the module plotted (that is, most module-specific interactions). The resulting top 150 gene pairs were plotted using Visant.

Gene ontology analyses. Functional enrichment was assessed using the DAVID database http://david.abcc.ncifcrf.gov/. For differentially expressed genes and co-expression modules, the background was set to the total list of genes expressed in the brain in the cortex data set. For genes containing differentially spliced exons, the background was set to the total set of genes showing evidence of alternative splicing in our RNA-seq data. The statistical significance threshold level for all gene ontology enrichment analyses was $P < 0.05$ (Benjamini and Hochberg corrected for multiple comparisons).

Statistical analyses. All gene set overlap analyses were performed by assessing the cumulative hypergeometric probability using the phyper function in R. The population size was defined as the total number of probes expressed in both data sets. If the comparison involved different platforms, the comparison was done at gene level.

Quantitative RT–PCR. One microgram of total RNA was treated with RNase-free DNase I (Invitrogen/Fermentas) and reverse-transcribed using Invitrogen Superscript II reverse-transcriptase and random hexanucleotide primers (Invitrogen). Real time PCR was performed on an ABI7900 cycler in 10 µl volume containing iTaq Sybrgreen (Biorad) and primers at a concentration of 0.5 µM each. The results shown in Supplementary Fig. 2b represent at least two independent cDNA synthesis experiments for each gene. *GAPDH* levels were used as an internal control. Statistical significance was assessed by a two-tailed t-test assuming unequal variance.

Semi-quantitative RT–PCR. Total RNA (600 ng) pooled from autism cases ($n = 2$–3) or controls ($n = 2$–3) was reverse-transcribed as described above. cDNA (50 ng) was subjected to 30 cycles of PCR amplification using the primers described in Supplementary Table 11. PCR products were separated on a 3% agarose gel stained with GelStar (Lonza).

RNA sequencing and data analysis.. 73-nucleotide reads were generated using an Illumina GAII sequencer according to the manufacturer's protocol. To generate sufficient read coverage for the quantitative analysis of alternative splicing events, reads for ASD and control brain samples were separately pooled and aligned to an existing database of EST and cDNA-derived alternative splicing junctions using

the Basic Local Alignment Tool (BLAT) as described previously[36,37]. Reads were considered properly aligned to a splice junction if at least 71 of the 73 nucleotides matched and at least 5 nucleotides mapped to each of the two exons forming the splice junction. Alternative exon inclusion values ('%inc'), representing the proportion of messenger RNA transcripts with the alternatively spliced exon included, were calculated for each mRNA pool as the ratio of reads aligning to the C1-A or A-C2 junctions against reads aligning against all three possible junctions as previously described[36] (C1-A, A-C2, C1-C2, see Supplementary Fig. 3). Calculated %inc values were considered reliable if at least one of the included junctions as well as the skipped junctions were covered by at least 20 reads. %inc values were compared across samples using Fisher's exact test and the Bonferroni–Hochberg correction to identify differentially spliced exons associated with autism. Differential splicing events were considered significant if they fulfilled both criteria of FDR < 0.1 and %inc difference between autism and controls $>15\%$.

GWAS set enrichment analysis. GWAS enrichment analysis was performed as previously described in ref. 38 with the main modification that we generated the null distribution, using permutation of gene labels rather than permutation of case/control labels, because the raw genotyping data was not available for all data sets. This approach has been proposed as an acceptable alternative to phenotype label permutation[38] and has been previously used for set enrichment analyses of GWAS data[39]. For all genes that met the robust expression criteria in our data set, we mapped the SNPs present on the Illumina 550k platform located within the transcript boundaries and an additional 20 kb on the 5′end and 10 kb on the 3′end. Each gene was assigned a GWAS significance value consisting of the lowest *P* value of all SNPs mapped to it. A gene set enrichment score (ES) based on the Kolmogorov–Smirnov statistic was calculated as previously described[38] using the $-\log(P\text{-value})$. The null distribution was generated by 10,000 random permutations of gene labels in the list of genes/*P*-value pairs and an enrichment score ESp was calculated for each permutation. To correct for the gene set size, the enrichment scores were scaled by subtracting the mean and dividing by the standard deviation of ESp. The resulting *z*-scores were used to calculate the significance p value.

28. Wintle, R. F. *et al.* (2010). A genotype resource for postmortem brain samples from the Autism Tissue Program. *Autism Res.* **4,** 89–97 (2011).
29. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).
31. 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).
32. Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28,** 264–278 (2008).
33. Albright, A. V. & Gonzalez-Scarano, F. Microarray analysis of activated mixed glial (microglia) and monocyte-derived macrophage gene expression. *J. Neuroimmunol.* **157,** 27–38 (2004).
34. Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nature Neurosci.* **11,** 1271–1282 (2008).
35. Miller, J. A., Horvath, S. & Geschwind, D. H. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc. Natl Acad. Sci. USA* **107,** 12698–12703 (2010).
36. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327,** 996–1000 (2010).
37. Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40,** 1413–1415 (2008).
38. Wang, K., Li, M. & Bucan, M. Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.* **81,** 1278–1283 (2007).
39. Zhang, K., Cui, S., Chang, S., Zhang, L. & Wang, J. *i*-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res.* **38** (suppl. 2), W90–W95 (2010).