

UCLA

UCLA Previously Published Works

Title

Identification of schizophrenia-associated loci by combining DNA methylation and gene expression data from whole blood

Permalink

<https://escholarship.org/uc/item/9wx1905f>

Journal

European Journal of Human Genetics, 23(8)

ISSN

1018-4813

Authors

van Eijk, Kristel R
de Jong, Simone
Strengman, Eric
et al.

Publication Date

2015-08-01

DOI

10.1038/ejhg.2014.245

Peer reviewed

ARTICLE

Identification of schizophrenia-associated loci by combining DNA methylation and gene expression data from whole blood

Kristel R van Eijk¹, Simone de Jong², Eric Strengman², Jacobine E Buizer-Voskamp¹, René S Kahn¹, Marco P Boks¹, Steve Horvath³ and Roel A Ophoff^{*,1,2,3}

Emerging evidence suggests that schizophrenia (SZ) susceptibility involves variation at genetic, epigenetic and transcriptome levels. We describe an integrated approach that leverages DNA methylation and gene expression data to prioritize genetic variation involved in disease. DNA methylation levels were obtained from whole blood of 260 SZ patients and 250 unaffected controls of which a subset with gene expression data was available. By assessing DNA methylation and gene expression in cases and controls, we identified 432 CpG sites with differential methylation levels that are associated with differential gene expression. We hypothesized that genetic factors involved in these methylation levels may be associated with the genetic risk of SZ susceptibility. To test this hypothesis, we used results from the Psychiatric Genomics Consortium SZ genome-wide association study (GWAS). We observe an enrichment of SZ-associated SNPs in the mQTLs of which the associated CpG site is also correlated with differential gene expression in SZ. While this enrichment was already apparent when using nominal significant thresholds, enrichment was even more pronounced when applying more stringent significance levels. One locus, previously identified as susceptibility locus in a SZ GWAS, involves SNP rs1191514:C>T, which regulates DNA methylation of calcium homeostasis modulator 1 that is also associated with differential gene expression in patients. Overall, our results suggest that epigenetic variation plays an important role in SZ susceptibility and that the integration of analyses of genetic, epigenetic and gene expression profiles may be a biologically meaningful approach for identifying disease susceptibility loci, even when using whole blood data in studies of brain-related disorders.

European Journal of Human Genetics (2015) 23, 1106–1110; doi:10.1038/ejhg.2014.245; published online 26 November 2014

INTRODUCTION

Although a number of susceptibility loci for psychiatric disorders have been identified through genome-wide association studies (GWAS), the vast majority of the estimated heritability of these traits remains unexplained. Schizophrenia (SZ), for example, is a common polygenic mental disorder affecting about 1% of the population and has an estimated heritability of 70–80%,^{1–3} but only a small fraction of the heritability can be attributed to known susceptibility loci.^{4,5} Very large numbers of samples are needed to comprehensively identify the hundreds or possibly thousands of genetic loci involved in SZ susceptibility.⁶ It has been suggested that epigenetic variation might be partly responsible for the missing heritability^{7,8} and be involved in phenotypic variation.^{9–11} Recent evidence shows that SZ susceptibility loci are enriched for gene expression QTLs,¹² and a similar enrichment exists for QTLs affecting DNA methylation and expression in bipolar disorder.¹³ Given these results, we hypothesized that intersecting disease-related gene expression data with disease-related methylation data might lead to identification of genetic susceptibility loci. To test this hypothesis, we combined available whole blood DNA methylation and gene expression data of SZ patients and healthy controls. Using the results of the Psychiatric Genomics Consortium (PGC) SZ mega

GWAS,¹⁴ we examined whether methylation QTLs (mQTLs) are enriched for disease susceptibility loci. While previous studies examined gene expression and DNA methylation separately and in different ways,^{12,13} we considered both DNA methylation and gene expression data simultaneously in the same SZ cases and controls. Previous studies investigating enrichment of QTL signals (involving either gene expression or DNA methylation) focused on SZ-related SNPs with genome-wide significant evidence of association. We combined epigenome and transcriptome levels of information and did not restrict ourselves to only those loci with prior evidence of genome-wide association. Our results show that the combined use of gene expression and methylation data outperforms either data modality when it comes to identifying disease susceptibility loci, even in a relatively small sample size compared with GWAS.

SUBJECTS AND METHODS

An outline of the approach is provided in Figure 1. Genome-wide genotype data and DNA methylation data were obtained from whole blood of 260 SZ patients and 250 control subjects; for a subset of 120 cases and 120 controls, whole blood gene expression data were available as well. A detailed description of the samples, procedures and quality control steps is provided in the Supplementary Material. The data is available via NCBI/GEO under accession

¹Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; ²UCLA Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles (UCLA), Los Angeles, CA, USA; ³Department of Human Genetics, University of California Los Angeles (UCLA), Los Angeles, CA, USA

*Correspondence: Professor RA Ophoff, UCLA Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles (UCLA), Gonda #4357C, Box 951761, 695 Charles E. Young Drive South, Los Angeles, CA 90095-1761, USA. Tel: +1 310 794 9602; Fax: +1 310 794 9613; E-mail: ROphoff@mednet.ucla.edu

Received 21 August 2013; revised 28 August 2013; accepted 30 September 2014; published online 26 November 2014

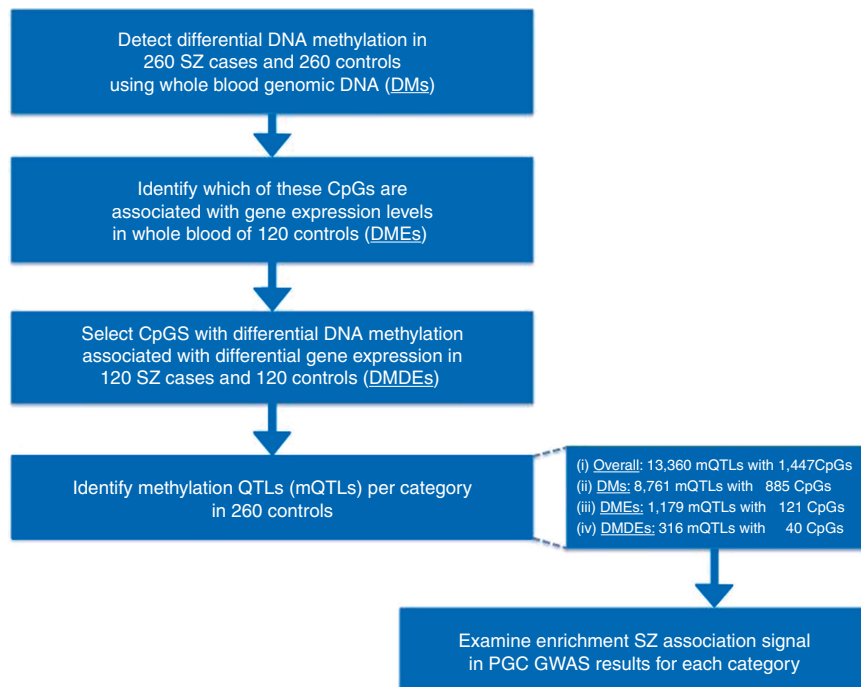


Figure 1 Outline of study with different filtering steps (see main text for details).

number GSE41037 and GSE38484. First, we examined differential methylation levels (after quantile normalization) between cases and controls. DNA methylation levels with low variability (a Beta range <0.1) were excluded to focus on CpGs with reasonably large biological variation. We used a linear model (the limma package in R¹⁵) to regress methylation values on disease status, gender and age.¹⁶ FDR correction at the 5% level was applied to correct for multiple testing.

Next, for methylation levels that differed between cases and controls (referred to as differentially methylated loci, or abbreviated DMs), association with gene expression was investigated using a genome-wide DNA methylation–gene expression association study as described previously.¹⁷ We focused on CpG sites with *cis* effects on expression within 500 kb interval. Locus-specific correction for multiple testing was applied for the number of expression probes within each *cis* region of that methylation probe (ie, 0.05 divided by the number of expression probes per *cis* area). This relatively lenient significance threshold was chosen to ensure a sufficient number of probes for subsequent filtering steps and analysis. The subset of DMs associated with gene expression is referred to as DMEs, as determined in control samples. A subset of the DMEs is not only associated with gene expression levels in general but is associated with transcript levels of differentially expressed genes when comparing SZ patients and controls as described previously.¹⁸ In short, we conducted linear regression to detect differences in expression levels between cases and controls after FDR correction at the 5% level, with age and gender as covariates.¹⁸ The DMEs associated with differential gene expression in SZ are referred to as Differential Methylation regions with Differential gene Expression (DMDEs).

To identify mQTLs, we used a multivariate linear regression model for regressing all methylation values (dependent variable) on the SNPs (independent variable) with disease status, age and gender as covariates using PLINK.¹⁹ We defined a SNP as *cis*-acting if it was significantly associated ($P < 5.0e-08$) with a methylation probe within 500 kb. Next, four groups of SNPs with a minimum minor allele frequency of 0.05 were generated with increasing functional relevance with regard to DNA methylation and gene expression: (i) SNPs representing all mQTLs regardless of differential methylation between patients and controls; (ii) SNPs from the previous step that represent mQTLs with differential methylation between cases and controls (DMs); (iii) SNPs from the previous step for which the associated methylation level is associated with gene expression (DMEs); and finally (iv) SNPs selected from the previous

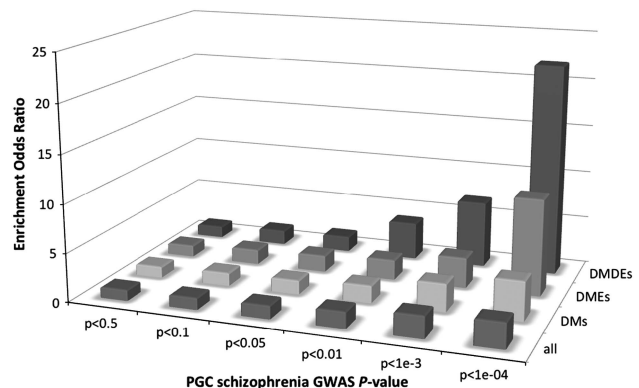


Figure 2 Enrichment of genetic association with schizophrenia. For each of the four categories of mQTLs in the whole blood case/control data set we observed enrichment of genetic association signal based on PGC schizophrenia genome-wide association study findings.¹⁴ The four categories are (i) all mQTLs without prior selection, (ii) DMs representing mQTLs with differential methylation between cases and controls, (iii) DMEs representing mQTLs with differential DNA methylation between cases and controls associated with gene expression and (iv) DMDEs representing mQTLs with differential DNA methylation that is associated with differential gene expression between cases and controls. Detailed values of odds ratios, significance levels and confidence intervals are provided in Supplementary Table 1.

step that are associated with differential methylation that is associated with differential gene expression between patients and controls (DMDEs). After linkage disequilibrium (LD)-pruning the SNPs with an R^2 of 0.2 using PLINK, to exclude possible signal bias and enrich for independent genetic signal, we extracted the SNPs from the PGC mega GWAS results with their association signal with SZ.

Based on the significance values observed in the PGC SZ GWAS, we categorized the four different SNP lists and examined the observed/expected ratios for different thresholds ranging from $P < 0.5$ to $< 1.0e-04$.

RESULTS

We examined whether genetic factors involved in differential methylation and gene expression in SZ are enriched for SZ susceptibility alleles. The results of the steps are shown in Figure 1. In total, 11 320 CpGs were differentially methylated in SZ cases *versus* controls (5% FDR correction). Of these, 1095 CpGs are associated with 1226 transcripts in *cis* (after locus-specific correction). The 1226 transcripts were examined for differential expression levels between patients and controls based on data from a previous study of our group.¹⁸ This step resulted in the identification of 391 transcripts (after 5% FDR correction); these 391 transcripts are associated with 432 CpG sites. Information about the differentially methylated probes is shown in Supplementary Table 1.

We subsequently calculated the mQTLs associated with CpGs in *cis*, and SNPs with a P -value $< 5.0e-08$ were grouped by the biological relevance of the associated CpG site (see Subjects and methods). Figure 1 shows the number of associated CpGs and SNPs, and the statistics of the associations are provided in Supplementary Table 2.

After calculating the observed/expected ratio of the mQTLs per PGC GWAS P -value threshold, our results show a significant pattern of enrichment of PGC SZ association signal by adding DNA methylation and gene expression data (Figure 2). The strongest

enrichment was observed for SNPs associated with DMDEs ($P = 0.0041$ with $OR = 22$ and $95\%CI = 2.60-83.7$; Fisher's exact test).

We identified one DMDE locus on chromosome 10 with genome-wide significant evidence of association with SZ. SNP rs11191514: C>T (chr10:hg19:c.104773364C>T) at this locus represents an mQTL associated with differential DNA methylation between cases and controls that is correlated with the expression of a gene that is also differentially expressed between cases and controls (Figures 3a-c). In addition, this SNP is in perfect LD with rs11191580: C>T (chr10:hg19:t.104906211C>T), a SNP with significant evidence of association with SZ according to the PGC study (PGC $P = 2.2e-08$). SNP rs11191514: C>T is also strongly associated with SZ (PGC $P = 8.7e-08$) and likely regulates DNA methylation in our data set ($P = 1.4e-14$, Figure 3d). SNP rs11191514: C>T is located in *CNNM2*, while the methylation probe associated with this SNP is located at the gene Calcium Homeostasis Modulator 1 (*CALHM1*), some 630 kb apart (Supplementary Figure 1). *CALHM1* regulates Ca^{2+} concentrations²⁰ and is highly expressed in the hippocampus,²⁰ which is implicated in SZ.²¹ Recently, a GWAS identified this whole locus as susceptibility locus for SZ, including the genes *CALHM1* and *CNNM2*, $P = 3.7e-13$.²²

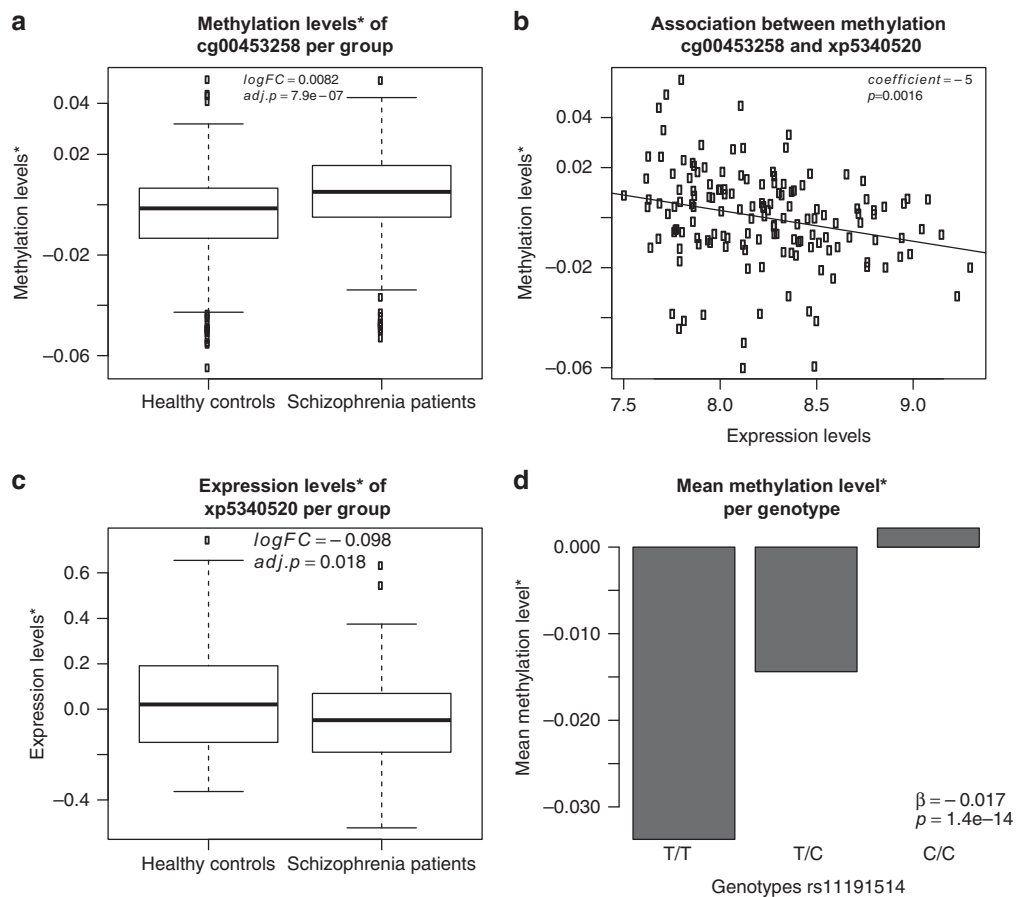


Figure 3 Correlations between rs11191514:C>T methylation/gene expression and disease of locus with previously reported genome-wide significant evidence of association with schizophrenia. (a) CpG site cg00453258 (representing gene *CALHM1*) is increased in schizophrenia cases *versus* controls; (b) this CpG site is negatively associated with gene expression, transcript probe xp5340520; (c) expression of this transcript is reduced in schizophrenia cases; (d) the mean methylation level per genotype. In the PGC schizophrenia GWAS result, the T allele is a protective allele with an OR of 0.82 (see Supplementary Table 1) with increased allele frequency in controls, and in our data set associated with a decreased methylation level. This fits with the increased methylation levels in schizophrenia. *To generate these plots the residuals of the methylation values were taken after correction for age and gender in a and b, and for expression in c.

In addition to rs11191514:C>T, we identified four additional independent DMDE SNPs associated with SZ with a significance level of $P < 0.01$ (Supplementary Table 3), that is, these SNPs are at best suggestive of a susceptibility locus for SZ.

These additional loci highlighted in our study include proline-rich transmembrane protein 1 (*PRRT1*), *HLA-C* and *MRPL41*. The CpG sites of these genes are all differentially methylated in patients, and are associated with genotype as well as with transcripts that are differentially expressed in patients. *PRRT1* and *HLA-C* are both located within the MHC region on the short arm of chromosome 6, a region with the most significant association with SZ.^{14,23} Little is known about the function of *PRRT1* and this finding suggests that further study of its involvement in SZ is warranted. *HLA-C* belongs to the HLA class I molecules that have a central role in the immune system and is located within the chromosomal region first implicated in SZ.¹¹ Finally, *MRPL41* is a mitochondrial ribosomal protein and has an important role in cell growth suppression in association with p53 and p27Kip1.²⁴ It is not clear how this ribosomal protein may be involved in the etiology of SZ. As the MHC region is overrepresented in the PGC SZ GWAS findings, we were concerned about a possible bias in our analysis. However, when we performed the same analyses without chromosome 6, we still observed an enrichment of genetic signal associated with SZ ($P = 0.0037$ with $OR = 23$ and $95\%CI = 2.8-89.2$; Fisher's exact test). This suggests that functional enrichment analysis by using gene expression and DNA methylation from whole blood is not dependent on the MHC region.

DISCUSSION

We investigated the enrichment of mQTLs for SZ susceptibility alleles using genotype and DNA methylation data obtained from whole blood of >500 SZ cases and controls, and of which gene expression data were available from a subset. We detected over 10 000 CpG sites that are differentially methylated in SZ patients. A subset of these sites ($n = 1095$) was associated with gene expression with 50% of these transcripts also showing differential gene expression in SZ. In addition, SNPs associated with differential methylation levels are enriched for SZ susceptibility alleles. This enrichment was even stronger if the differential methylation level associated with these SNPs were also associated with differential expression levels. While this enrichment was already apparent when using nominal significant thresholds, enrichment was even more pronounced when applying more stringent significance levels. Previous studies have shown that top GWAS findings are enriched for mQTLs and eQTLs in bipolar disorder¹⁵ and for eQTLs in SZ.¹² In accordance, here we show that mQTLs are enriched for SZ SNPs. In addition, we used different layers of genomic information to identify mQTLs associated with CpGs of functional relevance and found increasing enrichment. Our findings show that genetic variation affecting DNA methylation, that is associated with gene expression, has an important role in SZ susceptibility.

As SZ is primarily a brain-related disease, a limiting factor of our study may be the use of whole blood. However, our findings show that even when using blood, we observe enrichment of SZ-associated alleles, indicating that blood might be a reasonable surrogate for our approach. Replication of these analyses in brain tissue would be useful to understand the extent of enrichment of disease-specific signal when combining different genomic layers for prioritizing genomic loci. Another potential expansion of this study includes the obtaining of allele-specific methylation and gene expression information, which could provide more insight into the precise mechanism of *cis*-acting regulating SNPs and their effects on methylation and gene expression.

Finally, the relationship between genetic variation, DNA methylation, gene expression and disease susceptibility is complex and warrants further study.

In summary, we identified biologically plausible SZ susceptibility loci in whole blood in a relatively small sample of <600 subjects. We demonstrate that enrichment of genetic data using different layers of genomic information may be an efficient approach to identify disease susceptibility loci for neuropsychiatric traits. While our results are supportive of an important role of epigenetic regulation in SZ, we expect that this integrated approach based on blood DNA methylation and gene expression data from the same subjects may help prioritize SNPs from other GWAS as well.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the patients and controls for participating in the study. This study was supported by National Institutes of Health (NIH) grants R01 DA028526, R01 NS058980 and R01 MH 078075 (to RAO). We thank Dr Kim Staats for critical reading of the manuscript.

- 1 Lee SH, DeCandia TR, Ripke S *et al*: Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 2012; **44**: 247–250.
- 2 Sullivan PF, Daly MJ, O'Donovan M: Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet* 2012; **13**: 537–551.
- 3 Sullivan PF, Kendler KS, Neale MC: Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 2003; **60**: 1187–1192.
- 4 Purcell SM, Wray NR, Stone JL *et al*: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
- 5 So HC, Gui AH, Cherny SS, Sham PC: Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet Epidemiol* 2011; **35**: 310–317.
- 6 Manolio TA, Collins FS, Cox NJ *et al*: Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–753.
- 7 Petronis A: Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* 2010; **465**: 721–727.
- 8 Labrie V, Pai S, Petronis A: Epigenetics of major psychosis: progress, problems and perspectives. *Trends Genet* 2012; **28**: 427–435.
- 9 Feinberg AP, Irizarry RA: Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci USA* 2010; **107**(Suppl 1): 1757–1764.
- 10 Feinberg AP: Phenotypic plasticity and the epigenetics of human disease. *Nature* 2007; **447**: 433–440.
- 11 Bjornsson HT, Fallin MD, Feinberg AP: An integrated epigenetic and genetic approach to common human disease. *Trends Genet* 2004; **20**: 350–358.
- 12 Richards AL, Jones L, Moskva V *et al*: Schizophrenia susceptibility alleles are enriched for alleles that affect gene expression in adult human brain. *Mol Psychiatry* 2012; **17**: 193–201.
- 13 Gamazon ER, Badner JA, Cheng L *et al*: Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol Psychiatry* 2012; **18**: 340–346.
- 14 Ripke S, Sanders AR, Kendler KS *et al*: Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**: 969–976.
- 15 Smyth GK: Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004; **3**: Article3.
- 16 Horvath S, Zhang Y, Langfelder P *et al*: Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol* 2012; **13**: R97.
- 17 van Eijk KR, de Jong S, Boks MP *et al*: Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics* 2012; **13**: 636.
- 18 de Jong S, van Eijk KR, Zeegers DW *et al*: Expression QTL analysis of top loci from GWAS meta-analysis highlights additional schizophrenia candidate genes. *Eur J Hum Genet* 2012; **20**: 1004–1008.
- 19 Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 20 Dreses-Werringloer U, Lambert JC, Vingtdex V *et al*: A polymorphism in CALHM1 influences Ca²⁺ homeostasis, Aβ levels, and Alzheimer's disease risk. *Cell* 2008; **133**: 1149–1161.

- 21 Harrison PJ: The hippocampus in schizophrenia: a review of the neuropathological evidence and its pathophysiological implications. *Psychopharmacology* 2004; **174**: 151–162.
- 22 Ripke S, O'Dushlaine C, Chambert K *et al*: Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* 2013; **45**: 1150–1159.
- 23 Stefansson H, Ophoff RA, Steinberg S *et al*: Common variants conferring risk of schizophrenia. *Nature* 2009; **460**: 744–747.
- 24 Yoo YA, Kim MJ, Park JK *et al*: Mitochondrial ribosomal protein L41 suppresses cell growth in association with p53 and p27Kip1. *Mol Cell Biol* 2005; **25**: 6603–6616.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)