Molecular Psychiatry (2006) 11, 384–394

www.nature.com/mp

ORIGINAL ARTICLE

# Association analysis of the chromosome 4p-located *G protein-coupled receptor 78* (*GPR78*) gene in bipolar affective disorder and schizophrenia

SL Underwood[1,5], A Christoforou[1,5], PA Thomson[1], NR Wray[1], A Tenesa[2], J Whittaker[3], RA Adams[1], S Le Hellard[1], SW Morris[1], DHR Blackwood[4], WJ Muir[4], DJ Porteous[1] and KL Evans[1]

[1]*Medical Genetics Section, Molecular Medicine Centre, Western General Hospital, University of Edinburgh, Edinburgh, UK;* [2]*MRC Human Genetics Unit, Western General Hospital, Edinburgh, UK;* [3]*Department of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK and* [4]*Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, UK*

**The orphan *G protein-coupled receptor 78* (*GPR78*) gene lies within a region of chromosome 4p where we have previously shown linkage to bipolar affective disorder (BPAD) in a large Scottish family. *GPR78* was screened for single-nucleotide polymorphisms (SNPs) and a linkage disequilibrium map was constructed. Six tagging SNPs were selected and tested for association on a sample of 377 BPAD, 392 schizophrenia (SCZ) and 470 control individuals. Using standard $\chi^2$ statistics and a backwards logistic regression approach to adjust for the effect of sex, SNP rs1282, located approximately 3 kb upstream of the coding region, was identified as a potentially important variant in SCZ ($\chi^2$ *P* = 0.044; LRT *P* = 0.065). When the analysis was restricted to females, the strength of association increased to an uncorrected allele *P*-value of 0.015 (odds ratios (OR) = 1.688, 95% confidence intervals (CI): 1.104–2.581) and uncorrected genotype *P*-value of 0.015 (OR = 5.991, 95% CI: 1.545–23.232). Under the recessive model, the genotype *P*-value improved further to 0.005 (OR = 5.618, 95% CI: 1.460–21.617) and remained significant after correcting for multiple testing (*P* = 0.017). No single-marker association was detected in the SCZ males, in the BPAD individuals or with any other SNP. Haplotype analysis of the case–control samples revealed several global and individual haplotypes, with *P*-values < 0.05, all but one of which contained SNP rs1282. After correcting for multiple testing, two haplotypes remained significant in both the female BPAD individuals (*P* = 0.038 and 0.032) and in the full sample of affected female individuals (*P* = 0.044 and 0.033). Our results provide preliminary evidence for the involvement of *GPR78* in susceptibility to BPAD and SCZ in the Scottish population.**
*Molecular Psychiatry* (2006) **11**, 384–394. doi:10.1038/sj.mp.4001786; published online 3 January 2006

**Keywords:** bipolar disorder; schizophrenia; G protein-coupled receptors; association study; linkage disequilibrium

## Introduction

Bipolar affective disorder (BPAD) and schizophrenia (SCZ) are severe and debilitating mental illnesses, each of which affects approximately 1% of the population. Family, twin and adoption studies have supported a strong genetic component to the susceptibility of BPAD and SCZ, but have also indicated that the inheritance patterns are complex.[1] Several linkage regions reaching genome-wide significance have been reported.[2,3] Previously, we described a genome-wide linkage study in a large Scottish family (F22) segregating major affective disorder.[4] Two-point linkage analysis gave a maximum logarithm of odds (LOD) score of 4.1 to a region on chromosome 4p. A recent follow-up study of this family, in which five additional affected F22 individuals were identified and several additional microsatellite and single-nucleotide polymorphism (SNP) markers were incorporated, resulted in a maximum LOD score of 4.4.[5] Analysis of a small Scottish family also showed positive linkage to the region (F59, LOD = 0.9).[5] A number of other groups have found evidence of linkage of BPAD and/or SCZ to this 4p region as well. Asherson *et al.*[6] reported linkage in a schizoaffective family (F50, LOD = 2.0); Ewald *et al.*[7] reported linkage in BPAD families (LOD = 2.0); Detera-Wadleigh *et al.*[8] investigated families with major mental illness, and their largest family (F48) generated an LOD of 3.2; Williams *et al.*[9] found increased sharing in SCZ sibpairs (LOD = 1.7); Lerer *et al.*[10] found a

Correspondence: Ms A Christoforou, Medical Genetics Section, Molecular Medicine Centre, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, UK.
E-mail: A.Christoforou@ed.ac.uk
[5]These authors contributed equally to this work.
Received 22 August 2005; revised 3 October 2005; accepted 16 November 2005; published online 3 January 2006

non-parametric LOD score of 2.2 in families with SCZ and schizoaffective disorder; and Als *et al*.[11] found excess haplotype sharing (best *P*-value, $P = 0.00007$) in families with BPAD and SCZ.

Microsatellite haplotype analysis revealed that the linkage regions identified in the three families, F48, F50 and F59, overlap and dissect the significant F22 linkage region into four subregions (A–D), which can be prioritised on the basis of strongest linkage evidence and/or ancestral origin.[5] Inspection of a BAC clone contig of the F22-linked region revealed several positional candidate genes in region B,[12] including the orphan *G protein-coupled receptor 78* (*GPR78*). We selected *GPR78* for a case–control association study because of its location in the 4p-linked region and its predicted function as a G protein-coupled receptor (GPCR). The GPCR family is the most abundant member of the cell surface receptor families, recognising and transducing messages from sources as diverse as light, calcium, odorants and small molecules, and is well positioned to mediate subtle changes in cellular function. Several members of this family have already been implicated in the pathophysiology of psychiatric illness.[13–15]

*GPR78* was identified by virtue of its homology to orphan *GPR26*, which was identified in human and rat.[16,17] Rat *GPR26* encodes a 317-amino-acid protein distantly related to the serotonin 5-HT5α and gastrin-releasing hormone BB2 receptors.[17] The *GPR78* gene is approximately 6.5 kb long, encodes a 363-amino-acid protein and has three exons, which give rise to the classic seven transmembrane domain (TMD) receptor structure of GPCRs, with an extracellular *N*-terminus and an intracellular C terminus (Figure 1). *GPR78* shares highest sequence identity with *GPR26* (49% overall and 56% in the TMD regions). They have a similar intron–exon structure and a similar protein structure with a short amino-terminus, no asparagine-linked extracellular glycosylation site and cationic arginine and lysine residues in TMD 6 and 7, respectively.[16] Their overall structural homology suggests that they may encode receptors that bind to a common endogenous ligand, although no ligand has been identified for either of the two GPCRs[16,17] and nothing is known about their physiological role.

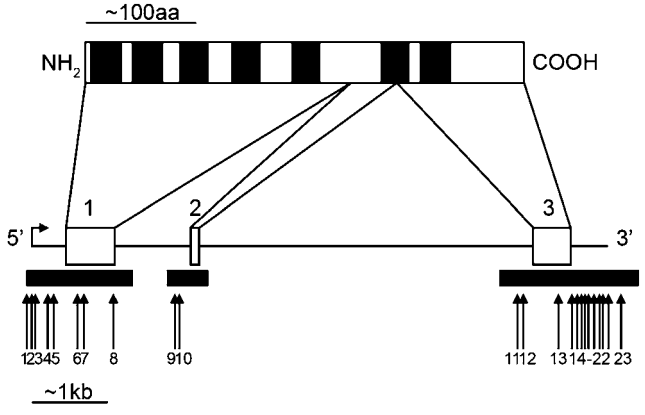


**Figure 1** The *GPR78* gene. The seven TMDs (black boxes) and extracellular N-terminal ($NH_2$) and intracellular C-terminal (COOH) domains arise from the three exons (numbered 1–3; white boxes). The black bars beneath the exons show the regions screened for polymorphisms using DNA from the linked families. The positions of the 23 SNPs identified are indicated by the arrows. The numbering of the SNPs corresponds to Table 1. The 'AOD' SNPs are not shown here.

*GPR78* mRNA is expressed in the pituitary and the placenta,[16] consistent with a potential role in the functioning of the hypothalamic-pituitary-adrenal (HPA) axis and in pregnancy. The HPA axis is involved in hormone and stress regulation,[18] and its dysregulation has been implicated in major affective disorder[19,20] and SCZ.[21] The prenatal period is a critical time for the development of systems such as the HPA axis, and there is evidence that maternal stress increases the likelihood of abnormal HPA functioning and abnormal responses to stress in the mature offspring.[22] Other prenatal insults transmitted via the mother, such as maternal infection or malnutrition, have also been implicated in the pathogenesis of psychiatric illness.[23] However, little is understood about the role of the prenatal environment and the involvement of the HPA axis in the pathogenesis of psychiatric illness.

The map location and putative function of *GPR78* combine to make it a good candidate gene for psychiatric illness. We have tested this hypothesis via a threestep process: First, we screened *GPR78* for SNPs from the 4p-linked families; second, we constructed a custom linkage disequilibrium (LD) map of the region and selected haplotype-tagging SNP (htSNPs) that best represented the variation in the region; and third, we performed a case–control association study in a sample of 377 BPAD, 392 SCZ and 470 control individuals.

## Materials and methods

This study was approved by the Multicentre Research Ethics Committee for Scotland and written informed consent was obtained from all participants.

*Association sample*

Individuals suffering from BPAD or SCZ were recruited from in-patient and outpatient services at the Royal Edinburgh and other Scottish psychiatric hospitals. All were screened using the semistructured Schedule for Affective Disorder and Schizophrenia-Lifetime version (SADS-L)[24] interview schedule by psychiatrists experienced in its use. This generated information that allowed diagnosis by both Diagnostic and Statistical Manual of Mental Disorders (4th Edition) (DSM-IV)[25] and International Classification of Diseases (10th Revision) (ICD-10)[26] criteria for BPAD or SCZ. All diagnoses were reviewed independently by two experienced psychiatrists and a consensus diagnosis was reached where necessary.

Control subjects were drawn from the same population in South East and South Central Scotland. The majority (391) were recruited through the Scottish National Blood Transfusion service. Although the blood donors were not screened by interview for personal or family history of psychiatric illness, donors are only allowed to donate blood if they are not currently on medication and have no chronic illness. The remaining controls (79) were recruited from the local population or from hospital staff. These controls were briefly screened by interview to exclude anyone currently on medication or with a history of treatment for psychiatric illness.

*Power calculations*

*A priori* power calculations, following the methods described in Risch,[27] showed that for the attainable sample size of 395 cases and 395 controls, we would have 80% power (Type 1 error = $5 \times 10^{-5}$) to detect an association of an allele that shows a heterozygote relative risk (multiplicative model) of 1.6 and has a control allele frequency is 0.40. When the control allele frequency is 0.10, the heterozygote relative risk of the allele would have to be at least 2 before we would have 80% power to detect an association.

*SNP discovery and genotyping*

*SNP discovery.* SNPs were identified from 46 individuals drawn from four families (F48, F50, F59 and F22) that show linkage to chromosome 4p. This provided 38 control chromosomes and four 'linked chromosomes'.[5] Each family contributed one 'linked chromosome', which represents the haplotype shared by all the individuals that contributed to the linkage signal in that family.

Primers were designed to cover the coding regions, the splice sites and approximately 1.2 kb of the 5′UTR and 500 bp of the 3′UTR of *GPR78*, using the Primer 3 primer design program (http://www.broad.mit.edu/cgi-bin/primer/primer3_www.cgi).[28] The STS sequences have been submitted to Genbank (Accession numbers: BV678010–BV678017). PCR conditions were optimised on control DNA. PCR was performed in total volume of 15 $\mu$l with 20 ng DNA, 1 × reaction buffer with 1.5 mM $MgCl_2$ (Perkin-Elmer), 100 $\mu$M of each dNTP (Sigma), 0.5 U *Taq* DNA polymerase (Sigma) and 0.33 $\mu$M of each primer (Invitrogen). PCR cycling was carried out on a PTC-225 thermal cycler (MJ Research). PCR cycling conditions consisted of denaturation at 93°C for 1 min, followed by 10 cycles of 93°C for 20 s, 65°C for 30 s, minus 1°C/cycle and 72°C for 1 min, followed by 30 cycles of 93°C for 20 s, 55°C for 30 s and 72°C for 1 min and a final extension of 72°C for 10 min. Three microlitres of the PCR product was resolved on a 2% agarose gel and DNA concentration was estimated with 250 and 500 ng of Ready Load™ PhiX174 RF DNA/*Hae*III fragments (Invitrogen).

DNA sequencing was performed using BigDye® Terminator v3.1 Cycle Sequencing Kit according to the manufacturer's protocols. Sequencing reactions were analysed using an ABI PRISM® 3730 Genetic Analyser as per the manufacturer's instructions. Sequence chromatograms were aligned using the phredPhrap software and visualised with the Consed program[29] (http://bozeman.mbt.washington.edu/index.html). SNPs were identified and genotypes noted by visual inspection of the sequence traces. All SNPs were identified from high-quality sequence reads. Genotypes were scored blind to phenotype by two individual investigators and then checked for Mendelian segregation.

*Genotyping.* The SNPs identified from the SNP discovery population were genotyped at the Sanger Institute by the MASSARRAY™ primer extension method (Sequenom®). Three additional SNPs (rs11736084, rs1282 and rs3756179) were supplied by Applied Biosystems as 'Assay on Demand' (AOD). Each of the commercial SNPs had a frequency greater than 0.10 in Caucasians and was genotyped by the Genetics Core of the Wellcome Trust Clinical Research Facility by the Taqman® method as per the manufacturer's instructions. The tagging SNPs selected for the association analysis were also genotyped by the Genetics Core of the Wellcome Trust Clinical Research Facility by the Taqman® method as per the manufacturer's instructions.

*Statistical analysis*

The standard $\chi^2$-test of independence was used to examine all successfully genotyped biallelic markers for deviations from Hardy–Weinberg equilibrium (HWE). An HW significance threshold of $P = 0.004$ (adjusted from 0.05 for the 14 successfully genotyped markers) was selected to account for testing multiple markers. Pairwise LD measure, $r^2$, was calculated by downloading the relevant genotype data into Haploview[30] (http://www.haploview.org). $r^2$ ranges from 0 (linkage equilibrium) to 1 (perfect LD). It measures the statistical correlation, squared, between two loci. $r^2 = 1$ if exactly two of the four possible haplotypes are present, which occurs only when the allele frequencies at the two loci are identical.[31] While a high $r^2$ can be obtained only if the allele frequencies are similar, too low a value of $r^2$ can result in insufficient power to detect an association at a marker locus assumed to be linked to the disease locus, due to the inverse relationship between $r^2$ and sample size.[32] The LD blocks were constructed by examining the $r^2$ between all pairs of markers, using an intermediate threshold of $r^2 > 0.65$. Tagging SNPs from each block were selected to represent haplotypes within the blocks with frequencies greater than 5%.

Linkage to chromosome 4p has been shown in families affected with BPAD and/or SCZ. Thus, we deemed it necessary to test the hypotheses that *GPR78* may be involved in the pathogenesis of BPAD, SCZ or both BPAD and SCZ, in which all affected individuals were grouped together, as it has been previously suggested that the two psychiatric illnesses may share

susceptibility loci.[33,34] Similarly, owing to evidence of differences between males and females in the manifestation of psychiatric illness[35] and from findings of sex-specific association in the analysis of other candidate genes,[15,36] analysis was also carried out on samples separated on the basis of gender.

Differences in allele and genotype frequencies between cases and controls were evaluated with the $\chi^2$-test of independence (with 1 and 2 degrees of freedom, respectively). Fisher's exact test, as implemented in *SISA* (http://home.clara.net/sisa/), was used when appropriate, for sparse contingency tables. Significance was initially declared at a nominal *P*-value of 0.05. The significance of the odds ratios (ORs) was determined using a *z*-statistic.[37]

Using SPSS (Release 12.0.0, September 2003), a backwards elimination logistic regression approach (as described in Cordell and Clayton[38]) was applied to test the significance of adjacent and non-adjacent marker genotypes, working individually or jointly, in predicting the diagnosis of each patient, while adjusting for the confounding effects of gender caused by differences in ascertainment. The patient diagnosis was declared as the dichotomous dependent variable (1 = affected, 0 = unaffected) and sex and the genotypes of the SNPs were declared as the independent variables. Sex was set as a categorical variable (1 = male, 0 = female), whereas the genotypes were tested as linear, which assumes that the logarithm of the OR for the heterozygotes is midway between that for the two homozygotes. The SPSS Backward Wald procedure fits all of the independent variables into the model and then removes the insignificant variables one-by-one based on the probability of Wald test statistic (removal $P > 0.10$). As each individual must be genotyped successfully on all markers in order to be included in the logistic regression analysis in SPSS, the final model identified from the backward Wald analysis was subsequently refitted using the most complete set of data available for the variables in that model, in order to utilise all of the available power. Only interactions between the variables emerging as significant from the backward analysis were tested, in order to avoid overfitting and to control multiple testing. As no prior knowledge of the interactions in GPCRs and their implication in psychiatric illness exists, there was no basis for selecting particular interactions to test and the sample size was insufficient to test all possible interactions. The significance of the model as a whole was assessed by the value of the log ratio statistic produced when the significant variables were included in the model.

Haplotype frequency estimation and comparison was carried out using the statistical analysis program COCAPHASE 2.43[39] (http://portal.litbio.org/Registered/Option/unphased.html). This software uses the EM algorithm to estimate the haplotype frequencies of unphased genotype data and standard unconditional logistic regression analysis, applying the likelihood ratio test under a log-linear model, to compare haplotype frequencies between cases and controls. In order to avoid misleading results caused by rare haplotypes, all haplotypes with a frequency less than or equal to 5% in both the cases and the controls were declared as rare and clumped together for the test of the null hypothesis, using the command line option '–rare 0.05'. *P*-values for both global and individual tests of haplotype frequencies were determined. The global test *P*-value assesses the significance of the overall difference in the distribution of haplotype frequencies between cases and controls. The *P*-value from the individual test represents the significance of the difference in frequency of an individual haplotype between cases and controls. A sliding-window approach was used to test all possible haplotypes of lengths two to six.

In order to account for the multiple SNPs and haplotypes tested, permutation analysis (1000 or 10 000 permutations) was performed also using Cocaphase. Cocaphase reassigns the diagnosis labels (case versus control) of the individuals. All of the markers within the specified window size are then tested and the most significant *P*-value is stored. Based on the distribution of the most significant *P*-values resulting from each test of the permuted data, a significance level is provided for the *P*-value of interest.

*Bioinformatics analysis*

Bioinformatics analysis was performed firstly to assess the functional significance of the non-synonymous SNPs identified. Amino-acid sequence conservation among other GPCRs and between species was investigated by querying the NCBI non-genomic sequences using the BLAST algorithm[40] (http://www.ncbi.nlm.nih.gov/BLAST/) and the GPR78 protein sequence (NP_543009) as the query. The program SIFT[41] (http://www.blocks.fhcrc.org/sift/SIFT.html) was then used to determine the probability that the observed amino-acid substitution is tolerated. The DAS-TM filter algorithm[42] (http://wooster.bip.bham.ac.uk/DAS.html) was applied to determine if the SNP in question alters the predicted TMD structure. The PROSITE[43] (http://us.expasy.org/prosite/) database of protein families and domains was searched to determine whether the alternative amino acids in the protein creates or abolishes a particular functional motif. Finally, NetPhos Version 2.0[44] (http://www.cbs.dtu.dk/services/NetPhos/), which uses a neural network-based method for predicting potential phosphorylation sites at serine, threonine or tyrosine residues in protein sequences, was applied, where appropriate, to predict whether a particular site is likely to be phosphorylated and if the substitution alters that process.

To assess the significance of any non-coding SNPs, cross-species conservation was examined using the 'Human/Chimp/Mouse/Rat/Chicken Multiz[45] Alignments & PhyloHMM[46] Cons' track on the UCSC Genome Browser.[47] The Multiz alignments were based on the human July 2003 (hg16), chimpanzee November 2003 (panTro1), mouse February 2003

(mm3), rat June 2003 (rn3) and chicken February 2004 (galGal2) assemblies.[48] The MATCH™ program[49] (http://www.gene-regulation.com/cgi-bin/pub/programs/match/bin/match.cgi), which is connected to the TRANSFAC® database,[50] was used to determine if the non-coding SNPs located in putative regulatory regions create or destroy any predicted transcription factor binding sites. Two sequences (50 bp) containing the alternate alleles for the SNP of interest were submitted. Beyond the default settings, MATCH was set to search only the vertebrate matrix groups and to minimise the false-positive matches.

## Results

*SNP discovery*

The coding region, splice sites and putative regulatory regions of the positional and functional candidate gene *GPR78* were screened for SNPs using DNA from the 4p-linked family members. Twenty-three SNPs were identifed (Figure 1), 18 of which had been previously described in dbSNP (www.ncbi.nlm.nih.gov.SNP/) (dbSNP build 121). The five SNPs identified in-house have since been submitted to dbSNP (see Table 1). One SNP (SNP 4, ih165) was dropped immediately from the study on account of its low minor allele frequency (MAF < 3%).

*LD map construction and tagging SNP selection*

In addition to the 22 SNPs discovered from the family panel, three commercial, 'AOD' SNPs were selected to cover further upstream and downstream of the coding region, with AOD SNPs rs11736084 and rs1282 almost 8.8 and 3.0 kb, respectively, upstream from the start codon and SNP rs3756179 approximately 3.6 kb downstream from terminating codon (Table 1). These 25 SNPs were genotyped in 89 control individuals. Of the 25 SNPs, only 14 amplified successfully, produced substantial (>70%) genotype data and were polymorphic. Of those 14 SNPs, nine met HW ($P > 0.004$) and had a minor allele frequency greater than 3%, making them suitable for the construction of an LD map (Table 1, in italics). One of the reasons that so many SNP assays were problematic was because the *GPR78* gene region is repetitive in parts (data not shown), making it difficult to construct primers to genotype specific SNPs reliably.

The purpose of constructing the custom haplotype map of *GPR78* was to select the set of htSNPs that best represented the haplotypic variation in the *GPR78*

**Table 1** The SNPs genotyped in the 95 control individuals

| *SNP number corresponding to* Figure 1 | *SNP ID* | *Distance (bp)* | *HW P-value* | *MAF (CTL)* | *Tagging SNP no.* |
|---|---|---|---|---|---|
| *AOD* | **rs11736084** | — | *0.943* | *0.302* | *snp1* |
| *AOD* | **rs1282** | *5811* | *0.487* | *0.186* | *snp2* |
| 1 | rs6447884 | 2337 | | | |
| *2* | **rs4385041** | *111* | *0.409* | *0.401* | |
| 3 | ih162 (rs17844775) | 28 | 0.000 | 0.143 | |
| 5 | rs4235270 | 279 | | | |
| *6* | **ih33 (rs17844776)** | *229* | *0.379* | *0.372* | |
| 7 | ih32 (rs17844777) | 114 | 0.000 | 0.448 | |
| *8* | **ih31 (rs17844778)** | *456* | *0.215* | *0.451* | *snp3* |
| 9 | rs9991820 | *777* | 0.002 | 0.380 | |
| 10 | rs10000720 | 6 | | | |
| 11 | rs3115388 | 4535 | | | |
| 12 | rs3115387 | 56 | | | |
| *13* | **rs9685931** | *338* | *0.517* | *0.128* | *snp4* |
| 14 | rs9683448 | 132 | 0.000 | 0.160 | |
| 15 | rs10938724 | 51 | | | |
| 16 | rs9799380 | 35 | | | |
| 17 | rs9799717 | 22 | | | |
| *18* | **rs9799807** | *19* | *0.084* | *0.122* | |
| *19* | **rs9799720** | *66* | *0.068* | *0.087* | *snp5* |
| 20 | rs9799809 | 68 | 0.000 | 0.040 | |
| 21 | rs9799721 | 20 | | | |
| 22 | ih38 (rs17844781) | 41 | 0.080 | 0.024 | |
| 23 | rs9799378 | 237 | | | |
| *AOD* | **rs3756179** | *2991* | *0.580* | *0.300* | *snp6* |

Numbering of SNPs corresponds to Figure 1. AOD refers to the 'Assay On Demand' SNPs selected to cover regions further upstream and downstream of the gene. dbSNP Accession numbers (rs*) are included (dbSNP build 121). For those identified in-house, an 'ih' number and the recently assigned 'rs' number (dbSNP build 123) are included. The distance (in base pairs) between each consecutive SNP, the HW test *P*-value and the minor allele frequency (MAF) of the SNPs in the control sample are provided. SNPs that were included in the LD analysis are bold.

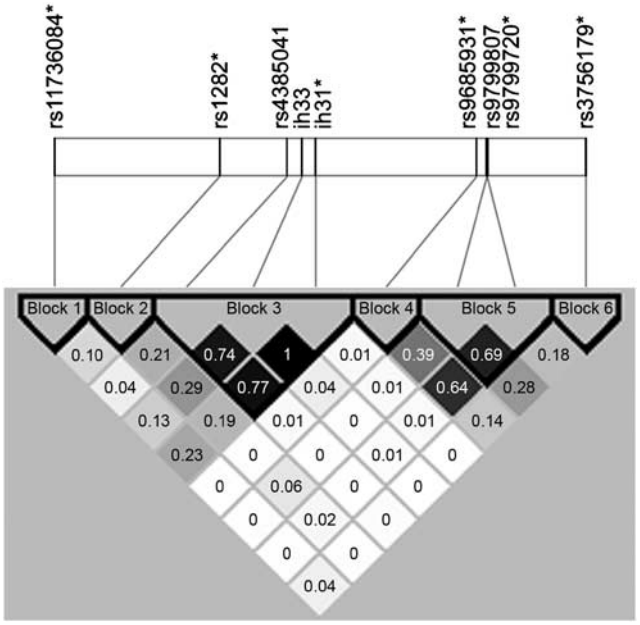


**Figure 2** LD map constructed using Haploview.[44] SNPs included in the analysis are at the top of the figure. Blocks are outlined by black bars. The values within boxes are the measures of $r^2$ between pairs of markers. Blocks were constructed based on $r^2 > 0.65$ threshold. The shading was provided by Haploview, with darker boxes representing higher $r^2$-values. SNPs selected to tag haplotypes are marked by an asterisk.

region in the Scottish sample under study, while minimising the genotyping requirements. Using the genotype data from the unaffected individuals, $r^2$ was calculated and used to determine the LD block boundaries. Figure 2 shows the results of the LD map construction. A threshold of $r^2 > 0.65$ revealed six LD blocks. The following six tagging SNPs were selected to represent all haplotypes, with a frequency greater than 5%, in each block: rs11736084 (snp1), rs1282 (snp2), ih31 (snp3), rs9685931 (snp4), rs9799720 (snp5) and rs3756179 (snp6), resulting in an average of one SNP per 3.13 kb (Table 1).

*Association analysis*

The six htSNPs were genotyped in a case–control sample of 377 individuals with BPAD (163 males, 214 females), 392 individuals with SCZ (280 males, 112 females) and 470 unrelated controls (244 males, 223 females, three unknowns). All six markers met HWE in the cases, as well as in the controls, at the adjusted HW threshold of 0.004 (lowest $P = 0.048$, in the controls; Supplementary Information, Table S1). The genotype success rates were rs11736084 (98.5%); rs1282 (99.1%); ih31 (94.3%); rs9685931 (97.1%); rs9799720 (97.6%) and rs3756179 (98.5%), resulting in an average success rate of 97.5%.

Marginally significant allelic association was detected between rs1282 (snp2) and SCZ ($P = 0.044$), with the patients more frequently carrying the T allele than the controls (OR = 1.300, 95% confidence interval (CI): 1.006–1.679) (Table 2). Only a trend toward significance ($P < 0.10$) was detected between snp2 and SCZ at the genotype level. When restricting the allele and genotype frequency analysis to the female patients, a stronger association was detected between snp2 and SCZ (allele $P = 0.015$; genotype $P = 0.015$), with the female SCZ individuals more frequently carrying the T allele (OR = 1.688, 95% CI: 1.104–2.581) and the TT genotype ($OR_{TT/CC} = 5.991$, 95% CI: 1.545–23.232; $OR_{CT/CC} = 1.272$, 95% CI: 0.752–2.152) (Table 2). A weaker association was detected in the analysis of all affected females (allele $P = 0.035$; genotype $P = 0.054$), reflecting a similar trend in the allele and genotype frequencies of the BPAD females as in those of the SCZ females (Supplementary Information, Table S1). SNP rs1282 is located almost 3 kb upstream of the methionine start codon. No single-marker association was detected with any other SNP, in the males or in the individuals with BPAD (Supplementary Information, Table S1). The backwards logistic regression approach, where genotypes were declared as the independent variables and the model was adjusted for the effects of sex, confirmed the findings of the single-marker analysis. It identified SNP rs1282 as the only SNP to survive the variable selection and only in the SCZ model (Wald $P = 0.066$, LRT $P = 0.065$). Likewise, when the analysis was restricted to the females, only SNP rs1282 emerged as significant (Wald and LRT $P = 0.020$).

Pairwise comparison of the ORs of the rs1282 (snp2) genotypes suggested that the OR of genotype TT was significantly different from that of CC when CC was set as the reference genotype, that is, OR = 1 ($z$-test $P = 0.005$) and from that of CT when CT was set as the reference genotype ($z$-test $P = 0.015$). However, the OR of CT was not significantly different from that of CC ($z$-test $P = 0.185$). This suggested a recessive mode of inheritance (TT versus CC + CT), which, when tested using the $\chi^2$-test, revealed a stronger association between SNP rs1282 and the SCZ females ($P = 0.005$, OR = 5.618, 95% CI: 1.460–21.617).

Haplotype frequencies were estimated and compared between cases and controls using unconditional logistic regression, as implemented in Cocaphase.[39] All possible haplotypes were tested for association using a sliding-window approach, for which the haplotypes that were estimated to be rare (frequency < 5%) in both the cases and the controls were clumped together in the test of the global null hypothesis. Many haplotypes at the global and individual level were found to be significant at the nominal significance level of 0.05, in all groups tested. Of the significant haplotypes, all but one contained SNP rs1282, with the T allele always conferring an increase in the risk of being affected (OR > 1) and the C allele a decrease in the risk of being affected (OR < 1) (Supplementary Information, Tables S2 and S3).

In order to account for testing multiple SNPs and haplotypes, permutation analysis of 10 000

**Table 2** Allele and genotype analysis of SNP rs1282 in all subgroups

| *SNP rs1282* | *Allele 1 (%)* | *Allele 2 (%)* | P-*value* ($\chi^2_1$) | $OR_{T/C}$ *(95% CI)* | *11 (%)* | *12 (%)* | *22 (%)* | P-*value* ($\chi^2_2$) | $OR_{TT/CC}$ *(95% CI)* |
|---|---|---|---|---|---|---|---|---|---|
| | *C* | *T* | | | *CC* | *CT* | *TT* | | |
| *All individuals* | | | | | | | | | |
| BPAD | 614 (83.2) | 124 (16.8) | 0.263 | 1.163 (0.892–1.516) | 258 (69.9) | 98 (26.6) | 13 (3.5) | 0.424 | 1.693 (0.731–3.923) |
| SCZ | 638 (81.6) | 144 (18.4) | **0.044** | 1.300 (1.006–1.679) | 264 (67.5) | 110 (28.1) | 17 (4.4) | 0.099 | 2.164 (0.974–4.804) |
| All cases | 1252 (82.4) | 268 (17.6) | 0.067 | 1.233 (0.985–1.543) | 522 (68.7) | 208 (27.4) | 30 (3.9) | 0.142 | 1.931 (0.932–4.002) |
| Controls | 789 (85.2) | 137 (14.8) | | | 336 (72.6) | 117 (25.3) | 10 (2.1) | | |
| *Male* | | | | | | | | | |
| BPAD | 265 (83.3) | 53 (16.7) | 0.876 | 1.031 (0.704–1.510) | 111 (69.8) | 43 (27.0) | 5 (3.1) | 0.989 | 1.088 (0.337–3.512) |
| SCZ | 462 (82.5) | 98 (17.5) | 0.592 | 1.093 (0.789–1.515) | 191 (68.2) | 80 (28.6) | 9 (3.2) | 0.862 | 1.138 (0.415–3.121) |
| All cases | 727 (82.8) | 151 (17.2) | 0.656 | 1.070 (0.794–1.444) | 302 (68.8) | 123 (28.0) | 14 (3.2) | 0.905 | 1.119 (0.443–2.827) |
| Controls | 402 (83.8) | 78 (16.2) | | | 169 (70.4) | 64 (26.7) | 7 (2.9) | | |
| *Female* | | | | | | | | | |
| BPAD | 349 (83.1) | 71 (16.9) | 0.152 | 1.314 (0.903–1.911) | 147 (70.0) | 55 (26.2) | 8 (3.8) | 0.229 | 2.975 (0.775–11.424) |
| SCZ | 176 (79.3) | 46 (20.7) | **0.015** | 1.688 (1.104–2.581) | 73 (65.8) | 30 (27.0) | 8 (7.2) | **0.015** | 5.991 (1.545–23.232) |
| All cases | 525 (81.8) | 117 (18.2) | **0.035** | 1.439 (1.025–2.021) | 220 (68.5) | 85 (26.5) | 16 (5.0) | 0.054 | 3.976 (1.140–13.871) |
| Controls | 381 (86.6) | 59 (13.4) | | | 164 (74.5) | 53 (24.1) | 3 (1.4) | | |

Columns include the counts of the alleles and genotypes with percentages in parentheses. Significance assessed with the $\chi^2$-test of independence, and declared at a nominal *P*-value of 0.05 (bold and underlined). ORs and 95% CIs are shown relative to the C allele and the CC genotype. Allele and genotype frequencies for all SNPs are located in Table S1 of the Supplementary Information.

**Table 3** Summary of significant haplotypes identified with Cocaphase after permutation analysis

| | | | |
|---|---|---|---|
| 1 | rs11736084 (C/T) | T | T |
| *2* | **rs1282 (C/T)** | *C* | *C* |
| 3 | ih31 (A/C) | C | C |
| 4 | rs9685931 (A/G) | G | G |
| 5 | rs9799720 (C/G) | | C |
| 6 | rs3756179 (C/T) | | |
| *All affected (F)* | | | |
| | Global test *P*-value (permuted *P*-value, s.e.) | 0.015 (0.044, 0.002) | 0.017 (0.033, 0.002) |
| | Individual test *P*-value (permuted *P*-value, s.e.) | 0.003 (0.048, 0.002) | 0.004 (0.070, 0.003) |
| | OR (95% CI) | 0.547 (0.367–0.815) | 0.562 (0.375–0.842) |
| | Haplotype frequency – case | 0.105 | 0.105 |
| | Haplotype frequency – control | 0.180 | 0.180 |
| *BPAD (F)* | | | |
| | Global test *P*-value (permuted *P*-value, s.e.) | 0.013 (0.038, 0.002) | 0.016 (0.032, 0.002) |
| | Individual test *P*-value (permuted *P*-value, s.e.) | 0.004 (0.066, 0.002) | 0.005 (0.101, 0.003) |
| | OR (95% CI) | 0.496 (0.313–0.785) | 0.514 (0.323–0.818) |
| | Haplotype frequency – case | 0.096 | 0.096 |
| | Haplotype frequency – control | 0.180 | 0.180 |

The six SNPs and the haplotypes are shown at the top, with the nominally significant SNP rs1282 in bold across. The results of the analysis of the female BPAD and SCZ individuals (all affected (F)) and of the female BPAD individuals (BPAD (F)) are shown under the haplotypes. Both global and individual test *P*-values are shown, along with their corresponding permuted *P*-values, ORs with 95% CIs and frequencies in the cases and in the controls. All haplotypes achieving nominal significance ($P<0.05$) in the study are shown in Table S2 and S3 of the Supplementary Information.

permutations was performed to correct the single markers and haplotypes that reached the nominal significance level. In the analysis of the SCZ females, SNP rs1282 only remained significant under the recessive model ($P=0.017$). No haplotype tested by either the global or individual test maintained its significance after permutation analysis. However, the global test *P*-values of two related haplotypes remained significant in both the BPAD females and all affected females. These haplotypes were the four-SNP haplotype, rs11736084–rs1282–ih31–rs9685931 (snp1–4) ($P=0.038$ and 0.044, respectively) and the five-SNP haplotype, rs11736084–rs1282–ih31–rs9685931–rs9799720 (snp1–5) ($P=0.032$ and 0.033, respectively). The individual test *P*-value of the four-SNP (snp1–4) haplotype, TCCG, also remained significant after permutation analysis ($P=0.048$, $R=0.547$, 95% CI: 0.367–0.815) (Table 3).

*Bioinformatics analysis*

Two SNPs, ih31 and rs9685931, both of which are in the significant four- and five-SNP haplotypes, change the amino-acid sequence of the GPR78 protein. SNP ih31 occurs in the first exon and changes amino acid 201 from an arginine to a serine, R201S. It occurs in the intracellular loop just after TMD V (Figure 1). It changes a charged residue to an uncharged residue in a string of charged residues. In the linked families, the minor allele, C, was observed on nine of the 38 control chromosomes (frequency = 0.237) and on the F50-linked chromosome. The arginine residue does not appear to be conserved among GPCRs or across species as other residues (aspartic acid, phenylalanine and valine) were found in that position, and SIFT[41] analysis indicated that the serine residue is tolerated. Furthermore, SNP ih31 was not predicted to alter the TMD structure nor to create or abolish a PROSITE-predicted protein motif, and NetPhos[44] did not predict the serine variant to be phosphorylated (data not shown).

SNP rs9685931 occurs in the third exon and changes amino acid 342 from an arginine to a histidine, R342H. It occurs in the intracellular C-terminal of GPR78. The variant A allele was observed on four of 38 control chromosomes (frequency = 0.105) and only the linked chromsome of F22. The C-terminal is highly divergent, making it difficult to align with other sequences. Subsequent bioinformatic analysis showed that the SNP allele does not alter the predicted TMD structure, but it does abolish a PROSITE-predicted protein kinase C phosphorylation site.

The other three SNPs from the significant haplotypes, including SNP rs1282, which is located in a putative regulatory region, were assessed by investigating cross-species conservation at these sites, using the Multiz and PhyloHMM track on the UCSC Genome Browser. None of the SNPs were located in or near conserved blocks of DNA (data not shown). When the MATCH™ program was used to search the alternate 50 b sequences centred about SNP rs1282, no transcription factor binding sites were predicted for either of the two alleles.

## Discussion

This is the first study reporting polymorphism screening, LD mapping and association analysis of the orphan *GPR78* gene in relation to major psychiatric illness. Our single-marker analysis identified one SNP to be significantly associated with SCZ in the females (corrected $P = 0.017$). Also, two haplotypes, both of which contain the significant SNP, were found to be significantly associated with BPAD ($P = 0.038$ and 0.032) and with both BPAD and SCZ ($P = 0.044$ and 0.033). These associations were only detected in the females.

At the time that this work was initiated, the International HapMap project[51] was very much in its infancy, and coverage of the *GPR78* gene consisted of two SNPs. Therefore, it was necessary to screen the gene for polymorphisms. This resulted in the identification of 23 SNPs, five of which had yet to be described in dbSNP (build 121). The nine SNPs suitable for LD analysis were distributed from the 5′UTR to exon 1 region and from exon 3 to the 3′UTR. Few SNPs were identified in the intervening intronic regions and exon 2 (Figure 1), and those that were failed to meet the inclusion criteria. As a result, LD in this intervening region has not been definitively characterised. A fully comprehensive screen of *GPR78* would need to include SNPs between the six LD blocks observed. The current release of the HapMap project (Release no. 16c.1, June 2005) shows that there are nine SNPs genotyped in the same region covered by the nine SNPs used for the LD analysis in our study (Chr4: 8766521–8785280, UCSC Genome Browser Release, July 2003). Only two of these SNPs are the same (rs1282 and rs9685931), making it difficult to compare directly the relative quality of coverage in the region. However, inspection of the location and LD patterns of these nine HapMap SNPs reveals that the same intervening region between exons 1 and 3 is still sparsely covered, offering no additional SNPs for improvement of the LD characterisation in that region.

The single-marker analysis of the case–control sample revealed an association between SNP rs1282 (snp2) and the females with SCZ (uncorrected allele $P = 0.015$, uncorrected genotype $P = 0.015$). The T allele conferred an OR of 1.688 (95% CI: 1.104–2.581) and the TT genotype an OR of 5.991 (95% CI: 1.545–23.232) (Table 2). Assuming a recessive mode of inheritance, the significance of rs1282 increased ($P = 0.005$, $OR_{TT} = 5.618$, 95% CI: 1.460–21.617) and remained significant after permutation testing ($P = 0.017$). The importance of SNP rs1282 is also emphasised by the fact that all but one of the haplotypes that reached nominal significance contained it. Also, its state, allele T versus allele C, dictated the direction of the OR associated with the haplotype; that is, the T allele always acted as a susceptibility variant ($OR > 1$), whereas the C allele acted as a protective variant ($OR < 1$). Furthermore, the backward logistic regression analysis also identified SNP rs1282 as the only significant factor in the SCZ females.

SNP rs1282 is located approximately 3 kb upstream of the *GPR78* start site, placing it in a putative regulatory region. However, bioinformatics analysis found that it was not in a conserved sequence block, nor did it alter any predicted transcription factor binding site. This suggests that SNP rs1282 is not itself functional, but instead may be in strong LD with a causative variant.

The sliding-window haplotype analysis identified two global haplotypes that remained significant ($P < 0.05$) after permutation analysis (10 000 permutations) in the sample of BPAD females and in the sample of all affected females. These haplotypes span the length of the gene and include SNP rs1282. They include the two non-synomous coding SNPs, ih31 and rs9685931, identified from the polymorphism screening process. Bioinformatics analysis does not predict that ih31 will affect the function of the protein in any way. However, rs9685931 was found to abolish a PROSITE-predicted protein kinase C phosphorylation site. PROSITE tends to overpredict protein motifs and therefore this cannot be assumed to be a functional motif. However, the intracellular position of the variant is consistent with an effect on intracellular signalling. Intracellular signalling cascades, such as cAMP, have been implicated in the pathophysiology of BPAD.[13] No association was detected with this SNP alone, arguing against rs9685931 itself being the causative variant, although it may be the case that the power gained in testing haplotypes instead of single markers[52,53] revealed this underlying association. Alternatively, a nearby SNP in strong LD with rs9685931 or a SNP working in conjunction with rs9685931 may be the causative variant.

The associations identified in this case–control study were all detected in the female individuals. A female-specific association between BPAD and/or SCZ has been previously observed in both *GPR50*[15] and *CREB1*.[36] However, we cannot definitively say that there is no effect in males because the 95% CIs of the ORs in the male groups overlap with those of the ORs in the female samples where the significance was observed, indicating that there is no significant difference between their respective ORs. Alternatively, it is possible that the association detected in the females is a false-positive. Therefore, replication studies in other, larger cohorts is warranted.

As ORs and 95% CIs are estimated using the sample from which the statistical significance of the associations was observed, it is likely that they are biased.[54] This means that, given the observed genotypic OR of 5.991 (95% CI: 1.545–23.232) for the TT genotype of SNP rs1282, the next study that is designed with the aim of replicating this should select a sample size that has the power to detect an OR of less than 1.5, as this is the upwardly biased lower bound of the 95% CI. This value is in keeping with the genotypic ORs of the

complex disease susceptibility genes that have been identified to date.[55]

*GPR78* was selected as a candidate gene for psychiatric illness by virtue of its genomic location in a region of confirmed linkage and its putative function. Expressed in the pituitary and the placenta,[16] *GPR78* may play a role in the functioning of the HPA axis and during pregnancy. The HPA axis is involved in stress regulation,[18] and there is evidence of its dysfunction being associated with BPAD[19] and SCZ.[21] The expression of *GPR78* in the placenta also links *GPR78* to the functioning of the HPA axis, as it has been shown that prenatal stress, resulting in the release of maternal hormones, can lead to the dysregulation of the HPA axis in the developing fetus.[22]

In summary, SNP rs1282 was found to be significantly associated in females with SCZ, under a recessive model. Two haplotypes spanning the length of the gene and including potentially functionally significant variants were also found to be significantly associated with both BPAD and all affected females in a Scottish sample. Our results implicate *GPR78* in the pathophysiology of these devastating disorders.

## Acknowledgments

We would like to thank the Sanger Institute and the Genetics Core of the Wellcome Trust Clinical Research Facility at the Western General Hospital for carrying out the SNP genotyping; M Walker and M van Beck for assistance with clinical aspects of the study and ST Cooper for assistance in the preparation of the manuscript. A Tenesa acknowledges funding from Cancer Research UK. This work was supported by grants from the Medical Research Council. This work was supported by the UK Medical Research Council.

## References

1 Craddock N, Khodel V, Van Eerdewegh P, Reich T. Mathematical limits of multilocus models: the genetic transmission of bipolar disorder. *Am J Hum Genet* 1995; **57**: 690–702.
2 Berry N, Jobanputra V, Pal H. Molecular genetics of schizophrenia: a critical review. *J Psychiatry Neurosci* 2003; **28**: 415–429.
3 Tsuang MT, Taylor L, Faraone SV. An overview of the genetics of psychotic mood disorders. *J Psychiatr Res* 2004; **38**: 3–15.
4 Blackwood DH, He L, Morris SW, McLean A, Whitton C, Thomson M *et al.* A locus for bipolar affective disorder on chromosome 4p. *Nat Genet* 1996; **12**: 427–430.
5 Le Hellard S, Lee AJ, Underwood S, Thomson PA, Morris SW, Torrance HS *et al.* Haplotype analysis and a novel allele sharing method refines a chromosome 4p locus linked to bipolar disorder and schizophrenia. 2005 (submitted).
6 Asherson P, Mant R, Williams N, Cardno A, Jones L, Murphy K *et al.* A study of chromosome 4p markers and dopamine D5 receptor gene in schizophrenia and bipolar disorder. *Mol Psychiatry* 1998; **3**: 310–320.
7 Ewald H, Degn B, Mors O, Kruse TA. Support for the possible locus on chromosome 4p16 for bipolar affective disorder. *Mol Psychiatry* 1998; **3**: 442–448.
8 Detera-Wadleigh SD, Badner JA, Berrettini WH, Yoshikawa T, Goldin LR, Turner G *et al.* A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proc Natl Acad Sci USA* 1999; **96**: 5604–5609.
9 Williams NM, Rees MI, Holmans P, Norton N, Cardno AG, Jones LA *et al.* A two-stage genome scan for schizophrenia susceptibility genes in 196 affected sibling pairs. *Hum Mol Genet* 1999; **8**: 1729–1739.
10 Lerer B, Segman RH, Hamdan A, Kanyas K, Karni O, Kohn Y *et al.* Genome scan of Arab Israeli families maps a schizophrenia susceptibility gene to chromosome 6q23 and supports a locus at chromosome 10q24. *Mol Psychiatry* 2003; **8**: 488–498.
11 Als TD, Dahl HA, Flint TJ, Wang AG, Vang M, Mors O *et al.* Possible evidence for a common risk locus for bipolar affective disorder and schizophrenia on chromosome 4p16 in patients from the Faroe Islands. *Mol Psychiatry* 2004; **9**: 93–98.
12 Evans KL, Le Hellard S, Morris SW, Lawson D, Whitton C, Semple CA *et al.* A 6.9 Mb high-resolution BAC/PAC contig of human 4p153–p161 a candidate region for bipolar affective disorder. *Genomics* 2001; **71**: 315–323.
13 Gould TD, Manji HK. Signaling networks in the pathophysiology and treatment of mood disorders. *J Psychosom Res* 2002; **53**: 687–697.
14 Gonzalez-Maeso J, Rodriguez-Puertas R, Meana JJ, Garcia-Sevilla JA, Guimon J. Neurotransmitter receptor-mediated activation of G-proteins in brains of suicide victims with mood disorders: selective supersensitivity of alpha(2A)-adrenoceptors. *Mol Psychiatry* 2002; **7**: 755–767.
15 Thomson PA, Wray NR, Thomson AM, Dunbar DR, Grassie MA, Condie A *et al.* Sex-specific association between bipolar affective disorder in women and GPR50, an X-linked orphan G protein-coupled receptor. *Mol Psychiatry* 2005; **10**: 470–478.
16 Lee DK, Nguyen T, Lynch KR, Cheng R, Vanti WB, Arkhitko O *et al.* Discovery and mapping of ten novel G protein-coupled receptor genes. *Gene* 2001; **275**: 83–91.
17 Lee DK, Lynch KR, Nguyen T, Im DS, Cheng R, Saldivia VR *et al.* Cloning and characterization of additional members of the G protein-coupled receptor family. *Biochim Biophys Acta* 2000; **1490**: 311–323.
18 Herman JP, Prewitt CM, Cullinan WE. Neuronal circuit regulation of the hypothalamo-pituitary-adrenocortical stress axis. *Crit Rev Neurobiol* 1996; **10**: 371–394.
19 Rybakowski JK, Twardowska K. The dexamethasone/corticotropin-releasing hormone test in depression in bipolar and unipolar affective illness. *J Psychiatr Res* 1999; **33**: 363–370.
20 Pariante CM, Miller AH. Glucocorticoid receptors in major depression: relevance to pathophysiology and treatment. *Biol Psychiatry* 2001; **49**: 391–404.
21 Altamura AC, Boin F, Maes M. HPA axis and cytokines dysregulation in schizophrenia: potential implications for the antipsychotic treatment. *Eur Neuropsychopharmacol* 1999; **10**: 1–4.
22 Weinstock M. Does prenatal stress impair coping and regulation of hypothalamic-pituitary-adrenal axis? *Neurosci Biobehav Rev* 1997; **21**: 1–10.
23 Bromet EJ, Fennig S. Epidemiology and natural history of schizophrenia. *Biol Psychiatry* 1999; **46**: 871–881.
24 Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry* 1978; **35**: 837–844.
25 Association AP. *Diagnostic and Statistical Manual of Mental Disorders* 4th edn. (DSM–IV). APA: Washington, DC, 1994.
26 World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. WHO: Geneva, 1992.
27 Risch NJ. Searching for genetic determinants in the new millennium. *Nature* 2000; **405**: 847–856.
28 Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000; **132**: 365–386.
29 Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**: 195–202.
30 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.

31 Wray NR. Allele frequencies and the $r^2$ measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Res Hum Genet* 2005; **8**: 87–94.
32 Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 2003; **4**: 587–597.
33 Berrettini W. Evidence for shared susceptibility in bipolar disorder and schizophrenia. *Am J Med Genet C* 2003; **123**: 59–64.
34 Craddock N, O'Donovan MC, Owen MJ. The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J Med Genet* 2005; **42**: 193–204.
35 Holden C. Sex and the suffering brain. *Science* 2005; **308**: 1574.
36 Zubenko GS, Hughes III HB, Stiffler JS, Brechbiel A, Zubenko WN, Maher BS *et al.* Sequence variations in CREB1 cosegregate with depressive disorders in women. *Mol Psychiatry* 2003; **8**: 611–618.
37 Kirkwood BR, Sterne JAC. *Essential Medical Statistics*, 2nd edn. Blackwell Science Ltd: Malden, MA, 2003.
38 Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 2002; **70**: 124–141.
39 Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003; **25**: 115–121.
40 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410.
41 Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**: 3812–3814.
42 Cserzo M, Eisenhaber F, Eisenhaber B, Simon I. On filtering false positive transmembrane protein predictions. *Protein Eng* 2002; **15**: 745–752.
43 Hulo N, Sigrist CJ, Le Saux V, Langendijk-Genevaux PS, Bordoli L, Gattiker A *et al.* Recent improvements to the PROSITE database. *Nucleic Acids Res* 2004; **32**: D134–D137.
44 Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 1999; **294**: 1351–1362.
45 Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* 2004; **14**: 708–715.
46 Siepel A, Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. *J Comput Biol* 2004; **11**: 413–428.
47 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM *et al.* The human genome browser at UCSC. *Genome Res* 2002; **12**: 996–1006.
48 Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT *et al.* The UCSC Genome Browser Database. *Nucleic Acids Res* 2003; **31**: 51–54.
49 Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* 2003; **31**: 3576–3579.
50 Wingender E, Chen X, Fricke E, Geffers R, Hehl R, Liebich I *et al.* The TRANSFAC system on gene expression regulation. *Nucleic Acids Res* 2001; **29**: 281–283.
51 Consortium TIH. The international HapMap project. *Nature* 2003; **18**: 789–796.
52 Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *Eur J Hum Genet* 2001; **9**: 291–300.
53 Jawaid A, Sham PC, Makoff AJ, Asherson PJ. Is haplotype tagging the panacea to association mapping studies? *Eur J Hum Genet* 2004; **12**: 259–262.
54 Goring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet* 2001; **69**: 1357–1369.
55 Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. *Nat Rev Genet* 2005; **6**: 109–118.

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