Citation: Transl Psychiatry (2012) 2, e113, doi:10.1038/tp.2012.42

www.nature.com/tp

Association between SNPs and gene expression in multiple regions of the human brain

S Kim[1], H Cho[2], D Lee[2] and MJ Webster[1]

Identifying the genetic *cis* associations between DNA variants (single-nucleotide polymorphisms (SNPs)) and gene expression in brain tissue may be a promising approach to find functionally relevant pathways that contribute to the etiology of psychiatric disorders. In this study, we examined the association between genetic variations and gene expression in prefrontal cortex, hippocampus, temporal cortex, thalamus and cerebellum in subjects with psychiatric disorders and in normal controls. We identified *cis* associations between 648 transcripts and 6725 SNPs in the various brain regions. Several SNPs showed brain regional-specific associations. The expression level of only one gene, *PDE4DIP*, was associated with a SNP, rs12124527, in all the brain regions tested here. From our data, we generated a list of brain *cis* expression quantitative trait loci (eQTL) genes that we compared with a list of schizophrenia candidate genes downloaded from the Schizophrenia Forum (SZgene) database (http://www.szgene.org/). Of the SZgene candidate genes, we found that the expression levels of four genes, *HTR2A*, *PLXNA2*, *SRR* and *TCF4*, were significantly associated with *cis* SNPs in at least one brain region tested. One gene, *SRR*, was also involved in a coexpression module that we found to be associated with disease status. In addition, a substantial number of *cis* eQTL genes were also involved in the module, suggesting eQTL analysis of brain tissue may identify more reliable susceptibility genes for schizophrenia than case–control genetic association analyses. In an attempt to facilitate the identification of genetic variations that may underlie the etiology of major psychiatric disorders, we have integrated the brain eQTL results into a public and online database, Stanley Neuropathology Consortium Integrative Database (SNCID; http://sncid.stanleyresearch.org).
Translational Psychiatry (2012) **2,** e113; doi:10.1038/tp.2012.42; published online 8 May 2012

Introduction

Schizophrenia, bipolar disorder and severe depression are common and highly disabling brain diseases caused by an interaction of genetic and environmental factors.[1,2] However, despite enormous efforts, the genetic variations that contribute to these diseases and their environmental risk factors remain elusive. Genome-wide association studies have frequently been employed to identify susceptibility genes and single-nucleotide polymorphisms (SNPs) that may be associated with these mental disorders.[3–5] A number of candidate genes for the disorders have been reported. For instance, a web resource for schizophrenia, the Schizophrenia Forum (SZgene) database (http://www.szgene.org/), includes results from 1727 genetic association studies and reports 1008 candidate genes and 8788 polymorphisms in the update on 15 April 2011.[6] Despite the numerous candidate genes reported for schizophrenia, the effect size of each variant is small or moderate and most associated SNPs have failed to be replicated. The need for independent and systematic validation to prioritize further examination of possible candidate genes for mental disease is widely acknowledged.

Identification of DNA sequence variants that regulate gene expression levels in a relevant tissue is one of the most promising approaches used to initially scan for candidate genes as well as to prioritize previously identified candidate genes that are associated with complex disease such as psychiatric disorders.[7–9] The identification of a *cis* association of a SNP with gene expression levels has been previously used to validate candidate genes for complex traits mapped to the same chromosomal locations.[10] Our recent study using an integrative approach that combined results from genome-wide SNP scans for the cytoarchitectural traits and *cis* expression quantitative trait loci (eQTL) analysis in the brain tissue revealed two novel candidate genes associated with cellular abnormalities in the prefrontal cortex of major psychiatric disorders.[11] Limited availability of human post-mortem brain tissues is a major obstacle to obtaining detailed brain expression complex trait loci (eQTL) mapping. Utilization of publicly available resources is an effective alternative strategy that may overcome such limitation. The Stanley Neuropathology Consortium Integrative Database (SNCID; http://sncid.stanleyresearch.org) is a publicly available and web-based tool that integrates expression microarray data sets from five brain regions including frontal cortex, temporal cortex, thalamus, cerebellum and hippocampus and genome-wide SNP genotype data sets of subjects in the Stanley Neuropathology Consortium (SNC) and the Array Collection (AC).[12] A total of 1749 neuropathology data sets using the

[1]Stanley Brain Research Laboratory, Stanley Medical Research Institute, Rockville, MD, USA and [2]Department of Bio and Brain Engineering, KAIST, Yuseong-gu, Daejeon, Republic of Korea
Correspondence: Dr MJ Webster, Stanley Medical Research Institute, 9800 Medical Center Drive, Rockville, MD 20850, USA or Dr D Lee, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea.
E-mail: websterm@stanleyresearch.org or dhlee@kaist.ac.kr
Keywords: *cis* SNP; eQTL; post-mortem brain; psychiatric disorders; schizophrenia; SNCID
Received 12 March 2012; accepted 10 April 2012

SNC are integrated into the database, which thereby enables one to further explore the correlations between gene expression levels and quantitative measures of neuropathological markers in the various brain regions. The specific aims of this study are twofold. First, we explore the candidate genes that may be functionally relevant for major psychiatric disorders by identifying *cis* associations between SNPs and gene expression in various brain tissues. Second, we examine the possible functional role of schizophrenia candidate genes that were previously identified in genetic association studies. Thus, we explored *cis* eQTLs in the four brain regions, frontal cortex, temporal cortex, thalamus and cerebellum, of SNC subjects and in hippocampus of AC subjects. We also repeated the analysis in frontal cortex data from the AC as a replication study to examine the overall consensus of *cis* eQTLs between the two frontal data sets. We then examined whether the expression levels of any candidate genes from the SZgene database meta-analysis (http://www.szgene.org/) were regulated by *cis* expressed SNPs (eSNPs) in brain tissues, in order to determine if there were any functional effects on gene expression of the previously identified schizophrenia susceptibility genes. Finally, we performed a coexpression network analysis between the genes in the frontal cortex that were differentially expressed between schizophrenia and normal controls and the *cis* eQTL genes in an attempt to identify the potential role of these genes in a disease-specific coexpression module.

Materials and methods

Data used in this study. Gene expression microarray data from frontal cortex,[13] cerebellum, thalamus and temporal cortex[14]were generated by multiple independent groups using samples from the SNC ($N=60$), which contains 15 well-matched cases in each of four groups: schizophrenia, bipolar disorder, major depression and unaffected controls.[15] Other sets of microarray data from frontal cortex[16,17]and hippocampus were generated using samples from the AC ($N=105$). The AC is an independent tissue collection containing 35 cases in each of three groups: schizophrenia, bipolar disorder and unaffected controls. The groups from both tissue collections are matched for descriptive variables such as age, gender, race, post-mortem interval, mRNA quality, brain pH and hemisphere. Outlier chip data were excluded in this analysis based on previous quality-control analyses for chip-level parameters such as scaling factor, gene call and average correlation.[18] Information for the microarray studies such as tissue collection, brain region and number of outlier chips is listed in the Supplementary Table S1 online. The confounding effects on the Frozen Robust Multiarray Analysis (fRMA)-normalized microarray gene expression data were identified using Surrogate Variable Analysis (SVA).[19] To adjust disease effect on the gene expression data, we randomly assign 0 or 1 for the primary variable in the SVA. All covariates from SVA were used in the linear regression to adjust the confounding effects on the gene expression data. The standardized residuals from the linear regression were used to evaluate the effectiveness of this method on removing confounding variables on two microarray data sets from both the SNC and AC. Transcripts correlated with potential confounding variables were identified using nonparametric analysis. The continuous variables such as age, brain pH, post-mortem interval and lifetime exposure to antipsychotics were examined by correlation analysis using R (open source program from Comprehensive R Archive Network (CRAN)). Two categorical variables such as microarray batch and sex were tested using variance analysis. Adjusted *P*-values, based on the Hochberg method that were <0.05, were considered significant. Although all cases and controls were included in the analysis, only the disorder cases were used for the correlation analysis for the effect of lifetime exposure to antipsychotics. SNP genotyping data using DNA samples from the SNC and the AC were generated by Dr Chun-Yu Liu and colleagues (University of Chicago, IL, USA) using the Human SNP Array 5.0 chips (Affymetrix, Santa Clara, CA, USA).[20]

eQTL analysis. Raw image files from SNP chips, quality-control analysis and identification of ethnic outliers were performed as previously described.[11] Briefly, genotypes were called using the BRLMM algorithm (Affymetrix). SNPs with a call rate of $<90\%$, minor allele frequency $<5\%$ or extreme deviation from Hardy–Weinberg equilibrium test ($P<0.05$) were filtered out for further eQTL analyses. A total of 309 531 SNPs passed this filter. For examination of population stratification, clustering was initially performed using the pairwise identity-by-state (IBS) calculator in the PLINK.[21] IBS pairwise distances were then plotted and examined by multidimensional scaling analysis and Z statistical analysis. Samples of >3 s.d. compared with the group mean were considered outliers. Four ethnic outliers from the SNC and three outliers from the AC were excluded in the eQTL analysis. One additional sample from AC was excluded because of a final diagnosis of CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy). We only used genotyped SNP data from chips for our association analysis rather than imputing genotypes because SNP imputation can often result in errors in genotyping and cause false-positive associations.[22] The standardized residuals from the linear regression were used as traits in PLINK for eQTL analyses. We defined *cis* eSNPs as those that were localized within 1 Mb of either the 5′ or the 3′ end of the gene. The *trans* eSNPs were defined as all SNPs that reached genome-wide significance level, except those in a *cis* position. We employed a conservative Bonferroni method to correct multiple testing for controlling false positives.[7] Adjusted *P*-values of <0.05 (unadjusted *P*-value; $1.6E{-}07 = 0.05/309\,531$) were considered genome-wide significant for eQTL analyses.

Coexpression network analysis. Unsupervised and supervised coexpression network analyses were performed using the Weighted Correlation Network Analysis (WGCNA) in R.[23] The coexpression network was generated using expression values of all genes in the frontal cortex of schizophrenia and normal controls from the AC (unsupervised WGCNA). A second coexpression network was generated using significant *cis* eQTL genes and genes that were differentially expressed in the frontal cortex between

schizophrenia and normal controls from the AC samples (supervised WGCNA).[24] A total of four microarray data sets (at www.stanleygenomics.org; study no. 1, 3, 5 and 7) were generated from prefrontal cortex. Three of these (study no. 1, 3 and 7)[16,17] were generated using the same platform, Affymetrix 133a, and hence to avoid variations between platforms we pooled the data from these three data sets. The pooled data were then subjected to median normalization with the biometric research branch (BRB)-array tools (http://linus.nci.nih.gov/BRB-ArrayTools.html) to remove systematic variations. After median normalization, confounding effects were adjusted using SVA and a linear regression method as described in the previous section. However, disease effect was not removed. Standardized residuals that were significantly associated with disease (nominal *P*-value <0.05) and standardized residuals of *cis* eQTL genes were then used as input for the WGCNA.[23] The minimum module size and the minimum height for merging modules were set at 30 and 0.25, respectively. The coexpression module was visualized using VisANT.[25]

Functional annotation. The *cis* eQTL genes and genes that were involved in the coexpression module were functionally annotated using the Database for Annotation, Visualization and Integrated Discovery (DAVID) database (http://david.abcc.ncifcrf.gov/home.jsp) and by the over-representational analysis method.[26] The biological processes of Gene Ontology Consortium (http://www.geneontology.org) were used for functional annotations. The *P*-values of <0.05 were considered significant.

Results

eQTL analysis in various human post-mortem brain tissues. Gene expression microarray data derived from post-mortem brain tissue are often confounded by uncontrolled biological, clinical and technical variables.[27] Batch effect is particularly problematic and has been shown to significantly affect gene expression levels in microarray data.[28,29] To remove the effect of batch and other confounding variables in our gene expression microarray data, we normalized the data using the newly developed method, fRMA, followed by the SVA.[19,30] We evaluated how effective this method was at removing confounding variables using two microarray data sets from both the SNC and AC (Supplementary Table S1 online). Using the data set from SNC temporal cortex (study 18) we found that microarray batch was the most significant confounding variable in both the RMA and fRMA-normalized data sets, with 947 and 1031 transcripts significantly correlated with batch, respectively (Supplementary Table S2 online). Using the data set from AC frontal cortex (study 1) we found that microarray batch and brain pH were both major confounding variables (Supplementary Table S3 online). The SVA successfully adjusted the effects of the confounding variables on both microarray data sets (Supplementary Table S2 and S3 online).

Using the SVA we obtained the standardized residuals from the linear regression with covariates and conducted a genome-wide eQTL analysis of various brain tissues. We used the standardized residuals as traits. We initially analyzed gene expression microarray data from frontal cortex, temporal cortex, thalamus and cerebellum from the SNC (Supplementary Table S1 online). Expression levels of a total of 53, 11, 84 and 27 genes were correlated with *cis* SNPs in the frontal cortex, temporal cortex, thalamus and cerebellum at genome-wide significance level, respectively (nominal $P<1.6E-07$; Figure 1a and Supplementary Table S4 online). Among the *cis* eQTL genes, expression levels of 16, 0, 20 and 5 genes were also significantly associated with *trans* SNPs in the frontal cortex, temporal cortex, thalamus and cerebellum, respectively (Supplementary Table S5 online). In addition, correlations between the expression levels of 31, 1, 69 and 15 genes and *cis* SNPs were unique in the frontal cortex, temporal cortex, thalamus and cerebellum, respectively (Figure 1a). The expression level of only one gene, *phosphodiesterase 4D interacting protein* (*PDE4DIP*), was associated with a SNP, rs12124527, in all the brain regions tested here. We then

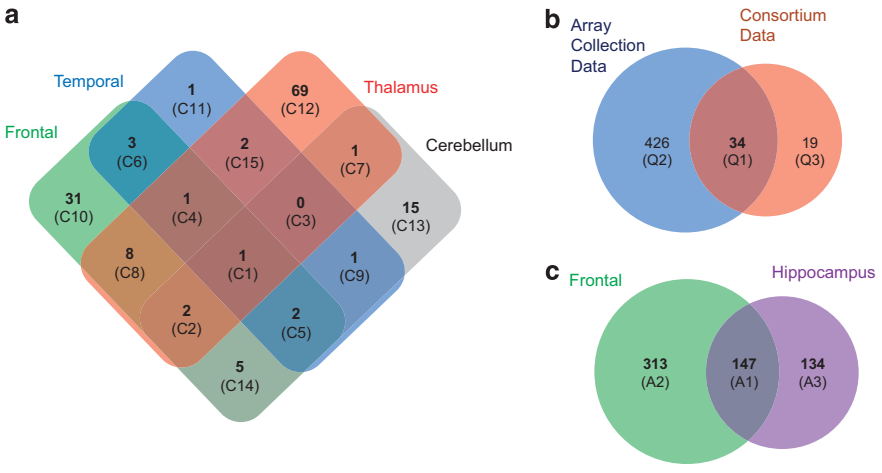

Figure 1 Number of *cis* expression quantitative trait loci (eQTL) genes in various brain regions. Venn diagram shows common and unique *cis* eQTL genes across multiple brain regions of the Stanley Neuropathology Consortium (SNC) samples (**a**) and of the Array Collection (AC) samples (**c**). Overlapped *cis* eQTL genes in the frontal cortex between the SNC samples and the AC samples are shown (**b**).

Table 1 Biological processes (Gene ontology) significantly associated with *cis* eQTL genes in the frontal cortex of the Stanley Neuropathology Consortium samples

Biological process categories	*Count*	*Fraction (%)*	P-*value*
GO:0007155–cell adhesion	7	1.361868	0.013
GO:0022610–biological adhesion	7	1.361868	0.013
GO:0007601–visual perception	4	0.77821	0.022
GO:0050953–sensory perception of light stimulus	4	0.77821	0.022
GO:0046907–intracellular transport	6	1.167315	0.036
GO:0035249–synaptic transmission, glutamatergic	2	0.389105	0.039

Abbreviations: eQTL, expression quantitative trait loci; GO, Gene Ontology.

replicated the *cis* eQTLs of the frontal cortex using the larger AC collection. The replication study revealed associations between *cis* SNPs and expression levels of 460 genes and replicated 34 *cis* eQTL genes out of 53 (64%) that were identified in the SNC study (Figure 1b and Supplementary Table S6 online). Moreover, 281 *cis* eQTL genes were identified in the AC hippocampus data and 147 *cis* eQTLs were common to both the frontal cortex and hippocampus (Figure 1c and Supplementary Table S7 online). Among the *cis* eQTL genes, expression levels of 43 and 46 genes were also significantly associated with *trans* SNPs in the frontal cortex and hippocampus, respectively (Supplementary Table S5 online). The association between *PDE4DIP* expression and the rs12124527 SNP was replicated in the AC frontal cortex and hippocampal data.

Next, we performed a functional annotation analysis to identify biological processes that were overrepresented in the brain *cis* eQTL genes. Whereas several processes such as cell adhesion, visual perception and glutamatergic transmission were overrepresented in the genes with *cis* eSNPs in the SNC frontal cortex (Table 1), metabolic processes such as glutamine metabolic process and protein transport and targeting and antigen processing were overrepresented in the AC frontal cortex (Table 2). Amino acid metabolic process, nucleotide biosynthesis and enzyme-linked receptor protein signaling pathways were significantly overrepresented in *cis* eQTL genes in the AC hippocampus (Supplementary Table S8 online).

Comparison between schizophrenia susceptibility candidate genes and brain *cis* eQTL genes. Genetic association studies have yielded numerous candidate genes that may increase the risk for schizophrenia. However, most candidate genes have not been replicated nor functionally validated. To examine the possible functional role of schizophrenia candidate genes, we compared the list of the candidate genes in the SZgene database meta-analysis (updated 12/1/2010) to our list of *cis* eQTL genes. The SZgene meta-analysis identified 45 genetic variants and 42 linked genes. After excluding the non-SNP variants from their data set, we were left with 39 SNPs and 39 linked candidate genes. Because only 6 SNP markers out of the 39 SNPs were included in our Affymetrix SNP 5.0 data set, we conducted a gene-level comparison instead of SNP-level comparison. We determined whether there were *cis* associations between the expression levels of the 39 candidate genes and SNPs within 1 Mb of the genes. Among the 39 candidate genes, we found that the expression levels of four genes, *HTR2A*, *PLXNA2*, *SRR* and *TCF4*, were significantly associated with *cis* SNPs in at least one brain region tested (Table 3). The expression levels of *HR2A* and PLXNA2 were associated with *cis* SNPs in the frontal cortex, whereas the expression levels of *SRR* (*serine racemase*) and *TCF4* were associated with *cis* SNPs in two brain regions. The *cis* eSNPs of these genes are located at least 25 kb from the SNPs that were significantly associated with schizophrenia in the SZgene meta-analysis. Thus the SZgene case–control genetic association analyses for schizophrenia may not have identified the most functionally relevant genetic variations that contribute to the etiology of psychiatric disorders.

Coexpression network analysis in the frontal cortex. To further examine whether or not the four schizophrenia candidate genes (*HTR2A*, *PLXNA2*, *SRR* and *TCF4*) and genes of which expression levels were regulated by *cis* SNPs may be involved in the etiology of schizophrenia, we performed both unsupervised and supervised gene coexpression network analyses using the AC frontal cortex data. We were unable to construct a coexpression module that was significantly associated with schizophrenia disease status using the unsupervised analysis. One module was associated with disease ($P = 0.05$); however, it was also associated with post-mortem interval ($P = 0.01$). We then constructed a supervised coexpression network using genes differentially expressed between schizophrenia and normal controls (Supplementary Table S9 online) and the *cis* eQTL genes obtained from the pooled data of three Affymetrix 133a microarray data sets that measured gene expression in the frontal cortex. We constructed one coexpression module that was significantly associated with schizophrenia disease status ($P = 2E-08$; Figure 2a). Age, sex, post-mortem interval, brain pH and lifetime antipsychotic treatment were not significantly associated with this module (all $P > 0.05$). Genes associated with apoptosis, chromatin organization, RNA splicing, cell cycle, regulation of nucleic acid metabolism and endocytosis were overrepresented in this module (Figure 2b and Supplementary Table 10 online). A previous coexpression network analysis that used gene expression microarray data from prefrontal cortex from schizophrenia subjects and controls[31] also identified a module (module 16) with similar overrepresentation of biological processes such as chromatin organization, cell cycle, endocytosis and regulation of nucleic acid metabolism. Apoptosis and endocytosis have previously been associated with the pathophysiology of the frontal cortex in schizophrenia,[32–44] and recent studies have also indicated that aberrant RNA splicing and epigenetic alterations may be involved in the pathophysiology

Table 2 Biological processes (Gene ontology) significantly associated with *cis* eQTL genes in the frontal cortex of the Array Collection samples

Biological process categories	*Count*	*Fraction (%)*	P-*value*
GO:0006541–glutamine metabolic process	6	0.13	5.00E−05
GO:0009064–glutamine family amino acid metabolic process	8	0.17	1.50E−04
GO:0006412–translation	19	0.41	5.10E−04
GO:0046907–intracellular transport	28	0.60	1.80E−03
GO:0006414–translational elongation	9	0.19	2.00E−03
GO:0009069–serine family amino acid metabolic process	5	0.11	2.50E−03
GO:0034613–cellular protein localization	19	0.41	5.50E−03
GO:0070727–cellular macromolecule localization	19	0.41	5.80E−03
GO:0002474–antigen processing and presentation of peptide antigen via MHC class I	4	0.09	6.00E−03
GO:0002483–antigen processing and presentation of endogenous peptide antigen	3	0.06	7.10E−03
GO:0019885–antigen processing and presentation of endogenous peptide antigen via MHC class I	3	0.06	7.10E−03
GO:0006508–proteolysis	37	0.79	7.60E−03
GO:0006605–protein targeting	12	0.26	9.50E−03
GO:0006886–intracellular protein transport	17	0.36	1.10E−02
GO:0019882–antigen processing and presentation	7	0.15	1.10E−02
GO:0046777–protein amino acid autophosphorylation	7	0.15	1.20E−02
GO:0015031–protein transport	28	0.60	1.30E−02
GO:0019883–antigen processing and presentation of endogenous antigen	3	0.06	1.30E−02
GO:0045184–establishment of protein localization	28	0.60	1.40E−02
GO:0051603–proteolysis involved in cellular protein catabolic process	23	0.49	1.70E−02
GO:0044257–cellular protein catabolic process	23	0.49	1.70E−02
GO:0043632–modification-dependent macromolecule catabolic process	22	0.47	2.00E−02
GO:0019941–modification-dependent protein catabolic process	22	0.47	2.00E−02
GO:0001510–RNA methylation	3	0.06	2.00E−02
GO:0044265–cellular macromolecule catabolic process	26	0.56	2.30E−02
GO:0030163–protein catabolic process	23	0.49	2.40E−02
GO:0009070–serine family amino acid biosynthetic process	3	0.06	2.40E−02
GO:0048002–antigen processing and presentation of peptide antigen	4	0.09	2.40E−02
GO:0008104–protein localization	30	0.64	2.50E−02
GO:0007143–female meiosis	3	0.06	2.90E−02
GO:0044271–nitrogen compound biosynthetic process	14	0.30	3.20E−02
GO:0006607–NLS-bearing substrate import into nucleus	3	0.06	3.30E−02
GO:0034660–ncRNA metabolic process	11	0.24	3.50E−02
GO:0006625–protein targeting to peroxisome	3	0.06	3.80E−02
GO:0043574–peroxisomal transport	3	0.06	4.40E−02
GO:0009057–macromolecule catabolic process	26	0.56	4.70E−02
GO:0006399–tRNA metabolic process	7	0.15	5.00E−02

Abbreviations: eQTL, expression quantitative trait loci; GO, Gene Ontology; MHC, major histocompatibility complex; NLS, nuclear localization signal; tRNA, transfer RNA.

Table 3 Schizophrenia candidate genes (from the *SZgene* database) with expression levels significantly associated with *cis* eSNPs in brain tissue

Symbol	*Chr.*	*Position*[a]	*Study_ID*	*Brain_Region*	*SNP*	*S_Position*	P-*value*[b]	*Adjusted* P-*value*
HTR2A	13	(R): 47407513-47470175	Study01	Frontal cortex	rs1923882	47411661	3.08E−10	9.45E−05
HTR2A	13	(R): 47407513-47470175	Study07	Frontal cortex	rs1923882	47411661	2.54E−08	7.80E−03
PLXNA2	1	(R): 208195587-208417665	Study01	Frontal cortex	rs6659522	208199100	3.26E−10	1.00E−04
PLXNA2	1	(R): 208195587-208417665	Study03	Frontal cortex	rs6659522	208199100	1.03E−09	3.15E−04
PLXNA2	1	(R): 208195587-208417665	Study07	Frontal cortex	rs6702082	208206946	4.04E−08	1.24E−02
SRR	17	(F): 2207248-2228553	Study01	Frontal cortex	rs16952025	2116798	3.24E−08	9.94E−03
SRR	17	(F): 2207248-2228553	Study03	Frontal cortex	rs16952025	2116798	8.86E−08	2.72E−02
SRR	17	(F): 2207248-2228553	Study07	Frontal cortex	rs16952025	2116798	1.65E−13	5.06E−08
SRR	17	(F): 2207248-2228553	Study17	Hippocampus	rs16952025	2116798	9.60E−11	2.95E−05
TCF4	18	(R): 52889562-53255860	Study01	Frontal cortex	rs1261085	52889967	1.08E−08	3.33E−03
TCF4	18	(R): 52889562-53255860	Study03	Frontal cortex	rs1261134	52931763	1.98E−15	6.09E−10
TCF4	18	(R): 52889562-53255860	Study05	Frontal cortex	rs1261073	52907820	2.18E−12	6.68E−07
TCF4	18	(R): 52889562-53255860	Study07	Frontal cortex	rs1261073	52907820	1.57E−15	4.82E−10
TCF4	18	(R): 52889562-53255860	Study16	Thalamus	rs1261134	52931763	3.86E−08	1.18E−02

Abbreviation: eSNP, expressed single-nucleotide polymorphism.
Genome build;hg19. [a]F and R represent the forward orientation or reverse orientation on a chromosome. [b]Adjusted *P*-value using Bonferroni method.

of schizophrenia.[35,36] Several genes associated with GABAergic neurons, including γ-aminobutyric acid (GABA) A receptor, δ (*GABRD*) and *parvalbumin* (*PVALB*), were found in the coexpression module. However, out of the four schizophrenia candidate genes common to both the SZgene meta-analysis and our *cis* eQTL gene list, only one, *SRR*, was found in the module. The biological process, response to drug, was overrepresented in the module and was enriched with 11 genes including *SRR*. A substantial number of *cis* eQTL genes were involved in the coexpression module

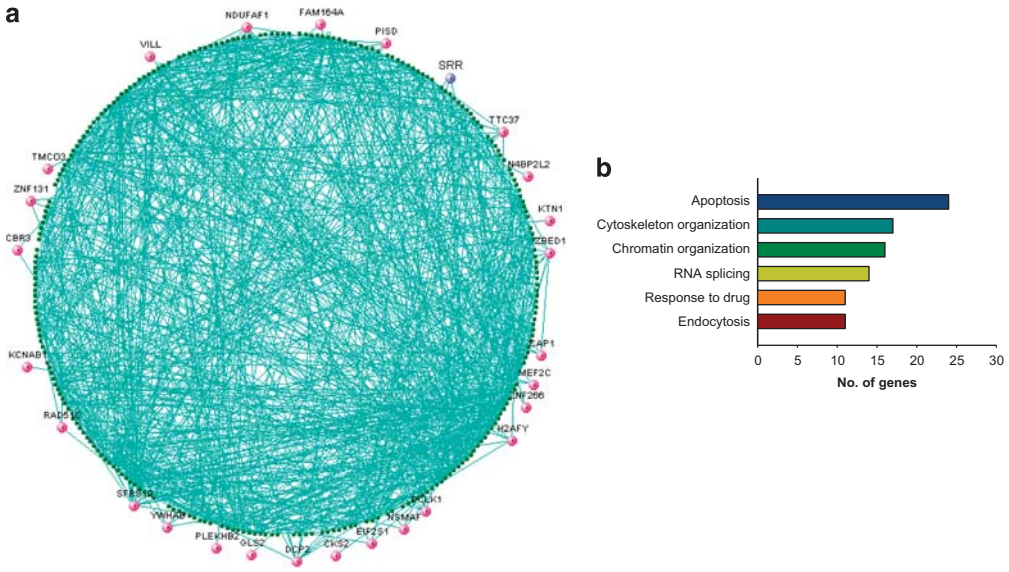

Figure 2 Coexpression network analysis in the frontal cortex. The coexpression module that is significantly associated with schizophrenia in frontal cortex of the Array Collection (AC) (**a**) and biological processes (Gene ontology) overrepresented in the genes in the coexpression module (**b**). Network connections with topological overlap above the threshold of 0.02 were visualized using VisANT.[25] The *cis* expression quantitative trait loci (eQTL) genes are pink. The candidate gene, *SRR* (*serine racemase*), derived from the meta-analyses of genetic studies in the SZgene database (http://www.szgene.org/) is in blue.

that was significantly associated with schizophrenia disease status and were also associated with the biological processes. This result indicates that *cis* eQTL analysis in brain tissue may more reliably identify susceptibility genes for schizophrenia as compared with the current case–control genetic association studies.

Discussion

Identifying genetic variations that affect gene expression in the brain may be a promising approach for finding molecular pathways that are functionally relevant to the etiology and/or treatment of mental disease. In this study, we conducted an eQTL analysis of 315 440 transcripts in 5 different brain regions from two different tissue collections and identified *cis* associations between 648 transcripts and 6725 SNPs. The expression of one gene, *PDE4DIP*, was associated with one SNP, rs12124527, in all brain regions examined. This association was also previously described in the frontal cortex.[20] The protein encoded by PDE4DIP serves to anchor phosphodiesterase 4D to the Golgi/centrosome region of the cell. A number of abnormalities in the phosphodiesterase signaling system have been described in the brains of subjects with schizophrenia, bipolar disorder and depression,[37–42] indicating that molecules within this system could be potential targets for therapeutic intervention.[38,43]

Approximately 14% of *cis* eQTL genes were also correlated with *trans* SNPs in various brain regions, suggesting that the expression levels of a subset of *cis* eQTL genes may be regulated by multiple variants. However, when we examined whether or not the expression levels of candidate genes from the SZgene database meta-analysis were significantly associated with *cis* SNPs, we found only 4 genes that overlapped between the SZgene database and our eQTL gene list. Furthermore, only one candidate gene, *SRR*, was involved in a coexpression module that was associated with schizophrenia. *SRR* maps to chromosome 17p13 and encodes an enzyme that synthesizes D-serine from L-serine.[44] The D-serine is an endogenous co-agonist of the *N*-methyl-D-aspartate (NMDA) receptor.[45] Hypofunction of the NMDA receptor is potentially a major underlying pathophysiology of schizophrenia.[46,47] Our results support this hypothesis and suggest that abnormal NMDA receptor-mediated signaling may be influenced by genetic variations. A SNP, rs16952025, localized in an intron of the neighboring gene, *SMG6*, was significantly associated with the expression level of *SRR*. However, there was no significant association between this SNP and the expression level of *SMG6*. Several post-mortem studies have examined levels of *SRR* mRNA and serine racemase protein in schizophrenia[48] and found abnormalities in schizophrenia, although the results have been inconsistent. Although *SRR* mRNA levels appear to be unchanged in frontal cortex of schizophrenia,[41] the protein levels have been reported to be either decreased,[49] increased[50] or unchanged.[51] The inconsistent results are most likely because of different methodologies, different cohorts (often with small numbers) and the different brain areas used. Consequently further study will be required in the future when larger cohorts become available to confirm changes in *SRR* levels in the brain of subjects with mental illness.

Our comprehensive brain eQTL analysis functionally validated only 4 genes out of 39 candidate genes positively identified in the SZgene meta-analysis. We were unable to identify any significant associations between the expression levels of the remaining genes and *cis* SNPs in any of the brain regions we tested. In fact, the 39 candidate genes were derived from 1008 candidate genes that were obtained from 1727 original genetic association studies. Such a low functional validation rate raises the possibility that the current case–control genetic association studies may not effectively

identify genetic variations that underlie the etiology of schizophrenia. However, there are other reasons that may contribute to a low functional validation rate. For example, the probes on the microarray platforms used to analyze gene expression in this study mainly bind to sequences in the 3′-untranslated regions and do not distinguish between various alternative splicing isoforms. Indeed, tissue-specific alternative RNA-splicing is very predominant in the brain.[52] Furthermore, intronic SNPs can be associated with the altered expression of specific alternative splicing isoforms of certain schizophrenia candidate genes, for example, *ErbB4* and *GRM3*.[53,54] Therefore, comprehensive expression profiling that includes various alternative splicing isoforms using deep mRNA-sequencing technology may aid in the identification of novel *cis* eQTL genes in human post-mortem brain tissues in the future.

The frontal cortex is one of the most thoroughly examined brain regions in post-mortem studies and many neuropathology abnormalities have been identified in this region in schizophrenia.[55,56] Previous gene expression microarray studies in the frontal cortex identified several biological processes that were overrepresented in the genes differentially expressed between schizophrenia and normal controls; for example, decreased presynaptic function, abnormal mitochondrial function and altered expression of apoptosis-related genes are all major findings from microarray studies of frontal cortex in schizophrenia.[33,34,57] However, glutamatergic transmission, amino acid metabolism, proteolysis and protein targeting were all overrepresented in the eQTL genes in the frontal cortex in our current study. Thus, the abnormalities described in the biological pathways from the eQTL study may be more directly related to genetic variation, whereas the pathways identified by gene expression studies are likely to be influenced by factors in addition to genetic variation, including epigenetics and environmental factors.

Although our study reveals a number of associations between *cis* SNPs and gene expression in multiple brain regions, the results should be interpreted with caution. First, the SNC, which we used for the initial eQTL analyses, contains a relatively small sample size ($N = 56$). Small sample size is known to generate higher false-positive associations as well as to be a cause of low detection power in genome-wide association analysis. Thus, the *cis* eQTL results from frontal cortex, cerebellum, thalamus, and temporal cortex using SNC samples should be viewed as exploratory. However, we subsequently performed a second analysis using an independent collection (AC) with a larger sample size ($N = 101$). Our previous power analysis using AC as well as a previous eSNP association study indicated that a relatively small sample size ($N = 100$) has $>80\%$ power to detect an association of gene expression traits with moderate effect size ($R^2 = 0.35$).[11,58] We therefore attempted to replicate the results of *cis* eQTLs in frontal cortex from SNC using the AC data. A total of 34 (64%) of the *cis* eQTL genes identified in the SNC frontal cortex data were also found in the AC frontal data. However, we identified 450 additional *cis* eQTL genes in the AC frontal cortex samples, which were not identified in the SNC frontal samples, suggesting that some significant *cis* eQTL genes may have been missed in the SNC analysis.

Second, using whole tissues for gene expression traits may dilute the effect of some genetic variants that may only act on cell type-specific gene expression. Although this phenomena has not been explored in the brain, there are numerous cell type-specific abnormalities in the brain of subjects with psychiatric disorders.[32,59–61] Thus, the use of cell type-specific expression traits in future studies may increase the power to identify *cis* eQTLs in the brain.

In this study, we investigated the associations between SNPs and gene expression in various human brain regions. Although previous brain eQTL studies focused on cortex,[8,62] we have extended the analysis to include the hippocampus, thalamus and cerebellum. These data can be used to identify genetic variations associated with psychiatric disorder and can be used to identify genetic variations that affect neuropathological abnormalities and gene expression changes. As we show in this study, the data can be used to functionally validate candidate genes to determine if they are affecting changes in gene expression in subjects with neuropsychiatric disorders. In order to facilitate further studies, we have integrated the genome-wide eQTL results from this study into the SNCID, which is a web-based database that also includes 1747 neuropathological markers measured in the same SNC samples. The update will allow users to investigate associations between SNPs and genes of interests in various brain regions and to further explore associations between SNPs and neuropathological markers and gene expression traits that are correlated with neuropathological markers in the various brain regions of subject with major psychiatric disorders.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements. We thank all the investigators who generated the original data in the SNCID, and their many collaborators, who made this database possible. We also thank all the technicians in the SMRI brain laboratory who prepared the brain tissues and extracted the RNA and DNA from the tissues. We specially thank the Keymind Company for their technical assistance with the database, in particular Marvin Suo. We thank Dr Horvath for helpful comments on network analysis using the WGCNA. HC and DL were supported by the World Class University program (R32-2008-000-10218-0) of the Ministry of Education, Science and Technology through the National Research Foundation of Korea.

1. Mueser KT, McGurk SR. Schizophrenia. *Lancet* 2004; **363**: 2063–2072.
2. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry* 2000; **157**: 1552–1562.
3. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
4. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D *et al.* Common variants conferring risk of schizophrenia. *Nature* 2009; **460**: 744–747.
5. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K *et al.* Whole-genome association study of bipolar disorder. *Mol Psychiatry* 2008; **13**: 558–569.
6. Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 2008; **40**: 827–834.
7. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC *et al.* A genome-wide association study of global gene expression. *Nat Genet* 2007; **39**: 1202–1207.
8. Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L *et al.* A survey of genetic human cortical gene expression. *Nat Genet* 2007; **39**: 1494–1499.
9. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C *et al.* Population genomics of human gene expression. *Nat Genet* 2007; **39**: 1217–1224.
10. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J *et al.* Genetics of gene expression and its effect on disease. *Nature* 2008; **452**: 423–428.

11. Kim S, Webster MJ. Integrative genome-wide association analysis of cytoarchitectural abnormalities in the prefrontal cortex of psychiatric disorders. *Mol Psychiatry* 2010; **16**: 452–461.
12. Kim S, Webster MJ. The stanley neuropathology consortium integrative database: a novel, web-based tool for exploring neuropathological markers in psychiatric disorders and the biological processes associated with abnormalities of those markers. *Neuropsychopharmacology* 2010; **35**: 473–482.
13. Tkachev D, Mimmack ML, Ryan MM, Wayland M, Freeman T, Jones PB *et al.* Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet* 2003; **362**: 798–805.
14. Aston C, Jiang L, Sokolov BP. Microarray analysis of postmortem temporal cortex from patients with schizophrenia. *J Neurosci Res* 2004; **77**: 858–866.
15. Torrey EF, Webster M, Knable M, Johnston N, Yolken RH. The stanley foundation brain collection and neuropathology consortium. *Schizophr Res* 2000; **44**: 151–155.
16. Iwamoto K, Bundo M, Kato T. Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Hum Mol Genet* 2005; **14**: 241–253.
17. Ryan MM, Lockstone HE, Huffaker SJ, Wayland MT, Webster MJ, Bahn S. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol Psychiatry* 2006; **11**: 965–978.
18. Higgs BW, Elashoff M, Richman S, Barci B. An online database for brain disease research. *BMC Genomics* 2006; **7**: 70.
19. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 2007; **3**: 1724–1735.
20. Liu C, Cheng L, Badner JA, Zhang D, Craig DW, Redman M *et al.* Whole-genome association mapping of gene expression in the human prefrontal cortex. *Mol Psychiatry* 2010; **15**: 779–784.
21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
22. Pei YF, Li J, Zhang L, Papasian CJ, Deng HW. Analyses and comparison of accuracy of different genotype imputation methods. *PLoS One* 2008; **3**: e3551.
23. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9**: 559.
24. Gargalovic PS, Imura M, Zhang B, Gharavi NM, Clark MJ, Pagnon J *et al.* Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. *Proc Natl Acad Sci USA* 2006; **103**: 12741–12746.
25. Hu Z, Mellor J, Wu J, DeLisi C. VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics* 2004; **5**: 17.
26. Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003; **4**: P3.
27. Mirnics K, Levitt P, Lewis DA. DNA microarray analysis of postmortem brain tissue. *Int Rev Neurobiol* 2004; **60**: 153–181.
28. Fare TL, Coffey EM, Dai H, He YD, Kessler DA, Kilian KA *et al.* Effects of atmospheric ozone on microarray data quality. *Anal Chem* 2003; **75**: 4672–4675.
29. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010; **11**: 733–739.
30. McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). *Biostatistics* 2010; **11**: 242–253.
31. Torkamani A, Dean B, Schork NJ, Thomas EA. Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res* 2010; **20**: 403–412.
32. Uranova NA, Vostrikov VM, Orlovskaya DD, Rachmanova VI. Oligodendroglial density in the prefrontal cortex in schizophrenia and mood disorders: a study from the Stanley Neuropathology Consortium. *Schizophr Res* 2004; **67**: 269–275.
33. Mirnics K, Middleton FA, Marquez A, Lewis DA, Levitt P. Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron* 2000; **28**: 53–67.
34. Kim S, Webster MJ. Correlation analysis between genome-wide expression profiles and cytoarchitectural abnormalities in the prefrontal cortex of psychiatric disorders. *Mol Psychiatry* 2010; **15**: 326–336.
35. McInnes LA, Lauriat TL. RNA metabolism and dysmyelination in schizophrenia. *Neurosci Biobehav Rev* 2006; **30**: 551–561.
36. Huang HS, Matevossian A, Whittle C, Kim SY, Schumacher A, Baker SP *et al.* Prefrontal dysfunction in schizophrenia involves mixed-lineage leukemia 1-regulated histone methylation at GABAergic gene promoters. *J Neurosci* 2007; **27**: 11254–11262.
37. Millar JK, Pickard BS, Mackie S, James R, Christie S, Buchanan SR *et al.* DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* 2005; **310**: 1187–1191.
38. Wong ML, Whelan F, Deloukas P, Whittaker P, Delgado M, Cantor RM *et al.* Phosphodiesterase genes are associated with susceptibility to major depression and antidepressant treatment response. *Proc Natl Acad Sci USA* 2006; **103**: 15124–15129.
39. Fatemi SH, King DP, Reutiman TJ, Folsom TD, Laurence JA, Lee S *et al.* PDE4B polymorphisms and decreased PDE4B expression are associated with schizophrenia. *Schizophr Res* 2008; **101**: 36–49.
40. Fatemi SH, Reutiman TJ, Folsom TD, Lee S. Phosphodiesterase-4A expression is reduced in cerebella of patients with bipolar disorder. *Psychiatr Genet* 2008; **18**: 282–288.
41. Fatemi SH, Folsom TD, Reutiman TJ, Vazquez G. Phosphodiesterase signaling system is disrupted in the cerebella of subjects with schizophrenia, bipolar disorder, and major depression. *Schizophr Res* 2009; **119**: 266–267.
42. Numata S, Iga J, Nakataki M, Tayoshi S, Taniguchi K, Sumitani S *et al.* Gene expression and association analyses of the phosphodiesterase 4B (PDE4B) gene in major depressive disorder in the Japanese population. *Am J Med Genet B Neuropsychiatr Genet* 2009; **150B**: 527–534.
43. Hennah W, Porteous D. The DISC1 pathway modulates expression of neurodevelopmental, synaptogenic and sensory perception genes. *PLoS One* 2009; **4**: e4906.
44. Wolosker H, Blackshaw S, Snyder SH. Serine racemase: a glial enzyme synthesizing D-serine to regulate glutamate-N-methyl-D-aspartate neurotransmission. *Proc Natl Acad Sci USA* 1999; **96**: 13409–13414.
45. Leeson PD, Iversen LL. The glycine site on the NMDA receptor: structure-activity relationships and therapeutic potential. *J Med Chem* 1994; **37**: 4053–4067.
46. Olney JW, Newcomer JW, Farber NB. NMDA receptor hypofunction model of schizophrenia. *J Psychiatr Res* 1999; **33**: 523–533.
47. Belforte JE, Zsiros V, Sklar ER, Jiang Z, Yu G, Li Y *et al.* Postnatal NMDA receptor ablation in corticolimbic interneurons confers schizophrenia-like phenotypes. *Nat Neurosci* 2010; **13**: 76–83.
48. Verrall L, Burnet PW, Betts JF, Harrison PJ. The neurobiology of D-amino acid oxidase and its involvement in schizophrenia. *Mol Psychiatry* 2010; **15**: 122–137.
49. Bendikov I, Nadri C, Amar S, Panizzutti R, De Miranda J, Wolosker H *et al.* A CSF and postmortem brain study of D-serine metabolic parameters in schizophrenia. *Schizophr Res* 2007; **90**: 41–51.
50. Verrall L, Walker M, Rawlings N, Benzel I, Kew JN, Harrison PJ *et al.* d-Amino acid oxidase and serine racemase in human brain: normal distribution and altered expression in schizophrenia. *Eur J Neurosci* 2007; **26**: 1657–1669.
51. Steffek AE, Haroutunian V, Meador-Woodruff JH. Serine racemase protein expression in cortex and hippocampus in schizophrenia. *NeuroReport* 2006; **17**: 1181–1185.
52. Grabowski PJ, Black DL. Alternative RNA splicing in the nervous system. *Prog Neurobiol* 2001; **65**: 289–308.
53. Sartorius LJ, Weinberger DR, Hyde TM, Harrison PJ, Kleinman JE, Lipska BK. Expression of a GRM3 splice variant is increased in the dorsolateral prefrontal cortex of individuals carrying a schizophrenia risk SNP. *Neuropsychopharmacology* 2008; **33**: 2626–2634.
54. Law AJ, Kleinman JE, Weinberger DR, Weickert CS. Disease-associated intronic variants in the ErbB4 gene are related to altered ErbB4 splice-variant expression in the brain in schizophrenia. *Hum Mol Genet* 2007; **16**: 129–141.
55. Knable MB, Barci BM, Bartko JJ, Webster MJ, Torrey EF. Molecular abnormalities in the major psychiatric illnesses: Classification and Regression Tree (CRT) analysis of post-mortem prefrontal markers. *Mol Psychiatry* 2002; **7**: 392–404.
56. Knable MB, Torrey EF, Webster MJ, Bartko JJ. Multivariate analysis of prefrontal cortical data from the Stanley Foundation Neuropathology Consortium. *Brain Res Bull* 2001; **55**: 651–659.
57. Prabakaran S, Swatton JE, Ryan MM, Huffaker SJ, Huang JT, Griffin JL *et al.* Mitochondrial dysfunction in schizophrenia: evidence for compromised brain metabolism and oxidative stress. *Mol Psychiatry* 2004; **9**: 684–697, 643.
58. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping determinants of human gene expression by regional and genome-wide association. *Nature* 2005; **437**: 1365–1369.
59. Beasley CL, Zhang ZJ, Patten I, Reynolds GP. Selective deficits in prefrontal cortical GABAergic neurons in schizophrenia defined by the presence of calcium-binding proteins. *Biol Psychiatry* 2002; **52**: 708–715.
60. Lewis DA, Hashimoto T, Morris HM. Cell and receptor type-specific alterations in markers of GABA neurotransmission in the prefrontal cortex of subjects with schizophrenia. *Neurotox Res* 2008; **14**: 237–248.
61. Vostrikov VM, Uranova NA, Orlovskaya DD. Deficit of perineuronal oligodendrocytes in the prefrontal cortex in schizophrenia and mood disorders. *Schizophr Res* 2007; **94**: 273–280.
62. Webster JA, Gibbs JR, Clarke J, Ray M, Zhang W, Holmans P *et al.* Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet* 2009; **84**: 445–458.

Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)