Molecular Psychiatry (2011) 16, 1117–1129

www.nature.com/mp

ORIGINAL ARTICLE

GWA study data mining and independent replication identify cardiomyopathy-associated 5 (*CMYA5*) as a risk gene for schizophrenia

X Chen[1,2], G Lee[1], BS Maher[1], AH Fanous[1,3,4], J Chen[1], Z Zhao[5], A Guo[5], E van den Oord[1,6], PF Sullivan[7], J Shi[8], DF Levinson[8], PV Gejman[9], A Sanders[9], J Duan[9], MJ Owen[10], NJ Craddock[10], MC O'Donovan[10], J Blackman[10], D Lewis[10], GK Kirov[10], W Qin[11], S Schwab[11], D Wildenauer[11], K Chowdari[12], V Nimgaonkar[12], RE Straub[13], DR Weinberger[13], FA O'Neill[14], D Walsh[15], M Bronstein[16], A Darvasi[16], T Lencz[17], AK Malhotra[17], D Rujescu[18], I Giegling[18], T Werge[19], T Hansen[19], A Ingason[19], MM Nöethen[20], M Rietschel[21,22], S Cichon[20,23], S Djurovic[24], OA Andreassen[24], RM Cantor[25], R Ophoff[25,26], A Corvin[27], DW Morris[27], M Gill[27], CN Pato[28], MT Pato[28], A Macedo[29], HMD Gurling[30], A McQuillin[30], J Pimm[30], C Hultman[31], P Lichtenstein[31], P Sklar[32,33], SM Purcell[32,33], E Scolnick[32,33], D St Clair[34], DHR Blackwood[35] and KS Kendler[1,2] and the GROUP investigators, the International Schizophrenia Consortium[36]

[1]*Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA;* [2]*Department of Molecular and Human Genetics, Virginia Commonwealth University, Richmond, VA, USA;* [3]*Washington VA Medical Center, Washington, DC, USA;* [4]*Department of Psychiatry, Georgetown University School of Medicine, Washington, DC, USA;* [5]*Departments of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA;* [6]*Department of Pharmacy, Virginia Commonwealth University, Richmond, VA, USA;* [7]*Departments of Genetics, Psychiatry, and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA;* [8]*Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA;* [9]*Center for Psychiatric Genetics, NorthShore University HealthSystem Research Institute, Evanston, IL, USA;* [10]*Department of Psychological Medicine, MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, UK;* [11]*School of Psychiatry, University of Western Australia and Western Australian Institute for Medical Research, Perth, WA, Australia;* [12]*Department of Psychiatry, University of Pittsburgh School of Medicine and Graduate School of Public Health, Pittsburgh, PA, USA;* [13]*Genes Cognition and Psychosis Program, Clinical Brain Disorders Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA;* [14]*The Department of Psychiatry, The Queens University, Belfast, Northern Ireland, UK;* [15]*The Health Research Board, Dublin, Ireland;* [16]*Department of Genetics, The Hebrew University of Jerusalem, Jerusalem, Israel;* [17]*Zucker Hillside Hospital, Psychiatry Research, North Shore-Long Island Jewish Health System, Glen Oaks, NY, USA;* [18]*Division of Molecular and Clinical Neurobiology, Ludwig-Maximilians-University, Munich, Germany;* [19]*Institute of Psychiatry, Mental Health Center Sct. Hans, Copenhagen University Hospital, Roskilde Denmark;* [20]*Department of Genomics, Life & Brain Center, Institute of Human Genetics, University of Bonn, Bonn, Germany;* [21]*Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health Mannheim, University of Heidelberg, Mannheim, Germany;* [22]*Department of Psychiatry, University of Bonn, Bonn, Germany;* [23]*Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany;* [24]*Institute of Psychiatry, University of Oslo and Oslo University Hospital, Oslo, Norway;* [25]*Department of Human Genetics, UCLA Center for Neurobehavioral Genetics, Los Angeles, CA, USA;* [26]*Department of Medical Genetics, University Medical Center Utrecht, CG Utrecht, The Netherlands;* [27]*Department of Psychiatry and Institute of Molecular Medicine, Neuropsychiatric Genetics Research Group, Trinity College Dublin, Dublin, Ireland;* [28]*Center for Genomic Psychiatry, University of Southern California, Los Angeles, CA, USA;* [29]*Department of Psychiatry University of Coimbra, Coimbra, Portugal;* [30]*Research Department of Mental Health Sciences, Molecular Psychiatry laboratory, Windeyer Institute of Medical Sciences, University College London Medical School, London, UK;* [31]*Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden;* [32]*Psychiatric and Neurodevelopmental Genetics Unit, Boston, MA, USA;* [33]*Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA;* [34]*Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK and* [35]*Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK*

Correspondence: Dr X Chen, Department of Psychiatry, Virginia Commonwealth University, 800 E. Leigh Street, Suite 390, Richmond, VA 23298-0424, USA.
E-mail: xchen@vcu.edu
[36]Members of the GROUP Investigators and the International Schizophrenia Consortium are listed in Appendix.
Received 10 April 2010; revised 3 August 2010; accepted 11 August 2010; published online 14 September 2010

We conducted data-mining analyses using the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) and molecular genetics of schizophrenia genome-wide association study supported by the genetic association information network (MGS-GAIN) schizophrenia data sets and performed bioinformatic prioritization for all the markers with *P*-values $\leqslant 0.05$ in both data sets. In this process, we found that in the *CMYA5* gene, there were two non-synonymous markers, rs3828611 and rs10043986, showing nominal significance in both the CATIE and MGS-GAIN samples. In a combined analysis of both the CATIE and MGS-GAIN samples, rs4704591 was identified as the most significant marker in the gene. Linkage disequilibrium analyses indicated that these markers were in low LD (3 828 611–rs10043986, $r^2 = 0.008$; rs10043986–rs4704591, $r^2 = 0.204$). In addition, *CMYA5* was reported to be physically interacting with the *DTNBP1* gene, a promising candidate for schizophrenia, suggesting that *CMYA5* may be involved in the same biological pathway and process. On the basis of this information, we performed replication studies for these three single-nucleotide polymorphisms. The rs3828611 was found to have conflicting results in our Irish samples and was dropped out without further investigation. The other two markers were verified in 23 other independent data sets. In a meta-analysis of all 23 replication samples (family samples, 912 families with 4160 subjects; case–control samples, 11 380 cases and 15 021 controls), we found that both markers are significantly associated with schizophrenia (rs10043986, odds ratio (OR) = 1.11, 95% confidence interval (CI) = 1.04–1.18, $P = 8.2 \times 10^{-4}$ and rs4704591, OR = 1.07, 95% CI = 1.03–1.11, $P = 3.0 \times 10^{-4}$). The results were also significant for the 22 Caucasian replication samples (rs10043986, OR = 1.11, 95% CI = 1.03–1.17, $P = 0.0026$ and rs4704591, OR = 1.07, 95% CI = 1.02–1.11, $P = 0.0015$). Furthermore, haplotype conditioned analyses indicated that the association signals observed at these two markers are independent. On the basis of these results, we concluded that *CMYA5* is associated with schizophrenia and further investigation of the gene is warranted.

Molecular Psychiatry (2011) **16,** 1117–1129; doi:10.1038/mp.2010.96; published online 14 September 2010

Keywords: association study; cardiomyopathy; GWA data mining; meta-analysis; schizophrenia

Introduction

Schizophrenia is a psychiatric disorder with a worldwide prevalence of 1%. It is characterized by delusions, hallucinations and deficits of cognition and emotion. There is sufficient data from family and twin studies, suggesting that genetic factors have significant functions in the etiology of the disease. In recent years, a large number of genetic association studies have identified many candidate genes for the disease; however, most of these genes do not have satisfactory replications. Most recently, several genome-wide associations (GWAs) have been reported.[1–6] Of the many potential leads discovered by these studies, the broad region in chromosome 6p is the most consistent finding.[2–4] Another gene, *ZNF804A*, has reached global significance when samples from both schizophrenia and bipolar disorder are combined.[6] Other genes, although not reaching genome-wide significance in initial samples, have consistent replications with many independent samples.[7,8]

These recent GWA studies of schizophrenia are not only promising, but also illustrate their limitations in detecting individual candidate genes with small effects on disease risk. Alternative approaches are also needed. In this study, we implemented a method that combines data mining of GWA data sets and bioinformatic prioritization to select promising candidate genes and follows by verification and meta-analyses of a large number of independent data sets. Specifically, we conducted GWA analyses of the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) and molecular genetics of schizophrenia GWA study supported by the genetic association information network (MGS-GAIN) samples and selected all candidate single-nucleotide polymorphisms (SNPs) with *P*-values $\leqslant 0.05$ in both CATIE and MGS-GAIN data sets. These markers were then analyzed by comprehensive bioinformatic prioritization procedures. Top candidates emerging from these analyses were further verified by independent samples. Using this approach, we analyzed 25 independent samples with a total of over 33 000 individuals and identified two SNPs, including a non-synonymous SNP, in and around the *CMYA5* gene to be significantly associated with schizophrenia. Here, we report the results from this study.

Materials and methods

Subjects and genotyping

In this study, we used 25 samples with a total of 33 834 subjects, including 912 families with 4160 subjects, 13 038 cases and 16 636 controls (the overlapping subjects between the CATIE and MGS-GAIN and MGS non-GAIN were excluded from these numbers). The CATIE and MGS-GAIN samples were used as our data-mining and hypothesis-generating samples in the first stage of our two-stage study. The other 23 samples were used as replication samples. Twenty-four of the 25 samples were of Caucasian ancestry, one sample, MGS-GAIN-AA, was of African American ancestry. Of these samples, 20 samples were used in GWA studies by individual groups and the subjects in these samples were typed by either the Affymetrix or Illumina microarray methods. Five samples, the Irish family (IFAM), Irish case–control

(ICC), Bonn, Pittsburgh and Ashkenazi were typed by the TaqMan method.[9] The quality of genotyping was assessed by individual groups to be satisfactory. The principle investigators, sample size and genotyping method were listed in Table 1.

Data mining and bioinformatic prioritization
We used the PLINK program[10] to conduct the GWA analyses. The GWA analyses were conducted with the quality-control filtered markers from the NIMH (http://nimhgenetics.org/) and GAIN (http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap) repositories for the CATIE and MGS-GAIN samples, respectively. In these analyses, only Caucasian subjects (CATIE, 492 cases and 523 controls; MGS-GAIN, 1166 cases and 1368 controls including the 236 overlapping controls between the two samples) were used and markers with a minor allele frequency $<1\%$ or a Hardy–Weinberg equilibrium P-value <0.0001 were excluded. For the CATIE data set, the seven principle components identified in the previous study[1] were used as covariates and a total of 446 225 markers were analyzed. For the MGS-GAIN sample, based on previous analyses that there was no significant stratification found in the sample,[2] no covariate was used. The number of markers analyzed for the MGS GAIN was 727 905. Note that we did not know at the time of GWA analyses that there were some overlapping subjects between the CATIE and MGS-GAIN samples; therefore, the two samples used in the data mining and bioinformatic prioritization were not completely independent. In the subsequent analyses for the common markers between the two data sets, the 236 overlapping subjects were excluded.

For bioinformatic prioritization, we first selected all markers with P-values $\leqslant 0.05$ in the two data sets, and matched them against each other. After the matching, there were 1128 SNPs with unadjusted P-values $\leqslant 0.05$ in both the CATIE and GAIN samples. We then conducted bioinformatic prioritization of these 1128 SNPs based on whether they are located in the evolutionarily conserved regions, genic regions (exons, introns, untranslated regions, or within 2 kb of a gene), transcription factor-binding sites, or whether they are located in known schizophrenia candidate genes (as listed in the sczgene database http://www.schizophreniaforum.org/res/sczgene/default.asp by June 2008) or whether the SNPs are non-synonymous. SNPs in each of these categories were assigned an empirical score: 2 for the non-synonymous and known schizophrenia candidate gene categories, 1 for the evolutionary conserved region, transcription factor-binding site, untranslated region and synonymous SNP category and 0.5 for the 'within 2 kb of a gene' category. Finally, SNPs were ranked by the sum of the scores.[11]

When the *CMYA5* gene was identified as the leading candidate, we performed LD structure analyses of the gene using the HAPLOVIEW program.[12] We extracted all markers in the gene plus 20 kb upstream and downstream sequences for the CATIE and MGS-GAIN samples, and selected the common markers between the two data sets. Data from the two data sets were combined. Association analysis for the combined samples was also conducted using UNPHASED program.[13]

Replication and meta-analyses of independent samples
On the basis of the prioritization, we initiated genotyping for three SNPs, rs3828611, rs10043986 and rs4704591, in our IFAM and ICC samples. For rs10043986 and rs4704591, the results from our Irish samples were consistent with that observed in the CATIE and MGS-GAIN data sets. The rs3828611 had inconsistent results between our Irish samples; therefore, was dropped without further investigation. To verify the association observed for rs10043986 and rs4704591, we requested genotyping of two additional samples (Bonn and Pittsburgh) and solicited data from GWA studies from 21 independent samples (see Table 1). The MGS-non-GAIN sample also had 208 overlapping control subjects with the CATIE data set. To maintain the independence among the samples used in the replication study, these overlapping subjects were removed from the MGS-non-GAIN sample.

Meta-analyses for all samples and replication samples only were conducted. We generated combined odds ratios (ORs) of the family-based and case–control samples using the information included in the primary analyses and standard meta-analytic techniques. For the IFAM sample, we used a PDT-like approach to generate the OR.[14] The PDT statistic compares the number of times a given parental allele ('risk' allele) is transmitted versus non-transmitted and examines allele sharing between affected and unaffected sibling pairs, whereas standard case–control approaches examine allele frequencies in cases versus controls. The parental transmission OR is constructed as (a/c)/(b/d), where a is the transmissions of the high-risk allele, c is the non-transmissions of the high-risk allele, b is the transmissions of all other alleles and d is the non-transmissions of all other alleles. In the sibling pair sample and the population-based samples, which compare case to control allele frequencies, we construct an OR as (a/c)/(b/d), where a is the number of major alleles present in cases, c is the number of minor alleles present in cases, b is the number of major alleles in controls and d is the number of minor alleles present in controls. In each of the EA case–control samples, we construct an OR as (a/c)/(b/d), where a is the number of major alleles present in cases, c is the number of minor alleles present in cases, b is the number of major alleles in controls and d is the number of minor alleles present in controls. In the AA sample, we fit a logistic regression model including the first principal component of population stratification as a covariate. The regression coefficient of the effect of the SNP allowed us to estimate an OR and variance for inclusion in the meta-analysis.

We used formal meta-analytic techniques to combine ORs across study types. We performed a

Table 1 Sample description

Sample	*Principle investigator*	*Ethnicity*	*Sample size (# family, case/control)*	*rs10043986 MAF*	*rs4704591 MAF*	*Genotyping method*	*Reference*
MGS GAIN	Pablo Gejman	European American	1166/1132	0.126	0.386	Affymetrix 6.0	[2]
CATIE	Patrick Sullivan	European American	492/523	0.119	0.385	Perlgen	[1]
MGS non-GAIN	Pablo Gejman	European American	1089/1065	0.132	0.384	Affymetrix 6.0	[2]
IFAM	Kenneth Kendler	Irish/English	455	0.108	0.387	TaqMan	[48]
ICC	Kenneth Kendler	Irish/English	792/625	0.132	0.356	TaqMan	[48]
Cardiff	Michael O'Donovan	UK	479/2530	n/a	0.394	Affymetrix 5.0	[6]
Pittsburgh	Vishwajit Nimgaonkar	European American	246	0.132	0.385	TaqMan	[49]
Bonn	Dieter Wildenauer	German	211	0.132	0.355	TaqMan	[50]
NIMH	Daniel Weinberger	European American	252/270	0.137	0.374	Illumina 550	[51]
ISC Dublin	Aiden Corvin	Irish/English	275/866	0.117	0.378	Affymetrix 6.0	[3]
ISC Edinburgh	Douglas Blackwood	English	369/287	0.104	0.373	Affymetrix 6.0	[3]
ISC Sweden1	Patrick Sullivan	Sweden	170/170	n/a	0.374	Affymetrix 5.0	[3]
ISC Sweden2	Patrick Sullivan	Sweden	390/230	0.134	0.387	Affymetrix 6.0	[3]
ISC Aberdeen	David St Clair	English	720/702	n/a	0.393	Affymetrix 5.0	[3]
ISC Portugal	Carlos Pato	Portuguese	347/216	n/a	0.372	Affymetrix 5.0	[3]
ISC London	Hugh Gurling	English	523/505	n/a	0.357	Affymetrix 5.0	[3]
ISC Bulgaria	Michael Owen	Bulgarian	528/611	0.113	0.375	Affymetrix 6.0	[3]
Copenhagen	Thomas Werge	Danish	494/343	0.125	n/a	Illumina 550	[52]
Oslo	Ole Adreassen	Norwegian	486/417	0.125	0.385	Affymetrix 6.0	[53]
Mannheim	Sven Cichon	German	464/1272	0.120	n/a	Illumina 550	Unpublished data
Munich	Dan Rujescu	German	916/2263	0.123	0.381	MALDI-TOF	[8]
Dutch	Roel Ophoff	Dutch	705/645	0.137	n/a	Illumina 550	[4]
ZHH	Anil Malhotra	European American	279/250	n/a	0.373	Affymetrix 5.0	[5]
Ashkenazi	Ariel Darvasi	Ashkenazi Jew	761/765	0.096	0.351	KASPar	[54]
MGS GAIN AA	Pablo Gejman	African American	915/949	0.025	0.129	Affymetrix 6.0	[2]
ALL samples			912 families, 13 038/16 636				

Abbreviations: CATIE, Clinical Antipsychotic Trials of Intervention Effectiveness; ICC, Irish case–control; IFAM, Irish family; MAF, minor allele frequency; MGS-GAIN, molecular genetics of schizophrenia genome-wide association study supported by the genetic association information network.

fixed-effects (Mantel–Haenszel) approach to meta-analysis.[15] Before pooling, we performed Cochran's (Q) χ^2 test of heterogeneity to ensure that each group of studies was suitable for meta-analysis. Generally, in meta-analysis, when significant heterogeneity is found, the studies are deemed unsuitable for pooling through a fixed-effects approach. In the summary meta-analysis of all studies, including the discovery and replication samples, there was a known overlap in controls between the MGS-GAIN and CATIE samples. We calculated the asymptotic correlation between the Z-scores of the two studies and performed a Z-score-based meta-analysis correcting for the correlation because of the shared controls to ensure appropriate type-I error rate.[16]

Testing independent effect between rs10043986 and rs4704951

As there were two SNPs showing association in the *CMYA5* gene, we evaluated whether the association signals observed at rs10043986 and rs4704591 were statistically independent. We took the approach implemented in the PLINK program[17] that compares the risk of haplotypes with identical alleles in the background locus, but different alleles at the locus to be evaluated. In this case, we inferred all four haplotypes for rs10043986–rs4704591, and tested the effects of haplotypes with the same allele at rs4704591, but different alleles at rs10043986. Our aim was to evaluate whether the effect of rs10043986 is independent of rs4704591. We use the UNPHASED program[13] to conduct this analysis as, unlike PLINK, it is able to combine family data and case–control data for such haplotype-based analyses.

Results

GWA studies data mining and bioinformatic prioritization

From the GWA analyses of the CATIE and MGS-GAIN data sets, there were 24 160 and 68 371 markers with unadjusted P-values $\leqslant 0.05$, respectively. Although none of the markers in the CATIE and MGS-GAIN reached genome-wide significance, the number of markers reaching nominal significance (that is 68 371) was significantly larger than the expected (that is 37 725) in the MGS-GAIN sample, suggesting that there were markers with true effects in this pool of nominally significant markers. Of these markers, there were 1228 markers having $P \leqslant 0.05$ in both data sets (Supplementary Table S1). These markers constituted the pool we used for further bioinformatic prioritization. From these markers, the informatics procedures revealed several top candidate genes (Table 2). Of these top candidates, *CMYA5* and *PTPN21* each had two non-synonymous SNPs. As the two non-synonymous markers in *CMYA5*, rs3828611 and rs10043986, had different frequencies, and were located in two different exons of the gene, we thought they may represent independent association signals. In contrast, the two markers in the *PTPN21* gene had very similar frequency. Therefore, we decided to focus on the *CMYA5* gene. There were other genes that had multiple markers with different frequencies (Table 2). These included *LRP1B*, *COLQ*, *SERINC1*, *PTPN21*, *EML5*, *NTRK3* and *NUTF2*. Further analyses of these genes may be necessary to verify their functions in schizophrenia.

We conducted literature search for the *CMYA5* gene and found that it was reported to be physically interacting with *DTNBP1*, a leading candidate for schizophrenia that was first reported in our IFAM sample[18] included in this study. We also analyzed single marker association for the shared SNPs between the CATIE and MGS-GAIN in this interval by combining CATIE and MGS-GAIN samples together. These analyses identified the most significant marker, rs4704591, which is located about 9 kb downstream of the gene. Note that at the time of our GWA analyses, we did not realize that there were overlapping subjects between the CATIE and MGS-GAIN studies. After removing the overlapping subjects between the CATIE and MGS-GAIN data sets, an analysis of the combined samples was performed. The P-values for the three markers were 0.0078 (OR = 1.31, 95% confidence interval (CI) = 1.07–1.60); 0.0050 (OR; 1.19, 95% CI = 1.06–1.30) and 0.00032 (OR = 1.17, 95% CI = 1.08–1.24) for rs3828611, rs10043986 and rs4704591, respectively (Figure 1a). The LD analyses of the 27 common markers shared by the CATIE and MGS-GAIN studies were performed using the HAPLOVIEW program[12] for the gene and 20 kb flanking sequences (Figure 1b). The LD between rs3828611 and rs10043986 was 0.008 (r^2), the LD between rs10043986 and rs4704591 was 0.208 (r^2) and the LD between rs3828611 and rs4704591 was 0.016 (r^2). As the LDs among these three markers were relatively low, it was likely that they represented different association signals. There were two other markers showing similar level of association as rs3828611. The rs6880680 was in high LD with rs3828611 ($r^2 = 0.713$), its effect may not be independent. The rs6870619 was in low LD with all other markers in this region. However, as its signal was not as strong as rs4704591 and it did not reach nominal significance in the CATIE sample, it was not pursued.

Verification of CMYA5 association in the Irish samples

On the basis of the data mining and bioinformatic prioritization, we initiated confirmation study using our IFAM and ICC samples for these three SNPs (rs3828611, rs10043986 and rs4704591). We used the UNPHASED program,[13] which was designed to combine case–control and family samples and to analyze our combined samples. The results of our Irish samples support the association of rs10043986 and rs4704951. For rs10043986, both the case–control and family samples showed the same direction of association for the same allele as that in the CATIE and MGS-GAIN data sets. However, neither the individual samples (case–control and family samples) nor the combined samples reached significance.

Table 2 Candidate genes selected from data mining and prioritization

SNP rs#	*Chr*	*Position (bp)*	*Allele*	*MAF*	*MGS-GAIN* P-*value*[a]	*CATIE* P-*value*	*Gene*	*Total score*	*Function*
rs5177	1	53 484 323	G/C	0.431	0.00021	0.00397	LRP8	1	Apolipoprotein e receptor
rs2297660	1	53 504 903	A/C	0.432	0.00041	0.00269	LRP8	2	
rs12745968	1	93 174 425	C/T	0.403	0.00000	0.00334	FAM69A	1	Transmembrane protein with unknown function
rs1123217	1	160 307 219	G/C	0.030	0.00028	0.01887	NOS1AP	1	Nitric oxide synthase 1 (neuronal) adaptor protein
rs3821153	2	36 606 626	C/A	0.412	0.00626	0.01738	CRIM1	2	Cysteine-rich transmembrane BMP regulator 1
rs2714212	2	141 824 989	G/A	0.419	0.00976	0.02163	LRP1B	1	Low-density lipoprotein-related protein 1B
rs4245855	2	141 865 798	G/A	0.324	0.00038	0.02462	LRP1B	1	
rs6429874	2	141 883 707	T/A	0.251	0.00412	0.01905	LRP1B	1	
rs12633820	3	15 469 006	T/C	0.379	0.00122	0.02583	COLQ	1	Collagen-like tail subunit (single strand of homotrimer) of asymmetric acetylcholinesterase
rs2305616	3	15 490 699	T/C	0.424	0.00079	0.03159	COLQ	2	
rs3828611	5	79 070 418	G/C	0.058	0.02248	0.01704	CMYA5	3	Cardiomyopathy associated 5
rs10043986	5	79 131 173	T/C	0.135	0.04148	0.02514	CMYA5	3	
rs9368649	6	31 046 862	G/A	0.112	0.00022	0.00020	—	1	N/A
rs197687	6	122 920 535	T/C	0.316	0.00863	0.00447	SERINC1	1	Serine incorporator 1
rs9398678	6	122 938 979	C/G	0.151	0.02391	0.00368	SERINC1	1	
rs7906952	10	15 025 308	T/C	0.271	0.04779	0.02120	DCLRE1C	1	DNA cross-link repair 1C
rs2457450	10	87 372 875	C/A	0.356	0.02467	0.01652	GRID1	3	Glutamate receptor, ionotropic, δ 1
rs7975818	12	55 844 772	G/C	0.329	0.02020	0.00192	LRP1	1	Low-density lipoprotein-related protein 1
rs2228225	12	56 145 698	G/A	0.356	0.01274	0.03582	GLI1	2	GLI family zinc-finger 1
rs2274736	14	88 008 405	G/A	0.354	0.00106	0.01705	PTPN21	3	Protein tyrosine phosphatase, non-receptor type 21
rs2401751	14	88 016 375	A/G	0.354	0.00084	0.01883	PTPN21	3	
rs10150311	14	88 046 225	G/A	0.317	0.00133	0.04164	PTPN21	2	
rs17260415	14	88 281 726	G/C	0.280	0.00650	0.04058	EML5	2	Echinoderm microtubule-associated protein like 5
rs7147796	14	88 298 322	C/G	0.469	0.03648	0.01437	EML5	1	
rs8182086	15	83 142 863	A/G	0.213	0.03924	0.04619	ZNF592	3	Zinc-finger protein 592
rs16941261	15	86 456 524	G/C	0.215	0.00008	0.03049	NTRK3	1	Neurotrophic tyrosine kinase, receptor, type 3
rs3784405	15	86 489 014	C/T	0.188	0.00041	0.03153	NTRK3	1	
rs7198357	16	66 442 120	A/G	0.160	0.00109	0.00143	NUTF2	1	Nuclear transport factor 2
rs2271293	16	66 459 571	A/G	0.104	0.00010	0.01518	NUTF2	1	

Abbreviations: CATIE, Clinical Antipsychotic Trials of Intervention Effectiveness; MAF, minor allele frequency from the MGS-GAIN sample; MGS-GAIN, molecular genetics of schizophrenia genome-wide association study supported by the genetic association information network; SNP, single-nucleotide polymorphism.
[a]MGS-GAIN *P*-values were obtained with some overlapping subjects with the CATIE data set.

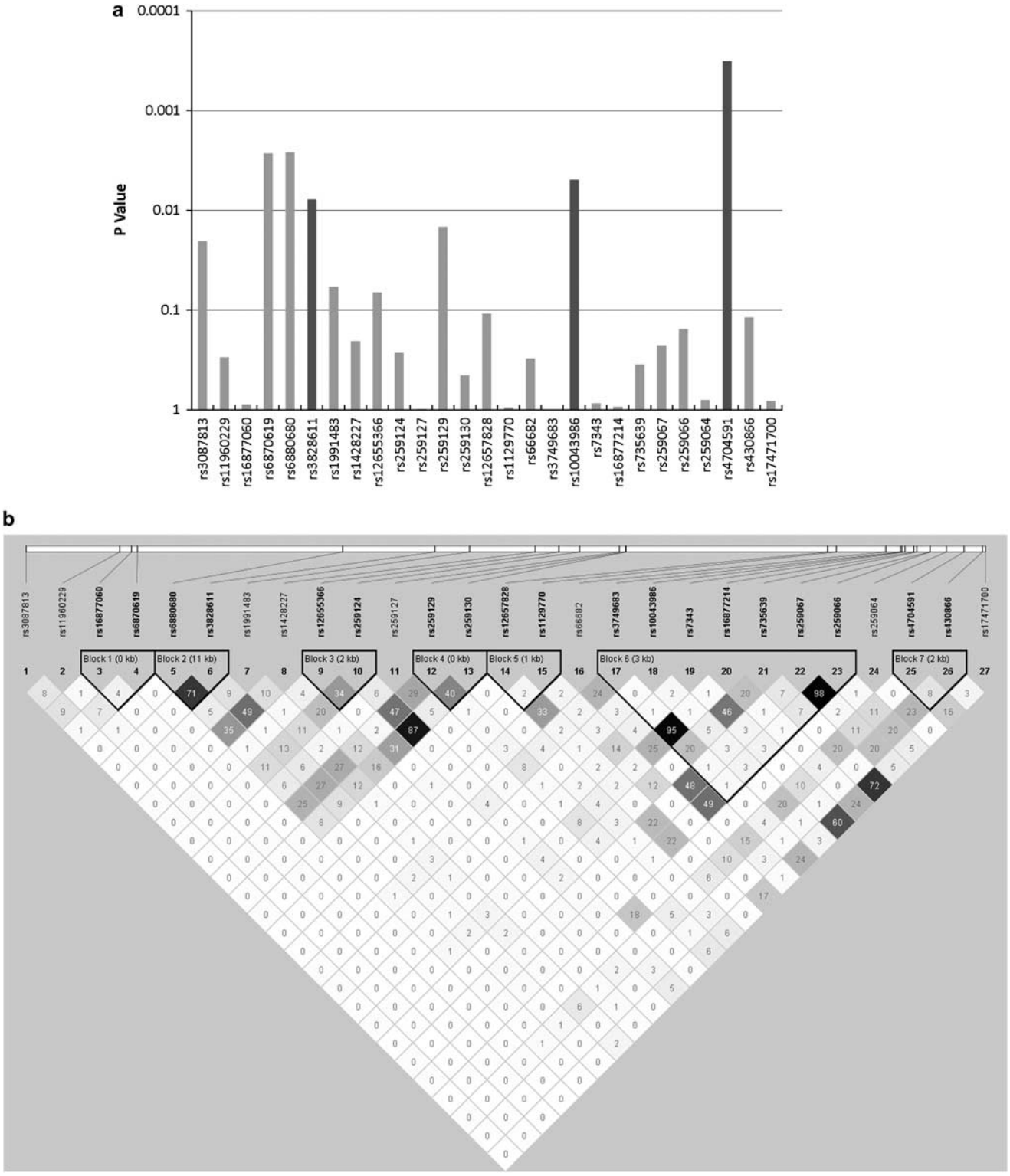

Figure 1 (**a**) Association analysis of the combined samples. The markers selected for replication were highlighted. (**b**) LD structure of the 27 markers typed in both CATIE and MGS-GAIN samples. Pair-wise LD values (r^2) were shown. The three markers studied were in low LD, suggesting that the association signals observed may be different at these markers.

For rs4704591, the case–control sample had a *P*-value of 0.2066 and the family sample had a *P*-value of 0.0083. The combined case–control and family samples had a *P*-value of 0.0041. The association of rs3828611 was in the opposite directions between our ICC and family samples (data not shown). Owing to

these conflict results, rs3828611 was dropped without further investigation.

Meta-analysis of rs10043986 and rs4704591

The results from ICC and family samples were encouraging. For further confirmation, we solicited data and replication from 23 more independent samples for rs10043986 and rs4704591. The information of all samples is summarized in Table 1, including the CATIE and MGS-GAIN used in our data-mining exercise. Overall, we had a total sample size of 33 834 subjects, including 912 families with 4160 subjects, 13 038 cases and 16 636 controls (the overlapping subjects were excluded from these numbers). Genotyping was conducted by individual groups using a variety of techniques (see Table 1). To ensure the quality, we examined the intensity plots for these two markers. Then, meta-analyses were performed. For the family samples, the counts for transmitted and untransmitted alleles were used. For the case–control samples, allele counts for cases and controls were used. In a meta-analysis of all 23 replication samples (family samples, 912 families with 4160 subjects; case–control samples, 11 380 cases and 15 021 controls), we found that both markers are significantly associated with schizophrenia (rs10043986, OR = 1.11, 95% CI = 1.04–1.18, $P = 8.2 \times 10^{-4}$ and rs4704591, OR = 1.07, 95% CI = 1.03–1.11, $P = 3.0 \times 10^{-4}$; Table 3). There was no significant heterogeneity among the samples (test of heterogeneity: rs10043986, Q = 13.88, d.f. = 16, $P = 0.61$; rs4704591, $Q = 17.15$, d.f. = 19, $P = 0.58$). The results were also significant for the 22 Caucasian replication samples (rs10043986, OR = 1.11, 95% CI = 1.03–1.17, $P = 0.0026$; rs4704591, OR = 1.07, 95% CI = 1.02–1.11, $P = 0.0015$). The results for the combined sample including CATIE and MGS GAIN and accounting for the overlap yields a combined *P*-value for rs4704591 of 1.11×10^{-5} ($Z = 4.39$) and for rs10043986 1.47×10^{-5} ($Z = 4.33$).

rs10043986 and rs4704591 are independently associated with schizophrenia

As we observed that two SNPs in the *CMYA5* gene are significantly associated with schizophrenia, we sought to evaluate whether these two association signals are independent. In the data-mining data sets, the LD between these two SNPs was relatively low ($r^2 = 0.208$) and similar results were obtained in our combined European samples ($r^2 = 0.212$), including the MGS-GAIN and CATIE samples. To test whether the effect of rs10043986 is independent of that of rs4704591, we inferred the haplotypes from these two markers for the combined European samples and evaluated whether those haplotypes sharing identical alleles at rs4704591, but different alleles at rs10043986, have different disease risk. If these haplotypes have significantly different risks, then the effects of these two markers would be at least partially independent. Table 4 summarized our results. From the table, it was clear that haplotypes sharing the same allele background at rs4704591 locus, that is C-C versus T-C and C-G versus T-G, showed significantly different risks to the disease. In other words, rs10043986 had an effect independent of that of rs4704591. We also checked the analyses with the PLINK program using all European case–control samples as PLINK could not combine family data with case–control data for such analysis. In this analysis, we checked the independent effect of rs10043986 by comparing the haplotypes sharing the same alleles at rs4704591 and the result was significant (OR = 1.07, $P = 0.0006$).

Table 3 Meta-analyses of rs10043986 and rs4704591 in all replication samples

Sample	*rs10043986 OR (95% CI)*	*rs4704591 OR (95% CI)*
MGS non-GAIN	1.16 (0.98–1.38)	0.99 (0.88–1.12)
IFAM	1.34 (0.83–2.17)	1.44 (1.08–1.92)
ICC	1.22 (0.96–1.57)	1.12 (0.94–1.33)
Cardiff	—	1.22 (1.06–1.41)
Pittsburgh	1.10 (0.72–1.66)	1.06 (0.80–1.41)
Bonn	1.28 (0.88–1.84)	1.28 (0.99–1.64)
NIMH	1.56 (1.09–2.23)	1.16 (0.91–1.50)
ISC Dublin	1.11 (0.82–1.50)	1.02 (0.84–1.24)
ISC Edinburgh	1.09 (0.76–1.55)	1.04 (0.83–1.30)
ISC Sweden 1	—	1.03 (0.75–1.40)
ISC Sweden 2	1.17 (0.84–1.64)	1.16 (0.91–1.46)
ISC Aberdeen	—	1.02 (0.88–1.19)
ISC Portugal	—	1.14 (0.89–1.46)
ISC London	—	1.12 (0.94–1.35)
ISC Bulgaria	0.97 (0.75–1.26)	1.01 (0.85–1.20)
Copenhagen	0.91 (0.69–1.21)	—
Oslo	0.99 (0.72–1.36)	0.96 (0.77–1.19)
Mannheim	0.94 (0.75–1.19)	—
Munich	1.12 (0.95–1.33)	1.02 (0.91–1.14)
Dutch	1.03 (0.83–1.28)	—
ZHH	—	1.08 (0.84–1.38)
Ashkenazi	1.18 (0.92–1.52)	0.93 (0.80–1.09)
MGS GAIN AA	1.56 (1.04–2.33)	1.13 (0.94–1.34)
All samples: *P*-value, OR (95% CI)	0.00082, 1.11 (1.04–1.18)	0.00030, 1.07 (1.03–1.11)

Abbreviations: CATIE, Clinical Antipsychotic Trials of Intervention Effectiveness; CI, confidence interval; ICC, Irish case–control; IFAM, Irish family; MAF, minor allele frequency from the MGS-GAIN sample; MGS-GAIN, molecular genetics of schizophrenia genome-wide association study supported by the genetic association information network; OR, odds ratio.

Discussion

In recent years, GWA studies have identified promising candidates in a number of complex disorders such as type 2 diabetes,[19–21] lung cancer,[22–24] Parkinson's disease,[25,26] rheumatoid arthritis[27] and systemic lupus erythematosus.[28] The results for schizophrenia have generally been less successful. Except the broad region in 6p and the *TCF4* and *NRGN* regions,[4] individual GWA studies have not produced candidates

Table 4 rs10043986 association conditioned on rs4704591 alleles

Haplotype[a]	*Frequency*		*OR*	*95% Lo*	*95% Hi*	χ^2	P-*value*
	Case and transmitted	*Control and untransmitted*					
C-C	0.247	0.266	1.00	1.00	1.00	0.10	0.74780
T-C	0.108	0.126	0.89	0.83	0.96	17.69	**0.00003**
C-G	0.642	0.603	1.04	0.99	1.09	11.23	**0.00081**
T-G	0.003	0.005	0.58	0.37	0.91	6.50	**0.01076**

Abbreviations: CI, confidence interval; OR, odds ratio.
[a]The haplotypes were inferred from rs10043986 to rs4704591. The first alleles were testing alleles (rs10043986), and the second alleles were conditioned alleles (rs4704591). Haplotype C-C is the reference haplotype. Given the same allele background at rs4704591, subjects carrying C-C and T-C haplotypes have different risks to the disease, and subjects carrying C-G and T-G halplotypes also have different risks to the disease, suggesting that rs10043986 has an effect independent of rs4704591.
Bold values signify $P \leqslant 0.05$.

reaching genome-wide significance yet. Of the many possible factors leading to the outcomes, insufficient power in these individual studies and the need to correct for a large number of markers tested may be important factors. However, as aggregated analyses indicated that there may be true findings among those markers passing nominal significance,[3] we believe that this is one of the most important contributions of GWA studies. Given the fact that there are markers/genes with true effects buried in the large number of tested markers, how to identify those markers with true effects is a practical issue facing the field. In this study, we adapted a two-stage approach, leading to the identification of two markers in the *CMYA5* gene. In the first hypothesis-generating stage, we conducted GWA analyses for two publicly available data sets, the CATIE and MGS-GAIN data sets, and selected and ranked markers by statistic and bioinformatic procedures. These procedures combined statistic and biological evaluations of markers with emphasis on the relevance of potential functions in disease. In this study, the finding of two non-synonymous markers in the *CMYA5* gene reaching nominal significance and the low LD between these markers had an important function in the selection of the gene for further verification and replication. The reported direct interaction between *CMYA5* and *DTNBP1*,[29] a leading candidate gene for schizophrenia, suggested that these genes may be involved in a common pathway or biological process. This piece of information moved the *CMYA5* gene to the top of our ranking list. In the second stage, a total of 23 independent data sets were used to evaluate the significance of these markers by standard meta-analyses. With these approaches, we were able to find that both markers in the gene are significantly associated with schizophrenia and there is no heterogeneity across the samples used in this study, including the MGS-GAIN-AA sample. Furthermore, the association signals observed in these two markers are independent. As we used a two-stage design in this study, the results should be evaluated by the number of markers tested in the second stage despite that we data mined two GWA data sets in our discovery stage. On the basis of this criterion and considering the large number of independent samples and the combined sample size, our results are significant. Importantly, one of our markers may have direct functional consequences as it changes the 4063rd amino acid of the protein from proline to leucin that would result in a change of residue size and hydrophobicity at the C-terminus of the protein, a region that was reported to interact with protein kinase regulator subunit.[30] The function of this non-synonymous SNP provides an opportunity to directly test its effect in the biology of schizophrenia.

Our motivation for this study was to find a way to reduce the penalty imposed by GWAs and enable us to identify markers with true effects, but not necessarily reaching conventional levels of genome-wide significance. GWA study is a great tool. Its systematic and hypothesis-free approach is objective and has great potential. However, in order to accomplish its aim, sufficient power and/or sample homogeneity are required to compensate for the steep penalty that has to be paid for testing hundreds of thousands markers. This creates a situation in which many markers may have true associations, but fail to reach GWA standards. On the basis of this rationale, we took the approach described in this study, leading to the identification of the *CMYA5* gene as a candidate for schizophrenia.

In retrospect, several aspects of the approach could have been improved. First, we did not take into account the differences in sample size and power between the MGS-GAIN and CATIE studies and used the same cutoff ($P \leqslant 0.05$) for both samples. Second, in matching the markers selected from the two data sets, we did not consider the sign (direction) of the association. A more objective approach might have been to perform a formal meta-analysis of the selected markers for the two data sets and take the meta-analysis *P*-values into consideration when ranking the markers. Third, in our bioinformatic prioritization, we focused on single SNP markers. We could

have extended these properties to markers in high LD with these markers, including imputation of untyped markers in the near neighborhood. Fourth, for a gene or region that had multiple markers associated with the disease, a haplotype analysis and testing of independent effects could have been conducted to select the best and independent markers for verification.

Our study provides an example that there are markers with true effects in the GWA studies, and given sufficiently large sample sizes, these markers can be identified. In this study, the observed ORs for rs10043986 and rs4704591 were 1.11 and 1.07, respectively, comparable with that observed in the *ZNF804A* gene.[6] For the *CMYA5* gene, there may be other association signals. The rs3828611 is the other non-synonymous marker selected by our data-mining procedures that has low LD with both rs10043986 and rs4704591. We did not pursue it further after the conflicting results from our ICC and family samples. In retrospect, our termination of rs3828611 may be premature.

The *CMYA5* gene, also known as myospryn, was first identified as a gene associated with cardiomyopathy.[31] The gene is highly expressed in skeletal muscle and heart, and is modestly expressed in brain (unpublished data). It is reported to be associated with left ventricular wall thickness in hypertension patients.[32] However, the function of the gene remains unknown. It has been reported to interact with *DTNBP1*,[29,33] the regulator subunit of protein kinase A[30] and desmin[34] in muscle cells. The interaction with *DTNBP1*, another leading candidate for schizophrenia,[35,36] is an interesting lead. *DTNBP1* was first reported to be associated with schizophrenia in our Irish sample.[18] Subsequently, many studies, including several studies that used samples[37–41] included in this paper, provided supporting evidence for the association. This interaction suggests that *CMYA5* may also be involved in the biogenesis of lysosome-related organelles complex 1 (BLOC-1) processes that have been suggested to be involved in schizophrenia.[42–45] The interaction with the regulatory subunit of protein kinase A suggests that *CMYA5* may be involved in the regulation of cAMP signal pathway, which is also implicated in schizophrenia.[46,47] These potential connections indicate that further studies may test epistatic interaction between these interacting partners, and examine their functions in the molecular, developmental and pathophysiological processes in schizophrenia.

In summary, using a two-stage design and with one of the largest sample sizes reported in recent literature, we report evidence that two SNPs with relatively low LD to each other in the *CMYA5* gene are independently associated with schizophrenia. These results suggest that there may be many markers in GWA data sets that have true but small effects. To identify these markers, a large sample size and collaborative work across many groups are essential.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We thank the volunteers, patients and their family members for participating in this study. This study was supported in part by a research grant (07R-1770) from the Stanley Medical Research Institute and an Independent Investigator Award from NARSAD to XC, and by grants to investigators involved in the collection and analyses of the samples from CATIE, GAIN, the international schizophrenia consortium and other independent samples (National Institutes of Health (MH41953, MH63480, MH56242, MH078075); NARSAD Young Investigator Award; Donald & Barbara Zucker Foundation, USA; the Medical Research Council and the Wellcome Trust Foundation, UK; the Research Council of Norway (Grant No. 163070/V50, 167153/V50); the South-Eastern Norway Health Authority (123/2004); Science Foundation Ireland and Health Research Board, Ireland; the Lundbeck Foundation and Danish National Advanced Technology Foundation, Denmark). The Ashkenazi samples are part of the Hebrew University Genetic Resource (HUGR). The principal investigators of the CATIE trial were Jeffrey A Lieberman, T Scott Stroup and Joseph P McEvoy. The CATIE trial was funded by a grant from the National Institute of Mental Health (N01 MH900001) along with MH074027 (PI PF Sullivan). Genotyping was funded by Eli Lilly and Company. The principle investigators for the MGS were Pablo Gejman and Douglas Levinson. MGS study was supported by funding from the National Institute of Mental Health and the National Alliance for Research on Schizophrenia and Depression. Genotyping of part of the sample was supported by GAIN and the Paul Michael Donovan Charitable Foundation. Genotyping was carried out by the Center for Genotyping and Analysis at the Broad Institute of Harvard and MIT with support from the National Center for Research Resources.

References

1 Sullivan PF, Lin D, Tzeng JY, van den OE, Perkins D, Stroup TS *et al.* Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 2008; **13**: 570–584.
2 Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009; **460**: 753–757.
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 Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane CR *et al.* Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol Psychiatry* 2007; **12**: 572–580.
6 O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 2008; **40**: 1053–1055.

7 O'Donovan MC, Norton N, Williams H, Peirce T, Moskvina V, Nikolov I *et al.* Analysis of 10 independent samples provides evidence for association between schizophrenia and a SNP flanking fibroblast growth factor receptor 2. *Mol Psychiatry* 2009; **14**: 30–36.
8 Ingason A, Giegling I, Cichon S, Hansen T, Rasmussen HB, Nielsen J *et al.* A large replication study and meta-analysis in European samples provides further support for association of AHI1 markers with schizophrenia. *Hum Mol Genet* 2010; **19**: 1379–1386.
9 Livak KJ. Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genet Anal* 1999; **14**: 143–149.
10 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.
11 Sun J, Jia P, Fanous AH, Webb BT, van den Oord EJ, Chen X *et al.* A multi-dimensional evidence-based candidate gene prioritization approach for complex diseases-schizophrenia as a case. *Bioinformatics* 2009; **25**: 2595–6602.
12 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
13 Dudbridge F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered* 2008; **66**: 87–98.
14 Martin ER, Monks SA, Warren LL, Kaplan NL. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 2000; **67**: 146–154.
15 Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959; **22**: 719–748.
16 Shyn SI, Shi J, Kraft JB, Potash JB, Knowles JA, Weissman MM *et al.* Novel loci for major depression identified by genome-wide association study of sequenced treatment alternatives to relieve depression and meta-analysis of three studies. *Mol Psychiatry* 2010 (in press).
17 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.
18 Straub RE, Jiang Y, MacLean CJ, Ma Y, Webb BT, Myakishev MV *et al.* Genetic variation in the 6p22.3 gene DTNBP1, the human ortholog of the mouse dysbindin gene, is associated with schizophrenia. *Am J Hum Genet* 2002; **71**: 337–348.
19 Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 2007; **316**: 1341–1345.
20 Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007; **445**: 881–885.
21 Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007; **316**: 1336–1341.
22 Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T *et al.* Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008; **40**: 616–622.
23 Spitz MR, Amos CI, Dong Q, Lin J, Wu X. The CHRNA5-A3 region on chromosome 15q24-25.1 is a risk factor both for nicotine dependence and for lung cancer. *J Natl Cancer Inst* 2008; **100**: 1552–1556.
24 Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008; **452**: 638–642.
25 Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M *et al.* Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet* 2009; **41**: 1303–1307.
26 Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet* 2009; **41**: 1308–1312.
27 Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burtt NP *et al.* Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat Genet* 2008; **40**: 1216–1223.
28 Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM *et al.* Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 2008; **40**: 1059–1061.
29 Benson MA, Tinsley CL, Blake DJ. Myospryn is a novel binding partner for dysbindin in muscle. *J Biol Chem* 2004; **279**: 10450–10458.
30 Reynolds JG, McCalmon SA, Tomczyk T, Naya FJ. Identification and mapping of protein kinase A binding sites in the costameric protein myospryn. *Biochim Biophys Acta* 2007; **1773**: 891–902.
31 Sarparanta J. Biology of myospryn: what's known? *J Muscle Res Cell Motil* 2008; **29**: 177–180.
32 Nakagami H, Kikuchi Y, Katsuya T, Morishita R, Akasaka H, Saitoh S *et al.* Gene polymorphism of myospryn (cardiomyopathy-associated 5) is associated with left ventricular wall thickness in patients with hypertension. *Hypertens Res* 2007; **30**: 1239–1246.
33 Talbot K, Cho DS, Ong WY, Benson MA, Han LY, Kazi HA *et al.* Dysbindin-1 is a synaptic and microtubular protein that binds brain snapin. *Hum Mol Genet* 2006; **15**: 3041–3054.
34 Kouloumenta A, Mavroidis M, Capetanaki Y. Proper perinuclear localization of the TRIM-like protein myospryn requires its binding partner desmin. *J Biol Chem* 2007; **282**: 35211–35221.
35 Owen MJ, Williams NM, O'Donovan MC. Dysbindin-1 and schizophrenia: from genetics to neuropathology. *J Clin Invest* 2004; **113**: 1255–1257.
36 Williams NM, O'Donovan MC, Owen MJ. Is the dysbindin gene (DTNBP1) a susceptibility gene for schizophrenia? *Schizophr Bull* 2005; **31**: 800–805.
37 Duan J, Martinez M, Sanders AR, Hou C, Burrell GJ, Krasner AJ *et al.* DTNBP1 (Dystrobrevin Binding Protein 1) and schizophrenia: association evidence in the 3′ end of the gene. *Hum Hered* 2007; **64**: 97–106.
38 Kirov G, Ivanov D, Williams NM, Preece A, Nikolov I, Milev R *et al.* Strong evidence for association between the dystrobrevin binding protein 1 gene (DTNBP1) and schizophrenia in 488 parent-offspring trios from Bulgaria. *Biol Psychiatry* 2004; **55**: 971–975.
39 Riley B, Kuo PH, Maher BS, Fanous AH, Sun J, Wormley B *et al.* The dystrobrevin binding protein 1 (DTNBP1) gene is associated with schizophrenia in the Irish Case Control Study of Schizophrenia (ICCSS) sample. *Schizophr Res* 2009; **115**: 245–253.
40 Williams NM, Preece A, Morris DW, Spurlock G, Bray NJ, Stephens M *et al.* Identification in 2 independent samples of a novel schizophrenia risk haplotype of the dystrobrevin binding protein gene (DTNBP1). *Arch Gen Psychiatry* 2004; **61**: 336–344.
41 Schwab SG, Knapp M, Mondabon S, Hallmayer J, Borrmann-Hassenbach M, Albus M *et al.* Support for association of schizophrenia with genetic variation in the 6p22.3 gene, dysbindin, in sib-pair families with linkage and in an additional sample of triad families. *Am J Hum Genet* 2003; **72**: 185–190.
42 Ghiani CA, Starcevic M, Rodriguez-Fernandez IA, Nazarian R, Cheli VT, Chan LN *et al.* The dysbindin-containing complex (BLOC-1) in brain: developmental regulation, interaction with SNARE proteins and role in neurite outgrowth. *Mol Psychiatry* 2010; **15**: 115, 204–15.
43 Iizuka Y, Sei Y, Weinberger DR, Straub RE. Evidence that the BLOC-1 protein dysbindin modulates dopamine D2 receptor internalization and signaling but not D1 internalization. *J Neurosci* 2007; **27**: 12390–12395.
44 Morris DW, Murphy K, Kenny N, Purcell SM, McGhee KA, Schwaiger S *et al.* Dysbindin (DTNBP1) and the biogenesis of lysosome-related organelles complex 1 (BLOC-1): main and epistatic gene effects are potential contributors to schizophrenia susceptibility. *Biol Psychiatry* 2008; **63**: 24–31.
45 Guo AY, Sun J, Riley BP, Thiselton DL, Kendler KS, Zhao Z. The dystrobrevin-binding protein 1 gene: features and networks. *Mol Psychiatry* 2009; **14**: 18–29.
46 Molteni R, Calabrese F, Racagni G, Fumagalli F, Riva MA. Antipsychotic drug actions on gene modulation and signaling mechanisms. *Pharmacol Ther* 2009; **124**: 74–85.
47 Siuciak JA. The role of phosphodiesterases in schizophrenia: therapeutic implications. *CNS Drugs* 2008; **22**: 983–993.
48 Chen X, Wang X, Hossain S, O'Neill FA, Walsh D, Pless L *et al.* Haplotypes spanning SPEC2, PDZ-G EF2 and ACSL6 genes are associated with schizophrenia. *Hum Mol Genet* 2006; **15**: 3329–3342.

49 Chowdari KV, Mirnics K, Semwal P, Wood J, Lawrence E, Bhatia T *et al.* Association and linkage analyses of RGS4 polymorphisms in schizophrenia. *Hum Mol Genet* 2002; **11**: 1373–1380.
50 Schwab SG, Hoefgen B, Hanses C, Hassenbach MB, Albus M, Lerer B *et al.* Further evidence for association of variants in the AKT1 gene with schizophrenia in a sample of European sib-pair families. *Biol Psychiatry* 2005; **58**: 446–450.
51 Egan MF, Goldberg TE, Kolachana BS, Callicott JH, Mazzanti CM, Straub RE *et al.* Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc Natl Acad Sci USA* 2001; **98**: 6917–6922.
52 Olsen L, Hansen T, Jakobsen KD, Djurovic S, Melle I, Agartz I *et al.* The estrogen hypothesis of schizophrenia implicates glucose metabolism: association study in three independent samples. *BMC Med Genet* 2008; **9**: 39.
53 Kahler AK, Djurovic S, Kulle B, Jonsson EG, Agartz I, Hall H *et al.* Association analysis of schizophrenia on 18 genes involved in neuronal migration: MDGA1 as a new susceptibility gene. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1089–1100.
54 Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ *et al.* Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet* 2008; **4**: e28.

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

Appendix

The members of the Genetic Risk and Outcome in Psychosis (GROUP):

René S Kahn[1], Don H Linszen[2], Jim van Os[3], Durk Wiersma[4], Richard Bruggeman[4], Wiepke Cahn[1], Lieuwe de Haan[2], Lydia Krabbendam[3] and Inez Myin-Germeys[3]

[1]Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Postbus 85060, 3508 AB, Utrecht, The Netherlands; [2]Department of Psychiatry, Academic Medical Centre University of Amsterdam, Amsterdam, NL326 Groot-Amsterdam, The Netherlands; [3]Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, P. Debyelaan 25, 6229 HX Maastricht, Maastricht, The Netherlands; [4]Department of Psychiatry, University Medical Center Groningen, University of Groningen, PO Box 30.001, 9700 RB Groningen, The Netherlands.

The members of the International schizophrenia consortium:

Cardiff University: Michael C O'Donovan[6], George K Kirov[6], Nick J Craddock[6], Peter A Holmans[6], Nigel M Williams[6], Lyudmila Georgieva[6], Ivan Nikolov[6], N Norton[6], H Williams[6], Draga Toncheva[16], Vihra Milanova[17], Michael J Owen[6]; *Karolinska Institutet/ University of North Carolina at Chapel Hill:* Christina M Hultman[11,12], Paul Lichtenstein[11], Emma F Thelander[11], Patrick Sullivan[7]; *Trinity College Dublin:* Derek W Morris[9], Colm T O'Dushlaine[9], Elaine Kenny[9], Emma M Quinn[9], Michael Gill[9], Aiden Corvin[9]; *University College London:* Andrew McQuillin[8], Khalid Choudhury[8], Susmita Datta[8], Jonathan Pimm[8], Srinivasa Thirumalai[18], Vinay Puri[8], Robert Krasucki[8], Jacob Lawrence[8], Digby Quested[19], Nicholas Bass[8], Hugh Gurling[8]; *University of Aberdeen:* Caroline Crombie[15], Gillian Fraser[15], Soh Leh Kuan[14], Nicholas Walker[20], David St Clair[14]; *University of Edinburgh:* Douglas HR Blackwood[10], Walter J Muir[10], Kevin A McGhee[10], Ben Pickard[10], Pat Malloy[10], Alan W Maclean[10], Margaret Van Beck[10]; *Queensland Institute of Medical Research:* Naomi R Wray[5], Stuart Macgregor[5], Peter M Visscher[5]; *University of Southern California:* Michele T Pato[13], Helena Medeiros[13], Frank Middleton[21], Celia Carvalho[13], Christopher Morley[21], Ayman Fanous[13,22,23,24], David Conti[13], James A Knowles[13], Carlos Paz Ferreira[25], Antonio Macedo[26], M Helena Azevedo[26], Carlos N Pato[13]; *Massachusetts General Hospital:* Jennifer L Stone[1,2,3,4], Douglas M Ruderfer[1,2,3,4], Andrew N Kirby[2,3,4], Manuel AR Ferreira[1,2,3,4], Mark J Daly[2,3,4], Shaun M Purcell[1,2,3,4], Pamela Sklar[1,2,3,4]; *Stanley Center for Psychiatric Research and Broad Institute of MIT and Harvard:* Shaun M Purcell[1,2,3,4], Jennifer L Stone[1,2,3,4], Kimberly Chambert[3,4], Douglas M Ruderfer[1,2,3,4], Finny Kuruvilla[4], Stacey B Gabriel[4], Kristin Ardlie[4], Jennifer L Moran[4], Mark J Daly[2,3,4], Edward M Scolnick[3,4], Pamela Sklar[1,2,3,4].

[1]Psychiatric and Neurodevelopmental Genetics Unit, [2]Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge Street, Boston, MA 02114, USA; [3]Stanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [4]The Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [5]Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia; [6]Department of Psychological Medicine, MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff C14 4XN, UK; [7]Departments of Genetics, Psychiatry, and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA; [8]Research Department of Mental Health Sciences, Molecular Psychiatry Laboratory, University College London Medical School, Windeyer Institute of Medical Sciences, 46 Cleveland Street, LondonW1T4JF, UK; [9]Department of Psychiatry and Institute of Molecular Medicine, NeuropsychiatricGenetics Research Group, Trinity College Dublin, Dublin 2, Ireland; [10]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh EH10 5HF, UK; [11]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-171 77 Stockholm, Sweden; [12]Department of Neuroscience, Psychiatry, Ulleråker, Uppsala University, SE-750 17 Uppsala, Sweden; [13]Center for Genomic Psychiatry, University of Southern California, Los Angeles, CA 90033, USA; [14]Institute of Medical Sciences, [15]Department of Mental Health, University of Aberdeen, Aberdeen AB25 2ZD, UK; [16]Department of Medical Genetics, University Hospital Maichin Dom, Sofia 1431, Bulgaria; [17]Department of Psychiatry, First Psychiatric Clinic, Alexander University Hospital, Sofia 1431, Bulgaria; [18]West Berkshire NHS Trust,

25 Erleigh Road, Reading RG3 5LR, UK; [19]Department of Psychiatry, University of Oxford, Warneford Hospital, Headington, Oxford OX3 7JX, UK; [20]Ravenscraig Hospital, Inverkip Road, Greenock PA16 9HA, UK; [21]State University of New York—Upstate Medical University, Syracuse, NY 13210, USA; [22]Washington VA Medical Center, Washington DC 20422, USA; [23]Department of Psychiatry, Georgetown University School of Medicine, Washington DC 20057, USA; [24]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA 23298, USA; [25]Department of Psychiatry, Sao Miguel, 9500-310 Azores, Portugal; [26]Department of Psychiatry University of Coimbra, 3004-504 Coimbra, Portugal.