Haplotype Analysis and a Novel Allele-Sharing Method Refines a Chromosome 4p Locus Linked to Bipolar Affective Disorder

Stephanie Le Hellard, Andrew J. Lee, Sarah Underwood, Pippa A. Thomson, Stewart W. Morris, Helen S. Torrance, Susan M. Anderson, Richard R. Adams, Pau Navarro, Andrea Christoforou, Lorna M. Houlihan, Sevilla Detera-Wadleigh, Michael J. Owen, Philip Asherson, Walter J. Muir, Douglas H.R. Blackwood, Naomi R. Wray, David J. Porteous, and Kathryn L. Evans

Background: Bipolar affective disorder (BPAD) and schizophrenia (SCZ) are common conditions. Their causes are unknown, but they include a substantial genetic component. Previously, we described significant linkage of BPAD to a chromosome 4p locus within a large pedigree (F22). Others subsequently have found evidence for linkage of BPAD and SCZ to this region.

Methods: We constructed high-resolution haplotypes for four linked families, calculated logarithm of the odds (LOD) scores, and developed a novel method to assess the extent of allele sharing within genes between the families.

Results: We describe an increase in the F22 LOD score for this region. Definition and comparison of the linked haplotypes allowed us to prioritize two subregions of 3.8 and 4.4 Mb. Analysis of the extent of allele sharing within these subregions identified 200 kb that shows increased allele sharing between families.

Conclusions: Linkage of BPAD to chromosome 4p has been strengthened. Haplotype analysis in the additional linked families refined the 20-Mb linkage region. Development of a novel allele-sharing method allowed us to bridge the gap between conventional linkage and association studies. Description of a 200-kb region of increased allele sharing prioritizes this region, which contains two functional candidate genes for BPAD, SLC2A9, and WDR1, for subsequent studies.

Key Words: Allele sharing, bipolar affective disorder, chromosome 4p, linkage, recurrent unipolar depression, schizophrenia

Family, twin, and adoption studies have shown that both BPAD (MIM 125480) and SCZ (MIM 181500) have a significant genetic component but that inheritance patterns are complex. Family and molecular studies also indicate that some genetic risk factors may contribute to predisposition to both SCZ and affective disorders, and there can be overlap in their clinical presentation (Berrettini 2000; Blackwood et al 1996; Maier et al 2002; Potash et al 2001). Previously, we described a large Scottish pedigree (F22) in which many family members experience symptoms of BPAD or recurrent major depressive disorder (unipolar depression [MIM 125480]). A whole genome scan of F22 found significant linkage of major affective disorder to chromosome 4p (maximum logarithm of the odds [LOD] score = 4.09; Blackwood et al 1996). Further support for this result comes from variance component analysis of the same data, which found significant evidence for a quantitative trait locus in the region (LOD of 3.7; Visscher et al 1999).

Subsequent to our report, a number of other groups also have found evidence for linkage of both BPAD and SCZ to the F22 linkage region. These reports include linkage in a Welsh family with SCZ and schizoaffective disorder (F50, LOD = 1.97; Asherson et al 1998), linkage in two Danish BPAD families (LOD = 2.00; Ewald et al 1998), linkage to major mental illness (including both SCZ and BPAD) in a large family from the United States (1 of 22 studied; F48, LOD = 3.24; Detera-Wadleigh et al 1999), linkage to BPAD with psychosis and suicidal behavior in a bipolar pedigree sample from the United States (LOD = 1.84; Cheng et al 2006), increased sharing in Welsh SCZ sib-pairs (LOD = 1.73; Williams et al 1999), a nonparametric LOD score of 2.2 in Arab Israeli families with SCZ and schizoaffective disorder (Lerer et al 2003), and excess haplotype sharing (best p value, $p = .00007$) in families from the Faroe Islands with BPAD and SCZ (Als et al 2004). A patient with schizophrenia and a balanced translocation t(4;13) (p16.1;q21.31) also highlights this region of chromosome 4p (Itokawa et al 2004). Finally, we have recently described preliminary evidence for association in GPR78, an orphan G-protein coupled receptor, in females with BPAD and SCZ (Underwood et al 2006). GPR78 maps to the F22 linkage region, but this finding has not yet been replicated, and we have no evidence for segregation of functional variants with the disorder in F22. Therefore, here we describe a systematic approach to narrowing the linkage region on chromosome 4p. Toward this objective, here we present an updated genetic analysis of two Scottish families with BPAD and recurrent major depressive disorder (F22, F59; Blackwood et al 1996); F50, a Welsh family with SCZ and schizoaffective disorder (Asherson et al 1998); and F48, a large U.S. family of Ashkenazi Jewish origin

From the Medical Genetics Section, School of Clinical and Molecular Medicine, Molecular Medicine Centre (SLH, AJL, SU, PAT, SWM, HST, SMA, RAA, AC, LMH, NRW, DJP, KLE), Institute of Evolutionary Biology, Kings Buildings (PN), and Psychiatry, Kennedy Tower, Royal Edinburgh Hospital, University of Edinburgh (WJM, DHRB), Edinburgh, United Kingdom; MAP Genetics, National Institute of Mental Health–National Institutes of Health (SD-W), Bethesda, Maryland; Department of Psychological Medicine, Wales College of Medicine, Henry Wellcome Building, Cardiff University (MJO), Cardiff, United Kingdom; and MRC Social Genetic Developmental Psychiatry, Institute of Psychiatry, King's College (PA), London, United Kingdom.

Address reprint requests to Kathryn L. Evans, PhD, Medical Genetics Section, School of Clinical and Molecular Medicine, Molecular Medicine Centre, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK, Scotland; E-mail: kathy.evans@ed.ac.uk.

Received March 12, 2006; revised June 9, 2006; accepted June 14, 2006.

0006-3223/07/$32.00
doi:10.1016/j.biopsych.2006.06.029

© 2007 Society of Biological Psychiatry

with BPAD, SCZ, recurrent major depressive disorder, and related conditions (Detera-Wadleigh et al 1999). In addition, we include clinical description, linkage and haplotype analysis of new cases of illness within F22, which strengthens our original finding (Blackwood et al 1996). Previously, F22, F59, and F50 had been subjected to low-resolution haplotype analysis with publicly available microsatellite markers at ~5 to 15 cM intervals (Asherson et al 1998; Blackwood et al 1996). F48 previously had been subject to a whole genome scan for linkage (Detera-Wadleigh et al 1999). This analysis found a chromosome 4p marker with a significant LOD score (3.24) and two markers from 4p14–p16 with elevated allele sharing, but this result had not been followed up by haplotype analysis. Our aim was to define the chromosome 4p regions linked to illness in the above four families at high resolution. We identified, mapped, and genotyped new polymorphic markers from the region, allowing us to define all recombination breakpoints accurately. High-resolution haplotype analysis is essential for accurate definition of a linked region, because any misspecification of the model used in linkage analysis will mean that the peak LOD score does not necessarily correspond to the most likely position of the susceptibility gene.

The definition of the linked haplotypes allowed us to define regions of overlap between them. These linkage results in the four families may reflect independent mutations at the same locus (allelic heterogeneity) or a common ancient origin (founder mutation). If there is allelic heterogeneity then it should be possible to identify independent polymorphisms at a single locus, as was the case in Crohn's disease (MIM 605966; Hugot et al 2001). If there is a founder mutation common to more than one family, then this will be flanked by a region of haplotype sharing that varies in extent between the families but includes a functional variant that is common to all cases. The presence of shared alleles on linked chromosomes from families with chromosome 8p–linked schizophrenia was used to identify a region of the NRG1 locus that was subsequently found to show association with SCZ in the Icelandic and Scottish populations (Stefansson et al 2002). We have developed a novel method to compare the extent of allele sharing in positional candidate genes between the four chromosomes that segregate with illness with that observed between the control chromosomes from the four families. This greatly improves the definition of the region most likely to contain a susceptibility gene for BPAD and, in view of the varied phenotypes seen in these families, possibly for other disorders, including schizoaffective disorder and SCZ.

Methods and Materials

This study was approved by the Multicentre Research Ethics Committee for Scotland, and appropriate informed consent was obtained from the human subjects.

Reevaluation of F22 Family Members

Since the original study, through continued contact with the family, we have identified five new cases in F22. All were interviewed using Schedule for Affective Disorders and Schizophrenia—Lifetime Version (SADS-L) (Blackwood et al 1996), and a consensus diagnosis using DSM-IV criteria was reached by two experienced psychiatrists (WJM and DHRB). The offspring in one family were removed from the analysis because major psychiatric illness was detected in first-degree relatives of the married in parent. This was not a fully systematic follow up, as some family members could not be interviewed.

Identification of New Microsatellite Markers

Initially, a minimum overlapping set of clones was chosen across the region to be investigated and subjected to partial digestion and subcloning, as described in the PAC subcloning section. Clones from the resultant library were hybridized with $(CA)_{15}$ and $(GT)_{15}$ oligos. The inserts of positive subclones were sequenced. Latterly, the sequences of BAC clones from the contig were retrieved from the EMBL database. All sequences were processed by Sputnik (http://www.abajian.com/sputnik/) and RepeatMasker (http://repeatmasker.genome.washington.edu/cgi-bin/RM2_req.pl). The Sputnik source code was altered (specifically, the ERROR_MATCH_POINTS symbolic constant was decreased from −6 to −100) to ensure that only perfect simple tandem-repeat arrays, that is, those made up of identical repeat units, were detected. Perfect simple tandem-repeat arrays longer than the following thresholds (12 repeat units for dinucleotide repeats, 7 for trinucleotides, 6 for tetranucleotides, and 5 for pentanucleotides) were identified. For each of these microsatellites, the 200-bp flanking sequence was excised from the relevant BAC sequence and used in primer design (Genbank accession numbers BV677285 to BV677322).

P1 Derived Artificial Chromosome (PAC) Subcloning

Clones were partially cleaved with Sau3AI under conditions that generate fragments ranging in size from 1.5 to .5 kb. Fragments were resolved on a 1% agarose gel (Invitrogen, Paisley, United Kingdom) run in 1× tris, acetic acid, EDTA (TAE). After electrophoresis, fragments of approximately .8 kb were excised, the gel slice was spun in a Spinex column (Corning BV, Schiphol-Riik, The Netherlands), and the DNA was concentrated by ethanol precipitation. The recovered fragments were ligated to pBluescript $SK^{+/-}$ (Stratagene Europe, Amsterdam, Zuiduost, The Netherlands) vector that had been cut with *Bam*HI, gel purified, and treated with heat killed (HK) phosphatase (Cambio, Cambridge, United Kingdom). Ligated DNA was electrotransformed into XL1-Blue cells (Stratagene), and recombinants were identified by plating approximately 400 colonies on a 12 × 8 cm L-agar plate containing ampicillin (80 μg/mL), isopropyl-1-thio-β-D-galactopyranoside (.1 mM) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (.1 mM). White (recombinant) colonies were picked, with a Flexys FLX11001 robot (PBA Technologies Cambridge, United Kingdom) fitted with a six-needle picking head, into five 384-well microtiter plates containing L-broth, ampicillin (80 μg/mL), and 7% glycerol.

Production of Filters

Clones from the above plates were gridded with a Flexys FLX11001 robot (PBA Technologies) onto Hybond N+ nylon membrane (GE Healthcare, Little Chalfont, Bucks, United Kingdom). The gridded filters were processed essentially as described in Cole et al (1996), with the exception that the spheroplasting stage was omitted.

Oligo Hybridization

Fifty nanograms of oligonucleotide was end-labeled with 30 μCi of γ-^{32}P adenosine triphosphate (>5000 Ci/mmol, Amersham) by using 10 U of polynucleotide kinase (Roche, Lewes, East Sussex, United Kingdom) at 37°C for 40 min. Unlabeled oligonucleotides were removed by passing the reaction mixture through a NAP5 column (Amersham Biosciences, Chalfont St. Giles, Bucks, United Kingdom), according to manufacturer's instructions. The eluate was hybridized to filters of gridded

clones overnight, at 72°C, in a hybridization mixture containing 5× sodium saline citrate (SSC), .1% sodium pyrophosphate, .5% sodium dodecyl sulfate (SDS), 5× Denhardt's, and 100 μg/mL of sonicated salmon sperm. Filters were washed in 6× SSC–.1% SDS at 72°C and were autoradiographed overnight.

DNA Preparation

DNA was prepared from positive subclones with a Wizard miniprep kit (Promega UK Ltd, Southampton, United Kingdom), according to manufacturer's instructions.

Sequencing

DNA from positive subclones was subjected to automatic sequencing on an ABI 377 Automated Sequencer by using Rhodamine or Big Dye Sequencing kits (Applera, Warrington, Cheshire, United Kingdom), in accordance with the manufacturer's instructions.

Primer Design

Sequence tagged site (STS) primers were designed from repeat-free regions of sequence (repetitive sequences were screened out by using Repeat Masker [http://repeatmasker.genome.washington.edu/cgi-bin/RM2_req.pl]) by using the primer design program Primer 3 (http://www.genome.wi.mit.edu/genome_software/other/primer3.html).

Polymerase Chain Reaction

Polymerase chain reactions were performed in a total volume of 25 μL, which included the following reagents: 2.5 μL of 10× polymerase chain reaction (PCR) buffer (Applera); 1.5 μL of 25 mM $MgCl_2$ (Cetus); .5 μL of 100 mM dGTP, dTTP, dATP, and dCTP (Applera, Warrington, Cheshire, UK); .2 μL of Taq polymerase (Cetus); and 1 μL of 20 mM (or 150 ng/μL) forward and reverse primers.

Microsatellite Genotyping

Microsatellite markers were amplified by PCR from 10–50 ng of genomic DNA and sized with an ABI3700 automated sequencer.

Single-nucleotide Polymorphism Identification

PCR primers were designed from sequence-flanking exons or predicted exons. Primers were used to amplify DNA from members of the four linked families. PCR products were checked on 2% agarose gel stained with ethidium bromide. PCR products were sequenced with the ABI Big Dye v3 kit, according to the manufacturer's instructions. The sequences were aligned by using the PhredPhrap program displayed with Consed (http://bozeman.mbt.washington.edu/index.html). The new single-nucleotide polymorphisms (SNPs) discovered in this way have been submitted to dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/index.html).

Linkage Analysis

Two-point parametric linkage analyses were performed by using the program MLINK from the LINKAGE package (Terwilliger and Ott 1994) for F22 and F59. Within each family, microsatellite markers were genotyped on those family members that were necessary to determine both the linked haplotype segregating in each family and to identify any recombination breakpoints (this was at least 25% of family members). Twenty-four markers were genotyped on F22, and 57, on F59. PedCheck (O'Connell et al 1998) was used to check for genotype inconsistencies. The models used are as described in Blackwood et al (1996); that is, narrow definition of phenotype (model 1) bipolar 1 and bipolar 2 disorder and broad definition of phenotype (model 2) bipolar 1 disorder, bipolar 2 disorder, and recurrent unipolar disorder. Cases with other psychiatric diagnoses were coded as unknown. Age-dependent penetrances were applied to controls. Given the structure of the families studied, only a dominant mode of inheritance was considered. Disease-allele frequencies, age-dependent penetrances, and phenocopy rates for the narrow- and broad-analysis models are as for model 1 and model 2 in Blackwood et al (1996). Equal allele frequencies were assumed, and modification of these did not have a significant effect on the results (data not shown).

Allele-sharing Analysis

The genotypes of 46 individuals from the four families were reconstructed into chromosomes by using the Merlin software (Abecasis et al 2002). Although allocation of specific alleles to haplotypes was generally possible, some ambiguity remained: 83 markers of 284 retained a degree of ambiguity, but on average, only 7.35% of the chromosomes in question were ambiguous. Within each family, the single haplotype that was shared by all affected individuals contributing to the linkage was denoted the linked haplotype for that family. There were 38 other chromosomes in region B and 37 in region D, which were designated control haplotypes. For F22 and F59, model 1, that is, bipolar 1 and bipolar 2 disorder, was used when determining the linked haplotype. This was expanded to include schizophrenia and schizoaffective disorder for F48 and F50. There were no individuals in these diagnostic categories who did not carry the linked haplotype. Further details on the families are provided in Table 1.

Pairwise comparisons were performed between all pairs of haplotypes within the linked group and separately for all pairs within the control group. For each pair of haplotypes, each marker was assigned a score on the basis of the size of the region of allele sharing in which it was found (based on the methods of

Table 1. Details of the Families Studied

Generation by Family	No. Affected (No. Genotyped)	No. Unaffected and Unknown (No. Genotyped) [No. with linked haplotype]
F22		
1	0 (0)	3 (0)
2	0 (0)	7 (0)
3	5 (5)	23 (10) [1]
4	13 (13)	50 (30) [12]
5	14 (14)	32 (22) [9]
F48		
1	1 (0)	17 (5) [4]
2	7 (5)	20 (10) [5]
3	8 (7)	7 (7) [2]
F50		
1	1 (1)	1 (1) [0]
2	2 (2)	3 (3) [1]
3	2 (2)	2 (2) [0]
F59		
1	1 (0)	3 (3) [1]
2	3 (3)	1 (0)
3	1 (1)	0 (0)

The table lists the number of generations and, for each generation, the number of affected (broad category) or unknown and unaffected individuals, with the number genotyped in parenthesis. The number of unaffected and unknown individuals with the linked haplotype is shown in square brackets.

Van der Meulen and te Meerman [1997] and Bourgain et al [2000]). Briefly, if an allele of a marker was shared between two haplotypes, then the number of alleles shared at adjacent consecutive markers determined the score for that marker (e.g., if alleles at three consecutive markers were shared, then each marker received a score of three). Scores at markers with ambiguous genotypes were weighted on the basis of the ambiguity (e.g., if it was known that a chromosome could carry one of two alleles at a particular marker, then that marker would contribute a score of .5). The scores were averaged across the pairwise tests to generate a length statistic for each marker for both the linked and control haplotypes. This then allowed definition of the extent of allele sharing between the linked haplotypes and between the control haplotypes.

Permutation Testing

Because it has been shown that the distributional properties of the length statistic are complex (Tzeng et al 2003), we used permutation analysis to test the null hypothesis, that is, that no difference was observed when the level of sharing among the linked haplotypes was compared with that among the control haplotypes. For the permutation analysis, all linked and control haplotypes were randomized and reallocated to the linked and control groups while retaining the proportion of haplotypes allocated to each group. The allele sharing analysis was then repeated with these redefined haplotypes. This process was performed 9999 times. The p value for the length statistic at each marker was calculated as $p = s/10{,}000$, where s was the number of times that the length statistic for the permutation replicates exceeded the length statistic using the real linked haplotypes. Statistical significance was investigated when $p < .05$. Where there was more than one consecutive marker with $p < .05$, the average p value and the number of linkage disequilibrium (LD) blocks were calculated for the region.

Many markers have been tested, and so some correction for multiple testing is required. Linkage disequilibrium between markers means that a Bonferroni correction would be overly conservative. Therefore the permutation replicates were used to account for multiple testing. Each of the 9999 permutated replicates was tested against all other sets of permutations in an approximation of a set of nested permutations (Becker and Knapp 2004). A p value for each marker was generated for each of the permuted replicates, as for the linked haplotypes described above. A corrected p value then was calculated from $r/10{,}000$, where r was the number of times that a permutation replicate covered as many LD blocks as the region found in the linked haplotypes and at a greater level of significance.

Investigation of Shared SNPs

To investigate the effect of nonsynonymous SNPs, we submitted multiple protein sequence alignments to the sorting intolerant from tolerant (SIFT) algorithm website (http://blocks.fhcrc.org/sift/SIFT.html), which uses a conservation-based approach to estimate the probability that a particular amino-acid change will be tolerated by the protein. Only predictions with enough sequence information to be meaningful (median sequence information of <3.25) were considered.

Accession Numbers

The following were the accession numbers used. For Genbank, BV677285 to BV677322; for dbSNP, ss46563120 to ss46563185.

Results

Markers for Linkage and Haplotype Analysis

Initially, we performed linkage and haplotype analysis with publicly available microsatellite markers. Once these were exhausted, we identified additional markers. At first, these were identified experimentally, by subcloning PAC and BAC clones from our contig across the majority of the linked region (Evans et al 2001). However, as more genomic sequence from clones in the region became available, we moved to in silico identification. In total, the two approaches resulted in the creation of 44 microsatellite markers. Where the supply of potential microsatellite markers had been exhausted, STSs were designed around publicly available SNPs which were amplified and sequenced in key family members, testing the informativeness of the polymorphism in those individuals.

Reevaluation of F22 Members

As expected, since the initial report of F22 (Blackwood et al 1996), there have been changes to the clinical status of several family members. Although we have been unable to follow up all family members, we have identified five new cases: one of BPAD and four of recurrent unipolar depression. These include one individual who was previously unaffected, two who previously had minor psychiatric illness, and two who were identified and assessed subsequent to our original report. In addition, the five offspring in one family were removed from the analysis because major psychiatric illness was present in first-degree relatives of both parents. This updated phenotype information was used in

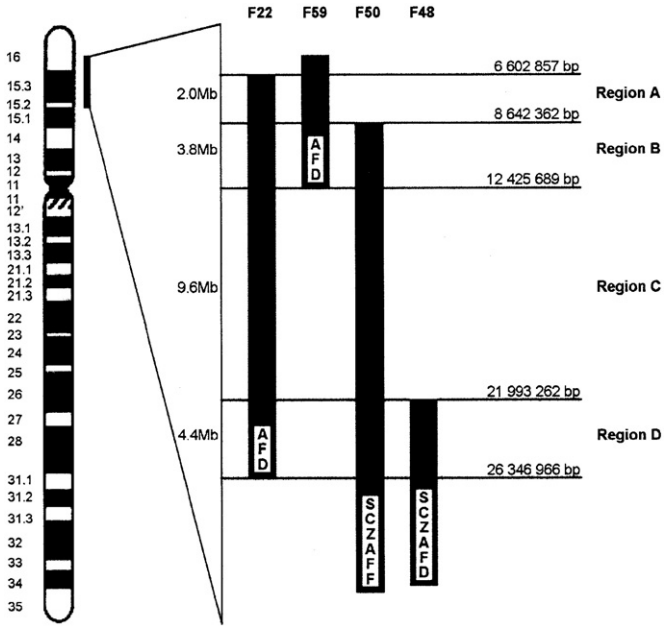

Figure 1. The overlap between the linked regions that segregate with illness in the four families. The sizes (in Mb) refer to the genomic distances between the points marked by the horizontal lines. The numbers on the horizontal lines are from National Center for Biotechnology Information build 35 (http://www.ncbi.nlm.nih.gov/) and are the map coordinates of each of the markers that define the boundaries of the linked haplotypes. Regions A to 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, bipolar affective disorder (BPAD) and recurrent major depressive disorder; SCZAFF, schizoaffective disorder and schizophrenia; SCZAFD, BPAD, recurrent major depressive disorder, schizophrenia, and others.

Table 2. List of Known Genes Studied

Abbreviated Gene Name	Full Gene Name	Genomic Region[a]
GPR78 (MIM 606921)	G protein–coupled receptor 78	B
CPZ (MIM 603105)	Carboxypeptidase Z	B
DRD5 (MIM 126453)	Dopamine receptor D5	B
SLC2A9 (MIM 606142)	Solute carrier family 2	B
WDR1 (MIM 604734)	WD repeat domain 1	B
MIST (MIM 233450)	Mast cell immunoreceptor signal transducer	B
HS3ST1 (MIM 603244)	Heparan sulfate D–glucosaminyl 3-*O*-sulfotransferase 1	B
GPR125	G protein–coupled receptor 125	D
GBA3 (MIM 606463)	Glucosidase beta acid 3	D
PPARGC1A (MIM 604517)	Peroxisome proliferative activated receptor gamma coactivator 1 alpha	D
DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	D
SOD3 (MIM 185490)	Superoxide dismutase 3	D
LG12 (MIM 608301)	Leucine rich gene, glioma inactivated 2	D
SLA/LP	Soluble liver antigen/liver pancreas antigen	D
P14K2B	Phosphatidylinositol 4 kinase type II beta	D
ANAPC4 (MIM 606947)	Anaphase-promoting complex subunit 4	D
SLC34A2 (MIM 604217)	Human type II sodium-dependent phosphate transporter (NaPi-3)	D
KIAA0746		D
RBPSUH (MIM 147183)	Recombination signal-binding protein suppressor of hairless (Drosophila), homologue of	D
CCKAR (MIM 118444)	Cholecystokinin A receptor	D

[a] To which gene maps.

the linkage and haplotype analysis described in the next section. Details of F22 and the other families studied are given in Table 1.

Linkage and Haplotype Analysis

We genotyped the polymorphic markers in key individuals from the four linked families and determined the haplotypes linked to disease susceptibility (the linked haplotypes). In F22, four of the five newly diagnosed individuals carry the linked haplotype. One individual with a diagnosis of unipolar depression did not carry the linked haplotype and appears to be a phenocopy. Two-point parametric linkage analysis on the updated F22 data produced a maximum LOD score of 4.41 at marker D4S394, under the narrow diagnostic model (previously 4.09). LOD scores were also calculated for F59, because this had not been performed previously. The highest LOD score was obtained for F59 at D4S431 (.88 for the broad model, four affected individuals with genotype data available). This LOD score, although small, contributes to the growing evidence for a predisposition locus in this chromosomal region.

Haplotypes were constructed for all four families. Figure 1 shows the overlapping linked haplotypes that determine the position of the candidate intervals and allow us to prioritize subregions. The boundaries of the candidate region are defined by the 20Mb linked haplotype in F22 and divide into four subregions as defined by the boundaries of the linked haplotypes in the other families (Figure 1, regions A to D). These genomic regions can be prioritized for further analysis by the strength of linkage evidence provided by the different families. On this basis, D (4.4 Mb) appears to be the best candidate region, followed by B, C, and A. However, region B (3.8 Mb) is also attractive for prioritizing for further study, because it is implicated in the three Celtic families, which potentially share a common ancestral origin.

Detection of SNPs for Allele-sharing Analysis

SNPs were identified by amplifying and sequencing exons, exon–intron boundaries, and the region approximately 1 kb up- and downstream of the 20 known genes (see Table 2) from intervals B and D using DNA from family members. We defined

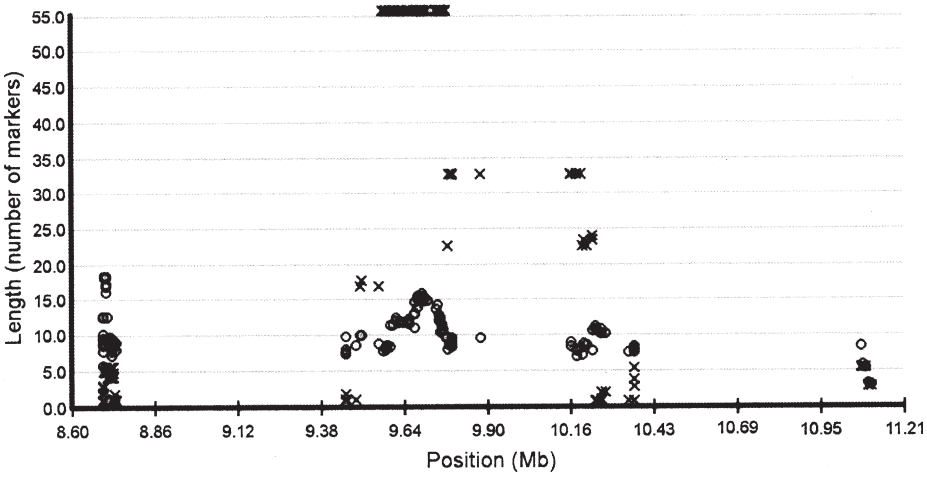

Figure 2. Allele sharing, measured by the number of consecutively shared markers, between chromosomes linked to psychiatric illness (linked chromosomes) and control chromosomes for region B. The allele sharing between linked chromosomes is represented by crosses, and between the control chromosomes, by circles. The position of the markers is given in Mb along the chromosome. There is one area of increased allele sharing between the linked chromosomes in region B.

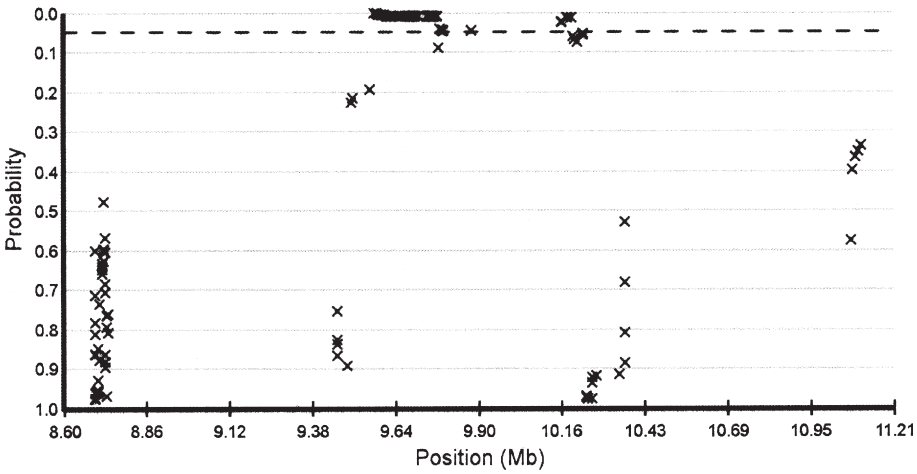

Figure 3. Permutation analysis for the three families linked to region B. The *p* value is plotted for each marker, and the dashed line indicates the $p = .05$ threshold. There are four significant regions of sharing in region B.

known genes as those genes with RefSeq status codes predicted, provisional, reviewed or validated and that had a protein described in Swissprot. We chose to sequence PCR products amplified from family members, rather than relying on publicly available SNPs, because this would result in the identification of SNPs specific to the populations under study and confirm the existence of database SNPs in our population. We identified 68 novel SNPs and characterized 216 publicly available SNPs in this sample (described in SOM Table 1).

Analysis of Allele and Haplotype Sharing

Three of the four families show linkage to region B and three to region D (Figure 1). In each region, this led to the identification of three haplotypes (one from each linked family) that were linked with the disease and 38 control haplotypes in region B and 37 in region D. In both regions, allele sharing between the three linked haplotypes was compared with that between the control haplotypes by counting the number of markers that make up a region of sharing. Within each group of haplotypes (linked or control), there may be an artificially high level of sharing where multiple markers have been genotyped within single LD blocks. However, the comparison between the linked haplotype and control haplotype sharing will adjust for this. The significance of the difference in sharing between linked and control haplotypes was assessed by permutation analysis.

For example, in region B, a haplotype consisting of 48 consecutive markers over five LD blocks (covering ~200 kb) showed excess allele sharing (Figure 2) between the three linked haplotypes (from F22, F59, and F50) when compared with the control haplotypes. To assess the significance of this result, we permutated the 38 control and the 3 linked haplotypes, randomly assigning chromosomes to either the linked or control group, and then calculated how often we would expect to see this result by chance. This indicated that the excess sharing in this subregion was significant ($p = .007$; Figure 3). Six other subregions also showed excess sharing (Figure 2, region B and Figure 4, region D). These are listed in Table 3. Three of them also are marginally significant after the permutation testing (Figure 3, region B; Figure 5, region D; and column 6 of Table 3). We then corrected for multiple testing by examining how often the permuted chromosomes generated an equivalent or greater significance. We did this by approximating nested permutations for each permutated set of chromosomes. For example, for the region described above (shared region 1 on Table 3), sharing across this number of LD blocks and with this level of significance was found in only a small percentage of permuted chromosome sets, leading to a corrected *p* value of .009. This was the only shared region that remained significant after correction for multiple testing (Table 3, column 7).

Investigation of Shared SNPs

We examined the 48 SNPs in the significant shared region (National Center for Biotechnology Information build 35 May 2004 assembly coordinates: 9,576,583–9,777,081; Table 4). Thir-

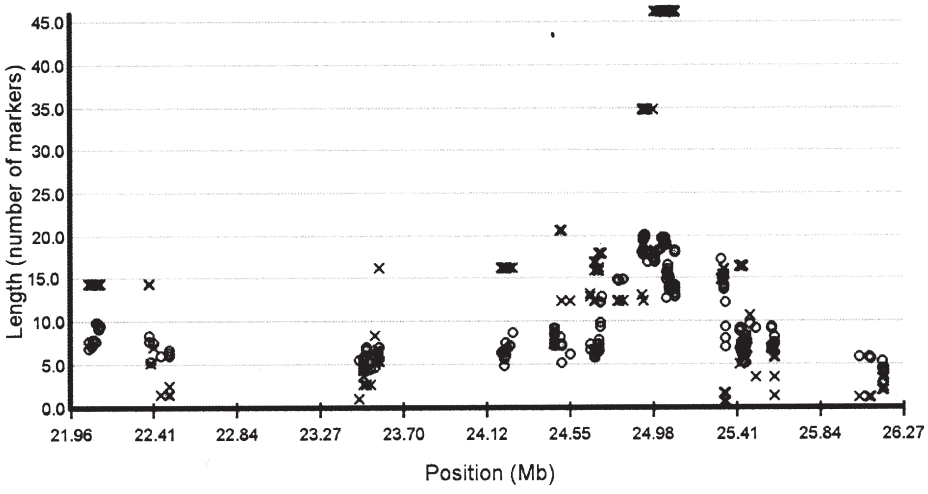

Figure 4. Allele sharing, measured by the number of consecutively shared markers, between chromosomes linked to psychiatric illness (linked chromosomes) and control chromosomes for region D. The allele sharing between linked chromosomes is represented by crosses, and between the control chromosomes, by circles. The position of the markers is in Mb along the chromosome. There are a number of areas of increased allele sharing between the linked chromosomes in region D.

Table 3. Shared Regions

Shared Region	Priority Region	Number of Shared Markers	Number of LD Blocks Shared	Length of Shared Region (kb)	*p* Value	Corrected *p* Value
1	B	48	5	200	.007	.009
2	B	9	1	9	.050	.193
3	B	1	1	n/a	.048	.195
4	B	5	4	30	.015	.086
5	D	10	2	55	.054	.520
6	D	3	1	9	.031	.416
7	D	41	9	103	.054	.070

Data include the properties of the regions of excess sharing, the subregion of the F22 linkage region to which they map, the number of shared markers within them, the number of linkage disequilibrium (LD) blocks they cover, their length, the *p* value after the test for significance, and the *p* value corrected for multiple testing.

ty-three of the SNPs are in SLC2A9, a member of the facilitative glucose transporter family. Twenty-eight are intronic and five, exonic. None of the intronic SNPs lie within known consensus splice sites. To investigate the possible functional consequences of the nonsynonymous exonic SNPs, we used the SIFT algorithm (Ng and Henikoff 2001), which uses multiple sequence alignments to predict tolerated amino acid substitutions. Only the R294H variant was predicted to affect protein structure or function ($p < .04$). Two SNPs are intergenic, and the remaining thirteen are in WDR1, a WD repeat domain protein. Ten of the WDR1 SNPs are in introns and three in exons. None of the intronic SNPs lie within known splice consensus sequences. The three exonic SNPs all are found in the 3′-UTR region. Because all of the shared SNPs were also present on at least 30% of control chromosomes, it is unlikely that any of them are the sole functional variant in this region. Instead, it is more likely that we have identified a shared haplotype that carries the functional SNPs.

Discussion

The additional cases of major affective disorder in family 22 described in this 10-year follow-up have confirmed and strengthened the linkage result to this region of chromosome 4p. This study demonstrates the value of large families when analyzing complex conditions such as BPAD, where in many cases there is phenotypic complexity and almost certainly locus heterogeneity and oligo or polygenicity. Like other complex disorders including breast cancer [MIM 114480], colon cancer [MIM 114500], diabetes [MIM 125823], and Alzheimer's disease [MIM 104300], there is growing evidence that some cases of BPAD and SCZ can arise from the inheritance of mutations with large effect and strong genotype–phenotype correlations. For example, there is support for such a subset from other extended BPAD pedigrees that show significant linkage of BPAD to loci on chromosomes 1q, 4q, 12q, 18q, 21q (Potash and DePaulo 2000).

We have generated new microsatellite and SNP markers and used these in genetic analysis of families with chromosome 4p linked BPAD and other, related phenotypes. The high-resolution haplotype analysis described allows comparison of the F22 linkage region with the linked haplotypes in other families. Although F48 also generates a LOD score of >3, the inclusion of families that have the capacity only to generate smaller LOD scores is a higher risk strategy, because of the increased likelihood of false positives. However, because these families contained affected individuals with recombination events that were complementary to those of the larger families, this allowed us to prioritize subregions of a large candidate region (Figure 1). On the basis of this linkage evidence, we then attempted to bridge the gap between the resolution provided by linkage analysis and that of association studies by assessing allele sharing in known genes from regions B and D. Our allele sharing method combined a modified version of a measure of allele sharing described elsewhere with two rounds of permutation analysis (Van der Meulen and te Meerman 1997; Bourgain et al 2000). The first assesses the significance of the finding, and the second accounts for multiple testing. As far as we are aware, these three aspects of the method have not previously been combined in this way with this aim.

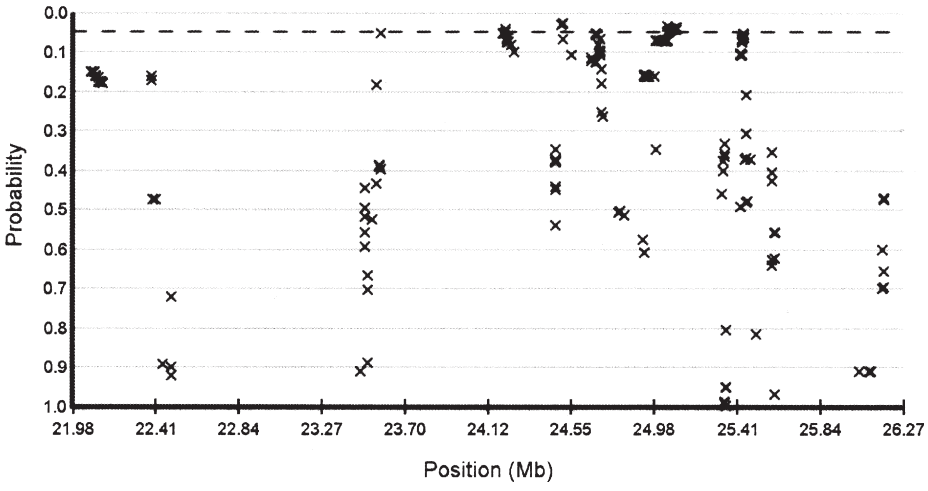

Figure 5. Permutation analysis for the three families linked to region D. The *p* value is plotted for each marker, and the dashed line indicates the $p = .05$ threshold. There are three significant regions of sharing in region D.

Table 4. Single-Nucleotide Polymorphisms (SNPs) in the Significant Shared Region

SNP	Location	Gene	SNP Location within Region	Amino Acid Change (Probability Change is Deleterious)	% of Control Chromosomes with 'Linked' Variant
C__11538907_10	9576833	SLC2A9	intronic		68
rs2280204	9586369	SLC2A9	intronic		84
rs2280205	9586442	SLC2A9	exonic	P350L (.3)	55
rs4697695	9592369	SLC2A9	intronic		71
rs3733591	9598649	SLC2A9	exonic	R294H (.04)	29
ih4 (ss46563120)	9598686	SLC2A9	exonic	V2821 (.25)	79
C__1216553_10	9611080	SLC2A9	intronic		79
rs2276965	9619994	SLC2A9	intronic		97
rs4292327	9620219	SLC2A9	intronic		74
C__1216527_10	9625348	SLC2A9	intronic		61
C__1216498_10	9632664	SLC2A9	intronic		61
C__1216472_10	9642152	SLC2A9	intronic		61
C__1216450_10	9658408	SLC2A9	intronic		82
ih7 (ss46563122)	9658710	SLC2A9	intronic		61
ih6 (ss46563121)	9658849	SLC2A9	exonic	synonymous	79
ih8 (ss46563123)	9675012	SLC2A9	intronic		79
rs7695555	9678175	SLC2A9	intronic		61
rs7662439	9678318	SLC2A9	intronic		32
rs1014290	9678380	SLC2A9	intronic		84
C__1216398_10	9683824	SLC2A9	intronic		66
rs3796841	9684423	SLC2A9	intronic		76
rs3796840	9684640	SLC2A9	intronic		95
C__1216379_10	9689982	SLC2A9	intronic		71
rs2240720	9696999	SLC2A9	intronic		66
rs2240721	9697083	SLC2A9	intronic		68
rs2240722	9697276	SLC2A9	intronic		61
rs2276963	9699358	SLC2A9	intronic		100
rs2276962	9699399	SLC2A9	intronic		97
rs2276961	9699500	SLC2A9	exonic	R25G (.51)	68
rs10030570	9703679	SLC2A9	intronic		82
rs6819833	9703873	SLC2A9	intronic		82
C__11546160_10	9707313	SLC2A9	intronic		84
C__1216308_10	9717623	SLC2A9	intronic		68
rs3775938	9746050		intergenic		32
rs4320137	9749488		intergenic		79
rs9732	9753177	WDR1	3′UTR		84
rs9926	9753379	WDR1	3′UTR		82
rs2241469	9756981	WDR1	3′UTR		50
rs12503195	9759291	WDR1	intronic		74
ih249 (ss46563157)	9759387	WDR1	intronic		92
rs2241470	9761089	WDR1	intronic		71
rs2241473	9762468	WDR1	intronic		84
rs2241475	9762707	WDR1	intronic		76
rs2241476	9763066	WDR1	intronic		76
rs2241477	9763189	WDR1	intronic		74
rs734122	9766384	WDR1	intronic		74
rs2241481	9775859	WDR1	intronic		45
rs2241485	9777331	WDR1	intronic		50

Data include the SNPs in the significant shared region (the SNPs discovered in house, ih, have their new db SNP accession numbers in brackets); their genomic location (NCBI Build 35 May 2004 assembly coordinates: 9,576,583–9,777,081); the gene in which they lie; the location within that gene; the amino acid change, where appropriate; the likelihood that change is deleterious, according to SIFT analysis; and the percentage of control chromosomes that have the shared allele.

The allele-sharing analysis identified a single significant region of excess sharing consistent with a founder mutation for susceptibility to BPAD in region B, where a haplotype is shared by three families of Celtic heritage. The haplotype sharing in this 200-kb region is significantly higher than that seen in control chromosomes (corrected $p = .009$). It is noteworthy that this 200-kb region is the only significant region of sharing observed of a total of more than 8 Mb. In addition, our region is contained within the most significant region ($p = .00007$) of haplotype sharing that is seen in distantly related patients with bipolar disorder and schizophrenia from the Faroe Islands (Als et al 2004).

The shared region identified includes part of the genes *SLC2A9* (MIM 606142) and *WDR1* (MIM 604734) and the 47 kb between these two genes. *SLC2A9* is known to code for a brain-expressed member of the SLC2A facilitative glucose transporter family that plays a significant role in maintaining glucose

homeostasis. An imbalance in glucose homeostasis leading to, for example, altered energy provision to neurons could be invoked as a predisposing mechanism to psychiatric illness. A putative link between energy provision and bipolar disorder has already been made; there are a number of lines of evidence implicating mitochondrial dysfunction in bipolar disorder (Kato and Kato 2000). *WDR1* encodes a brain expressed protein containing nine WD repeat domains that provide protein–protein interactions thought to induce the disassembly of actin filaments, offering a plausible mechanism for neuronal dysfunction.

No excess allele sharing was identified in genes from region D. This is not unexpected because, in contrast to the case in region B, the families are from ethnically diverse populations.

In conclusion, definition of the linked haplotypes and analysis of the extent of allele sharing between families has allowed us to highlight positional candidate genes and subregions of the chromosome 4p BPAD candidate region that are worthy of intense study.

AJL and SU contributed equally to this work.

This work was supported by grants from the Chief Scientist Office, the Medical Research Council, and the Wellcome Trust. We thank John Beekman, Diane Lawson, John Maule, Ling Mo, Jane Savill, Mahes Sivarajah, and Claire Whitton for technical assistance; Maura Walker and Margaret van Beck for assistance with clinical aspects of the study; the Wellcome Trust Clinical Research Facility Genetics Core for help with sample preparation; and Simon T. Cooper for assistance in the preparation of the manuscript.

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002): Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97–101.

Als TD, Dahl HA, Flint TJ, Wang AG, Vang M, Mors O, et al (2004): Possible evidence for a common risk locus for bipolar affective disorder and schizophrenia on chromosome 4p16 in patients from the Faroe Islands. *Mol Psychiatry* 9:93–98.

Asherson P, Mant R, Williams N, Cardno A, Jones L, Murphy K, et al (1998): A study of chromosome 4p markers and dopamine D5 receptor gene in schizophrenia and bipolar disorder. *Mol Psychiatry* 3:310–320.

Becker T, Knapp M (2004): A powerful strategy to account for multiple testing in the context of haplotype analysis. *Am J Hum Genet* 75:561–570.

Berrettini WH (2000): Are schizophrenic and bipolar disorders related? A review of family and molecular studies. *Biol Psychiatry* 48:531–538.

Blackwood DH, He L, Morris SW, McLean A, Whitton C, Thomson M, et al (1996): A locus for bipolar affective disorder on chromosome 4p. *Nat Genet* 12:427–430.

Bourgain C, Genin E, Quesneville H, Clerget-Darpoux F (2000): Search for multifactorial disease susceptibility genes in founder populations. *Ann Hum Genet* 64:255–265.

Cheng R, Juo SH, Loth JE, Nee J, Iossifov I, Blumenthal R, et al (2006): Genome-wide linkage scan in a large bipolar disorder sample from the National Institute of Mental Health genetics initiative suggests putative loci for bipolar disorder, psychosis, suicide, and panic disorder. *Mol Psychiatry* 11:252–260.

Cole CG, Collins JE, Dunham I (1996): YAC library screening. I. Preparation of hybridization filters and PCRpools. *Methods Mol Biol* 54:23–31.

Detera-Wadleigh SD, Badner JA, Berrettini WH, Yoshikawa T, Goldin LR, Turner G, et al (1999): 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 U S A* 96:5604–5609.

Evans KL, Le Hellard S, Morris SW, Lawson D, Whitton C, Semple CA, et al (2001): A 6.9-Mb high-resolution BAC/PAC contig of human 4p15.3-p16.1, a candidate region for bipolar affective disorder. *Genomics* 71: 315–323.

Ewald H, Degn B, Mors O, Kruse TA (1998): Support for the possible locus on chromosome 4p16 for bipolar affective disorder. *Mol Psychiatry* 3:442–448.

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, et al (2001): Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411:599–603.

Itokawa M, Kasuga T, Yoshikawa T, Matsushita M (2004): Identification of a male schizophrenic patient carrying a de novo balanced translocation, t(4; 13)(p16.1; q21.31). *Psychiatry Clin Neurosci* 58:333–337.

Kato T, Kato N (2000): Mitochondrial dysfunction in bipolar disorder. *Bipolar Disord* 2:180–190.

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

Maier W, Lichtermann D, Franke P, Heun R, Falkai P, Rietschel M (2002): The dichotomy of schizophrenia and affective disorders in extended pedigrees. *Schizophr Res* 57:259–266.

Ng PC, Henikoff S (2001): Predicting deleterious amino acid substitutions. *Genome Res* 11:863–874.

O'Connell JR, Weeks DE (1998): PedCheck: A program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266.

Potash JB, DePaulo JR Jr (2000): Searching high and low: A review of the genetics of bipolar disorder. *Bipolar Disord* 2:8–26.

Potash JB, Willour VL, Chiu YF, Simpson SG, MacKinnon DF, Pearlson GD, et al (2001): The familial aggregation of psychotic symptoms in bipolar disorder pedigrees. *Am J Psychiatry* 158:1258–1264.

Stefansson H, Sigurdsson E, Steinthorsdottir V, Bjornsdottir S, Sigmundsson T, Ghosh S, et al (2002): Neuregulin 1 and susceptibility to schizophrenia. *Am J Hum Genet* 71:877–892.

Terwilliger JD, Ott J (1994): *Handbook of Human Genetic Linkage*. Baltimore, MD: John Hopkins University Press.

Tzeng JY, Devlin B, Wasserman L, Roeder K (2003): On the identification of disease mutations by the analysis of haplotype similarity and goodness of fit. *Am J Hum Genet* 72:891–902.

Underwood SL, Christoforou A, Thomson PA, Wray NR, Tenesa A, Whittaker J, et al (2006): Association analysis of the chromosome 4p-located G protein-coupled receptor 78 (GPR78) gene in bipolar affective disorder and schizophrenia. *Mol Psychiatry* 11:384–394.

Van der Meulen MA, te Meerman GJ (1997): Haplotype sharing analysis in affected individuals from nuclear families with at least one affected offspring. *Genet Epidemiol* 14:915–920.

Visscher PM, Haley CS, Heath SC, Muir WJ, Blackwood DH (1999): Detecting QTLs for uni- and bipolar disorder using a variance component method. *Psychiatr Genet* 9:75–84.

Williams NM, Rees MI, Holmans P, Norton N, Cardno AG, Jones LA, et al (1999): A two-stage genome scan for schizophrenia susceptibility genes in 196 affected sibling pairs. *Hum Mol Genet* 8:1729–1739.