npg

ORIGINAL ARTICLE

Polygenic determinants of white matter volume derived from GWAS lack reproducibility in a replicate sample

S Papiol[1,2,3,7], M Mitjans[1,3,4,7], F Assogna[5], F Piras[5], C Hammer[1], C Caltagirone[6], B Arias[3,4], H Ehrenreich[1,2] and G Spalletta[5]

A recent publication reported an exciting polygenic effect of schizophrenia (SCZ) risk variants, identified by a large genome-wide association study (GWAS), on total brain and white matter volumes in schizophrenic patients and, even more prominently, in healthy subjects. The aim of the present work was to replicate and then potentially extend these findings. According to the original publication, polygenic risk scores—using single nucleotide polymorphism (SNP) information of SCZ GWAS—(polygenic SCZ risk scores; PSS) were calculated in 122 healthy subjects, enrolled in a structural magnetic resonance imaging (MRI) study. These scores were computed based on *P*-values and odds ratios available through the Psychiatric GWAS Consortium. In addition, polygenic white matter scores (PWM) were calculated, using the respective SNP subset in the original publication. None of the polygenic scores, either PSS or PWM, were found to be associated with total brain, white matter or gray matter volume in our replicate sample. Minor differences between the original and the present study that might have contributed to lack of reproducibility (but unlikely explain it fully), are number of subjects, ethnicity, age distribution, array technology, SNP imputation quality and MRI scanner type. In contrast to the original publication, our results do not reveal the slightest signal of association of the described sets of GWAS-identified SCZ risk variants with brain volumes in adults. Caution is indicated in interpreting studies building on polygenic risk scores without replication sample.

INTRODUCTION

The Psychiatric Genome-Wide Association Study (GWAS) Consortium has recently published a large GWAS, including 9394 schizophrenia (SCZ) cases and 12 462 healthy controls, identifying common variants that contribute to SCZ susceptibility with relatively small odds ratios.[1] Besides these genome-wide significant loci, evidence has been accruing that a significant proportion of the risk for SCZ may lie in markers not achieving genome-wide significance in GWAS. For instance, a quantitative polygenic SCZ risk score (PSS) was calculated based on the nominally associated alleles in a discovery sample. This polygenic score explained up to 3% of variance in SCZ in a number of independent samples.[2] Several authors have subsequently explored whether such a polygenic effect might be associated not only with the disease but also with disease-relevant phenotypes. Although some studies described associations for example, with cognitive aging[3] or with a functional imaging substrate of working memory processing,[4] others reported a lack of association with psychosis dimensions[5] or intelligence.[6]

Along these lines, a recent study investigated the polygenic effect of SCZ-associated single nucleotide polymorphisms (SNPs) on brain volume (total brain, white matter and gray matter).[7] The proportion of variance explained by the PSS was around 5% for both total brain and white matter volumes. The authors subsequently generated a polygenic white matter score (PWM) out of those 2020 SCZ-related SNPs showing the most significant

associations with white matter volume in their sample. This PWM, that is, a final subset of 186 SNPs, influenced white matter volume most strongly. Importantly, effects were not only detected in patients but also in healthy control subjects, leading to the author's assumption that 'a relatively small subset of SCZ genetic risk variants is related to the normal development of white matter'.[7]

Considering the potentially high general importance of these findings for the complex genetics and the disease-unrelated biological grounds of adult brain dimensions, the present study has been designed to replicate (i) the PSS and (ii) the PWM effects on total brain and white matter volume in healthy subjects.

MATERIALS AND METHODS

Participants and inclusion criteria

The study was approved by the Santa Lucia Foundation Ethical Committee and performed in accordance with the Helsinki Declaration. After signing an informed consent form, 122 healthy subjects of Italian origin were included. Participants were consecutively recruited by local advertisement from universities, community recreational centers and hospitals (personnel). Inclusion criteria were age 20–80 years and suitability for magnetic resonance imaging (MRI) scanning. Exclusion criteria included: (i) suspicion of cognitive impairment (score ⩽26) or dementia based on Mini Mental State examination,[8] the Mental Deterioration Battery[9] and the NINCDS-ADRDA criteria for dementia;[10] (ii) subjective complaint of memory difficulties or of any other cognitive deficit, interfering with daily living;

[1]Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Göttingen, Germany; [2]DFG Research Center for Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB), Göttingen, Germany; [3]Centro de Investigaciones Biomédicas en Red de Salud Mental (CIBERSAM), Barcelona, Spain; [4]Unitat Antropologia, Departament de Biologia Animal, Facultat de Biologia and Institut de Biomedicina (IBUB), Universitat de Barcelona, Barcelona, Spain; [5]Fondazione Santa Lucia, Laboratorio di Neuropsichiatria, Roma, Italy and [6]Department of Neuroscience, IRCCS Fondazione Santa Lucia and Tor Vergata University, Roma, Italy. Correspondence: Professor H Ehrenreich, Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Hermann-Rein Strasse 3, 37075 Göttingen, Germany.
E-mail: ehrenreich@em.mpg.de
[7]These authors contributed equally to this work.
Received 13 September 2013; revised 4 November 2013; accepted 25 November 2013

2

(iii) presence of major non-stabilized medical illnesses (that is, non-stabilized diabetes, obstructive pulmonary disease or asthma, hematologic/oncologic disorders, vitamin B12 or folate deficiency, pernicious anemia, clinically significant and unstable active gastrointestinal, renal, hepatic, endocrine or cardiovascular disorders and recently treated hypothyroidism); (iv) known or suspected history of alcoholism, drug dependence and abuse, head trauma and mental disorders according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition–Text Revision (DSM-IV-TR) criteria (all subjects were interviewed by Structured Clinical Interview for DSM disorders–NonPatient edition (SCID-NP), DSM-IV-TR);[11] (v) presence of vascular brain lesions, brain tumor and/or marked cortical and subcortical atrophy on MRI scan. In particular, the presence, severity and location of vascular lesions were rated according to a protocol designed for the Rotterdam Scan Study.[12] Generally, they were considered present in cases of hyperintense lesions on both proton-density and T2-weighted (see image acquisition) and rated semiquantitatively as 0 (none), 1 (pencil-thin lining), 2 (smooth halo) or 3 (large confluent) for three separate regions; adjacent to frontal horns (frontal caps), adjacent to the wall of the lateral ventricles (bands) and adjacent to the occipital horns (occipital caps). The total vascular lesion load was calculated by adding the region-specific scores (range 0–9). In the present study, only participants rated 0–1 were included.

Image acquisition and processing

All participants underwent the same imaging protocol, which included standard clinical sequences (Fluid Attenuated Inversion Recovery (FLAIR), Proton Density-T2-weighted) and a whole-brain high-resolution T1-weighted sequence obtained in the sagittal plane using a modified driven equilibrium Fourier transform sequence (TE/TR = 2.4/7.92 ms, flip angle: 15°, voxel size: 1 mm^3) using a 3 T Allegra MR imager (Siemens, Erlangen, Germany) with a standard quadrature head coil. All planar sequence acquisitions were obtained in the plane of the anterior commissure––posterior commissure line. Particular care was taken to center the subject in the head coil and to restrain the subject's movements with cushions and adhesive medical tape. MRI-based quantification of cerebral volumes was performed using Freesurfer (v.4.05) software package (http://surfer.nmr.mgh.harvard.edu). Freesurfer includes a sophisticated automated segmentation algorithm, which delineates gross brain anatomy into a series of cortical and subcortical labels. The stream consists of five different stages, fully described elsewhere.[13,14] Initially, the MRI volumes were registered to the Talairach space and the output images were intensity normalized. At the next stage, the skull was automatically stripped off the three-dimensional anatomical data set by employing a hybrid method that uses both watershed algorithms and deformable surface models. At this stage, manual intervention is needed to visualize and edit areas of skull and the areas of cortex or cerebellum that should be corrected. After skull stripping, the output brain mask was labeled using a probabilistic atlas[15] and a complex algorithm combining information on image intensity, probabilistic atlas location and the local spatial relationships between structures.[16–18] For the purpose of this study, calculated volumes (in mm^3, subsequently converted to ml) for these labels were summed up to derive estimates of total gray and white matter volume as well as total brain volume. The FreeSurfer software and its documentation can be downloaded from http://surfer.nmr.mgh.harvard.edu.

Discovery sample data

Summary results including risk variants, their P-values and associated odds ratios from the recent international collaborative GWAS on 9394 SCZ cases and 12 462 healthy controls were collected from the Psychiatric GWAS Consortium.[1] Relevant information on the methods used by the consortium is described elsewhere.[1]

Genetic analysis of the target sample

Genotyping of the Italian target sample was performed using a semi-custom Axiom myDesign genotyping array (Affymetrix, Santa Clara, CA, USA), based on a CEU (Caucasian residents of European ancestry from UT, USA) marker backbone including 518 722 SNPs, and a custom marker set including 102 537 SNPs. The array was designed using the Axiom Design center (www.affymetrix.com), applying diverse selection criteria.[19] Genotyping was performed by Affymetrix on a GeneTitan platform. Several quality control steps were applied (SNP call rate >97%, Fisher's linear discriminant, heterozygous cluster strength offset, homozygote ratio offset). In a subsequent step, some SNPs were filtered out based on minor allele frequency < 0.02 or if the χ^2-test for Hardy–Weinberg equilibrium was < 1×10^{-6}.[19] For the present study, SNPs on chromosomes X and Y and mitochondrial DNA were excluded, leaving 574 505 SNPs for analyses. This SNP set was then used to calculate multidimensional scaling components in order to control for population stratification. Similarly, the inbreeding coefficient was calculated by making use of a pruned version of this SNP set. MDS components and inbreeding coefficients were calculated using PLINK.[20]

Derivation of polygenic scores

Polygenic SCZ risk scores. Markers directly genotyped in the target sample were used for the generation of this score. As described in the Psychiatric Genomics Consortium guidelines regarding polygenic risk profile analyses (http://pgc.unc.edu), all SNPs in the extended major histocompatibility complex region were removed at this stage, except one representing the best hit in this region. PSS were calculated following the methods described by Purcell et al.[2] In brief, sets of SNPs with P-values below different cutoffs (0.0001; 0.0005; 0.001; 0.005; 0.01; 0.05; 0.1; 0.2; 0.3; 0.4) in the discovery sample were defined. In order to identify polygenic effects

Table 1. Variance explained (R^2) and P-values of the association of PSS (based on different P-value thresholds) and of PWM (based on the complete or partial SNP set) with total brain, white matter and gray matter volumes in the healthy target sample (N = 122)

	No. SNPs	Total brain		White matter		Gray matter	
		R^2 change	P-value	R^2 change	P-value	R^2 change	P-value
PSS (polygenic SCZ risk score)							
P < 0.0001	106	−0.007	0.613	−0.008	0.740	−0.007	0.590
P < 0.0005	311	−0.009	0.928	−0.010	0.903	−0.009	0.801
P < 0.001	482	0.007	0.180	−0.009	0.776	0.023	0.054
P < 0.005	1480	−0.006	0.559	−0.009	0.983	0.000	0.320
P < 0.01	2528	−0.005	0.507	−0.006	0.558	−0.005	0.570
P < 0.05	8555	−0.006	0.582	−0.007	0.613	−0.008	0.710
P < 0.1	14 491	−0.009	0.819	−0.008	0.678	−0.009	0.999
P < 0.2	24 870	−0.008	0.689	−0.009	0.856	−0.007	0.624
P < 0.3	33 660	−0.007	0.635	−0.010	0.941	−0.005	0.489
P < 0.4	41 398	−0.007	0.661	−0.007	0.941	−0.003	0.445
PWM (polygenic white matter score)							
Complete SNP set	185	−0.008	0.719	−0.009	0.744	−0.001	0.333
Partial set	142	−0.002	0.391	−0.009	0.872	0.006	0.187

Abbreviations: SCZ, schizophrenia; SNP, single-nucleotide polymorphism. R^2 refers to R^2 adjusted for the number of predictors in the model.

due to independent SNPs in linkage equilibrium, each SNP set was pruned based on a pairwise r^2 threshold of 0.25 and a sliding window of 50 SNPs shifting 5 SNPs at each step (see Table 1 for the total number of SNPs finally included under each P-value threshold). For each subject in the target sample, a PSS was calculated for the different P-value thresholds. For each SNP, the number of risk variants (0, 1, 2) an individual carries was multiplied by the logarithm of the odds ratio for that particular variant. Both pruning and scoring were performed using PLINK.[20]

Polygenic white matter scores. A total of 66 out of the 186 SNPs that constitute the PWM[7] were directly genotyped in the Axiom array. The remaining 120 SNPs were imputed using IMPUTE v2.2.[21] Imputation was performed with the 1000 Genomes Project Phase 1 (March 2012) reference panel, and best-guess genotypes were used for polygenic scoring. One of the SNPs (rs9880959) could not be imputed. Mean ± s.d. imputation quality score was 0.906 ± 0.127. First, PWM including the complete set of 185 SNPs was calculated following the previously described methods. Second, an alternative PWM was calculated using a partial set of 142 SNPs, including only those directly genotyped and those with imputation quality score ⩾ 0.9 (Figure 1).

Statistical analysis

Standardized values of total brain, white matter and gray matter volumes were obtained after correction for age, gender and intracranial volume. These corrected brain dimensions were used as dependent variables in a linear regression model. Ten multidimensional scaling components, the inbreeding coefficient and the number of SNPs used to calculate the polygenic scores were selected as covariates of potential relevance. R^2 values derived from a model including all of these covariates were subtracted from R^2 values from a model including covariates plus the respective polygenic score. The difference between the R^2 adjusted for the number of predictors in the model represents the increase in the variance explained attributable to the polygenic score. All these calculations as well as sign tests were carried out with SPSS 17.0 (IBM-Deutschland GmbH, Munich, Germany).[22] PLINK[20] was used for testing association of each of the 185 SNPs belonging to PWM with the different brain volumetric variables. QUANTO software[23] was used for power calculations in the target sample.

RESULTS

The target sample comprised a total of 122 Italian subjects (56 males, 66 females), aged 45.0 ± 17.0 (mean ± s.d.) years (range 20–80), with an education level of 14.4 ± 3.7 (mean ± s.d.) years (range 5–24 years). Mean volumes (± s.d.) (ml) were 978.1 ± 112.7, 407.1 ± 55.7 and 551.5 ± 65.5 for total brain, white matter and gray matter, respectively.

PSS were not associated ($P > 0.05$ in all comparisons) with any of the brain dimensions in the present sample of healthy subjects, irrespective of the P-value cutoffs selected for analysis (Table 1).
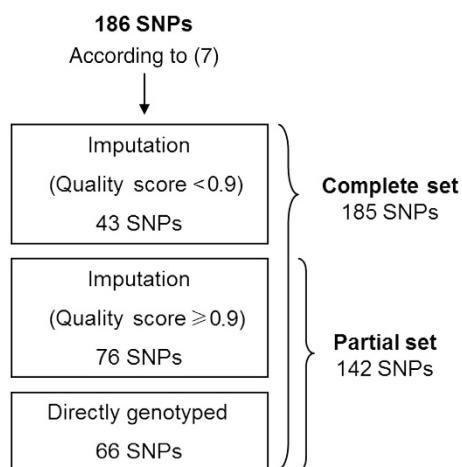

Figure 1. Scheme of source and number of single nucleotide polymorphisms (SNPs) used to calculate polygenic white matter scores (PWM).

The 'best result' obtained was a trend toward association ($P = 0.054$) with gray matter volume at P cutoff < 0.001 leading to a change in adjusted R^2 of 0.023. Gray matter, however, was not associated with PSS in the original study.[7]

Similarly, PWM analysis did not yield any significant association ($P > 0.05$ in all comparisons) either with white matter volume, the main variable of interest or the other brain volumes studied. This lack of significance was true for both the complete and the partial set of SNPs (Table 1; Figure 1). When the individual effects of these markers on volumetric variables were evaluated, none of the P-values survived Bonferroni correction (Supplementary Table S1). Sign tests for consistency between the original study[7] and the present one, based on the sign of the beta coefficients on imaging variables, did not yield significant results ($P > 0.05$).

As the mean age of our target sample was significantly ($P = 0.0001$) higher than the mean age of the sample analyzed in the original publication (mean ± s.d.: 45.0 ± 17.0 versus 32.3 ± 12.2 years, respectively),[7] we performed an exploratory analysis in our sample, excluding subjects aged > 55 years. Also, the resulting smaller subset ($N = 80$), with a mean age of 34.5 ± 9.67 years (no longer different from the original sample; $P = 0.166$), did not reveal the slightest trend toward an association (all comparisons $P > 0.05$) (Supplementary Table S2).

DISCUSSION

A recent original publication[7] reported an effect of GWAS-identified SCZ risk variants, when compiled to polygenic scores, on total brain and white matter volumes in SCZ and even more pronounced in healthy individuals. The present study has not been able to replicate these effects in an independent sample of healthy subjects.

Differences between the original publication and the present study that have to be discussed as potential causal contributors to the lack of reproducibility are: the sample size ($N = 142$ versus $N = 122$), the ethnic background, the age distribution (mean age 32 versus 45 years), the array technology (Affymetrix versus Illumina, San Diego, CA, USA), the SNP imputation quality and the magnetic field strength of the MRI scanner underlying volumetry (1.5 versus 3 T).

The healthy sample analyzed in our study ($N = 122$) had 80% power to detect $R^2 > 6.2\%$ between PSS and the different brain volumes selected for analysis. In the original report, the amount of maximum variance explained ranged from 4.8 to 5.1%, regarding total brain and white matter volumes, respectively. Thus, the power of the present sample was below 80%. Therefore, a reduced power cannot be entirely excluded as one reason for the non-replication results presented here. However, the results of the sign tests regarding the comparability between samples, together with the complete lack of any 'signal' in the polygenic approach, suggest that a lack of power may not account for the lack of replication. Of course, a winner's curse effect in the original study cannot be entirely excluded.

Regarding the ethnic background, correction for population stratification by introduction of multidimensional scaling components as covariates should have controlled for potential stratification within our replicate sample. We included as many as 10 multidimensional scaling components as covariates to diminish bias from population stratification. Reducing them in an exploratory fashion to four did not significantly alter the results, arguing against overcorrection bias (data not shown). Both samples are of European Caucasian origin, pointing to a high degree of similarity in their common genomic variation involved in disease or phenotype susceptibility.

The significant age difference between the original study sample and our cohort may also have a part in the failure to reproduce the associations described in the original paper. However, exploratory reduction of the mean age of our study

group by removing individuals that were >55 years old failed to provide even the least signal of an association between PSS or PWM and total brain or white matter dimensions.

The original study[7] used the Illumina HumanHap550 beadchip (Illumina) for SNP genotyping, whereas we used semi-custom Axiom (Affymetrix). As both arrays are based on very similar principles regarding linkage disequilibrium thresholds for genome-wide coverage, have a similar number of markers and are identical regarding about 30% of the SNPs, this technological difference does not seem to account for non-reproducibility of the original publication.

SNP imputation quality was nearly perfect in 76 SNPs, whereas 43 SNPs did not reach the highest quality threshold. However, as described in the Materials and Methods section, the overall mean quality score was >0.9. Nevertheless, with a total of 142/185 SNPs, at least an approximation of the PWM results should have been reached.

The scanner in which MRI sequences were acquired is different in the original study (1.5 T) compared with the present work (3 T). In fact, 3 T scanners increase baseline magnetization, leading to an about twofold increase in signal-to-noise ratio, which, in turn, improves accuracy and reproducibility of tissue classification results and thus sensitivity of volumetry regarding morphometric differences.[24] Therefore, also this difference should not be critical for our failure to replicate the previous results. Moreover, the deep methodological (software) differences between the two segmentation processes used in the original[7] versus our study might account for the mismatch between the two samples. In this regard, it has to be pointed out that the method used here is nowadays the gold standard in brain segmentation, whereas the method employed by Terwissscha van Scheltinga *et al.*[7] is based on older software.

Taken together, even with the limitations of our replication sample discussed above (and always bearing in mind the possibility of a winner's curse effect in the original study), at least a 'signal in the right direction' would have to be expected if the PSS or PWM associations with brain dimensions were of general validity. In light of the obvious interest of studying effects of polygenic risk scores on specific subphenotypes of relevance for complex psychiatric disorders, our results admonish that replication studies are absolutely essential for this kind of analyses. This all the more as polygenic settings of interest cannot easily be explored in animal models to confirm their specific importance. It remains to be established whether by including even larger numbers of individuals in case–control GWAS, the heterogeneity problem of SCZ or other mental diseases (and of health) will be solved. Although much more labor intense than GWAS, large-scale phenotype-based genetic association studies will be pivotal for further investigating the genotype contribution to complex disease phenotypes, thereby extending and complementing the GWAS efforts.

CONFLICT OF INTEREST

ACKNOWLEDGMENTS

REFERENCES

1 Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA *et al.* Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**: 969–976.

2 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.

3 McIntosh AM, Gow A, Luciano M, Davies G, Liewald DC, Harris SE *et al.* Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. *Biol Psychiatry* 2013; **73**: 938–943.

4 Walton E, Turner J, Gollub RL, Manoach DS, Yendiki A, Ho BC *et al.* Cumulative genetic risk and prefrontal activity in patients with schizophrenia. *Schizophr Bull* 2013; **39**: 703–711.

5 Derks EM, Vorstman JA, Ripke S, Kahn RS, Ophoff RA. Schizophrenia Psychiatric Genomic Consortium, Investigation of the genetic association between quantitative measures of psychosis and schizophrenia: a polygenic risk score analysis. *PLoS One* 2012; **7**: e37852.

6 Terwissscha van Scheltinga AF, Bakker SC, van Haren NEM, Derks EM, BuizerVoskamp JE, Cahn W W. *et al.* Schizophrenia genetic variants are not associated with intelligence. *Psychol Med* 2013; **43**: 2563–2570.

7 Terwissscha van Scheltinga AF, Bakker SC, van Haren NE, Derks EM, Buizer-Voskamp JE, Boos HB *et al.* Genetic schizophrenia risk variants jointly modulate total brain and white matter volume. *Biol Psychiatry* 2013; **73**: 525–531.

8 Folstein MF, Folstein SE, McHugh PR. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; **12**: 189–198.

9 Carlesimo GA, Caltagirone C, Gainotti G. The Mental Deterioration Battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. The Group for the Standardization of the Mental Deterioration Battery. *Eur Neurol* 1996; **36**: 378–384.

10 McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 1984; **34**: 939–944.

11 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th edn, American Psychiatric Association: Washington DC, USA, 1994.

12 Ikram MA, van der Lugt A, Niessen WJ, Krestin GP, Koudstaal PJ, Hofman A *et al.* The Rotterdam Scan Study: design and update up to 2012. *Eur J Epidemiol* 2011; **26**: 811–824.

13 Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 1999; **9**: 179–194.

14 Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci USA* 2000; **97**: 11050–11055.

15 Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 2010; **53**: 1–15.

16 Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 2002; **33**: 341–355.

17 Fischl B, Salat DH, van der Kouwe AJ, Makris N, Segonne F, Quinn BT *et al.* Sequence-independent segmentation of magnetic resonance images. *Neuroimage* 2004; **23**(Suppl 1): S69–S84.

18 Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH *et al.* Automatically parcellating the human cerebral cortex. *Cereb Cortex* 2004; **14**: 11–22.

19 Hammer C, Stepniak B, Schneider A, Papiol S, Tantra M, Begemann M*et al.* Neuropsychiatric disease relevance of circulating anti-NMDA receptor autoantibodies depends on blood brain barrier integrity. *Mol Psychiatry* advance online publication, 3 September 2013; doi:10.1038/mp.2013.110 (in press).

20 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.

21 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.

22 *SPSS Statistics for Windows*, Version 17.0. SPSS Inc: Chicago, USA.

23 Gauderman W, Morrison J. QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies, available at http://hydra.usc.edu/gxe, 2006.

24 Tardif CL, Collins DL, Pike GB. Regional impact of field strength on voxel-based morphometry results. *Hum Brain Mapp* 2010; **31**: 943–957.

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