DOCK4 and CEACAM21 as novel schizophrenia candidate genes in the Jewish population

Anna Alkelai[1]*, Sara Lupoli[2,4]*, Lior Greenbaum[1], Yoav Kohn[1], Kyra Kanyas-Sarner[1],
Edna Ben-Asher[3], Doron Lancet[3], Fabio Macciardi[4,5] and Bernard Lerer[1]

[1] *Biological Psychiatry Laboratory, Department of Psychiatry, Hadassah – Hebrew University Medical Center, Jerusalem, Israel*
[2] *INSPE, Scientific Institute San Raffaele, Milan, Italy*
[3] *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel*
[4] *Department of Medicine, Surgery and Dentistry, University of Milan, Italy*
[5] *Department of Psychiatry and Human Behavior, University of California Irvine, USA*

Abstract

It is well accepted that schizophrenia has a strong genetic component. Several genome-wide association studies (GWASs) of schizophrenia have been published in recent years; most of them population based with a case-control design. Nevertheless, identifying the specific genetic variants which contribute to susceptibility to the disorder remains a challenging task. A family-based GWAS strategy may be helpful in the identification of schizophrenia susceptibility genes since it is protected against population stratification, enables better accounting for genotyping errors and is more sensitive for identification of rare variants which have a very low frequency in the general population. In this project we implemented a family-based GWAS of schizophrenia in a sample of 107 Jewish-Israeli families. We found one genome-wide significant association in the intron of the *DOCK4* gene (rs2074127, p value$=1.134 \times 10^{-7}$) and six additional nominally significant association signals with $p < 1 \times 10^{-5}$. One of the top single nucleotide polymorphisms ($p < 1 \times 10^{-5}$) which is located in the predicted intron of the *CEACAM21* gene was significantly replicated in independent family-based sample of Arab-Israeli origin (rs4803480: p value$=0.002$; combined p value$=9.61 \times 10^{-8}$), surviving correction for multiple testing. Both *DOCK4* and *CEACAM21* are biologically reasonable candidate genes for schizophrenia although generalizability of the association of *DOCK4* with schizophrenia should be investigated in further studies. In addition, gene-wide significant associations were found within three schizophrenia candidate genes: *PGBD1*, *RELN* and *PRODH*, replicating previously reported associations. By application of a family-based strategy to GWAS, our study revealed new schizophrenia susceptibility loci in the Jewish-Israeli population.

Introduction

It is well established that in addition to environmental risk factors, schizophrenia has a strong genetic component, with an estimated heritability of ∼80% (Sullivan *et al.* 2003). Nevertheless, identifying the specific genetic variants which contribute to susceptibility to the disorder has proved a challenging task.

In recent years it has become increasingly accepted that the genetic architecture of schizophrenia involves a large number of common and rare variants with small and large effects, respectively (Owen *et al.* 2010; Schwab & Wildenauer, 2009; Wray & Visscher, 2010). More than one model of inheritance is probably involved (Mitchell & Porteous, 2011). Several genome-wide association studies (GWASs) of schizophrenia have been published in recent years; almost all of them have been population based with a case-control design (Kirov *et al.* 2009; Lencz *et al.* 2007; Mah *et al.* 2006; O'Donovan *et al.* 2008; Shi *et al.* 2009; Shifman *et al.* 2008; Stefansson *et al.* 2009; Sullivan *et al.* 2008). These studies revealed a number of

Address for correspondence: Professor B. Lerer, Biological Psychiatry Laboratory, Department of Psychiatry, Hadassah – Hebrew University Medical Center, Ein Karem, Jerusalem 91120, Israel.
Tel.: 972-2-6777185   *Fax*: 972-2-6439294
Email: lerer@cc.huji.ac.il
* These authors contributed equally to this work.

novel non-overlapping schizophrenia susceptibility genes.

In the current project, we implemented a GWAS of schizophrenia with a family-based design in a sample consisting of Jewish-Israeli families. Family-based association studies are less common in the field of psychiatric genetics, perhaps due to the more challenging sample recruitment. However, the family-based approach has several marked advantages compared to the more common case-control design. First, using family-based samples provides robust protection against population stratification which may lead to spurious results due to difference in allele frequencies among different subpopulations included in the sample. In addition, the family-based method enables better accounting for genotyping errors by detection of Mendelian errors. Therefore, implementation of a family-based GWAS strategy may be helpful in the identification of schizophrenia susceptibility genes (Laird & Lange, 2006). Second, successful population-based searching for common risk alleles with small effect sizes requires very large samples, or enrichment of these risk alleles in moderately sized samples by chance or by using an ethnically homogeneous sample (Schwab & Wildenauer, 2009). Rare variants with large effect have a very low frequency in the general population and therefore will not be detected by the currently widely used population-based GWAS strategy. However, they could be detected in families and in ethnically homogeneous populations (Owen *et al.* 2010). On the other hand, these high impact genetic variants may be ethnically specific, with low generalization to other populations.

In this project, we applied a family-based GWAS strategy, using a sample which belongs to the relatively ethnically homogeneous, Jewish-Israeli population. Major support for the rationale of this approach was provided by a recently published comprehensive study (Behar *et al.* 2010), which showed that different Jewish populations in the world genetically resemble each other, as well as other populations in the Levant region (e.g. Druze and Cypriot), more than they resemble the host populations of the countries in which they reside (Behar *et al.* 2010). We may therefore expect affected individuals to share fewer disease causative variants than a general Caucasian population. To identify schizophrenia susceptibility genes in our Jewish-Israeli family sample, we performed genome-wide single nucleotide polymorphism (SNP) genotyping, using the HumanCNV-370 BeadArray (Illumina, USA) platform and family-based analysis. A validation step followed, applying the same family-based procedure in an independent sample of Arab origin.

Methods

Family samples

The Jewish-Israeli family sample consisted of 107 nuclear families [331 individuals with DNA of whom 155 are affected, 43 Ashkenazi and 64 non-Ashkenazi families (both parents being of Ashkenazi or non-Ashkenazi origin), mixed families (one parent Ashkenazi and one non-Ashkenazi) were not included]; 70 were 'triad' families (parents and one affected offspring), while the rest had two or more offspring, with at least one affected with schizophrenia. Clinical evaluation included a semi-structured interview, the Schedule for Affective Disorders and Schizophrenia – Lifetime Version (SADS-L; Spitzer & Endicott, 1977), a Family History Diagnostic Interview (FHRDC; Andreasen *et al.* 1977) and assessment of medical records. Primary diagnoses were established by a best-estimate procedure (Baron *et al.* 1994) according to Research Diagnostic Criteria (RDC; Spitzer *et al.* 1978).

For validation, we used available genotyping data from our previous GWAS (Alkelai *et al.* unpublished data) in a homogeneous Arab-Israeli family sample (Alkelai *et al.* 2009; Amann-Zalcenstein *et al.* 2006; Lerer *et al.* 2003; Levi *et al.* 2005), consisting of 58 nuclear families (198 genotyped individuals of whom 99 were affected). Further details of the clinical samples are provided in our previous publications (Alkelai *et al.* 2009; Amann-Zalcenstein *et al.* 2006; Lerer *et al.* 2003; Levi *et al.* 2005).

Research protocols were approved by the Helsinki Committee (Internal Review Board) of the Hadassah – Hebrew University Medical Center and all participants gave written informed consent.

Genotyping

DNA was isolated from blood samples or immortalized cell lines by standard methods. Genotyping of the main Jewish sample (as well as the Arab validation sample) was performed at the Platform of Genomics and Bioinformatics, University of Milan. For the GWAS, ~750 ng genomic DNA was used to genotype each subject for 370 404 Phase I Hap Map tagging SNPs on the HumanCNV-370 BeadArrays (Illumina, USA). Samples were processed according to the Illumina Infinium 2 assay. Each sample was whole-genome amplified, fragmented, precipitated and hybridized overnight for a minimum

of 16 h at 48 °C to locus-specific probes on the BeadArray. Non-specifically hybridized fragments were removed by washing while the remaining specifically hybridized DNA fragments were processed for the single base extension reaction, stained and imaged on an Illumina BeadArray Reader. Normalized bead intensity data obtained for each sample were analysed with Illumina GenomeStudio 1.0.2 software, which generated SNP genotypes from fluorescent intensities using the manufacturer's default cluster settings (Fan *et al.* 2006; Steemers & Gunderson, 2005).

Quality control

Quality control (QC) of the Jewish and Arab samples was performed using PLINK version 1.06 (Purcell *et al.* 2007) (http://pngu.mgh.harvard.edu/~purcell/plink/) and included evaluation of call rate, check of SNPs with (1) no calls, (2) with MAF <0.05 and (3) genotyping rate <0.9; when below these standards, SNPs were removed from further analyses. In addition, individuals with missing genotyping of >10% were not included in the analyses. Hardy–Weinberg equilibrium (HWE) testing was also performed and SNPs that showed a significant deviation from HWE ($p < 0.00001$) in parents were excluded. Additional QC steps taken were to check for the assessment of genetic homogeneity according to the family, sex-check and Mendelian transmission rate. Respectively SNPs with >10% and families with >5% Mendelian error rate were discarded. For the genetic homogeneity analysis of the families, we used the estimation of pair-wise identity by descent (IBD). For each individual pair, we evaluated the probability of sharing zero, one or two alleles and the proportion of IBD between them. This way, we were able to correctly 're-construct' each family's pedigree and also to exclude individuals whose genotype did not match our family records. The sex-check was performed using GenomeStudio software. On the basis of the genotyping calls of the sex chromosomes GenomeStudio software estimated the sex for each subject.

Association analyses

PBAT version 3.6 (Lange *et al.* 2004; Van Steen & Lange, 2005) (http://www.biostat.harvard.edu/) which incorporates an extended and improved transmission disequilibrium test (TDT) for family samples with more than one offspring and various structure, was used for association analysis of the main family sample. PBAT statistics were calculated under the null hypothesis of 'no-linkage-and-no-association'. The

mode of inheritance of schizophrenia is complex and the process of GWAS analysis with PBAT is highly labour- and computer-intensive; therefore we chose the additive model for the data analysis. Although other models of schizophrenia inheritance (recessive or dominant) are plausible, applying them also increases the burden of multiple testing. The additive model is one of the most common methods used to analyse GWAS data when no previous assumption about model of inheritance can be made. The minimal number of informative families was restricted to 10. The PBAT software is not compatible with sex-linked SNPs, therefore the analysis included only autosomal SNPs. We used QVALUE software (http://genomine.org/qvalue/) (Storey & Tibshirani, 2003) with the default smoother method parameters to calculate false discovery rate (FDR)-based q values, in order to measure and estimate the statistical significance of the association results at the genome-wide level. The cut-off for significant association at the genome-wide level was set at FDR q value <0.05; thus, we expected no more than 5% of declared discoveries to be false. We also used the rigorous Bonferroni correction for multiple testing at the GWAS level, although it may be over-conservative, since it does not take into account the intrinsic correlations between tests because of linkage disequilibrium among SNPs. We then used PLINK software (Purcell *et al.* 2007) to estimate effect sizes for the implicated loci in a subset of the sample that included only trios.

PBAT version 3.6 (Lange *et al.* 2004; Van Steen & Lange, 2005) was also used for the analysis of the best seven SNPs (from the Jewish-Israeli sample) in the Arab validation sample. The TDT analysis was performed while using the null hypothesis of 'no-linkage-and-no-association', exactly as in the discovery sample. In the combined analysis of the two family-based samples, 520 individuals (331 Jews, 189 Arabs) were analysed by applying the same PBAT software, which is robust against effects of population stratification and admixture.

GWAS results for a selected set of candidate genes for schizophrenia

We focused separately on a selected set of schizophrenia candidate genes according to (1) SchizophreniaGene Top Results from the SZGene database (Allen *et al.* 2008) (http://www.schizophreniaforum.org/res/sczgene/default.asp), (2) the set selected by Sullivan *et al.* (2008) and (3) previous GWASs (Kirov *et al.* 2009; Lencz *et al.* 2007; Mah *et al.* 2006; O'Donovan *et al.* 2008; Shi *et al.* 2009; Shifman *et al.*

Table 1. Replication study results

Chr	Marker	bp	Minor allele	Freq	HWE parents	Direction of the affect	p value	Combined analysis[a] p value
13	rs2152700	68 582 849	3	0.453	0.902	Positive	0.483	3.288×10^{-5}
15	rs2468756	31 224 276	1	0.443	0.822	Positive	0.433	8.27×10^{-5}
10	rs1412115	34 128 059	1	0.366	0.954	Negative	0.148	5.709×10^{-6}
7	rs38827	117 501 273	1	0.108	0.243	Positive	0.572	0.001
18	rs2194631	68 871 803	1	0.155	0.419	Negative	0.265	2.63×10^{-5}
19	rs4803480	46 758 119	1	0.130	0.864	Negative	0.003	9.612×10^{-8}

HWE, Hardy–Weinberg equilibrium; Chr, chromosome; Freq, minor allele frequency.

The single nucleotide polymorphisms which pass the correction for multiple testing are highlighted in light grey.

[a] The combined analysis of the Jewish and Arab samples.

2008; Stefansson *et al.* 2009; Sullivan *et al.* 2008) (Table 3). All the genes from these three sources were analysed. Overall, 46 genes were studied. In order to study association of these candidate genes with schizophrenia systematically, we analysed all the SNPs (863) genotyped within them. Bonferroni correction for multiple testing was used to consider the results experiment-wide (863 tests) and gene-wide (according to number of SNPs per gene) significant.

Results

Quality control

In the Jewish-Israeli family sample, the 331 genotyped individuals had a mean genotyping rate of 0.991. In total, 3340 SNPs failed the locus missingness test, 21 280 SNPs had a minor allele frequency <0.05 and 415 SNPs failed the HWE test. Eleven markers showed a Mendelian error rate above 10% and were removed; no family showed more than 5% Mendelian errors. After frequency and genotyping pruning 311 517 autosomal SNPs remained available for the association analysis.

In the Arab-Israeli family sample, GWAS data QC measures were taken as described by Alkelai *et al.* (unpublished data). Briefly, the same Ilumina GWAS platform was used. We excluded three individuals who showed a low genotyping rate (<0.9). The remaining 195 genotyped individuals had a mean genotyping rate of 0.989. In total, 7906 SNPs failed the locus missingness test, 22 096 SNPs had a minor allele frequency <0.05, 170 SNPs failed the HWE test and nine markers showed a Mendelian error rate >10%. Looking at family transmissions we removed two individuals. No family showed more than 5% Mendelian errors. On the basis of genetic homogeneity

assessment, we excluded four individuals, leaving us with 189 genotyped individuals and 57 nuclear families. A total of 307 472 autosomal SNPs remained available for the association analysis; including the top six SNPs of the Jewish-Israeli family sample GWAS (Alkelai *et al.* unpublished data).

GWAS

The top results of our study ($p<1\times10^{-5}$) are presented in Table 2 and Fig. 1. We found genome-wide significant association (q value <0.05, the best p value$=1.134\times10^{-7}$) for a SNP (rs2074127) positioned within the *DOCK4* gene (intron 6). Six additional association signals were detected with $p<1\times10^{-5}$ and q value <0.4. Five of the top seven SNPs are located on different chromosomes (chr13: rs2152700; chr15: rs2468756; chr10: rs1412115; chr18: rs2194631; chr19: rs4803480), and two (rs2074127, rs38827) are in the 7q31.1-q31.31 region, but at a distance of ∼3 million bp and therefore may be assumed to represent independent signals. Results of all top ($p<0.0001$) associations are given in Supplementary Table S1 (available online). Interestingly, we could identify several association clusters: rs2468756, rs2444953 and rs4780082 near the *FMN1* gene; rs1412115 and rs2253574 near the *PARD3* gene; and, rs2194631 and rs7242928 near the *NETO1* gene. These association clusters encouraged us to postulate that the findings are due to true positives rather than genotyping errors or chance findings.

Replication study

To determine generalizability of our best findings, we checked the association of the top seven Jewish sample SNPs with schizophrenia in a validation sample of Arab families (see Methods section), using the same

Table 2. The top seven results of GWAS for association with schizophrenia ($p < 1 \times 10^{-5}$)

Chr	Marker	bp	Allele	Freq	HW_parents	Gen_rate	nbr_info_fam	Direction of the main effect	p value	q value	OR	Gene
7	rs2074127	111 411 758	3	0.217	0.010	0.994	35	Positive	1.134×10^{-7}	0.034	3.000	*DOCK4*
13	rs2152700	68 582 849	3	0.450	0.692	0.976	55	Positive	2.493×10^{-6}	0.360	2.654	PCDH9, **KLHL1**
15	rs2468756	31 224 276	1	0.500	0.010	0.976	55	Positive	5.923×10^{-6}	0.360	3.278	**FMN1**, RYR3
10	rs1412115	34 128 059	1	0.236	0.780	0.991	60	Negative	6.556×10^{-6}	0.360	0.353	NRP1, **PARD3**
7	rs38827	117 501 273	1	0.113	0.947	0.988	36	Negative	7.420×10^{-6}	0.360	0.237	CTTNBP2, *NAA38*
18	rs2194631	68 871 803	1	0.056	0.979	1	33	Negative	9.547×10^{-6}	0.360	0.193	**NETO1**, FBXO15
19	rs4803480	46 758 119	1	0.158	0.060	0.982	42	Negative	9.649×10^{-6}	0.360	0.273	ATP5SL, **CEACAM21**

nbr_info_fam, Number of informative families; HW, Hardy–Weinberg equilibrium; Chr, chromosome; Allele, minor allele; Freq, minor allele frequency; Gen_rate, genotyping rate; OR, odds ratio.

p values were obtained using PBAT. The FDR cut-off was 0.05.

The single nucleotide polymorphisms (SNPs) which pass the correction for multiple testing are highlighted in light grey. The closest gene to the SNP is highlighted in bold.

TDT association test. This validation sample had been used by us for a GWAS in a previous study (Alkelai *et al.* unpublished data). All SNPs of interest were previously genotyped, and data were available for the association study. Applying the TDT procedure, one of the top SNPs, which is located on chromosome 19, was significantly associated with schizophrenia (rs4803480: *p* value = 0.002), surviving correction for multiple testing for seven tests (Table 1). Importantly, the risk allele, A, was consistently the same in both discovery and validation sample. Combined analysis of the two samples strengthened the association of rs4803480 with schizophrenia (*p* value = 9.612×10^{-8}).

GWAS results for a selected set of candidate genes for schizophrenia

The best SNP associations in our sample were not found in genes which are considered best schizophrenia candidate genes. We studied separately association of all genotyped SNPs (863) within 46 genes of interest (see Methods section for selection criteria). Gene-wide significant associations were found within three investigated genes: *PGBD1* (rs1150724, *p* = 0.007), *RELN* (rs39339, rs262342, *p* = 0.0005) and *PRODH* (rs2238732, *p* = 0.003). Full results are given in Table 3.

Discussion

We report here the results of a GWAS performed for schizophrenia in a sample of 107 Jewish-Israeli families. Of the top findings, the *DOCK4* intronic SNP rs2074127 withstood Bonferroni correction for multiple testing (*p* value = 1.134×10^{-7}) as well as the less rigorous FDR correction (*q* value < 0.05). The *DOCK4* gene on 7q31.1 encodes the dedicator of cytokinesis 4 protein. This protein is highly expressed in the developing brain and has been shown to have a role in regulating dendritic growth in rat hippocampal neurons (Ueda *et al.* 2008). *DOCK4* is involved in the Wnt/beta-catenin pathway, which has been associated with schizophrenia (Freyberg *et al.* 2010) and is required for Wnt/beta-catenin activity (Upadhyay *et al.* 2008). *DOCK4* was found in two large-scale studies to be associated with dyslexia (Pagnamenta *et al.* 2010) and autism (Maestrini *et al.* 2010) which have previously been reported to share common genetic factors with schizophrenia (Carroll & Owen, 2009). The *DOCK4* gene is widely expressed in various human tissues, and has high expression ratios in different brain regions (especially in caudate nucleus, amygdala

(*a*)

(*b*)
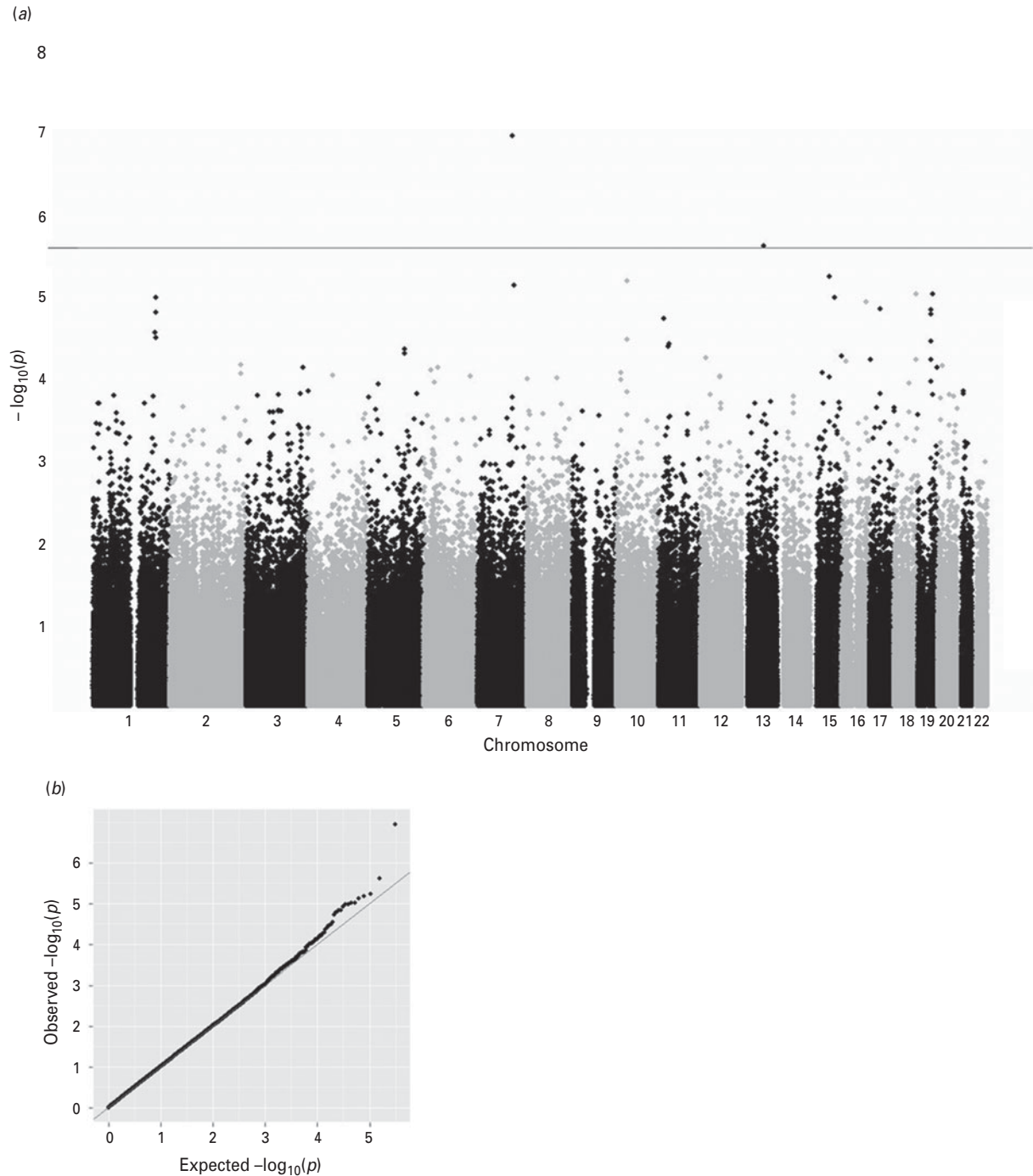
Fig. 1. GWAS results. (*a*) 'Manhattan' plot showing the $-\log_{10}$ (*p* values) of single nucleotide polymorphism (SNPs) from the single SNP association analysis, according to the position of the SNPs on each chromosome. The horizontal grey line indicates the genome-wide significance cut-off. (*b*) QQ plot of the observed $-\log_{10}$ (*p* values) *vs.* expected $-\log_{10}$ (*p* values).

and prefrontal cortex) (according to UCSC Genome Browser: http://genome.ucsc.edu/ and GeneNote browser: http://bioinfo2.weizmann.ac.il/cgi-bin/genenote/home_page.pl).

It was proposed by Owen *et al.* that attention should not be restricted to the few strongest findings of a GWAS (Owen *et al.* 2010); therefore, the possible role of other nominally significant SNPs in the

Table 3. GWAS results for a selected set of candidate genes for schizophrenia

Gene name	Gene product	Number of SNPs per gene	Chr	Start	End	Gene size (bp)	p values < 0.05	p values < 0.001	Minimal p value	Bonferroni gene-wide significance cut-off
MTHFR	5,10-methylenetetrahydrofolate reductase	6	1	11 768 374	11 788 747	20 374	0	0	0.24	0.0083
GRIK3	Glutamate receptor, ionotropic, kainate 3	20	1	37 033 715	37 272 431	238 716	0	0	0.11	0.0025
PDE4B	Phosphodiesterase 4B	61	1	66 030 781	66 612 850	582 069	3	0	0.005	0.0008
RGS4	Regulator of G-protein signalling 4	1	1	161 305 020	161 313 216	8197	0	0	0.68	0.0500
PLXNA2	Plexin A2	36	1	206 262 211	206 484 288	222 078	1	0	0.04	0.0014
DISC1	Disrupted in schizophrenia 1	41	1	229 829 184	230 069 075	239 892	1	0	0.03	0.0012
IL1B	Interleukin-1beta	0	2	113 303 808	113 310 827	7019	n.a	n.a	n.a	n.a
ZNF804A	Zinc finger protein 804A	22	2	185 171 338	185 512 459	341 121	0	0	0.27	0.0023
DRD3	Dopamine receptor D_3	10	3	115 330 247	115 380 589	50 343	0	0	0.14	0.0050
CCKAR	Cholecystokinin A receptor	2	4	26 092 116	26 101 140	9024	0	0	0.09	0.0250
GABRB2	Gamma-aminobutyric acid (GABA) A receptor beta	24	5	160 648 014	160 907 708	259 695	0	0	0.12	0.0021
DRD1	Dopamine receptor D_1	1	5	174 800 281	174 803 769	3489	0	0	0.6	0.0500
NOTCH4	Notch homolog 4	7	6	3 377 585	3 406 812	29 227	0	0	0.19	0.0071
DTNBP1	Dystrobrevinbinding protein 1	15	6	15 631 018	15 771 250	140 233	3	0	0.03	0.0033
HIST1H2BJ	Histone cluster 1	0	6	27 208 074	27 208 554	480	n.a	n.a	n.a	n.a
PRSS16	Thymus-specific serine protease precursor	0	6	27 323 481	27 332 377	8896	n.a	n.a	n.a	n.a
PGBD1	*PiggyBac transposable element derived 1*	*4*	*6*	*28 357 293*	*28 378 304*	*21 011*	*2*	*0*	*0.007*	*0.0125*
RPP21	Ribonuclease P protein subunit p21	1	6	30 420 916	30 422 611	1695	0	0	0.87	0.0500
MDGA1	MAM domain containing glycosylphosphatidylinositol anchor 1	21	6	37 708 262	37 773 744	65 482	0	0	0.05	0.0024
RELN	Reelin	104	7	102 899 469	103 417 199	517 731	7	2	0.0005	0.0005
SLC18A1	Solute carrier family 18 (vesicular monoamine)	8	8	20 046 647	20 084 997	38 351	0	0	0.09	0.0063
PPP3CC	Serine/threonine-protein phosphatase 2B	5	8	22 354 541	22 454 582	100 041	0	0	0.09	0.0100
NRG1	Neuregulin 1	142	8	31 616 810	32 720 312	1 103 503	3	0	0.009	0.0004
GRIN1	NMDA receptor 1	1	9	139 153 430	139 183 029	29 600	0	0	0.78	0.0500
DRD4	Dopamine receptor D_4	0	11	627 305	630 703	3398	n.a	n.a	n.a	n.a
TPH1	Tryptophan hydroxylase 1	2	11	17 998 660	18 018 911	20 252	0	0	0.29	0.0250
DRD2	Dopamine receptor D_2	9	11	112 785 527	112 851 211	65 685	1	0	0.04	0.0056
NRGN	Neurogranin	1	11	124 115 039	124 122 312	7273	0	0	0.74	0.0500
OPCML	Opioid binding protein/cell adhesion molecule-like	93	11	131 790 085	132 318 247	528 162	6	0	0.01	0.0005
GRIN2B	N-methyl-D-aspartate receptor subunit 2B	84	12	13 605 677	14 024 289	418 613	5	0	0.008	0.0006
DAO	D-amino-acid oxidase	8	12	107 797 986	107 818 839	20 853	0	0	0.25	0.0063
CCDC60	Coiled-coil domain containing 60	34	12	118 256 900	118 463 234	206 334	0	0	0.11	0.0015
HTR2A	Serotonin receptor 2A	15	13	46 305 514	46 369 170	63 657	2	0	0.03	0.0033
DAOA	D-Amino acid oxidase activator	8	13	104 916 566	104 941 383	24 818	0	0	0.17	0.0063
AKT1	V-akt murine thymoma oncogene homolog 1	1	14	104 306 732	104 333 125	26 394	0	0	0.48	0.0500
RPGRIP1L	RPGRIP1-like isoform a	7	16	52 191 319	52 295 272	103 954	0	0	0.34	0.0071

[*continued overleaf*

Table 3 (*cont.*)

Gene name	Gene product	Number of SNPs per gene	Chr	Start	End	Gene size (bp)	p values <0.05	p values <0.001	Minimal p value	Bonferroni gene-wide significance cut-off
HP	Haptoglobin	0	16	70 646 009	70 652 456	6447	n.a	n.a	n.a	n.a
SRR	Serine racemase	1	17	2 153 998	2 175 303	21 305	0	0	0.6	0.0500
SLC6A4	Serotonin transporter	4	17	25 547 506	25 587 080	39 575	0	0	0.4	0.0125
TCF4	Transcription factor 4	46	18	51 040 560	51 406 858	366 298	2	0	0.03	0.0011
APOE	Apolipoprotein E	0	19	50 100 879	50 104 490	3611	n.a	n.a	n.a	n.a
PRODH	Proline dehydrogenase 1	10	22	17 280 294	17 303 814	23 521	1	0	0.003	0.0050
COMT	Catechol-*O*-methyltransferase	8	22	18 309 263	18 337 496	28 234	0	0	0.06	0.0063
ZDHHC8	Zinc finger, DHHC-type containing 8	0	22	18 499 365	18 513 974	14 610	n.a	n.a	n.a	n.a
CSF2RA	Colony stimulating factor 2 receptor A	0	X	1 347 693	1 388 828	41 136	n.a	n.a	n.a	n.a
IL3RA	Interleukin 3 receptor A	0	X	1 415 509	1 461 582	46 074	n.a	n.a	n.a	n.a

Chr, Chromosome; n.a., not available.
Genes from previous GWASs are highlighted in light grey. Genes selected by Sullivan *et al.* (2008) are highlighted in dark grey.

pathogenesis of schizophrenia should be considered. Multiple true associations may lie below the genome-wide significance level, and are erroneously neglected. For the validation step we chose SNPs with *p* value $<1 \times 10^{-5}$ in agreement with the National Human Genome Research Institute (NIHGRI) threshold of eligibility for inclusion in the NHGRI Catalog of Published GWASs (http://www.genome.gov/gwastudies/). A *p* value of $<1 \times 10^{-5}$ threshold is considered as moderately strong for association according to O'Donovan *et al.* (2008). Six additional intergenic SNPs were nominally associated with schizophrenia in our GWAS sample ($p < 1 \times 10^{-5}$, *q* value <0.4). Although none of them withstood the GWAS significance threshold, we consider them as interesting candidates for further study. Accordingly, we performed a replication study for the top seven SNPs in a geographically isolated sample of Arab-Israeli families from an isolated area in Israel which we previously genotyped for a GWAS using the same Illumina platform (Alkelai *et al.* unpublished data). We chose to perform the replication study in a family-based rather than a case-control sample due to the possibility of implementing the same TDT used in the original study in the Arab-Israeli validation sample. One of the studied SNPs, rs4803480, was associated with schizophrenia in the validation stage ($p = 9.649 \times 10^{-6}$ in the Jewish-Israeli family sample, $p = 0.0028$ in the Arab-Israeli family sample, $p = 9.612 \times 10^{-8}$ in the combined analysis of the two samples). The 'A' allele was consistently associated with increased risk of schizophrenia in both samples. This association survives the Bonferroni correction for multiple testing in the validation stage (seven tests), emphasizing the robustness of the finding and encouraging us to regard it as true positive replication. The rs4803480 SNP is located in a predicted intron (the region contains a large number of human spliced ESTs and mRNAs) of the *CEACAM21* gene, which encodes the carcino-embryonic antigen-related cell adhesion molecule 3. This molecule is an innate immune receptor, expressed on granulocytes and targeted against human-specific pathogens (Pils *et al.* 2008). Immune system dysregulation has been reported in schizophrenia, such as elevated levels of IL-6 and IL-2 and alternation in T-helper cell activation (Strous & Shoenfeld, 2006). Autoimmune processes have been suggested to play role in the pathophysiology of the disease, at least in a subgroup of patients (Strous & Shoenfeld, 2006). In fact, many genes associated with schizophrenia in genetic studies like *DISC1*, *NRG1*, *RGS4*, and *TPH1* are implicated in host–pathogen interactions. In a large schizophrenia GWAS, the most significant

results were found in the major histocompatibility complex (MHC) region on chromosome 6p21.3-22.1 (Stefansson *et al.* 2009). These findings, together with evidence that prenatal infections with viral or bacterial pathogens may contribute to the aetiology of schizophrenia (Brown, 2006), are in an agreement with the association of the *CEACAM21* gene with schizophrenia. *CEACAM21* is a low-level widely expressed gene, with higher ratios in bone marrow, hypothalamus and liver [UCSC Genome Browser (http://genome.ucsc.edu/) and GeneNote browser (http://bioinfo2.weizmann.ac.il/cgi-bin/genenote/home_page.pl)].

We were not able to replicate the association with schizophrenia of the five other SNPs studied in the Arab validation study. In addition, the *DOCK4* SNP, rs2074127 could not be tested due to significant deviation from HWE in the replication sample. Perfect surrogates for this SNP were not found. At this stage we cannot address the generalizability of the *DOCK4* gene association with schizophrenia in populations other than the Jewish one and further studies are required (although association with autism has been reported; Maestrini *et al.* 2010).

Focusing separately on selected sets of 46 well-recognized schizophrenia candidate genes (see the Methods section for selection criteria), we found gene-wide significant associations within three genes: *PGBD1* (rs1150724), *RELN* (rs39339, rs262342) and *PRODH* (rs2238732), supporting previously found associations (Liu *et al.* 2002, 2010; Shifman *et al.* 2008; Stefansson *et al.* 2009). Previously identified common schizophrenia susceptibility alleles are characterized by small effect sizes, which are proportional to the power to detect these alleles. Therefore, it is highly unlikely that true significant results (which are expected to represent only a small fraction of all schizophrenia susceptibility loci) of one study with a few thousand cases and controls will be found significant or will be among the top hits in another study of a moderate size sample (Owen *et al.* 2010). The finding of the *RELN* association is not surprising, since this gene has been already found to be associated with schizophrenia in the Jewish population (Liu *et al.* 2010; Shifman *et al.* 2008). We used gene-wide cut-off in the candidate gene sub-analysis of the GWAS data. This cut-off may be too lenient when a small number of SNPs which do not properly cover the gene are analysed. However, applying GWAS cut-off for this purpose is in our view overly conservative. For risk alleles with a small effect size, achieving a genome-wide significance cut-off could be an unrealistic task while studying relatively small samples (Neale & Sham, 2004).

Some limitations of the current study should be acknowledged. We undertook a GWAS in a relatively small family sample (107 families) using the HumanCNV-370 BeadArrays with only 370 404 Phase I Hap Map tagging SNPs which might not be dense enough to identify all the variants associated with schizophrenia in our sample. However, the relatively small sample size should be balanced by the genetically homogeneous nature of the sample, which is less diverse than the general Caucasian population, and by the sample being family-based, allowing application of the TDT association test. Since the platform we used is possibly not sufficiently dense to identify all the disease-associated variants, further research with denser platforms or next-generation sequencing is required.

In conclusion, by application of a family-based strategy to GWAS, our study revealed new schizophrenia susceptibility loci in the Jewish population. The most interesting finding concern *DOCK4* (withstands GWAS significance) and *CEACAM21* (replicated in an independent sample), both biologically reasonable candidate genes for schizophrenia. Further study in additional populations is required to address the generalizability of the findings.

Note

Supplementary material accompanies this paper on the Journal's website (http://journals.cambridge.org/pnp).

Acknowledgements

Statement of Interest

None.

References

Alkelai A, Kohn Y, Olender T, Sarner-Kanyas K, *et al.* (2009). Evidence for an interaction of schizophrenia susceptibility loci on chromosome 6q23.3 and 10q24.33-q26.13 in Arab Israeli families. *American Journal of Medical Genetics, B: Neuropsychiatric Genetics* **150B**, 914–925.

Allen NC, Bagade S, McQueen MB, Ioannidis JP, *et al.* (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nature Genetics* **40**, 827–834.

Amann-Zalcenstein D, Avidan N, Kanyas K, Ebstein RP, *et al.* (2006). AHI1, a pivotal neurodevelopmental gene, and C6orf217 are associated with susceptibility to schizophrenia. *European Journal of Human Genetics* **14**, 1111–1119.

Andreasen NC, Endicott J, Spitzer RL, Winokur G (1977). The family history method using diagnostic criteria. Reliability and validity. *Archives of General Psychiatry* **34**, 1229–1235.

Baron M, Endicott J, Lerer B, Loth JE, *et al.* (1994). A pedigree series for mapping disease genes in bipolar affective disorder: sampling, assessment, and analytic considerations. *Psychiatric Genetics* **4**, 43–55.

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, *et al.* (2010). The genome-wide structure of the Jewish people. *Nature* **466**, 238–242.

Brown AS (2006). Prenatal infection as a risk factor for schizophrenia. *Schizophrenia Bulletin* **32**, 200–202.

Carroll LS, Owen MJ (2009). Genetic overlap between autism, schizophrenia and bipolar disorder. *Genome Medicine* **1**, 102.

Fan JB, Gunderson KL, Bibikova M, Yeakley JM, *et al.* (2006). Illumina universal bead arrays. *Methods in Enzymology* **410**, 57–73.

Freyberg Z, Ferrando SJ, Javitch JA (2010). Roles of the Akt/ GSK-3 and Wnt signaling pathways in schizophrenia and antipsychotic drug action. *American Journal of Psychiatry* **167**, 388–396.

Kirov G, Zaharieva I, Georgieva L, Moskvina V, *et al.* (2009). A genome-wide association study in 574 schizophrenia trios using DNA pooling. *Molecular Psychiatry* **14**, 796–803.

Laird NM, Lange C (2006). Family-based designs in the age of large-scale gene-association studies. *Nature Reviews Genetics* **7**, 385–394.

Lange C, DeMeo D, Silverman EK, Weiss ST, *et al.* (2004). PBAT: tools for family-based association studies. *American Journal of Human Genetics* **74**, 367–369.

Lencz T, Lambert C, DeRosse P, Burdick KE, *et al.* (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences USA* **104**, 19942–19947.

Lerer B, Segman RH, Hamdan A, Kanyas K, *et al.* (2003). Genome scan of Arab Israeli families maps a schizophrenia susceptibility gene to chromosome 6q23 and supports a locus at chromosome 10q24. *Molecular Psychiatry* **8**, 488–498.

Levi A, Kohn Y, Kanyas K, Amann D, *et al.* (2005). Fine mapping of a schizophrenia susceptibility locus at chromosome 6q23: increased evidence for linkage and reduced linkage interval. *European Journal of Human Genetics* **13**, 763–771.

Liu H, Heath SC, Sobin C, Roos JL, *et al.* (2002). Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia. *Proceedings of the National Academy of Sciences USA* **99**, 3717–3722.

Liu Y, Chen PL, McGrath J, Wolyniec P, *et al.* (2010). Replication of an association of a common variant in the Reelin gene (RELN) with schizophrenia in Ashkenazi Jewish women. *Psychiatric Genetics* **20**, 184–186.

Maestrini E, Pagnamenta AT, Lamb JA, Bacchelli E, *et al.* (2010). High-density SNP association study and copy number variation analysis of the AUTS1 and AUTS5 loci implicate the IMMP2L-DOCK4 gene region in autism susceptibility. *Molecular Psychiatry* **15**, 954–968.

Mah S, Nelson MR, Delisi LE, Reneland RH, *et al.* (2006). Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Molecular Psychiatry* **11**, 471–478.

Mitchell KJ, Porteous DJ (2011). Rethinking the genetic architecture of schizophrenia. *Psychological Medicine* **41**, 19–32.

Neale BM, Sham PC (2004). The future of association studies: gene-based analysis and replication. *American Journal of Human Genetics* **75**, 353–362.

O'Donovan MC, Craddock N, Norton N, Williams H, *et al.* (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nature Genetics* **40**, 1053–1055.

Owen MJ, Craddock N, O'Donovan MC (2010). Suggestion of roles for both common and rare risk variants in genome-wide studies of schizophrenia. *Archives of General Psychiatry* **67**, 667–673.

Pagnamenta AT, Bacchelli E, de Jonge MV, Mirza G, *et al.* (2010). Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biological Psychiatry* **68**, 320–328.

Pils S, Gerrard DT, Meyer A, Hauck CR (2008). CEACAM3: an innate immune receptor directed against human-restricted bacterial pathogens. *International Journal of Medical Microbiology* **298**, 553–560.

Purcell S, Neale B, Todd-Brown K, Thomas L, *et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–575.

Schwab SG, Wildenauer DB (2009). Update on key previously proposed candidate genes for schizophrenia. *Current Opinion in Psychiatry* **22**, 147–153.

Shi J, Levinson DF, Duan J, Sanders AR, *et al.* (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757.

Shifman S, Johannesson M, Bronstein M, Chen SX, *et al.* (2008). Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genetics* **4**, e28.

Spitzer RL, Endicott J (eds) (1977). *The Schedule for Affective Disorders and Schizophrenia, Lifetime Version*, 3rd edn. New York: New York State Psychiatric Institute.

Spitzer RL, Endicott J, Robins E (1978). Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry* **35**, 773–782.

Steemers FJ, Gunderson KL (2005). Illumina, Inc. *Pharmacogenomics* **6**, 777–782.

Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, *et al.* (2009). Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747.

Storey JD, Tibshirani R (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences USA* **100**, 9440–9445.

Strous RD, Shoenfeld Y (2006). Schizophrenia, autoimmunity and immune system dysregulation: a comprehensive model updated and revisited. *Journal of Autoimmunity* **27**, 71–80.

Sullivan PF, Kendler KS, Neale MC (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of General Psychiatry* **60**, 1187–1192.

Sullivan PF, Lin D, Tzeng JY, van den Oord E, *et al.* (2008). Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Molecular Psychiatry* **13**, 570–584.

Ueda S, Fujimoto S, Hiramoto K, Negishi M, *et al.* (2008). Dock4 regulates dendritic development in hippocampal neurons. *Journal of Neuroscience Research* **86**, 3052–3061.

Upadhyay G, Goessling W, North TE, Xavier R, *et al.* (2008). Molecular association between beta-catenin degradation complex and Rac guanine exchange factor DOCK4 is essential for Wnt/beta-catenin signaling. *Oncogene* **27**, 5845–5855.

Van Steen K, Lange C (2005). PBAT: a comprehensive software package for genome-wide association analysis of complex family-based studies. *Human Genomics* **2**, 67–69.

Wray NR, Visscher PM (2010). Narrowing the boundaries of the genetic architecture of schizophrenia. *Schizophrenia Bulletin* **36**, 14–23.