Molecular Psychiatry (2007) 12, 1011–1025

www.nature.com/mp

ORIGINAL ARTICLE

Association analysis of the chromosome 4p15–p16 candidate region for bipolar disorder and schizophrenia

A Christoforou[1,4], S Le Hellard[1,4,5], PA Thomson[1], SW Morris[1], A Tenesa[2], BS Pickard[1], NR Wray[1,6], WJ Muir[1,3], DH Blackwood[1,3], 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 and* [3]*Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK*

Several independent linkage studies have identified chromosome 4p15–p16 as a putative region of susceptibility for bipolar disorder (BP), schizophrenia (SCZ) and related phenotypes. Previously, we identified two subregions (B and D) of the 4p15–p16 region that are shared by three of four 4p-linked families examined. Here, we describe a large-scale association analysis of regions B and D (3.8 and 4.5 Mb, respectively). We selected 408 haplotype-tagging single nucleotide polymorphisms (SNPs) on a block-by-block basis from the International HapMap project and tested them in 368 BP, 386 SCZ and 458 control individuals. Nominal significance thresholds were determined using principal component analysis as implemented in the program SNPSpD. In region B, overlapping SNPs and haplotypes met the region-wide threshold ($P \leqslant 0.0005$) at the global and individual haplotype test level and clustered in two regions. In region D, no individual SNPs were nominally significant, but multiple global and individual haplotypes were associated with BP and/or SCZ (region-wide threshold, $P \leqslant 0.0003$). These overlapping haplotypes fell into two regions. Within each of these four clusters, at least one globally significant haplotype withstood permutation testing ($P_{gp} \leqslant 0.05$). Five predicted genes were found within these associated regions, while *Known*/*RefSeq* genes, including *KIAA0746* and *PPARGC1A*, mapped nearby. There were also nine other clusters within regions B and D with nominally significant haplotypes, but only at the individual haplotype level. *KIAA0746*, *PPARGC1A*, *GPR125*, *CCKAR* and *DKFZp761B107* overlapped with these regions. This study has identified significant associations between BP and SCZ within the chromosome 4p linkage region, resulting in candidate regions worthy of further investigation.
Molecular Psychiatry (2007) **12,** 1011–1025; doi:10.1038/sj.mp.4002003; published online 24 April 2007

Keywords: bipolar disorder; schizophrenia; chromosome 4p; association study; haplotype analysis

Introduction

Bipolar disorder (BP) (MIM 125480) and schizophrenia (SCZ) (MIM 181500) are major psychiatric disorders, each of which affects 1% of the world population. In recent surveys, the World Health Organization has identified BP and SCZ as the sixth and ninth leading causes of disability worldwide, respectively.[1] BP and SCZ are major burdens for patients, their families and society and incur large economic costs. Family, twin and adoption studies have shown that susceptibility to both illnesses is likely to involve interaction between a substantial genetic component and environmental factors. Progress towards identifying these genetic elements is now being made, including, for example, the *neuregulin* (*NRG1*)[2,3] and *disrupted in schizophrenia 1* (*DISC1*)[4] genes in susceptibility to SCZ and BP.

Previously, we described a large Scottish family that showed significant linkage of BP and recurrent major depression (MIM 125480) to chromosome 4p15–p16 (log of odds (LOD) = 4.1).[5] Further support for this result came from variance component analysis, which found significant evidence for a quantitative trait locus in the region (LOD = 3.7).[6] Recently, we reported a follow-up, which strengthened the original linkage result (LOD = 4.4).[7] A number of independent groups have also provided support for the involvement of this candidate region in susceptibility to major mental illness (as previously reviewed in Le Hellard *et al.*[7]). These findings include linkage to major mental illness in a large family from the USA (one of 22 studied) (F48, LOD = 3.24)[8] and excess

Correspondence: Dr A Christoforou, Medical Genetics, Centre of Molecular Medicine, Western General Hospital, Crewe Road South, Edinburgh, EH4 2XU, UK.
E-mail: a.christoforou@ed.ac.uk
[4]These authors contributed equally to this work.
[5]Current address: Dr Einar Martens' Research Group for Biological Psychiatry and Bergen Mental Health Research Center, Section of Medical Genetics and Molecular Medicine, University of Bergen, Bergen, Norway.
[6]Current address: Queensland Institute of Medical Research, Post Office Royal Brisbane Hospital, Brisbane, QLD, Australia.
Received 27 October 2006; revised 1 January 2007; accepted 6 February 2007; published online 24 April 2007

haplotype sharing in families from the Faroe Islands with BP and SCZ (best *P*-value, $P = 0.00007$).[9] This region of chromosome 4p has also been highlighted by a balanced translocation t(4;13) (p16.1;q21.31) discovered in a patient with SCZ.[10]

Recently, we refined the 20-Mb candidate region on 4p15–p16 by defining the linked haplotypes in family F22 and in three other families (F48,[8] F50[11] and F59[5]) that also show linkage to this region.[7] This delineated two subregions (B and D) of 3.8 and 4.5 Mb, respectively, that both showed linkage in three of the four 4p-linked families studied (Figure 1, adapted with kind permission from Le Hellard *et al.*[7]). In addition, we have recently described preliminary evidence for association in *GPR78*, a functional candidate gene that maps to region B.[12]

As the regions identified by linkage analysis are still relatively large, we chose to combine studies of extended pedigrees with a population-based approach. Our strategy is based on the hypothesis that genes that are implicated by mutations with large effects and strong genotype–phenotype correlations in pedigrees with multiple affected cases may also harbor variants that predispose to the disease in the general population. Genome-wide linkage analyses of families with multiple affected cases have already identified candidate regions where association studies have subsequently led to the identification of genes involved in susceptibility to a number of complex disorders. These include the *complement factor H* gene[13–15] and *LOC387715*[16] in age-related macular degeneration as well as the *NRG1*[2] and *DISC1*[17,18] genes in SCZ. Therefore, in line with the strategy of first linkage and then association, we screened our candidate regions, B and D, for association with BP and SCZ in a Scottish case–control sample. To carry out a comprehensive survey, we selected single nucleotide polymorphisms (SNPs) from the International HapMap Project[19] (http://www.hapmap.org/). A linkage disequilibium (LD) map of the two regions was constructed, using the data from the CEPH trios, and haplotype-tagging SNPs (htSNPs) were selected on the basis of their ability to differentiate common haplotypes (frequency $\geqslant 10\%$). Here, we report significant associations with BP and/or SCZ in the chromosome 4p15–p16 candidate region.

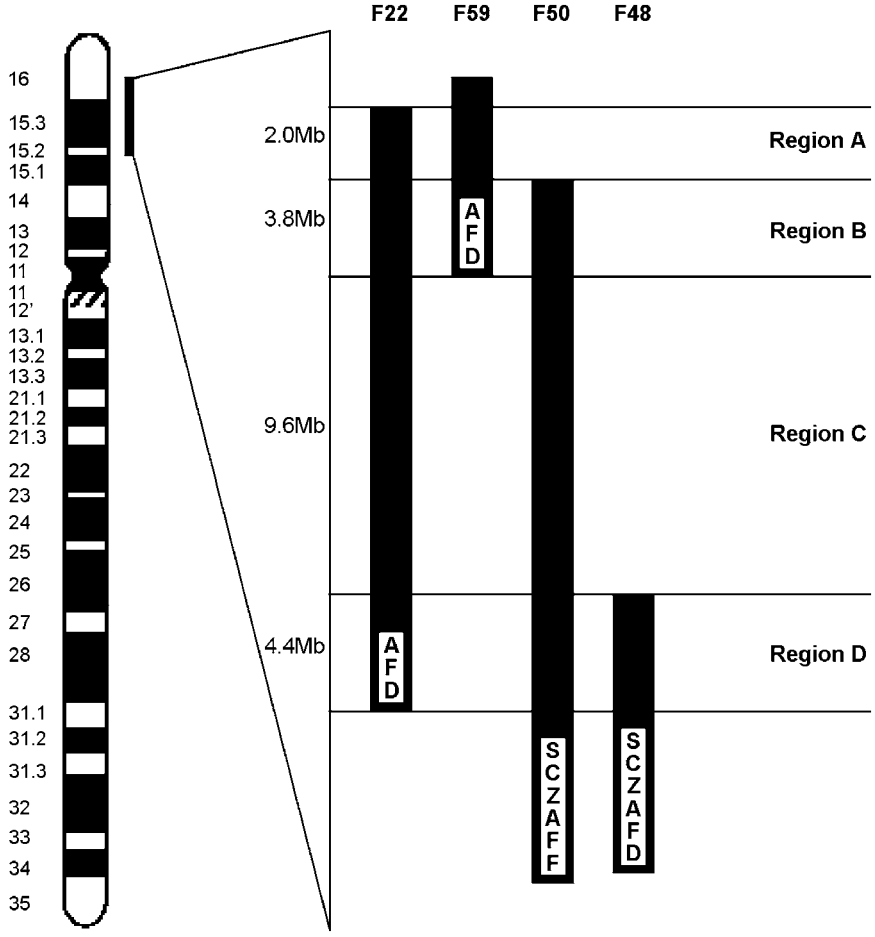

Figure 1 The overlap between the linked regions that segregate with illness in the four families. The sizes (in Mb) refer to the genomic distances between the points marked by the horizontal lines. Regions A–D indicate subregions of the F22 linkage region that show linkage in at least one other family. The illnesses observed in the families are indicated on the figure as follows: 'AFD', BP and recurrent major depressive disorder; 'SCZAFF', schizoaffective disorder and SCZ and 'SCZAFD', BP, recurrent major depressive disorder, SCZ and others. (Adapted with kind permission from Le Hellard *et al.*[7]).

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 with BP or SCZ were recruited from inpatient and outpatient services at the Royal Edinburgh Hospital and other Scottish psychiatric hospitals. Patients were screened by experienced psychiatrists using the semistructured interview: the Schedule for Affective Disorder and Schizophrenia – Lifetime version (SADS-L).[20] Based on the information gathered from the interview and on case-note review, diagnoses were reached by consensus between the two experienced psychiatrists (DHB and WJM) and made according to the Diagnostic and Statistical Manual of Mental Disorders (4th edn) (DSM–IV).[21]

Control subjects were drawn from the same population in South East and South Central Scotland. The majority (~80%) was recruited through the Scottish national blood transfusion service, which only accepts donors that are not currently on medication and have no chronic illness. The remaining controls were recruited from the local population or from hospital staff. Each of the additional controls was briefly screened by interview to exclude anyone currently on medication or with a history of treatment for psychiatric illness. Both the cases and the controls were sampled from the general Scottish population, 98% of which is of (self-reported) white background (89.9% 'white Scottish', 7.5% 'other white British', 1% 'white Irish' and 1.6% 'any other white background')[22] (http://www.scotland.gov.uk/Publications/2004/02/18876/32917).

Selection of markers

SNP genotype data for the 30 CEPH trios (Utah residents with ancestry from northern and western Europe) (CEU) were downloaded in overlapping segments of approximately 1 Mb from HapMap Release 7 (May 2004) (http://www.hapmap.org). The data were subsequently uploaded into Haploview v 2.5 (http://www.broad.mit.edu/mpg/haploview/index.php)[23] to construct LD maps for the regions. Pair-wise comparisons of markers more than 500 kb apart were ignored (Haploview default). Only markers with a minor allele frequency (MAF) greater than or equal to 0.10, a Hardy–Weinberg (HW) *P*-value greater than 0.001 (Haploview default) and a genotyping success rate of 0.75 or better (Haploview default) were included in the LD analysis. The haplotype blocks were defined using the solid spine of LD approach, which creates blocks only when the first and last SNPs are in strong LD ($|D'| > 0.80$) with all of the intermediate SNPs.[23] Then, for haplotypes with a frequency of at least 0.01, adjacent haplotype blocks with a Hedrick's multiallelic D' (MAD′)[24] greater than or equal to 0.95 were merged manually. This process was repeated until the MAD′ between any two adjacent blocks was less than 0.95. Finally, using Haploview's internal tagging program, htSNPs were selected on a block-by-block basis to represent haplotypes of frequencies greater than or equal to 0.10. Individual SNPs (singletons) that fell between blocks were also included in the set of htSNPs. However, the LD between them and adjacent blocks was not determined.

In a previous study,[18] we compared the LD map and haplotype frequencies between our samples and the HapMap CEPH trios for a region on chromosome 1. Those results indicated that the two populations have similar LD structure and haplotype frequencies. Also, in a study of the consistency of the tagging SNP approach in multiple samples, Ke *et al.*[25] observed a high level of consistency between the tagging SNPs selected in their UK sample and the US CEPH samples. The results of these two studies validate our use of HapMap data to select htSNPs for our population.

Genotyping

Genotyping was performed by Illumina Inc. (San Diego, CA, USA), using the high-throughput BeadArray platform technology. A total of 421 SNPs were submitted for genotyping. Four hundred and eight of these SNPs were successfully genotyped at an average locus success rate of 99% (range: 94.3–99.9%) in a total of 1212 individuals (93% sample success rate).

Statistical analysis

Regions B and D were analyzed separately as they are independent linkage regions, possibly representing two different susceptibility loci, with three families of Celtic origin showing linkage to B and three with broader ancestry showing linkage to D.[7]

The standard χ^2 test of independence was used to examine all markers for deviations from HW equilibrium in the control sample. The Haploview[23] default HW significance threshold of $P = 0.001$ was selected as a cutoff to balance the possibility of multiple testing artifacts against the occurrence of methodological problems or structural abnormalities.

Single-marker analysis

Differences in single-marker allele and genotype frequencies between cases and controls were evaluated with the χ^2 test of independence (with 1 and 2 degrees of freedom, respectively). The level at which individual markers were declared nominally significant was determined using SNPSpD[26] (http://genepi.qimr.edu.au/general/daleN/SNPSpD/) on the full sample of cases and controls. SNPSpD performs a simple multiple testing correction procedure for SNPs in LD with each other, determining the effective number of independent markers (M_{eff})[26,27] in the entire set. Then, using the estimate of the M_{eff} (M_{effLi}[28]), the program provides the Sidak-corrected[29] significance threshold required to keep the Type I error rate at 5%. We determined thresholds both for each region separately (region-wide thresholds) and for regions B and D together (experiment-wide thresh-

old). The adjusted significance thresholds are referred to as 'Nyholt-corrected' thresholds.

Nyholt's multiple testing correction has been shown to provide a good approximation to permutation-based corrections, which are the 'gold standard' for multiple, correlated tests.[26] To help assess whether this was the case in our study, the markers that met the Nyholt-corrected threshold at either the relevant region- or experiment-wide significance level were subsequently corrected by permutation analysis (10 000 shuffles), using the statistical analysis program Cocaphase 2.43[30] (unphased suite of programs: http://www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/). The permutation analysis provides a *P*-value that is adjusted at the 'region-wise' or 'experiment-wise' significance level[31] for the most significant single-marker *P*-value observed for a set of markers. This corrects for the multiple markers tested in the particular region of interest. Individual markers that met the Nyholt-corrected threshold, but were not the most significant, were subjected to the permutation analysis once the SNP(s) with better nominal *P*-values were removed from the data set (as carried out, for example, in the Holm 'step-down' procedure[32]). Although the Cocaphase permutation option does not provide the threshold at which the Type I error rate would be maintained at 0.05, a significant permutation-corrected *P*-value ($P_p \leqslant 0.05$) supports using the Nyholt-corrected thresholds for our data set.

Haplotype analysis

Haplotype frequency estimation and comparison was carried out using the program Cocaphase 2.43.[30] Using a sliding-window approach, haplotypes of lengths 2–5 were tested. *P*-values for both global and individual tests of haplotype frequencies were calculated. The global test *P*-value (P_g) determines the significance of the overall difference in the distribution of haplotype frequencies between cases and controls. To avoid misleading results caused by multiple rare haplotypes in the global test of the null hypothesis,[33] haplotypes with a frequency less than or equal to 5% in both the cases and the controls were designated rare and clumped together into one haplotype. The *P*-value from the individual test (P_i) represents the significance of the difference in case versus control frequency of an individual haplotype when compared with all the other possible haplotypes in that window.

Haplotypes were declared nominally significant if they met the region-wide Nyholt-corrected thresholds determined in the single-marker analysis for regions B or D (as carried out in Fallin *et al.*[34] and Greenwood *et al.*[35]). Since the Nyholt-corrected thresholds were determined by accounting for the correlation between single markers, it is unclear how accurately these apply to sliding-window haplotype analysis, which also involves many highly correlated tests. Therefore, a subset of haplotype *P*-values were also corrected by permutation analysis (1000 shuffles) at the region-wide level using Cocaphase 2.43.[30] The haplotypes that were corrected were those with nominally significant global *P*-values and their corresponding haplotypes (only for lengths $\leqslant$ three SNPs) with nominally significant individual *P*-values. *P*-values (P_{gp} or P_{ip}) that met the 0.05 threshold after permutation correction were declared significant. Not all haplotypes could be corrected because of the intensive computational requirements of the permutation-based analysis. Also, only the most significant haplotype of a particular sliding-window length tested in a sample of cases and controls could be corrected.

All tests were performed on the BP and SCZ samples separately and combined (All), and on these samples separated on the basis of gender (male (M) and female (F)). *P*-values were not further adjusted for these additional hypothesis-based tests (see Discussion). Odds ratios (ORs) and 95% confidence intervals (CIs) were estimated for the single markers and haplotypes that met the relevant region-wide Nyholt-corrected thresholds.

Delineating the associated regions

Each of the individual haplotypes that met the region-wide Nyholt-significant threshold was followed along the sliding windows in both the 5′ and 3′ directions, using the output from the single-marker and haplotype analyses, to determine how far each haplotype extended while $P_i \leqslant 0.01$. The htSNPs that covered the extended haplotypes were then matched up with their corresponding block. Finally, in order to keep the blocks intact, the original LD maps were consulted to determine how far each block extended on the basis of the SNPs available in HapMap Release 7 (May 2004). The regions spanned by the blocks containing the clusters of extended haplotypes were defined as the associated regions.

Gene identification

The genomic region spanning the associated region was queried in UCSC (browser version May 2004) for (i) *Known* genes (June 2005) based on protein data from UniProt (SWISS-PROT and TrEMBL) and mRNA data from RefSeq and GenBank; (ii) NCBI *RefSeq* genes from the NCBI mRNA reference sequence collection and (iii) main prediction class AceView genes that do not correspond to *Known* or *RefSeq* genes. AceView genes are constructed from mRNA, expressed sequence tag (EST) and genomic evidence using the Acembly program (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly)[36] and are categorized into a prediction class of either main or putative genes. Main prediction class genes must encode a putative protein with a coding sequence of at least 100 amino acids or they must be supported by at least one identical EST or mRNA clone and match at least eight bases on either side exactly to the genome. For sequences that do not meet either of the above criteria to be included as main class predictions, they must either be identified with an NCBI RefSeq or OMIM number or they must encode a protein with significant BlastP homology ($E < 10^{-3}$) to a cDNA-supported nematode AceView protein.

If no *Known* or *RefSeq* genes were mapped to the associated region, the closest *Known*/*RefSeq* gene, telomeric or centromeric of the region and the intervening distance, was noted.

Results

Selection of htSNPs

SNP genotype data was downloaded from a version of HapMap Data Release 7 (May 2004) for the regions B and D on chromosome 4p15–p16, covering 3.8 and 4.5 Mb, respectively. Details of the SNP coverage are presented in Table 1. A total of 1370 SNPs with a MAF$\geqslant$0.10 were available for the two regions and, using Haploview v2.5, a total of 235 LD blocks were defined in the two regions (region B: 85 and region D: 150). htSNPs were selected on a block-by-block basis to represent the genetic variation present in the common haplotypes ($\geqslant$10%). In total, 421 htSNPs were selected, resulting in an average density of 1 SNP per 20 kb and approximately 2 (range: 1–5) SNPs per block.

Association analysis

A total of 408 of the 421 htSNPs (149 in 83 blocks in region B, 259 in 146 blocks in region D) were successfully genotyped in 368 BP individuals (160 males, 208 females), 386 SCZ individuals (276 males, 110 females) and 458 controls (237 males, 218 females, three unknown). Thirteen SNPs were excluded because of assay design or genotyping failure. All of the remaining 408 markers met HW equilibrium in the controls at the default Haploview HW *P*-value threshold of 0.001 (HW *P*-value $\geqslant$0.0016) and had an average genotyping success rate of 99%. The SNPs used in the association study, including their project number, rs number, position on chromosome 4 (NCBI Bld 34), the block that they represent, their alleles and control HW *P*-values are shown in the supporting online material (SOM) Supplementary Table 1 (region B) and Supplementary Table 2 (region D).

Nyholt-corrected threshold

Correction for multiple testing is not straightforward in this study, as we have tested multiple, highly correlated SNPs and overlapping or nested haplotypes. Therefore, we applied Nyholt's spectral decomposition method, as implemented in SNPSpD, to determine an appropriate nominal significance level based on the overall correlation between the SNPs in the full sample of cases and controls. Of the 149 htSNPs spanning region B, 108 were estimated to be independent, resulting in a region-wide Nyholt-corrected threshold of $P\leqslant0.0005$. For region D, of the 259 htSNPs successfully genotyped, 191 were estimated to be independent, resulting in a region-wide threshold of $P\leqslant0.0003$. Of the 408 SNPs in both regions, 295 were estimated to be independent, resulting in an experiment-wide threshold of $P\leqslant0.0002$.

Region B

The distribution of single-marker allele (χ^2_1) *P*-values along region B is shown in Figure 2: 5.7% of the *P*-values are less than or equal to 0.05 and 1.6% are less than or equal to 0.01, suggesting that there is no overall inflation of type I error in the data set. Three SNPs (56, 126 and 131) met the region-wide Nyholt-corrected threshold of $P\leqslant0.0005$ in the analysis of allele frequencies (Table 2). They point to two separate, apparently sex-specific regions. However, none of the three SNPs are predicted to be functionally significant (see SOM Bioinformatics). The sliding-window haplotype analysis identified a total of 36 haplotypes that met either the region-wide (15 haplotypes; $P\leqslant0.0005$) or experiment-wide (21 haplotypes; $P\leqslant0.0002$) Nyholt-corrected threshold at the global and/or individual haplotype test level (SOM Supplementary Table 3). These haplotypes cluster in four distinct regions (B-1–B-4), two of which (B-1 and B-4) encompass the two regions identified in the single-marker analysis. B-1 and B-4 also contain haplotypes with nominally significant global *P*-values as well as nominally significant individual *P*-values (Figure 3, Table 3). They are described in more detail below. B-2 and B-3 each consist of a single haplotype with a significant individual *P*-value ($P_i=0.00046$ and 0.00020, respectively) and are further described in the SOM (SOM Results and Discussion and SOM Supplementary Tables 3 and 4, top). In addition, we identified haplotypes that overlap with B-1 and B-4 in other subgroups of the study cohort, including the full sample of BP cases and the full sample of affected individuals (SOM Supplementary Table 3). They had nominally significant individual, but not global, *P*-values (range $P_i=0.000065$–0.00049).

Association to cluster B-1 was seen in BP males. B-1 spans from SNP 48 to 57 and contains the significant SNP 56 (rs6820330) ($P=0.00023$), which remained

Table 1 Details of block-based SNP coverage for regions B and D on chromosome 4p15–p16

Chromosome 4 *Region*	*Coordinates covered* *(NCBI Bld34) (size, Mb)*	*Available SNPs (HapMap May 2004)*	*Blocks*[a] *(Singletons)*	*Average block size (kb) (Excluding singletons (range))*	*htSNPs*[a] *(Singletons)*	*htSNP density (kb/SNP)*
B	8769508–12567891 (3.8)	544	85 (18)	27 (34 (1–137))	154 (18)	25
D	22034705–26549096 (4.5)	826	150 (40)	22 (29 (1–191))	267 (40)	17
Total	(8.3)	1370	235 (58)	24 (31 (1–191))	421 (58)	20

Abbreviations: Mb, megabase; SNP, single nucleotide polymorphism; ht, haplotype tagging.
[a]Includes singletons, which are SNPs that fell between blocks in Release 7 (May 2004) of the HapMap project.

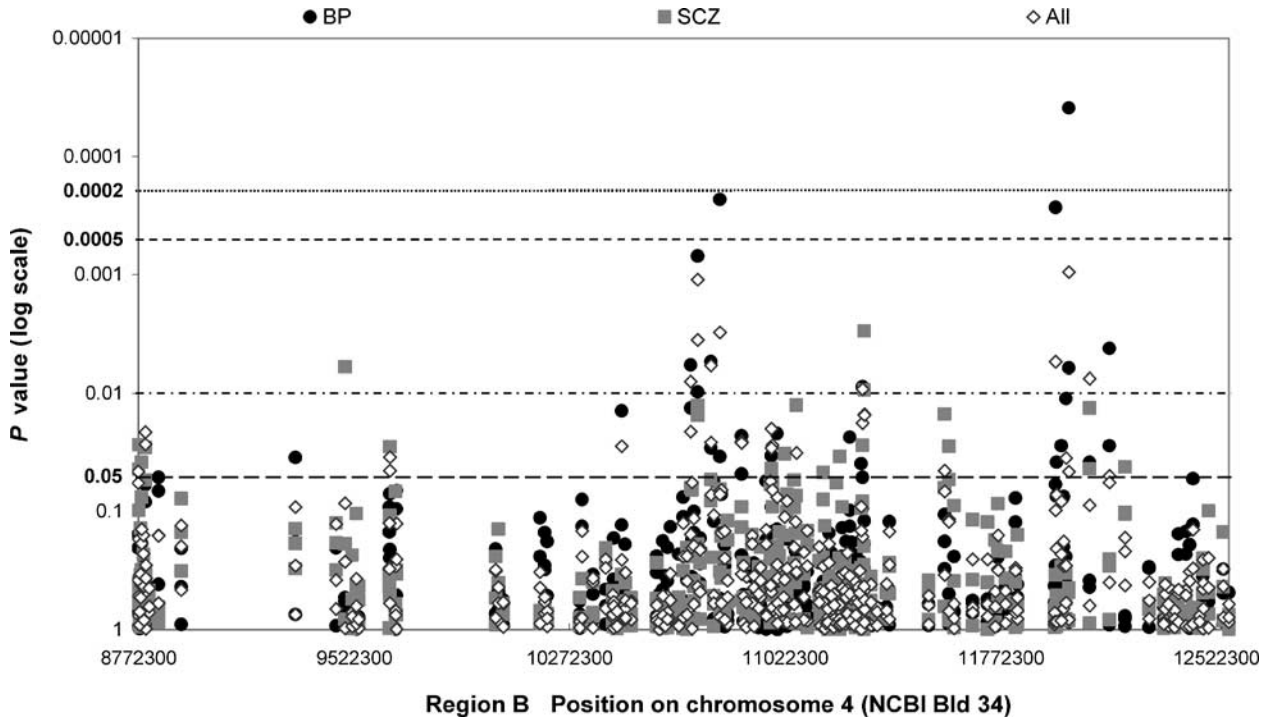

Figure 2 Region B: distribution of single-marker allele *P*-values. The *P*-values are plotted against the position of the SNP tested along chromosome 4 (NCBI Bld. 34) and represents the results of the entire sample and the sample separated by diagnosis: BP (black circles), SCZ (gray squares) and all cases (All, white diamonds). Four thresholds are indicated with dashed horizontal lines, starting from the top down: $P = 0.0002$ (experiment-wide Nyholt-corrected threshold), $P = 0.0005$ (region-wide Nyholt-corrected threshold), $P = 0.01$ and $P = 0.05$.

significant after permutation analysis ($P_p = 0.036$) (Table 3, top). The G allele at SNP 56 conferred an increase in OR of 1.8 (95% CI: 1.3–2.6), and the genotypic analysis revealed that either one or two copies of allele G at SNP 56 significantly increase the OR for BP in males to the same extent (OR = 1.7, 95% CI: 1.1–2.7; OR = 3.9, 95% CI: 1.6–9.6, respectively), suggesting a dominant pattern of inheritance. However, the significance under the dominant model did not improve ($P_{dom} = 0.0013$ versus genotype, $P = 0.0011$, Table 2).

Three haplotypes within B-1 had global *P*-values that met the region-wide Nyholt-corrected threshold ($P \leqslant 0.0005$) in the BP males, the most significant of which was the 3-SNP haplotype from SNP 51 to 53 ($P_g = 0.000073$) (Table 3, top). It remained significant after permutation correction ($P_{gp} = 0.012$). The 3-SNP individual haplotype, G-C-T (SNP 51–53) had the most significant individual *P*-value ($P_i = 0.000066$) and it showed a trend towards significance after permutation correction ($P_{ip} = 0.060$). Aligning the local overlapping individual haplotypes that met the region-wide threshold and that also conferred an increase in susceptibility to BP in the males (maximum OR = 2.8, 95% CI: 1.7–4.4; overall 95% CI: 1.2–4.4) identified an underlying 7-SNP extended haplotype (G-C-T-G-T-G-C, SNP 51–57, not tested).

The region of association for B-1 – delineated by following each of the significant individual haplotypes in both directions along the sliding windows (while $P_i \leqslant 0.01$) – was 202 kb. No *Known*/*RefSeq* genes or AceView-predicted genes overlapped with cluster B-1 (UCSC May 2004 Browser). The nearest gene, *mast cell immunoreceptor signal transducer* (*MIST*), is located 239kb telomeric of B-1 (Table 4, top).

Association to cluster B-4 was identified in BP females. B-4 spans from SNP 126 to 134 and contains SNPs 126 (rs1982655) and 131 (rs1454883), which met the region-wide Nyholt-corrected threshold ($P = 0.00027$ and 0.000039, respectively) (Table 3, bottom). Both SNPs remained significant after permutation analysis at the region-wide level ($P_p = 0.038$ and 0.0066, respectively). For SNP 126, the C allele conferred an increase in risk (OR = 3.0; 95% CI: 1.7–5.5) and the genotype analysis revealed that risk for BP in females only increases with two copies of allele C (OR = 3.0, 95% CI: 1.7–5.5), suggesting a recessive pattern of inheritance, which is further supported by a stronger association ($P_{rec} = 0.00010$ versus $P = 0.00027$, Table 2). For SNP 131, which was the only individual marker that met the experiment-wide Nyholt-corrected threshold ($P \leqslant 0.0002$) and which also remained significant after permutation analysis at that experiment-wide level ($P_p = 0.016$), the T allele conferred an increase in risk (OR = 3.2; 95% CI: 1.8–5.7). Genotype analysis supported a codominant mode of inheritance for SNP 131 (genotype $P = 0.00019$, Table 2).

Six haplotypes within B-4 had global *P*-values that met the region-wide Nyholt-corrected threshold (Table 3, bottom). The most significant of these

Table 2 Allele and genotype analysis of significant SNPs in region B

SNP project # (rs#)	*Diagnosis*	*Sex*	*Allele1*[a] *(%)*	*Allele 2*[a] *(%)*	P-*value* (χ_1^2) *(permuted P-value, s.e.)*	$OR_{2/1}$ *(95% CI)*	*Genotype 11 (%)*	*Genotype 12 (%)*	*Genotype 22 (%)*	*Total*[b]	P-*value* (χ_2^2)	$OR_{22/11}$ *(95% CI)*
56 (rs6820330)	Controls	M	377 (80)	93 (20)	0.00023		150 (64)	77 (33)	8 (3)	235		
	BP	M	220 (69)	100 (31)	(0.036, 0.002)	1.8 (1.3–2.6)	76 (48)	68 (43)	16 (10)	160	0.0011	3.9 (1.6–9.6)
126 (rs1982655)	Controls	F	208 (48)	224 (52)	0.00027		40 (19)	128 (59)	48 (22)	216		
	BP	F	251 (61)	163 (39)	(0.038, 0.002)	0.60 (0.46–0.79)	73 (35)	105 (51)	29 (14)	207	0.00027	0.33 (0.18–0.60)
131 (rs1454883)	Controls	F	253 (59)	179 (41)	0.000039		72 (33)	109 (51)	35 (16)	216		
	BP	F	182 (44)	228 (56)	(0.0066, 0.001) (0.016, 0.001)[c]	1.8 (1.3–2.3)	40 (20)	102 (50)	63 (31)	205	0.00019	3.2 (1.8–5.7)

Abbreviations: BP, bipolar disorder; CI, confidence interval; F, females; M, males; OR, odds ratio; s.e., standard error; SNP, single nucleotide polymorphism.
Underlined *P*-value indicates *P*-value that met the region-wide Nyholt-corrected threshold ($P \leqslant 0.0005$).
[a]Alleles 1 and 2 are specified for each SNP in SOM Table S1.
[b]Refers to total number of genotypes available for analysis in that sample subgroup.
[c]Permuted *P*-value and standard error for this SNP was also determined at the experiment-wide level (10 000 shuffles).

was the 2-SNP haplotype from SNP 130 to 131 ($P_g = 0.00013$) in the BP females, which remained significant after permutation analysis at the region-wide level ($P_{gp} = 0.016$). The 4-SNP individual haplotype, T-T-T-G (SNP 131–134) had the most significant individual *P*-value ($P_i = 0.000014$, not corrected; OR = 3.4, 95%CI: 2.1–5.6). Aligning the local overlapping individual haplotypes that met the region-wide threshold revealed two underlying haplotypes: T-T-A-G-T-C (SNP 126–131, not tested) and T-T-T-G (SNP 131–134). The individual overlapping haplotypes that make up the first of these, which have a 'C' allele at SNP 131, all confer a decrease in susceptibility to, or a protective effect against, BP in females (minimum OR = 0.6, 95%CI: 0.4–0.7; overall 95%CI: 0.4–0.8). Whereas, the haplotypes making up the latter, which carry a 'T' allele at SNP 131, all confer an increase in susceptibility to BP in females (maximum OR = 3.4, 95%CI: 2.1–5.6; overall 95% CI: 1.5–5.6).

The individual haplotype T-T-T-G (SNP 131–134) met the region- and experiment-wide Nyholt thresholds in the full sample of affected females ($P_i = 0.00019$), but was less significant than in the BP females ($P_i = 0.000014$) (SOM Supplementary Table 3). The region of nominal significance ($P_i \leqslant 0.01$) for the individual haplotypes in the BP females stretched over 603 kb. Again, current annotation of this region fails to identify *Known* or *RefSeq* genes, but there are two main class Aceview gene predictions, one of which falls between the two significant individual SNPs, 126 and 131. *Heparan sulfate (glucosamine) 3-O-sulfotransferase 1* (*HS3ST1*) is the nearest known gene, located 577 kb telomeric of B-4 (Table 4, top).

Region D

The distribution of single-marker allele (χ_1^2) *P*-values for region D is shown in Figure 4: 5.5% of the *P*-values are less than or equal to 0.05 and 1.0% are less than or equal to 0.01, suggesting that there is no overall inflation of Type I error in the data set. No individual markers were nominally significant at $P \leqslant 0.0003$. However, the sliding-window haplotype analysis identified 35 haplotypes that met either the region- (10 haplotypes; $P \leqslant 0.0003$) or experiment-wide (25 haplotypes; $P \leqslant 0.0002$) Nyholt-corrected threshold at the global and/or individual haplotype test level (SOM Supplementary Table 5). These haplotypes fall into nine nonoverlapping regions (D-1–D-9), only two of which (D-2 and D-7) contained haplotypes with nominally significant global *P*-values (Figure 5, Table 5). They are described in more detail below. The other regions, five of which overlap with known genes (SOM Supplementary Table 4, bottom), consist of one or two haplotypes with only significant individual *P*-values (range $P_i = 0.00011$–0.00030). They are described further in the SOM (SOM Results and Discussion and SOM Supplementary Tables 4 and 5).

Cluster D-2 spans from SNP 48 to 58, with the strongest associations emerging in the full sample of all affected females at both the global (SNP 50–52, $P_g = 0.000019$) and individual (T-A-G, SNP 50–52,

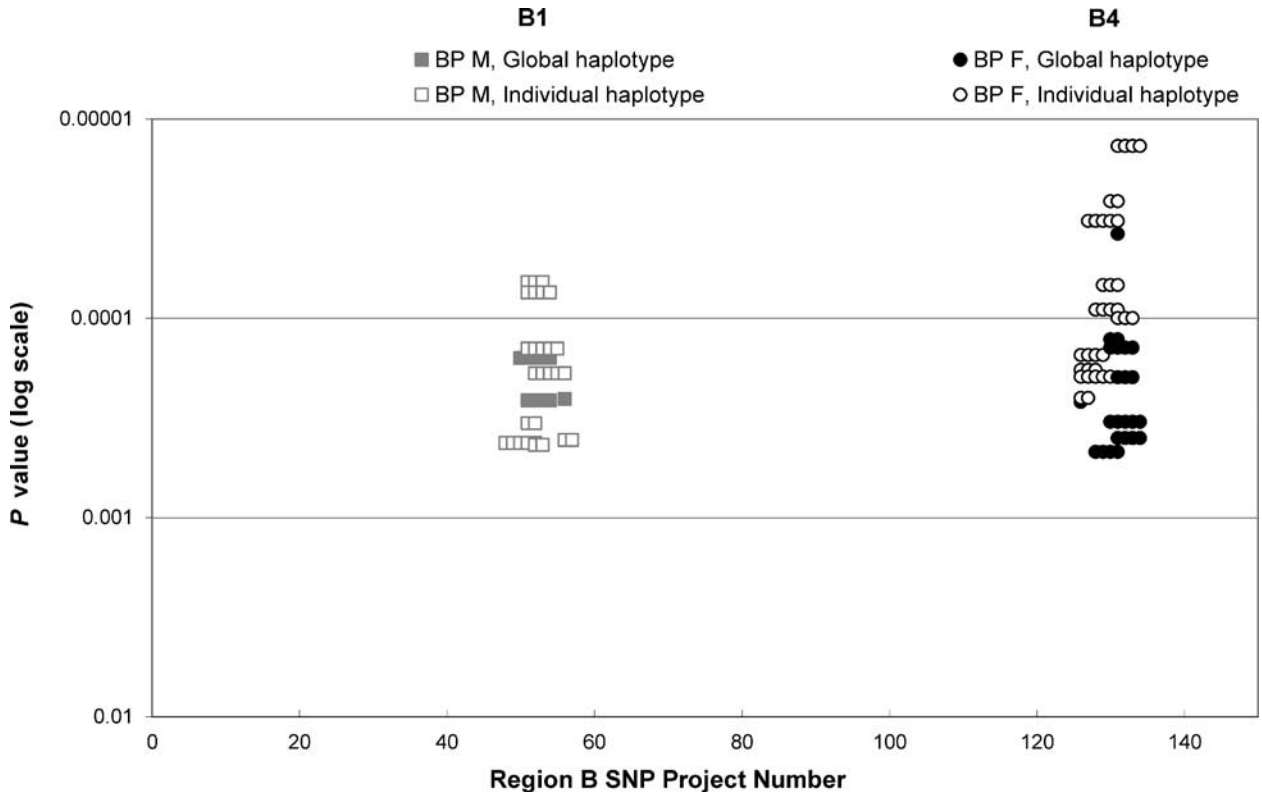

Figure 3 Region B: single markers and global and individual haplotypes making up clusters B-1 and B-4. The *P*-values of the significant markers and haplotypes ($P \leqslant 0.0005$) in region B are plotted against the SNP project number.

$P_i = 0.0000033$, OR = 3.2, 95% CI: 2.0–5.1) levels (Table 5, top). Both the 3-SNP global and 3-SNP individual haplotypes remained significant after permutation-based correction ($P_{gp} = 0.0080$ and $P_{ip} = 0.0040$, respectively). Aligning all the local overlapping individual haplotypes that met the region-wide Nyholt-corrected threshold, and which confer an increase to susceptibility to BP or SCZ in women (maximum OR = 3.2, 95% CI: 2.0–5.1; overall 95% CI range: 1.4–5.1), revealed an underlying 11-SNP haplotype (A-A-T-A-G-T-A-C-A-G-T, SNP 48–58, not tested).

The two 3-SNP global and individual haplotypes described above were also the most significant in the sample of BP females ($P_g = 0.00011$ and $P_i = 0.0000047$, respectively) and in the full sample of BP individuals ($P_g = 0.00030$ and $P_i = 0.00012$, respectively). Significance was maintained after permutation-based correction in the sample of BP females ($P_{gp} = 0.033$ and $P_{ip} = 0.0080$, respectively), but not in the full sample of BP individuals ($P_{gp} = 0.075$ and $P_{ip} = 0.16$, respectively). However, three haplotypes, spanning SNPs 53–57 along the 11-SNP underlying haplotype described above, had nominally significant individual, but not global, *P*-values in the SCZ females and in the full sample of affected females ($P \leqslant 0.0003$). One of these haplotypes (T-A-C-A, SNP 53–56) showed a stronger association in the SCZ females ($P_i = 0.000032$) than in the full sample of affected females ($P_i = 0.000045$) (SOM Supplementary Table 5). The 201 kb region of association contains one main AceView-predicted gene. The nearest known gene to D-2 is the *peroxisome proliferator-activated receptor gamma coactivator 1 alpha* (*PPARGC1A*) gene, located just 25 kb telomeric of D-2 (Table 4, bottom).

Cluster D-7 spans from SNP 214–218 (Table 5, bottom). The most significant global *P*-value was seen in the full sample of SCZ individuals with a 5-SNP haplotype extending from SNP 214 to 218 ($P_g = 0.000070$), which remained significant after permutation analysis ($P_{gp} = 0.033$). The same haplotype was also nominally significant in the SCZ males ($P_g = 0.00018$) and it showed a trend towards significance after permutation correction ($P_{gp} = 0.088$). The most significant individual haplotype was the 5-SNP haplotype, A-G-T-C-C, in the SCZ males (uncorrected $P_i = 0.000079$, OR = 0.1, 95% CI: 0.0–0.3). The 121 kb-associated region contains two main class Aceview gene predictions. The nearest known gene to this associated region is the hypothetical gene *KIAA0746*, which is located just 17 kb telomeric of the region (Table 4, bottom).

Discussion

We have performed an association study across the two priority regions that together span 8.3 Mb of the 20 Mb chromosome 4p15–p16 linkage region for BP and related phenotypes. This is the first report of an association study across these two candidate regions and we have identified a total of 13 subregions – four in region B and nine in region D – that are associated with BP and/or SCZ in our Scottish cohort. Four of

Table 3 Region B clusters supported by significant global haplotypes

SNP no.	*48*	*49*	*50*	*51*	*52*	*53*	*54*	*55*	*56*	*57*	*Case frequency*	*Control frequency*	P_i (P_{ip}, *s.e.*)	P_g (P_{gp}, *s.e.*)	*OR (95% CI)*
Block no.	*30*	*30*	*31*	*32*	*32*	*33*	*33*	*34*	*34*	*34*					
B-1															
BP M	C	A	A	G	C						0.13	0.05	0.00042	0.0018	3.2 (1.9–5.6)
BP M			**50**	**51**	**52**	**53**	**54**				NA – global	NA – global	NS	**0.00016 (0.059, 0.007)**	NA – global
BP M				G	C						0.25	0.14	0.00034	0.0017	2.2 (1.5–3.2)
BP M				**G**	**C**	**T**					0.19	0.08	0.000066 (0.060, 0.008)	**0.000073 (0.012, 0.003)**	2.6 (1.6–4.1)
BP M				**G**	**C**	**T**	**G**				0.18	0.08	0.000075	**0.00026 (0.048, 0.007)**	2.8 (1.7–4.4)
BP M				G	C	T	G	T			0.18	0.08	0.00014	0.0006558	2.7 (1.7–4.4)
BP M					C	T					0.17	0.08	0.00043	0.0018	2.4 (1.5–3.8)
BP M					C	T	G	T	G		0.17	0.08	0.00019	0.0010	2.5 (1.6–4.1)
BP M									**G**		0.31	0.20	NA – SM	**0.00023 (0.036, 0.002)**	1.8 (1.3–2.6)
BP M									G	C	0.31	0.20	0.00041	0.0014	1.7 (1.2–2.4)
SNP no.	*126*	*127*	*128*	*129*	*130*	*131*	*132*	*133*	*134*		*Case frequency*	*Control frequency*	P_i (P_{ip}, *s.e.*)	P_g (P_{gp}, *s.e.*)	*OR (95% CI)*
Block no.	*71*	*71*	*72*	*72*	*72*	*72*	*73*	*74*	*75*						
B-4															
BP F	**T**										0.39	0.52	NA – SM	**0.00027 (0.038, 0.002)**	0.6 (0.5–0.8)
BP F	T	T									0.40	0.52	0.00025	0.0008407	0.6 (0.5–0.8)
BP F	T	T	A								0.38	0.50	0.00018	0.0044	0.6 (0.4–0.8)
BP F	T	T	A	G							0.37	0.50	0.00015	0.0073	0.6 (0.4–0.8)
BP F	T	T	A	G	T						0.38	0.50	0.00020	0.0059	0.6 (0.4–0.8)
BP F		T	A	G	T	C					0.35	0.50	0.000033	0.0008618	0.6 (0.4–0.7)
BP F			**A**	**G**	**T**	**C**					0.44	0.57	0.000091	**0.00047**[a]	0.6 (0.4–0.8)
BP F				G	T	C					0.44	0.58	0.000069	0.00056	0.6 (0.4–0.8)
BP F					**T**	**C**					0.44	0.59	0.000026 (0.012, 0.001)	**0.00013 (0.016, 0.001)**	0.6 (0.4–0.7)
BP F					**130**	**131**	**132**	**133**			NA – global	NA – global	NS	**0.00014 (0.028, 0.002)**	NA – global
BP F					**130**	**131**	**132**	**133**	**134**		NA – global	NA – global	NS	**0.00033 (0.069, 0.008)**	NA – global
BP F						**T**					0.56	0.41	NA – SM	**0.000039 (0.0066, 0.001)**	1.8 (1.3–2.3)
BP F						**T**	**T**	**T**			0.30	0.20	0.00010[a]	**0.00020 (0.028, 0.002)**	2.4 (1.5–3.6)
BP F						**T**	**T**	**T**	**G**		0.24	0.11	0.000014	**0.00040**[a]	3.4 (2.1–5.6)

Abbreviations: BP, bipolar disorder; CI, confidence interval; F, females; M, males; P_g, global haplotype test *P*-value; P_{gp}, permuted P_g; P_i, individual haplotype test *P*-value; P_{ip}, permuted P_i; NA–global, not applicable because case–control frequencies and ORs do not apply to global haplotypes; NA–SM, not applicable because the allele *P*-value of a single marker is the same at the global and individual level; NS, no corresponding individual haplotypes were **s**ignificant at the region-wide level; OR, odds ratio; SNP, single nucleotide polymorphism.
Bold values indicate haplotypes that were significant at the global level and their P_g-value; bold and underlined values highlight the significant single markers and their *P*-value.
[a]This haplotype *P*-value was not corrected by Cocaphase because it was not the most significant in the particular sliding-window size and subgroup tested.

Table 4 Details of the four clusters with significant global haplotypes and the corresponding delineated associated regions

Cluster (Blocks)[a]	*Associated region (Blocks)*[a]	*UCSC coordinates (July 2003*[b]*,* NCBI Bld 34) of associated region	*Size (kb)*	*Known/RefSeq genes in associated region*	*If no Known/RefSeq genes, then nearest Known/RefSeq gene (distance away: '−' if telomeric; '+' if centromeric)*	*Number of AceView main class gene predictions*
Region D						
B-1 SNPs 48–57 (30–34)	SNPs 48–61 (30–36)	chr4:10675941–10878006	202	None	*MIST* (−239 kb)	None
B-4 SNPs 126–134 (69–75)	SNPs 121–138 (68–76)	chr4:11758342–12361604	603	None	*HS3ST1* (−577 kb)	2
Region D						
D-2 SNPs 48–58 (30–34)	SNPs 46–60 (29–35)	chr4:23317763–23519188	201	None	*PPARGC1A* (+25 kb)	1
D-7 SNPs 214–218 (118–121)	SNPs 211–221 (118–122)	chr4:25632760–25753356	121	None	*KIAA0746* (−17kb)	2

Abbreviations: SNPs, single nucleotide polymorphisms.
[a]Block numbers correspond to the LD block construction based on HapMap May 2004 data.
[b]Refers to UCSC Genome Browser Release July 03 (http://genome.ucsc.edu/cgi-bin/hgGateway?clade=vertebrate&org=Human&db=hg16&hgsid=75196110).

these 13 regions are supported by significant SNPs and/or at least one global haplotype that remained significant after permutation-based correction for multiple testing, as did a number of the corresponding individual haplotypes ($P_p \leqslant 0.05$).

Regions of association supported by significant common individual haplotypes (frequency > 5%), but not significant global haplotype *P*-values, were also considered because such haplotypes have previously been shown to represent true associations.[37] Similarly, regions defined by significant rare individual haplotypes were considered because the association of rare variants, including haplotypes, has previously led to the identification of susceptibility loci for complex diseases.[38,39] These associated regions and the genes that they contain are further discussed in the SOM Results and Discussion.

To facilitate further studies, we delineated the region of association identified by the four clusters containing the globally significant haplotypes described above. No *Known* or *RefSeq* genes were identified in the 1.1 Mb of genomic DNA covered by the four regions of association (B-1, B-4, D-2 and D-7). However, there are five predicted genes in these regions. Four of the five are weak candidates, while the fifth predicted gene, *LOC389203*, (found in associated region D-7) is an interesting candidate, encoding a 105-amino-acid protein. It is defined by 153 cDNA clones and shows significant cross-species homology (not shown).

The known genes *KIAA0746*, *PPARGC1A*, *MIST* and *HS3ST1* map 17, 25, 239 and 577 kb, respectively, from a region of association, suggesting the possibility that the functional variant is disrupting a regulatory element located in noncoding sequence, as has been seen in Hirschsprung disease,[40] cardiac repolarization[41] and age-related macular degeneration,[42] for example. As regulatory regions are known to exist as far as 1 Mb up- or downstream of the transcription unit,[43] any of the four associated regions may contain a putative regulatory element. Indeed, there are numerous conserved elements (not shown) and, therefore, putative regulatory or otherwise functional regions within each of these associated regions. The nature of psychiatric illness, namely its onset in adulthood and relatively subtle nature of the phenotypic abnormalities, and its apparent polygenic nature may very well point to the involvement of regulatory, rather than coding, polymorphisms. The candidacy of these nearby genes as susceptibility genes for psychiatric illness is discussed in the SOM Results and Discussion.

Previously, we reported a candidate gene association study where we demonstrated preliminary evidence for association with the (region B-located) *G protein-couple receptor 78* (*GPR78*) gene in this sample.[12] Only one SNP, rs1282 (region B SNP 1), was in common between the two studies because of the differences in the SNP selection methods and resources between them. In this study, SNP1 and haplotypes spanning the GPR78 genomic region met the traditional nominal significance threshold of

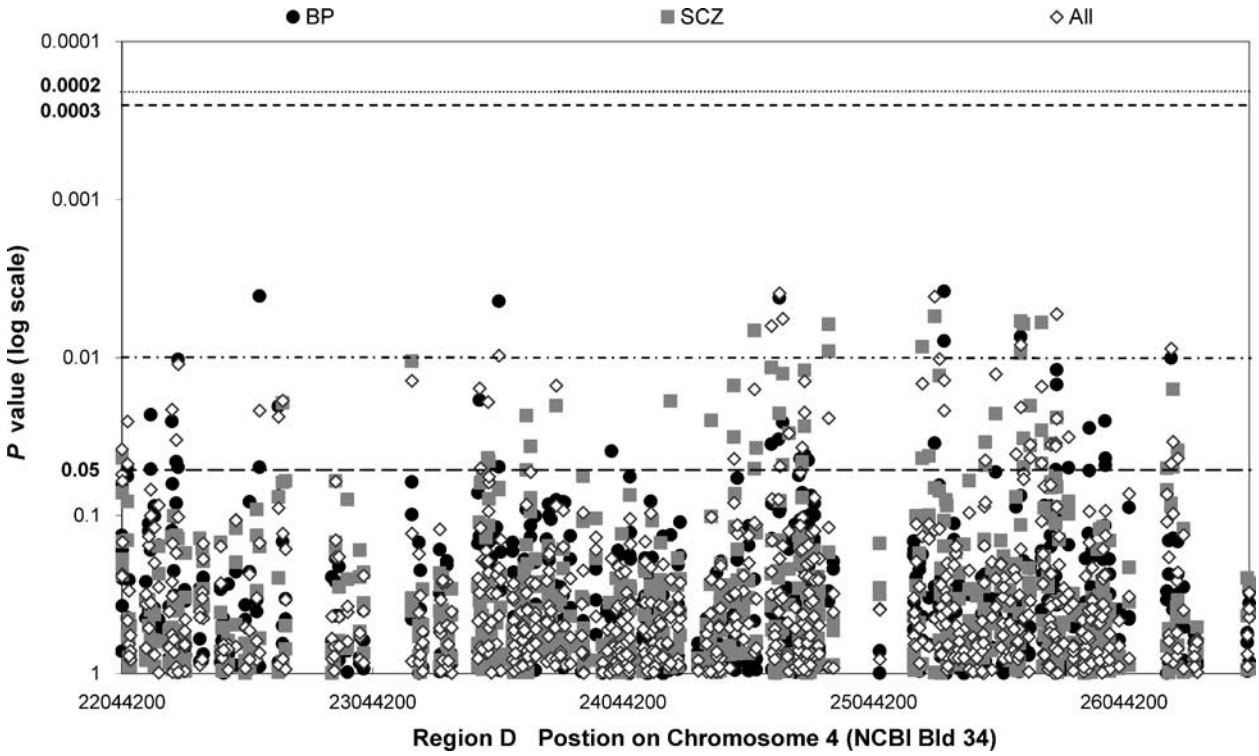

Figure 4 Region D: distribution of single-marker allele *P*-values. The *P*-values are plotted against the position of the SNP tested along chromosome 4 (NCBI Bld. 34) and represents the results of the entire sample and the sample separated by diagnosis: BP (black circles), SCZ (gray squares) and all cases (All, white diamonds). Four thresholds are indicated with dashed horizontal lines, starting from the top down: $P=0.0002$ (experiment-wide Nyholt-corrected threshold), $P=0.0003$ (region-wide Nyholt-corrected threshold), $P=0.01$ and $P=0.05$.

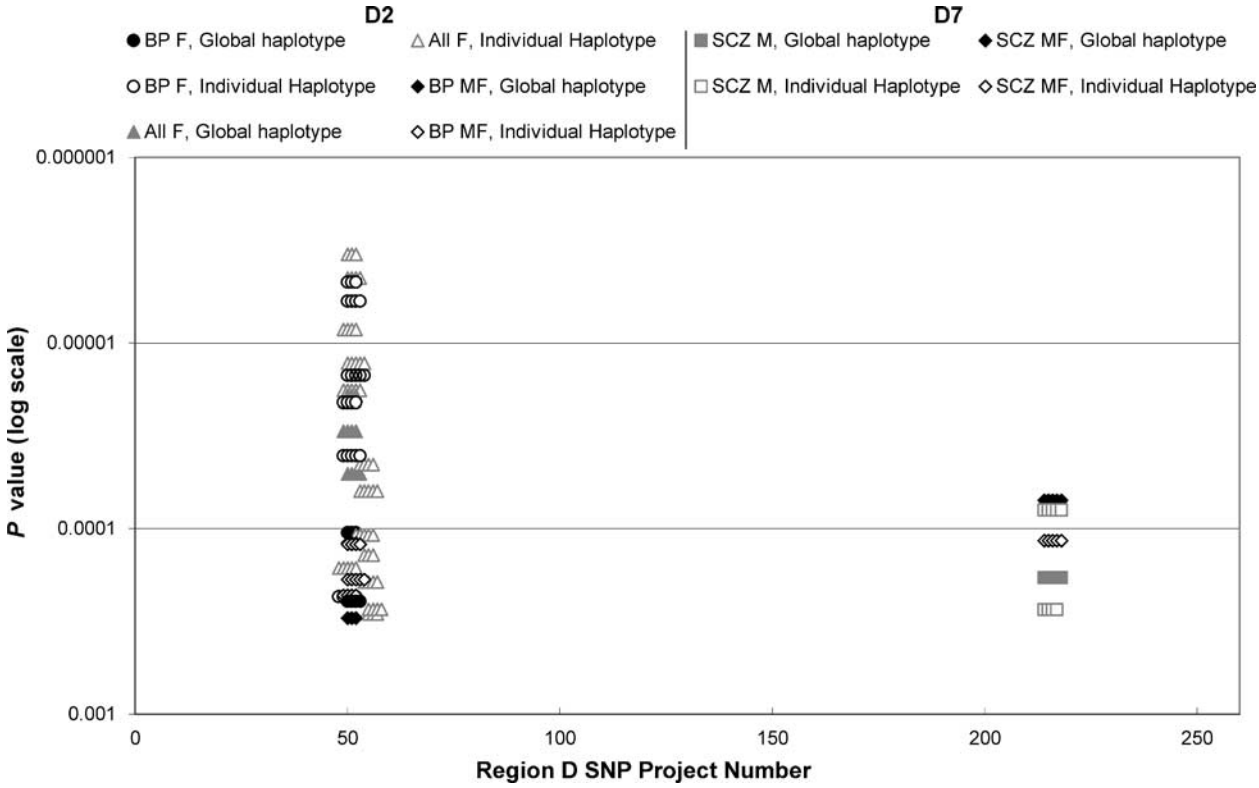

Figure 5 Region D: global and individual haplotypes making up clusters D-2 and D-7. The *P*-values of the significant haplotypes ($P \leqslant 0.0003$) in region D are plotted against the SNP project number.

Table 5 Region D clusters supported by significant global haplotypes

SNP no.	*48*	*49*	*50*	*51*	*52*	*53*	*54*	*55*	*56*	*57*	*58*	*Case frequency*	*Control frequency*	P_i (P_{ip}, *s.e.*)	P_g (P_{gp}, *s.e.*)	*OR (95% CI)*
Block no.	*30*	*30*	*30*	*31*	*31*	*31*	*32*	*32*	*32*	*33*	*34*					
D-2																
All F	A	A	T	A	G							0.12	0.05	0.00016	0.00087	2.7 (1.6–4.6)
All F		**A**	**T**	**A**	**G**							0.14	0.05	0.0000085	**0.000030 (0.017, 0.004)**	3.1 (1.9–5.1)
All F		A	T	A	G	T						0.13	0.05	0.000018	0.00054	3.0 (1.8–5.0)
All F			**T**	**A**	**G**							0.15	0.06	0.0000033 (0.0040, 0.002)	**0.000019 (0.0080, 0.003)**	3.2 (2.0–5.1)
All F			**T**	**A**	**G**	**T**						0.15	0.06	0.0000045	**0.000050**[a]	3.0 (1.8–4.8)
All F			T	A	G	T	A					0.14	0.06	0.000013	0.00034	2.9 (1.7–4.7)
All F					G	T	A	C	A			0.11	0.04	0.00011	0.0023	2.8 (1.6–5.0)
All F						T	A	C	A			0.18	0.09	0.000045	0.0035	2.2 (1.5–3.4)
All F						T	A	C	A	G		0.18	0.09	0.000063	0.0042	2.2 (1.4–3.5)
All F							A	C	A			0.18	0.09	0.00014	0.0060	2.2 (1.4–3.3)
All F							A	C	A	G		0.17	0.09	0.00019	0.011	2.2 (1.4–3.4)
All F								C	A	G		0.18	0.09	0.00029	0.0095	2.2 (1.5–3.4)
All F								C	A	G	T	0.16	0.08	0.00027	0.020	2.2 (1.4–3.4)
BP F	A	A	T	A	G							0.12	0.05	0.00023	0.0033	2.9 (1.6–5.1)
BP F		A	T	A	G							0.14	0.05	0.000021	0.0003419	3.2 (1.9–5.5)
BP F		A	T	A	G	T						0.14	0.05	0.000040	0.0017	2.9 (1.7–5.0)
BP F			**T**	**A**	**G**							0.16	0.06	0.0000047 (0.0080, 0.003)	**0.00011 (0.033, 0.006)**	3.2 (2.0–5.3)
BP F			**T**	**A**	**G**	**T**						0.16	0.06	0.0000060	**0.00025 (0.076, 0.008)**	3.0 (1.8–5.0)
BP F			T	A	G	T	A					0.15	0.06	0.000015	0.0011	2.9 (1.7–4.9)
BP MF		A	T	A	G							0.13	0.08	0.00023	0.0016	1.9 (1.3–2.7)
BP MF			**T**	**A**	**G**							0.15	0.08	0.00012 (0.16, 0.012)	**0.00030 (0.075, 0.008)**	1.9 (1.4–2.6)
BP MF			T	A	G	T						0.15	0.08	0.00012	0.0016	1.9 (1.3–2.7)
BP MF			T	A	G	T	A					0.14	0.08	0.00019	0.025	1.9 (1.3–2.6)
SNP no.	*214*	*215*	*216*	*217*	*218*							*Case frequency*	*Control frequency*	P_i (P_{ip}, *s.e.*)	P_g (P_{gp}, *s.e.*)	*OR (95% CI)*
Block no.	*118*	*119*	*120*	*120*	*121*											
D-7																
SCZ M	A	G	T	C								0.06	0.14	0.00027	0.00046	0.4 (0.2–0.6)
SCZ M	**A**	**G**	**T**	**C**	**C**							0.01	0.06	0.000079	**0.00018 (0.088, 0.009)**	0.1 (0.0–0.3)
SCZ MF	**A**	**G**	**T**	**C**	**C**							0.01	0.06	0.00012	**0.000070 (0.033, 0.006)**	0.2 (0.1–0.5)

Abbreviations: BP, bipolar disorder; CI, confidence interval; F, females; M, males; P_g, global haplotype test *P*-value; P_{gp}, permuted P_g; P_i, individual haplotype test *P*-value; P_{ip}, permuted P_i; OR, odds ratio; SCZ, schizophrenia; SNP, single nucleotide polymorphism.
Bold values indicate haplotypes that were significant at the global level and their P_g-value.
[a]This haplotype *P*-value was not corrected by Cocaphase because it was not the most significant in the particular sliding-window size and subgroup tested.

$P \leqslant 0.05$, but not the region-wide Nyholt-corrected threshold (region B, $P \leqslant 0.0002$) set to account for the multiple markers tested in this study.

To perform this large-scale association study, we carried out careful selection of the case sample and included only definite cases of BP or SCZ, which were recruited from the general Scottish population. The control sample was taken from the same population as the cases; however, stratification may still exist.[44] Thus, a preliminary analysis of stratification using 66 SNPs[45] and the population stratification program Structure[46] was performed. This analysis showed no evidence of population stratification in our sample (Thomson *et al.*, unpublished). Also, the distribution of the allele *P*-values for regions B and D (Figures 2 and 3), which suggests that there is no overall inflation of Type I error, supports the evidence for no stratification in the sample.

We used a systematic approach to SNP selection that aimed to cover the two linked subregions (B and D) as thoroughly as possible. Our selection method relied on the construction of LD blocks, which allowed us to select htSNPs that efficiently captured a predetermined percentage of the known variation within the block[47] and some inestimable portion of the unidentified variation. Given that the definition of a block is a much debated area,[48,49] it is important to point out that our subsequent method of analysis was not dependent on the block structure being wholly accurate or complete. The primary purpose of the blocks was to facilitate the selection of the htSNPs, whose power in an association study is based on their overall ability to represent the genetic variation in a predefined area.[50] Selecting SNPs on a block-by-block basis also offers a more thorough coverage of the region;[51] is more likely to provide positional information on an associated region than a block-free approach[48,50,52] and provides coverage that is more robust to marker failure owing to the inherent redundancy. The SNP selection was carried out using Release 7 (May 2004) of the HapMap Project and therefore improved coverage of the linked candidate regions would now be possible.[53]

We carried out both single-marker and haplotype-based analysis of the data because haplotype-based analysis can be more powerful in some cases.[54] Specifically, as the frequency of the causative variant is unknown, it is more likely to be matched, and possibly detected, when a range of haplotype frequencies is tested.[55,56] This, however, results in an increase in the number of (highly-correlated) tests performed and in an additional multiple-testing concern, for which there is no straightforward solution. The 'gold standard' for multiple-testing correction is permutation analysis because it determines the empirical distribution and significance threshold of the data, while accounting for its correlation structure. Permutation analysis, however, is computationally intensive and extremely time-consuming, particularly in its application to multiple, large haplotypes. Thus, we employed a two-step procedure that enabled us first to mine the association data for the most significant markers and haplotypes, based on a significance (Nyholt-corrected) threshold preadjusted for the number of SNP tested. Then, having validated this threshold at the single-marker level with permutation analysis, we ranked the significant haplotypes based on their level of support (global versus individual) and performed permutation-based corrections on the most promising haplotype candidates – that is, those with support at the global level. Each of the four high-ranking associated regions (B-1, B-4, D-2 and D-7) was supported by global and, in many cases, individual haplotypes that withstood permutation testing ($P_p \leqslant 0.05$), providing more reliable support for these regions and suggesting that the region-wide Nyholt-corrected thresholds were appropriate cutoffs for the haplotype tests as well.

We did not adjust our nominal significance threshold to take into account the extra tests that were performed as a result of pooling the cases and splitting the sample by gender, because adjusting the *P*-value, using a Bonferroni-type correction, is overly conservative given the nonindependence and the relevance and importance of these tests, as discussed below. Furthermore, our findings should be taken in the context of the prior probability conferred on the chromosome 4p linkage region, particularly in regions B and D, through the multiple, positive linkage results.[57,58]

In this study, we have identified several significant regions, some showing association to BP (B-1 and B-4), others to SCZ (D-7) and still others to BP and SCZ together (D-2). These findings reflect what has already been observed in family-based studies for the 4p locus, where linkage has been shown to BP and/or SCZ and related illnesses.[5,8,11] This phenomenon has also been observed in a proportion of other susceptibility loci for psychiatric illness,[59–61] suggesting that these illnesses may, in some cases, share genetic predisposition across diagnostic boundaries. Family studies and the overlap in the clinical presentation between BP and SCZ also support this hypothesis.[60,62–64] It is also possible that regions B and D each harbor separate loci (one or more in each) for BP and SCZ, respectively.

It is also noteworthy that some of the associations are sex-specific, occurring only in males (e.g. B-1) or only in females (e.g. B-4). Although there is no difference in the lifetime risk for BP or SCZ between males and females, there is substantial evidence of gender differences in the onset, course and outcome of psychiatric illness.[65] For example, women with BP are more likely to experience rapid cycling between the depressed and manic states, mixed mania and antidepressant-induced mania.[66] Women are more likely to have a depressive episode before their first manic episode, while men tend to present with mania followed by depression.[67] Also, women have been found to be more likely than men to have a comorbid psychiatric or physical diagnosis, such as substance abuse or migraines.[66] Finally, there is also evidence of later age of onset of BP for women.[68] In SCZ, men show the first signs of the illness significantly earlier than women and have a poorer course.[69,70] In terms of treatment, differences between men and women

in both clinical and pharmacokinetic responses to psychotropic drugs have been reported.[71] Both BP and SCZ are heterogeneous phenotypes; thus, these illness-specific gender differences suggest the possibility of a genetically more homogenous subgroup in the male- or female-specific BP and/or SCZ samples, potentially increasing the power to detect association.[72] Indeed, experimental[40] and empirical evidence has indicated that modelling for sex-specific architecture may increase the power to identify the underlying genetic variants, particularly in the light of the recent findings that the expression patterns of hundreds to thousands of genes in mice are sexually dimorphic. Finally, sex-specific associations have been identified in the analysis of other candidate genes in psychiatric illness.[17,60,73]

Finally, Type I error is a concern in any association study. Therefore, an important target for future genetic studies should be replicating these findings in independent samples, carrying out functional studies on the genes identified and homing in on the functional variant.

This large-scale association study covers two priority regions totaling 8.3 Mb of the 20 Mb candidate linkage region on chromosome 4p. It has identified four significant regions of association that include predicted gene transcripts and potential regulatory regions for further investigation.

Acknowledgments

This work was supported from grants from the Chief Scientists Office, Scottish Executive; the Health Foundation, London; the Medical Research Council, UK and the Wellcome Trust. We thank the patients, their families and volunteers for their participation in this study. We thank Maura Walker, Margaret van Beck, Sally Roe and Susan Jackson for the collation of patient data. We thank Kirsty Millar, Helen Torrance, Susan Anderson, Alison Condie, John Beekman, Pat Malloy, Alan MacLean, Rosalind Launchbury, Sebastienne Buchanan and the Wellcome Trust CRF Genetics Core for their help in the preparation of the samples. We thank Illumina, San Diego, for genotyping our samples. We thank Richard Adams for bioinformatics advice and Simon T Cooper for assistance in the preparation of the paper. Albert Tenesa acknowledges funding from Cancer Research UK.

References

1 Lopez AD, Murray CC. The global burden of disease, 1990–2020. *Nat Med* 1998; **4**: 1241–1243.
2 Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S *et al.* Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 2002; **71**: 877–892.
3 Thomson PA, Christoforou A, Morris SW, Adie E, Pickard BS, Porteous DJ *et al.* Association of neuregulin 1 with schizophrenia and bipolar disorder in a second cohort from the Scottish population. *Mol Psychiatry* 2007; **12**: 94–104.
4 Hennah W, Thomson P, Peltonen L, Porteous D. Genes and schizophrenia: beyond schizophrenia: the role of DISC1 in major mental illness. *Schizophr Bull* 2006; **32**: 409–416.
5 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.
6 Visscher PM, Haley CS, Heath SC, Muir WJ, Blackwood DH. Detecting QTLs for uni- and bipolar disorder using a variance component method. *Psychiatr Genet* 1999; **9**: 75–84.
7 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 affective disorder. *Biol Psychiatry* 2007; **61**: 797–805.
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 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.
10 Itokawa M, Kasuga T, Yoshikawa T, Matsushita M. Identification of a male schizophrenic patient carrying a *de novo* balanced translocation, t(4; 13)(p16.1; q21.31). *Psychiatry Clin Neurosci* 2004; **58**: 333–337.
11 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.
12 Underwood SL, Christoforou A, Thomson PA, Wray NR, Tenesa A, Whittaker J *et al.* Association analysis of the chromosome 4p-located G protein-coupled receptor 78 (GPR78) gene in bipolar affective disorder and schizophrenia. *Mol Psychiatry* 2006; **11**: 384–394.
13 Edwards AO, Ritter III R, Abel KJ, Manning A, Panhuysen C, Farrer LA. Complement factor H polymorphism and age-related macular degeneration. *Science* 2005; **308**: 421–424.
14 Haines JL, Hauser MA, Schmidt S, Scott WK, Olson LM, Gallins P *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* 2005; **308**: 419–421.
15 Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* 2005; **308**: 385–389.
16 Rivera A, Fisher SA, Fritsche LG, Keilhauer CN, Lichtner P, Meitinger T *et al.* Hypothetical LOC387715 is a second major susceptibility gene for age-related macular degeneration, contributing independently of complement factor H to disease risk. *Hum Mol Genet* 2005; **14**: 3227–3236.
17 Hennah W, Varilo T, Kestila M, Paunio T, Arajarvi R, Haukka J *et al.* Haplotype transmission analysis provides evidence of association for DISC1 to schizophrenia and suggests sex-dependent effects. *Hum Mol Genet* 2003; **12**: 3151–3159.
18 Thomson PA, Wray NR, Millar JK, Evans KL, Hellard SL, Condie A *et al.* Association between the TRAX/DISC locus and both bipolar disorder and schizophrenia in the Scottish population. *Mol Psychiatry* 2005; **10**: 657–668, 616.
19 Consortium TIH. The international HapMap project. *Nature* 2003; **18**: 789–796.
20 Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry* 1978; **35**: 837–844.
21 Association AP. *Diagnostic and Statistical Manual of Mental Disorders (4th edn) (DSM–IV).* APA: Washington, DC, 1994.
22 Office of the Chief Statistician SE. *Analysis of Ethnicity in the 2001 Census – Summary Report.* Scottish Executive: Edinburgh, 2004.
23 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
24 Hedrick PW. Gametic disequilibrium measures: proceed with caution. *Genetics* 1987; **117**: 331–341.
25 Ke X, Durrant C, Morris AP, Hunt S, Bentley DR, Deloukas P *et al.* Efficiency and consistency of haplotype tagging of dense SNP maps in multiple samples. *Hum Mol Genet* 2004; **13**: 2557–2565.
26 Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004; **74**: 765–769.

27 Cheverud JM. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* 2001; **87**: 52–58.
28 Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* 2005; **95**: 221–227.
29 Sidak Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc* 1967; **62**: 626–633.
30 Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003; **25**: 115–121.
31 Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. *Genetics* 1994; **138**: 963–971.
32 Holm S. A simple sequentially rejective multiple test procedure. *Scand J Statist* 1979; **6**: 65–70.
33 Koeleman BP, Dudbridge F, Cordell HJ, Todd JA. Adaptation of the extended transmission/disequilibrium test to distinguish disease associations of multiple loci: the Conditional Extended Transmission/Disequilibrium Test. *Ann Hum Genet* 2000; **64**: 207–213.
34 Fallin MD, Lasseter VK, Avramopoulos D, Nicodemus KK, Wolyniec PS, McGrath JA *et al.* Bipolar I disorder and schizophrenia: a 440-single-nucleotide polymorphism screen of 64 candidate genes among Ashkenazi Jewish case–parent trios. *Am J Hum Genet* 2005; **77**: 918–936.
35 Greenwood TA, Schork NJ, Eskin E, Kelsoe JR. Identification of additional variants within the human dopamine-transporter gene provides further evidence for an association with bipolar disorder in two independent samples. *Mol Psychiatry* 2006; **11**: 125–133, 115.
36 Thierry-Mieg D, Thierry-Mieg J, Potdevin M, Sienkiewicz M. AceView: identification and functional annotation of cDNA-supported genes in higher organisms. unpublished. The AceView genes: http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly.
37 Fallin D, Cohen A, Essioux L, Chumakov I, Blumenfeld M, Cohen D *et al.* Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. *Genome Res* 2001; **11**: 143–151.
38 Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM *et al.* Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci USA* 2006; **103**: 1810–1815.
39 Fearnhead NS, Wilding JL, Winney B, Tonks S, Bartlett S, Bicknell DC *et al.* Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc Natl Acad Sci USA* 2004; **101**: 15992–15997.
40 Emison ES, McCallion AS, Kashuk CS, Bush RT, Grice E, Lin S *et al.* A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* 2005; **434**: 857–863.
41 Arking DE, Pfeufer A, Post W, Kao WH, Newton-Cheh C, Ikeda M *et al.* A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat Genet* 2006; **38**: 644–651.
42 Maller J, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ *et al.* Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 2006; **38**: 1055–1059.
43 Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am J Hum Genet* 2005; **76**: 8–32.
44 Redden DT, Allison DB. The Effect of Assortative Mating upon Genetic Association Studies: Spurious Associations and Population Substructure in the Absence of Admixture. *Behav Genet* 2006; **36**: 678–686.
45 Turakulov R, Easteal S. Number of SNPS loci needed to detect population structure. *Hum Hered* 2003; **55**: 37–45.
46 Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003; **164**: 1567–1587.
47 Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G *et al.* Haplotype tagging for the identification of common disease genes. *Nat Genet* 2001; **29**: 233–237.
48 Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; **29**: 229–232.
49 Ding K, Kullo IJ. Methods for the selection of tagging SNPs: a comparison of tagging efficiency and performance. *Eur J Hum Genet* 2007; **15**: 228–236.
50 Goldstein DB, Ahmadi KR, Weale ME, Wood NW. Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet* 2003; **19**: 615–622.
51 Website. Haploview Documentation: http://www.broad.mit.edu/mpg/haploview/using.php#lddisplay. 2006.
52 Goldstein DB. Pharmacogenetics in the laboratory and the clinic. *N Engl J Med* 2003; **348**: 553–556.
53 Pe'er I, Chretien YR, de Bakker PI, Barrett JC, Daly MJ, Altshuler DM. Biases and reconciliation in estimates of linkage disequilibrium in the human genome. *Am J Hum Genet* 2006; **78**: 588–603.
54 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.
55 Wray NR. Allele frequencies and the r2 measure of linkage disequilibrium: impact on design and interpretation of association studies. *Twin Res Hum Genet* 2005; **8**: 87–94.
56 Schaid DJ. Genetic epidemiology and haplotypes. *Genet Epidemiol* 2004; **27**: 317–320.
57 Sullivan PF, Eaves LJ, Kendler KS, Neale MC. Genetic case-control association studies in neuropsychiatry. *Arch Gen Psychiatry* 2001; **58**: 1015–1024.
58 Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004; **96**: 434–442.
59 Craddock N, O'Donovan MC, Owen MJ. Genes for schizophrenia and bipolar disorder? Implications for psychiatric nosology. *Schizophr Bull* 2006; **32**: 9–16.
60 Hamshere ML, Bennett P, Williams N, Segurado R, Cardno A, Norton N *et al.* Genomewide linkage scan in schizoaffective disorder: significant evidence for linkage at 1q42 close to DISC1, and suggestive evidence at 22q11 and 19p13. *Arch Gen Psychiatry* 2005; **62**: 1081–1088.
61 Van Den Bogaert A, Del-Favero J, Van Broeckhoven C. Major affective disorders and schizophrenia: a common molecular signature? *Hum Mutat* 2006; **27**: 833–853.
62 Berrettini W. Evidence for shared susceptibility in bipolar disorder and schizophrenia. *Am J Med Genet C Semin Med Genet* 2003; **123**: 59–64.
63 Maier W, Lichtermann D, Franke P, Heun R, Falkai P, Rietschel M. The dichotomy of schizophrenia and affective disorders in extended pedigrees. *Schizophr Res* 2002; **57**: 259–266.
64 Potash JB, Willour VL, Chiu YF, Simpson SG, MacKinnon DF, Pearlson GD *et al.* The familial aggregation of psychotic symptoms in bipolar disorder pedigrees. *Am J Psychiatry* 2001; **158**: 1258–1264.
65 Holden C. Sex and the suffering brain. *Science* 2005; **308**: 1574.
66 Burt VK, Rasgon N. Special considerations in treating bipolar disorder in women. *Bipolar Disord* 2004; **6**: 2–13.
67 Barnes C, Mitchell P. Considerations in the management of bipolar disorder in women. *Aust N Z J Psychiatry* 2005; **39**: 662–673.
68 Kennedy N, Boydell J, Kalidindi S, Fearon P, Jones PB, van Os J *et al.* Gender differences in incidence and age at onset of mania and bipolar disorder over a 35-year period in Camberwell, England. *Am J Psychiatry* 2005; **162**: 257–262.
69 Riecher-Rossler A, Hafner H. Gender aspects in schizophrenia: bridging the border between social and biological psychiatry. *Acta Psychiatr Scand Suppl* 2000; **407**: 58–62.
70 Hafner H. Gender differences in schizophrenia. *Psychoneuroendocrinology* 2003; **28**(Suppl 2): 17–54.
71 Viguera AC, Tondo L, Baldessarini RJ. Sex differences in response to lithium treatment. *Am J Psychiatry* 2000; **157**: 1509–1511.
72 Hori H, Ohmori O, Matsumoto C, Shinkai T, Nakamura J. NAD(P)H: quinone oxidoreductase (NQO1) gene polymorphism and schizophrenia. *Psychiatry Res* 2003; **118**: 235–239.
73 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.

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