 Human Molecular Genetics, 1999, Vol. 8, No. 9 **1729–1739**

A two-stage genome scan for schizophrenia susceptibility genes in 196 affected sibling pairs

N. M. Williams[+], M. I. Rees[+], P. Holmans, N. Norton, A. G. Cardno, L. A. Jones, K. C. Murphy, R. D. Sanders, G. McCarthy, M. Y. Gray, I. Fenton, P. McGuffin and M. J. Owen[§]

Neuropsychiatric Genetics Unit, Departments of Psychological Medicine and Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF14 4XN, UK

Received April 16, 1999; Revised and Accepted June 22, 1999

We undertook a systematic search for linkage in 196 affected sibling pairs (ASPs) with DSMIV schizophrenia. In stage 1 we typed 97 ASPs with 229 microsatellite markers at an average inter-marker distance of 17.26 cM. Multipoint affected sib pair analysis identified seven regions with a maximum lod score (MLS) at or above the level associated with a nominal pointwise significance of 5%, on chromosomes 2q, 4p, 10q, 15q, 18p, 20q and Xcen. In stage 2 we genotyped a further 54 markers in 196 ASPs together with parents and unaffected siblings. This allowed the regions identified in stage 1 to be typed at an average spacing of 5.15 cM, while the region of interest on chromosome 2 was typed to 9.55 cM. Analysis was performed on the whole data set. Simulation studies suggested that we would expect one multipoint MLS of 1.5 per genome scan in the absence of linkage. An MLS of 3 would be expected only once in every 20 genome scans and thus corresponds to a genome-wide significance of 0.05. We obtained three multipoint MLSs >1.5 and, on this basis, the results on chromosomes 4p, 18q and Xcen can be considered suggestive. However, none approached a genome-wide significance of 0.05. The power of this study was >0.95 to detect a susceptibility locus of $\lambda_s = 3$ with a genome-wide significance of 0.05, but only 0.70 to detect a locus of $\lambda_s = 2$. Our results suggest that common genes of major effect ($\lambda_s > 3$) are unlikely to exist for schizophrenia.

INTRODUCTION

Schizophrenia is a severe, disabling mental disorder with a lifetime risk in the general population of ~1%. Family studies over many years have confirmed an increased risk in relatives of probands with the disorder (1) and the recurrence risk in siblings is ~10 times the risk to the general population ($\lambda_s = 10$). Schizophrenia has also been the focus of a large number of twin and adoption studies which have allowed the relative contributions of nature and nurture to be determined. The evidence here is relatively clear cut and compelling: shared genes rather than shared environments underlie the familial aggregation of the disorder (2). As yet unspecified non-shared environmental factors seem to play a role, but twin and adoption studies have revealed a substantial genetic contribution with even conservative estimates of heritability >60% (3,4). Twin and adoption studies have also shown that schizophrenia shares familial, and probably genetic, liability with a range of other psychotic illnesses (5,6) and personality disorders, such as schizotypal personality disorder (7), collectively known as schizophrenia spectrum disorders.

Studies on the recurrence risks in various classes of relative allow us to exclude the possibility that schizophrenia is a single gene disorder or collection of single gene disorders even when incomplete penetrance is taken into account. Rather, the mode of transmission is complex and non-Mendelian (8). The commonest mode of transmission is likely to be oligogenic, polygenic or a mixture of the two (1,9). However, the number of susceptibility loci, the disease risk conferred by each locus and the degree of interaction between loci all remain unknown. The contribution of individual genes to the familiality of a disorder can be expressed in terms of the locus-specific λ_s, i.e. the relative risk to siblings resulting from possession of the disease allele (10). Risch (10) has calculated that the data for schizophrenia are incompatible with the existence of a single locus of $\lambda_s > 3$ and, unless extreme epistasis exists, models with two or three loci of $\lambda_s \leq 2$ are more plausible. Despite these uncertainties, the strength of the genetic effect makes schizophrenia a compelling candidate for molecular genetic approaches.

The most appropriate approach to detecting susceptibility genes of moderate effect in complex disorders is by allele sharing linkage methods in large samples of nuclear families (11–13). We report here the results of a systematic search for linkage in 196 affected sibling pairs (ASPs) with schizophrenia. Based on a cost function analysis (14), we adopted a highly efficient two-stage procedure for genotyping, which greatly increased the ratio of power to number of genotypes. Stage 1 consisted of genotyping 97 ASPs, without parents or unaffected sibs, using a fairly loose grid of 229 microsatellite markers with an average inter-marker distance of 17.26 cM. Liberal statistical criteria for identification of regions of interest were used taking multipoint lod scores that corresponded to a nominal *P*-value of 0.05 or less. This resulted in the identification of seven regions for further follow-up. In stage 2, a series of much finer

[+]These authors contributed equally to this work
[§]To whom correspondence should be addressed. Tel: +44 1222 743248; Fax: +44 1222 746534; Email: owenmj@cardiff.ac.uk

grids with inter-marker distances of 5 cM were studied in the seven regions identified in stage 1 using a total sample of 196 ASPs (consisting of the 97 from stage 1 plus a further 99) and including parents and affected siblings, where available. Genome-wide significance levels (15) were calculated from detailed simulation studies of our entire two-stage procedure. As well as obtaining evidence for linkage as measured by the multipoint maximum lod score, we also undertook exclusion mapping to determine what effect sizes at each point on the genome were consistent with the observed marker data. Genetic effect sizes were specified using the parameter λ_s (10). This enabled us to determine whether a region with a low lod score was really inconsistent with a proposed genetic model or whether the study merely lacked power.

RESULTS

In stage 1 we typed 229 microsatellite markers with a mean heterozygosity of 0.78 (range 0.42–0.93) and an average marker density of 17.26 cM in 97 ASPs (average number of pairs genotyped 87.6, range 72–96). Multipoint affected sib pair analysis using MAPMAKER/SIBS identified a total of seven regions in stage 1 with a maximum lod score (MLS) at or above the level associated with a nominal pointwise significance of 5% (MLS ≥ 0.74). These peaks were on chromosomes 2q, 4p, 10q, 15q, 18p, 20q and Xcen.

In stage 2 we genotyped a further 54 markers with a mean heterozygosity of 0.77 (0.51–0.89) in our full sample of 196 (average number genotyped 175.52) ASPs together with their parents and unaffected siblings, where available (Table 1). This allowed the regions on chromosomes 4, 10, 15, 18, 20 and X to be typed at an average spacing of 5.15 cM, while the region of interest on chromosome 2 was typed to 9.55 cM because of its size. Multipoint affected sib pair analysis was performed on the combined data set of the stage 1 and stage 2 samples, rather than treating stage 2 simply as a replication set (14).

The results of simulation studies suggested that the power of this study was >0.95 to detect a susceptibility locus of $\lambda_s = 3$ with a genome-wide significance of 0.05 (14), but only 0.70 to detect a locus of $\lambda_s = 2$ with the conservative assumption that a locus lies midway between two phase 2 markers.

A description of the families from which the 196 ASPs were drawn is presented in Table 1. The results of the multipoint lod scores for both stage 1 and stage 2, together with the results of exclusion analyses based on a λ_s of 3, 2 and 1.5 as calculated using MAPMAKER/SIBS are shown in Figure 1. The results of single point analysis for markers typed in stage 2 with an IBD sharing ≥56% are presented in Table 2. Under the assumption of no dominance variance we were able to exclude (lod score < –2.0) susceptibility genes with an effect of $\lambda_s = 3$ and 2 from 82.8 and 48.7% of the genome, respectively, while genes with an effect size of $\lambda_s = 1.5$ could only be excluded from 9.3% of the genome.

Our most significant result in stage 1 came from chromosome 10p24 (multipoint MLS = 2.26), but after typing a higher density of markers in stage 2 this region only achieved an MLS of 0.48. However, it contained the highest single point MLS of all markers genotyped in our stage 2 sample [D10S542 = 1.92; 62% IBD sharing using SPLINK when maximized over Holmans' 'possible triangle' (16)]. The observed drop in the multipoint MLS was a result of lower sharing at the flanking markers.

Table 1. Summary description of sample studied

Breakdown of sample	Stage 1	Stage 2	Total
Families			
Sibling pair families	73	65	138
Sibling trio families	8	6	14
Sibling quartets/quintets	0	2	2
Total number of affected sib pairs	97	99	196
Number with both parents			50
Number with one parent			56
Number with no parents			90
Number with at least one unaffected sibling			54
Individuals			
Total number of affected subjects	170	157	327
Male	117	100	216
Female	53	57	111
Diagnostic classification			
Schizophrenia/schizophrenia	82	81	163
Schizophrenia/schizoaffective bipolar	7	10	17
Schizophrenia/schizoaffective depressed	5	7	12
Schizoaffective bipolar/schizoaffective depressed	2	1	3
Schizoaffective depressed/schizoaffective depressed	1	0	1

Simulation studies of our entire two-stage procedure suggested that we would have expected to have obtained on average one multipoint MLS of 1.5 per genome scan in the absence of linkage. An MLS of 3 would have been expected only once in every 20 genome scans in the absence of linkage and thus corresponds to a genome-wide significance of 0.05. We obtained three peaks achieving a multipoint MLS >1.5 and, on this basis, the results on chromosomes 4, 18 and X can be considered to satisfy Lander and Kruglyak's definition of 'suggestive' linkage (15). However, none of our findings approached a genome-wide significance of 0.05 corresponding to Lander and Kruglyak's definition of 'significant' linkage.

DISCUSSION

The two stage strategy that we employed required far fewer genotypes than would have been required had we screened the entire sample with markers at 5 cM intervals (70 000 genotypes compared with 300 000) and yet the power of the two approaches was comparable (we had 70 and 95% power to detect at a genome-wide 0.05 significance for $\lambda_s = 2$ and 3, respectively, compared with ~80 and 99% for a complete 5 cM genome screen). Thus, although the average marker interval in stage 1 was relatively high (17 cM), the use of a lax criterion for admission of a region to stage 2 ensured that power was not greatly compromised. Another noteworthy feature of our study is that it was designed to reduce the incidence of false positives. Thus we followed a predetermined study design and did not test for linkage with multiple definitions of the phenotype or multiple genetic models. Moreover, we performed simulation analysis according to our experimental design which allowed us to estimate the genome-wide significance of our results.

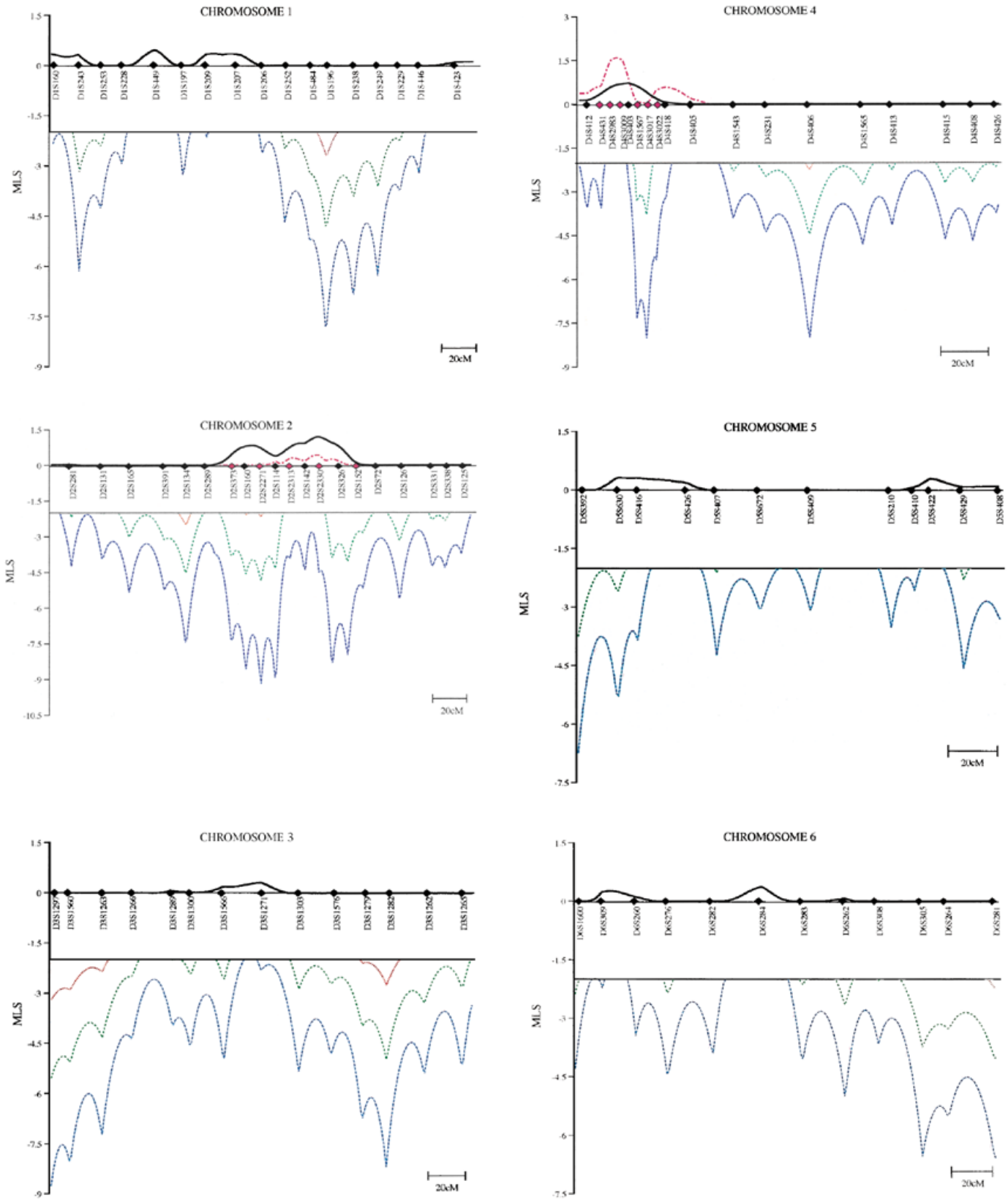

Figure 1. Multipoint maps for each chromosome, generated by MAPMAKER/SIBS. The black and magenta lines represent the maximum lod scores after genotyping in the stage 1 and stage 2 samples, respectively. The positions of the markers typed in stage 1 of the genome screen are shown as black diamonds, with the extra markers genotyped in stage 2 shown as magenta diamonds. Exclusion plots for loci of λ_s = 1.5, 2 and 3 are included for all chromosomes.

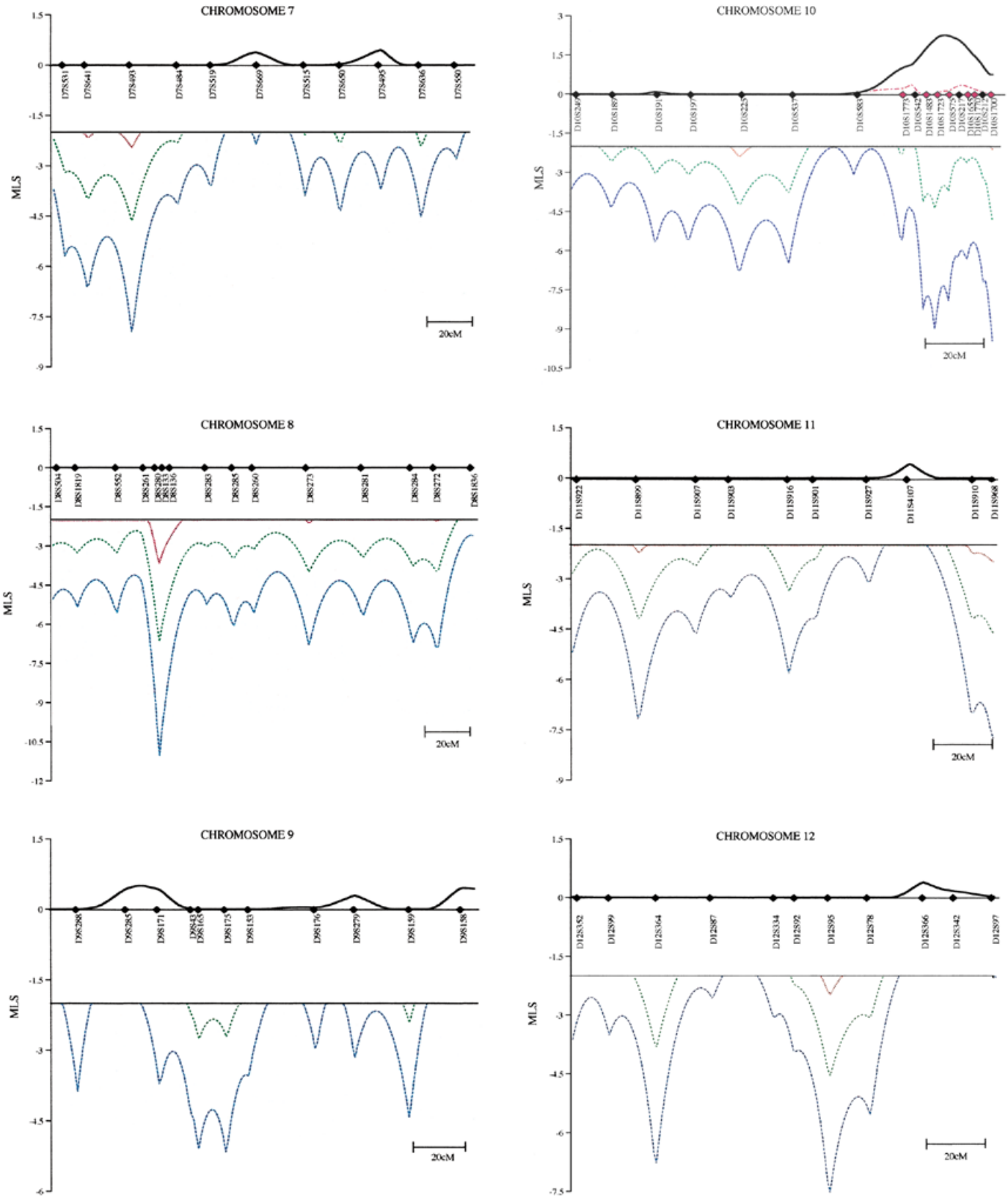

Figure 1. *Continued*

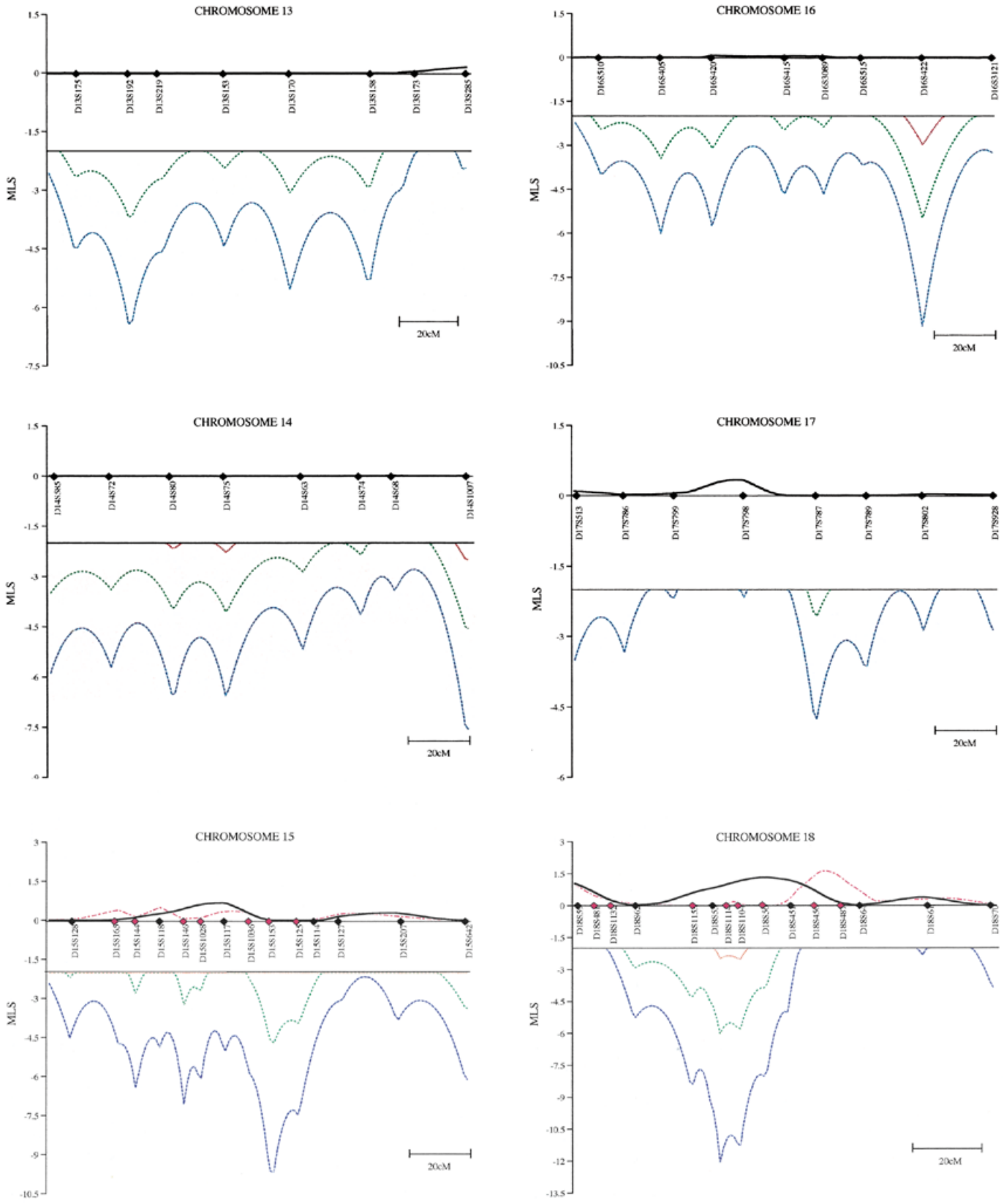

Figure 1. *Continued*

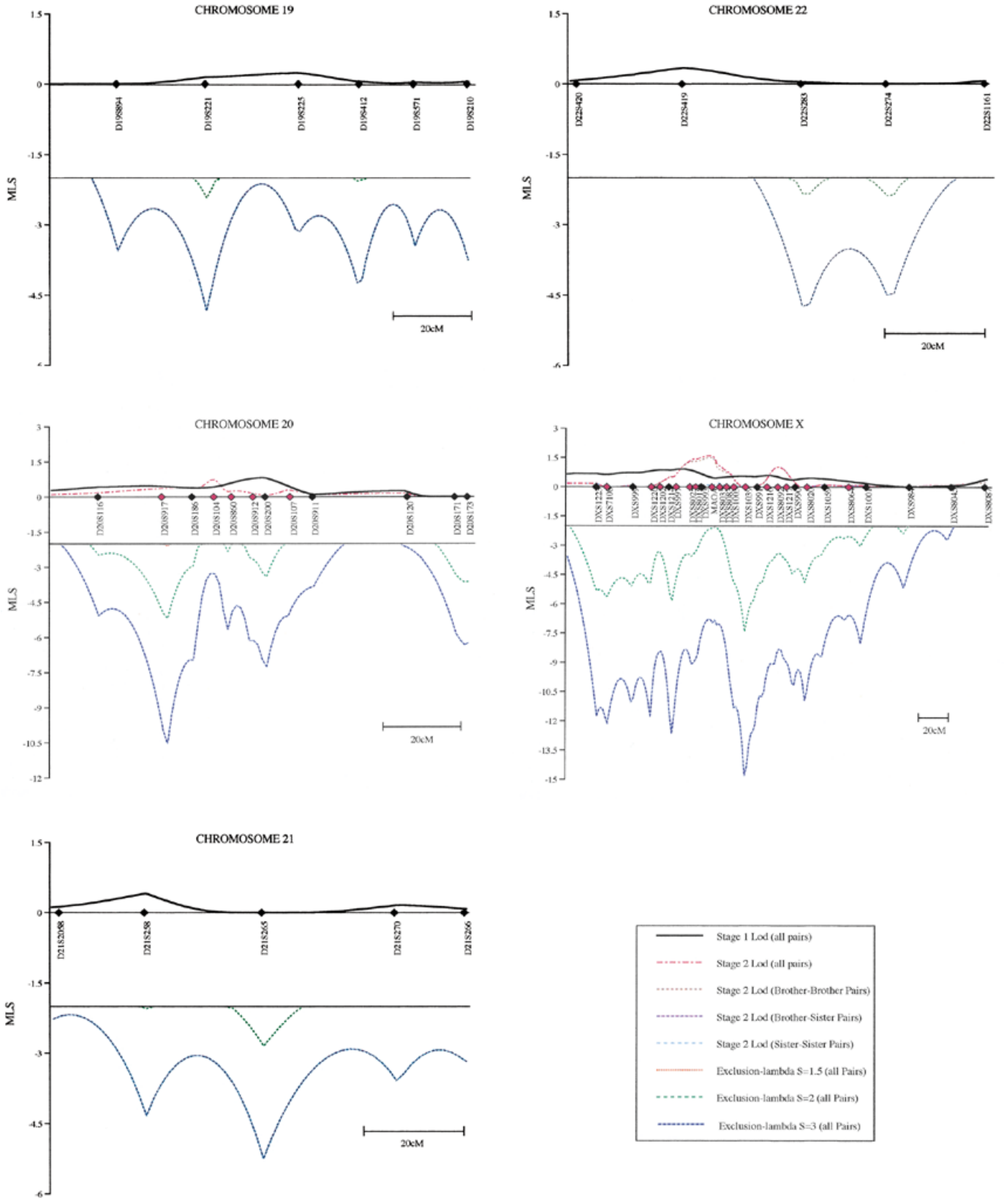

Figure 1. *Continued*

Table 2. The results of single point analysis for markers typed with an IBD sharing ≥56%

Marker	Distance from pter (cM)	IBD (%)	SPLINK MLS	MAPMAKER/SIBS MLS
Chromosome 2	**181**			**0.49**
D2S2313	164	58	0.93	
D2S142	172	56	0.63	
D2S2330	181	56	0.65	
Chromosome 4	**22**			**1.73**
D4S2983	16	60	1.7	
D4S3009	22	60	1.46	
D4S403	25	56	0.74	
D4S3022	40	57	0.89	
D4S418	44	58	0.77	
Chromosome 10	**149**			**0.48**
D10S542	149	62	1.92	
D10S217	168	57	0.97	
Chromosome 15	**33**			**0.55**
D15S165	21	56	0.95	
D15S118	33	57	1.03	
D15S117	52	56	0.78	
D15S1036	58	56	0.96	
Chromosome 18	**73**			**1.62**
D18S451	64	60	1.76	
D18S450	71	59	1.33	
D18S487	80	56	0.7	
Chromosome 20	**38**			**0.99**
D20S186	33	57	0.67	
D20S104	38	59	1.4	
D20S860	42	59	0.89	
Chromosome X	**70**			Total = **1.6**
DXS1214	49	62	0.78(B-B)	(B-B = **1.5**)
DXS997	50	59	0.36(B-B)	
DXS8025	59	63	0.86(B-B)	
DXS8015	62	61	0.53(B-B)	
DXS993	70	67	1.45(B-B)	
MAOA	72	59	0.37(B-B)	
DXS8035	75	62	0.75(B-B)	
DXS8083	79	61	0.45(B-B)	
DXS1003	82	59	0.38(B-B)	
DXS1216	93	56	0.13(B-B)	
DXS8092	98	66	1.33(B-B)	
DXS1217	103	59	0.27(B-B)	

The absence in our results of any locus with overwhelming evidence of linkage suggests that no common genes of major effect are likely to exist for schizophrenia. Moreover, our exclusion mapping supports the observed recurrence risks in relatives (10) which suggest that schizophrenia is unlikely to result from genes of $\lambda_s > 3$, but instead is more likely to result from the action of several or even many genes of moderate or smaller effect. While none of our regions studied in stage 2 reached statistical significance they cannot be entirely discounted, because if these regions contain susceptibility genes of only small to moderate effect then a much larger sample size than our 196 ASPs would be required to replicate these results through linkage.

We identified three regions with maximum lod scores which met the criteria defined by Lander and Krugylak as 'suggestive' (15) by having a probability of occurring by chance less than once per genome scan. While these regions on chromosomes 4, 18 and X fall far short of a level that warrants positional cloning, further families, including singleton cases, are currently being ascertained in order to replicate these findings.

The peak identified on 4p had a multipoint MLS of 1.73 at a region flanked by the markers D4S431 and D4S403. Furthermore, despite the high density of markers we were unable to exclude the presence of a gene of $\lambda_s = 3$ from this region. This region has previously been implicated in a study of a sample of 24 families multiply affected with schizophrenia and related disorders. The overall MLS was 1.12 with D4S403 but an MLS of 1.96 was reported for this marker in a single family which contained cases of schizophrenia and schizoaffective disorder

(17). This is of considerable interest because Blackwood *et al.* (18) have reported a lod score of 3.3 to markers in this region in families multiply affected with bipolar affective disorder. Finally, an independent study has reported an MLS of 2.0 from markers within this region in a sample of two large Danish pedigrees segregating bipolar affective disorder (19). Most family studies have suggested that schizophrenia and bipolar disorder are genetically distinct (20,21). However, the two psychoses share many features in common and can be hard to distinguish clinically. Indeed, cluster analyses have suggested that there is a continuum between the two disorders in clinical terms (22–25) and it is possible that they have susceptibility loci in common, perhaps with different alleles predisposing to each disorder.

A 30 cM region on the X chromosome defined by the markers DXS1214 and DXS8083 yielded a multipoint MLS of 1.6. After stratification according to sex, the 93 brother–brother pairs contributed most of the allele sharing, with marker DXS993 yielding a single point MLS of 1.45 (67% IBD in the brother–brother pairs). There have been previous reports of modest lod scores between schizophrenia and markers in this region of the X chromosome but none of these has even achieved the threshold for suggestive genome-wide significance (26–28). Moreover, interpretation of these findings is complicated by the evidence that this region of the X chromosome shows grandpaternal-specific transmission ratio distortion in males (29). If this finding is confirmed it will mean that the null hypothesis of 0.50 alleles shared IBD in sibling pairs is no longer valid and make interpretation of linkage findings in this region problematical. However, linkage studies in some other disorders have not shown increased allele sharing in this region (30–33), which suggests that this may not be an important source of artefact in studies of this kind.

The final region to show suggestive evidence for linkage was on chromosome 18q near markers D18S450 and D18S487, which gave a multipoint MLS of 1.62. This region appears to coincide with the reports of suggestive linkage to schizophrenia and bipolar disorder by Merette *et al.* (34) (MLS = 3.18 for a combined sample of schizophrenia and bipolar affective disorder) and Ewald *et al.* (35) (MLS = 1.83 for bipolar affective disorder). Previous work by Schwab *et al.* (36) and colleagues has provided suggestive evidence for linkage on 18p22–21 in schizophrenia. While this region lay within the region of positivity obtained in stage 1 of our study, our stage 2 data were able to exclude a locus of $\lambda_s = 2$ from this region. Schwab *et al.* also reported evidence in schizophrenia for transmission distortion of the 124 bp allele of an intronic CA repeat polymorphism in the gene encoding G-olf α on 18p22–21. This is located between D18S53 and D18S1114, in a region from which our data exclude a locus of $\lambda_s \geq 1.5$.

To date, most of the linkage studies looking for schizophrenia susceptibility genes have focused on large multiply affected families. Early studies of large families segregating schizophrenia and related phenotypes initially produced positive findings (37), but unfortunately these could not be replicated. The reasons for this have become clear as data from systematic genome scans have accumulated; highly penetrant mutations causing schizophrenia are at best extremely rare and quite possibly non-existent (38). The 'false positives' were due largely to a combination of multiple testing and the use of statistical methodology and significance levels derived from work on single gene disorders.

In spite of the failure to identify genes of major effect in multiply affected families, moderately significant evidence for linkage has been found in more than one dataset in several chromosomal regions. Areas implicated include chromosomes 6p24–22 (39,40), 8p22–21 (41) and 22q11–12 (42), for which supportive data have also been obtained from international collaborative studies (43,44) and more recent findings, including those on 13q14.1–q32 (45,46), 5q21–q31 (47,48), 18p 22–21 (36) and 10p15–p11 (49–51). The strongest evidence for linkage comes from the findings on 6p24–22 and 13q14.1–q32, but even here the most compelling findings only just achieved genome-wide significance levels of 0.05 and in the case of 6p24–22 this does not survive correction for the use of multiple diagnostic models. Moreover, in no case has a clear pattern of replication emerged with both positive and negative findings reported for each region of interest. Indeed, we failed to find support for any of the above regions in the present study. We have already published data on 6p24–22 from our stage 1 sample using additional markers to those employed in the two-stage genome screen reported here (52). This failed to support findings of linkage in this area and excluded a gene of $\lambda_s = 3$ from the entire region. Similarly, the region 13q14.1–q32 failed meet the criteria for inclusion in stage 2 of our genome screen (MLS $\geq$ 0.74). However, we were unable to exclude such a locus $\lambda_s \geq 3$ from this region. We are therefore currently studying data from a 5 cM grid of markers from this region typed in our full sample of 197 ASPs. It is also possible that we failed to demonstrate linkage in these regions because of heterogeneity. Susceptibility alleles of larger effect may be more common in other samples due to diagnostic, ethnic or geographical differences or because in many instances families containing multiple affected members were studied.

At the present time our findings and those from studies of multiply affected families (44,46,53,54) support the predictions made by Risch (10) from genetic epidemiological data: it is highly unlikely that a commonly occurring locus of effect size $\lambda_s > 3$ exists, but there is suggestive evidence implicating a number of regions which is consistent with the existence of at least some susceptibility alleles of moderate effect ($\lambda_s = 1.5$–3). Simulation studies predict that true linkages may be hard to replicate under oligogenic inheritance (55), unfortunately in no case is the statistical evidence for linkage in schizophrenia yet sufficiently compelling nor the putatively linked areas circumscribed enough to warrant large-scale efforts aimed at cloning disease genes and some or possibly all these findings could be false positives. Allele sharing methods are robust, requiring no assumptions about the mode of transmission. They can also detect genes of small effect, of the magnitude possibly operating in schizophrenia ($\lambda_s = 1.5$–3) in sample sizes that are realistic (600–800 ASPs) (56,57) but considerably larger than those used to date. Therefore, the current challenge for psychiatric genetics is to collect adequately sized samples using a uniform, robust and reliable clinical methodology.

MATERIALS AND METHODS

Families

Families in which two or more siblings met criteria for DSMIV schizophrenia or schizoaffective disorder (58) were ascertained through mental health services and relatives' support groups in

England, Wales, Scotland and Southern Ireland. Any ASPs with diagnoses of schizoaffective bipolar/schizoaffective bipolar were excluded. The sample comprised 327 individuals, grouped as 196 sibling pairs (138 pairs, 14 sibships with three affected, one sibship with four affected and one with five affected). The gender distribution was 216 males and 111 females. All individuals were Caucasian and had been born in the UK or Eire. The mean age at interview was 40.46 years. Sibling pairs that had missing parental genotypes were checked for being half-sibs using the program RELATIVE (59).

Diagnosis

Consensus diagnoses were made by two independent, trained raters (G.McC., R.D.S., A.G.C., L.A.J., K.C.M. or M.Y.G.) based on all available clinical information including a semi-structured interview [PSE-9 (60) or SCAN (61)], examination of case notes and information from relatives and mental health professionals. All interviews were conducted by psychiatrists and psychologists after written consent was obtained following local ethical approval guidelines. Vignettes were prepared on background, interview and case note data.

Clinical information was collated using the OPCRIT checklist (62), the Scales for the Assessment of Positive and Negative Symptoms (SAPS/SANS) and the Global Assessment Scale (GAS). The rating scales were completed by the interviewer and individual members of pairs were rated by separate raters in 69.4% of cases. Each case was separately rated by two independent raters and a consensus diagnosis was reached. If there were any discrepancies the case was brought to a full team meeting so consensus could be reached or the case was excluded. Forty cases were rated against consensus for each rater to obtain their κ score reliability. Inter-rater reliability for diagnosis was excellent, with an average κ score of 0.9 for all raters for the duration of the study.

Markers

All markers were simple sequence repeats. The great majority were dinucleotide repeats (n = 276) with a small number of trinucleotide (n = 2) and tetranucleotide (n = 2) repeats. Most (n = 140) were drawn from the index set developed by Reed *et al.* (63) with further markers being derived from the Genethon (http://www.genethon.fr) (n = 138), Co-operative Human Linkage Centre (http://www.chlc.org/) (n = 2) and Marshfield (http://www.marshmed.org) (n = 3) human genetic linkage maps (64,65). The order of markers in stage 2 was checked by searching for possible double recombinants using Genehunter 2.0 (66) and the estimated inter-marker distances used for multipoint linkage analysis.

Genotyping

After informed consent, blood samples were collected and DNA extracted by standard procedures. When a blood sample could not be obtained, DNA was extracted from 25 ml of saline mouthwash samples by centrifugation at 2000 *g* for 5 min followed by incubation with 50 mg/ml proteinase K and 10% SDS at 50°C for 12 h. The DNA was then isolated by standard phenol/chloroform extraction. PCR was performed in 12 µl reaction volumes containing 48 ng of genomic DNA, 5 pmol of each primer, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris–HCl (pH 8.3), 200 µM dNTPs and 1 U *Taq* polymerase (Amersham Life Science). All PCRs were performed on Hybaid Omnigene thermocyclers with an initial denaturation stage at 94°C for 5 min followed by 30 cycles (30 s at 94°C, 30 s at 52–68°C, 30 s at 72°C) and a final extension step at 72°C for 10 min. Reactions for each marker were performed separately, with products being multiplexed into size-specific sets prior to gel electrophoresis. Markers were fluorescently labelled and typed on ABI373 sequencers using the Genescan and Genotyper software (Applied Biosystems). Seventeen markers were amplified using the conditions described above with radioactively ^{33}P-labelled 5' primers. Radioactive PCR products were resolved on 6% denaturing polyacrylamide gels (Sequagel-6; National Diagnostics) and then autoradiographed for 24 h (BioMax MR-1; Kodak). The subsequent sizes of the alleles were determined by comparison with an M13mp18 sequencing ladder. All genotype and phenotype data was stored in MegaBase (67), which was also used to code the allele sizes into whole numbers, check the genotypes for non-Mendelian inheritance and produce the appropriate files for linkage analysis.

Statistical analysis

The program SPLINK (68) used maximum likelihood methods to calculate the single point lod scores for each marker, which were maximized under the 'possible triangle' restrictions (15). Multipoint analyses were performed using the MAPMAKER/SIBS package (69), which calculates the MLS at each point in the genome by estimating the maximum likelihood sharing probabilities (IBD) for each sib pair. For the pairs which had both parents missing the allele frequencies influence the inferred sharing and if specified incorrectly they can increase the chance of false positive results. We used the allele frequencies estimated by SPLINK from our dataset using maximum likelihood methods. In addition, we used MAPMAKER/SIBS to perform exclusion mapping for genes with a λ_s of 3, 2 and 1.5. A disease locus of fixed effect was placed at each point on the genome in turn and the multipoint lod score calculated. Lods < –2 were taken as evidence that a disease locus at that position with that effect size (or greater) was inconsistent with the data. Genetic effect sizes were specified using the parameter λ_s (10).

Calculation of genome-wide significance levels

Replicates of the genome were simulated under the null hypothesis of there being no disease susceptibility locus for a sample of families identical in structure to the full (stage 2) sample. The marker map used was defined as follows: loci present in the stage 1 or stage 2 analyses of the actual dataset were included in their actual locations, with their observed allele frequencies. Since regions testing positive in stage 1 in the simulated data will not necessarily correspond to those observed in the real data, a stage 2 map was approximated for the regions of the genome not implicated in our actual stage 2 analyses by simulating loci with four equifrequent alleles (heterozygosity = 0.75) in locations which split the distances between existing stage 1 marker loci into equal segments of 4–7 cM. For example, if the inter-marker distance was 12 cM, one extra locus would be added, giving two inter-marker distances of 6 cM each. If the distance was 15 cM, two extra loci would be added, giving three distances of 5 cM each.

Analyses were performed as follows. The individuals and marker loci actually typed in stage 1 were extracted from each

replicate simulated genome and analysed using MAPMAKER/SIBS, to produce a stage 1 analysis. The entire simulated dataset was then analysed using MAPMAKER/SIBS to approximate a stage 2 analysis. The locations of each region of the genome giving a stage 1 lod of 0.7 or greater were determined. The stage 2 lod scores for all locations within a given region were examined and the 'peak' approximated by the highest of these scores. This procedure was repeated on 1000 simulated genomes and the expected number of hits at a given lod score level approximated by the average number of simulated stage 2 peaks reaching or exceeding that level per genome.

A similar method was used to calculate power. Since power to detect a disease locus of given size will depend upon where that locus was in relation to the actual stage 1 map, an approximation was made by assuming that each marker had four equifrequent alleles. A 20 cM grid was used in stage 1 and a 5 cM grid in stage 2. The disease locus was placed midway between stage 2 marker loci, 7.5 cM from the nearest stage 1 marker locus. This was intended to represent the most unfavourable position for detection of linkage and, therefore, to give a conservative estimate of true power. Power was estimated as the proportion of replicates giving both stage 1 and stage 2 lods at the position of the disease locus which satisfied the respective criteria.

ACKNOWLEDGEMENTS

We thank the many mental health professionals and the NSF who assisted with recruitment and all the subjects who took part. This work was supported by an MRC programme grant no. G9309834 to M.O. and P.McG. and by a grant from Glaxo-Wellcome R&D.

REFERENCES

1. McGuffin, P., Owen, M.J. and Farmer, A.E. (1995) Genetic-basis of schizophrenia. *Lancet*, **346**, 678–682.
2. McGuffin, P., Asherson, P., Owen, M.J. and Farmer, A.E. (1994) The strength of the genetic effect—is there room for an environmental influence in the etiology of schizophrenia? *Br. J. Psychiat.*, **164**, 593–599.
3. Cannon, T.D., Kaprio, J., Lonnqvist, J., Huttunen, M. and Koskenvuo, M. (1998) The genetic epidemiology of schizophrenia in a Finnish twin cohort—a population-based modeling study. *Arch. Gen. Psychiat.*, **55**, 67–74.
4. Cardno, A.G., Marshall, E.J., Coid, B., Macdonald, A.M., Ribchester, T.R., Davies, N.J., Venturi, P., Jones, L.A., Lewis, S.W., Sham, P.C., Gottesman, I.I., Farmer, A.E., McGuffin, P., Reveley, A.M. and Murray, R.M. (1999) Heritability estimates for psychotic disorders—the Maudsley twin psychosis series. *Arch. Gen. Psychiat.*, **56**, 162–168.
5. Farmer, A.E., McGuffin, P. and Gottesman, I.I. (1997) Twin concordance for DSM-III schizophrenia: scrutinising the validity of the definition. *Arch. Gen. Psychiat.*, **44**, 634–641.
6. Kendler, K.S., McGuire, M., Gruenberg, A.M., Spellman, M., O'Hare, A. and Walsh, D. (1993) The Roscommon Family Study. II. The risk of nonschizophrenic nonaffective psychoses in relatives. *Arch. Gen. Psychiat.*, **50**, 645–652.
7. Kendler, K.S., McGuire, M., Gruenberg, A.M., O'Hare, A., Spellman, M. and Walsh, D. (1993) The Roscommon Family Study. III. Schizophrenia-related personality disorders in relatives. *Arch. Gen. Psychiat.*, **50**, 781–788.
8. McGue, M. and Gottesman, I.I. (1989) A single dominant gene still cannot account for the transmission of schizophrenia. *Arch. Gen. Psychiat.*, **46**, 478–479.
9. Gottesman, I.I. and Shields, G. (1967) A polygenic theory of schizophrenia. *Proc. Natl Acad. Sci. USA*, **58**, 199–205.
10. Risch, N. (1990) Linkage strategies for genetically complex traits. 2. The power of affected relative pairs. *Am. J. Hum. Genet.*, **46**, 229–241.
11. Owen, M.J. and McGuffin, P. (1993) Association and linkage—complementary strategies for complex disorders. *J. Med. Genet.*, **30**, 638–639.
12 Lander, E.S. and Schork, N.J. (1994) Genetic dissection of complex traits. *Science*, **265**, 2037–2048.
13. Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science*, **273**, 1516–1517.
14. Holmans, P. and Craddock, N. (1997) Efficient strategies for genome scanning using maximum-likelihood affected-sib-pair analysis. *Am. J. Hum. Genet.*, **60**, 657–666.
15. Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits—guidelines for interpreting and reporting linkage results. *Nature Genet.*, **11**, 241–247.
16. Holmans, P. (1993) Asymptotic properties of affected-sib-pair linkage analysis. *Am. J. Hum. Genet.*, **52**, 362–374.
17. Asherson, P., Mant, R., Williams, N., Cardno, A., Jones, L., Murphy, K., Collier, D.A., Nanko, S., Craddock, N., Morris, S., Muir, W., Blackwood, B., McGuffin, P. and Owen, M.J. (1998) A study of chromosome 4p markers and dopamine D5 receptor gene in schizophrenia and bipolar disorder. *Mol. Psychiat.*, **3**, 310–320.
18. Blackwood, D.H., He, L., Morris, S.W., McLean, A., Whitton, C., Thomson, M., Walker, M.T., Woodburn, K., Sharp, C.M., Wright, A.F., Shibasaki, Y., St Clair, D.M., Porteous, D.J. and Muir, W.J. (1996) A locus for bipolar affective-disorder on chromosome 4p. *Nature Genet.*, **12**, 427–430.
19. Ewald, H., Degn, B., Mors, O. and Kruse, T.A. (1998) Support for the possible locus on chromosome 4p16 for bipolar affective disorder. *Mol. Psychiat.*, **3**, 442–448.
20. Maier, W., Lichtermann, D., Minges, J., Hallmayer, J., Heun, R., Benkert, O. and Levinson, D.F. (1993) Continuity and discontinuity of affective-disorders and schizophrenia—results of a controlled family study. *Arch. Gen. Psychiat.*, **50**, 871–883.
21. McGuffin, P., Owen, M.J., O'Donovan, M.C., Thapar, A. and Gottesman, I. (1994) *Seminars in Psychiatric Genetics*. Gaskell, London, UK.
22. Kendell, R.E. and Gourlay, J. (1970) The clinical distinction between the affective psychosis and schizophrenia. *Br. J. Psychiat.*, **117**, 261–266.
23. Taylor, M.A. (1992) Are schizophrenia and affective-disorder related—a selective literature-review. *Am. J. Psychiat.*, **149**, 22–32.
24. Taylor, M.A., Berenbaum, S.A., Jampala, V.C. and Cloninger, C.R. (1993) Are schizophrenia and affective-disorder related—preliminary data from a family study. *Am. J. Psychiat.*, **150**, 278–285.
25. Crow, T.J. (1995) A continuum of psychosis, one human gene, and not much else—the case for homogeneity. *Schizophrenia Res.*, **17**, 135–145.
26. Dann, J., DeLisi, L.E., Devoto, M., Laval, S., Nancarrow, D.J., Shields, G., Smith, A., Loftus, J., Peterson, P., Vita, A., Comazzi, M., Invernizzi, G., Levinson, D.F., Wildenauer, D., Mowry, B.J., Collier, D., Powell, J., Crowe, R.R., Andreasen, N.C., Silverman, J.M., Mohs, R.C., Murray, R.M., Walters, M.K., Lennon, D.P. and Crow, T.J. (1997) A linkage study of schizophrenia to markers within Xp11 near the MAOB gene. *Psychiat. Res.*, **70**, 131–143.
27. DeLisi, L.E., Laval, S., Stewart, J., Wellman, N., Smith, A., Razi, K., Dann, J. and Crow, T.J. (1998) Progress in locating a gene for schizophrenia on the sex chromosomes. *Am. J. Med. Genet.*, **81**, 470.
28. Hovatta, I., Varilo, T., Suvisaari, J., Terwilliger, J.D., Ollikainen, V., Arajarvi, R., Juvonen, H., Kokko-Sahin, M.-L., Visnen, L., Mannila, H., Lnnqvist, J. and Peltonen, L. (1998) A genome-wide search for schizophrenia genes in an internal isolate of Finland suggesting multiple susceptibility loci. *Am. J. Med. Genet.*, **81**, 453–454.
29. Naumova, A.K., Leppert, M., Barker, D.F., Morgan, K. and Sapienza, C. (1998) Parental origin-dependant, male offspring-specific transmission-ratio distortion at loci on the human X chromosome. *Am. J. Hum. Genet.*, **62**, 1493–1499.
30. Reich, T., Goate, A., Edenberg, H., Rice, J., Foroud, T., Hesselbrock, V., Schuckit, M., Porjesz, B., Nurenberg, J., Crowe, R. and Begleiter, H. (1998) Replicating genetic linkage in the collaborative study on the genetics of alcoholism (COGA). *Am. J. Med. Genet.*, **81**, 478–479.
31. Davies, J.L., Kawaguchi, Y., Bennett, S.T., Copeman, J.B., Cordell, H.J., Pritchard, L.E., Reed, P.W., Gough, S.C.L., Jenkins, S.C., Palmer, S.M., Balfour, K.M., Rowe, B.R., Farrall, M., Barnett, A.H., Bain, S.C. and Todd, J.A. (1994) Genome-wide search for human type-1 diabetes susceptibility genes. *Nature*, **371**, 130–136.
32. International Molecular Genetic Study of Autism Consortium (1998) A full genome screen for autism with evidence for linkage to a region on chromosome 7q. *Hum. Mol. Genet.*, **7**, 571–578.

33. Hanis, C.L., Boerwinkle, E., Chakraborty, R., Ellsworth, D.L., Concannon, P., Stirling, B., Morrison, V.A., Wapelhorst, B., Spielman, R.S., Gogolin-Ewens, K.J., Shepard, J.M., Williams, S.R., Risch, N., Hinds, D., Iwasaki, N., Ogata, M., Omori, Y., Petzold, C., Rietzch, H., Schroder, H.E., Schulze, J., Cox, N.J., Menzel, S., Boriraj, V.V. and Chen, X. (1996) A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nature Genet.*, **13**, 161–166.
34. Merette, C., Bissonnette, L., Rouillard, E., Roy, M.A. and Maziade, M. (1997) Positive linkage results for bipolar disorder on 18q in large kindreds from eastern Quebec. *Am. J. Med. Genet.*, **74**, 674.
35. Ewald, H., Mors, O., Koed, K., Eiberg, H. and Kruse, T.A. (1997) Susceptibility locus for bipolar affective disorder on chromosome 18? *Am. J. Med. Genet.*, **74**, 672–673.
36. Schwab, S.G., Hallmayer, J., Lerer, B., Albus, M., Borrmann, M., Honig, S., Strauss, M., Segman, R., Lichtermann, D., Knapp, M., Trixler, M., Maier, W. and Wildenauer, D.B. (1998) Support for a chromosome 18p locus conferring susceptibility to functional psychoses in families with schizophrenia, by association and linkage analysis. *Am. J. Hum. Genet.*, **63**, 1139–1152.
37. Sherrington, R., Brynjolfsson, J., Petursson, H., Potter, M., Dudleston, K., Barraclough, B., Wasmuth, J., Dobbs, M. and Gurling, H. (1988) Localization of a susceptibility locus for schizophrenia on chromosome-5. *Nature*, **336**, 164–167.
38. McGuffin, P. and Owen, M.J. (1996) Molecular genetic studies of schizophrenia. *Cold Spring Harbor Symp. Quant. Biol.*, **61**, 815–822.
39. Wang, S., Sun, C.E., Walczak, C.A., Ziegle, J.S., Kipps, B.R., Goldin, L.R. and Diehl, S.R. (1995) Evidence for a susceptibility locus for schizophrenia on chromosome 6pter–p22. *Nature Genet.*, **10**, 41–46.
40. Straub, R.E., MacLean, C.J., O'Neill, F.A., Burke, J., Murphy, B., Duke, F., Shinkwin, R., Webb, B.T., Zhang, J. and Walsh, D. (1995) A potential vulnerability locus for schizophrenia on chromosome 6p24–22—evidence for genetic-heterogeneity. *Nature Genet.*, **11**, 287–293.
41. Pulver, A.E., Lasseter, V.K., Kasch, L., Wolyniec, P., Nestadt, G., Blouin, J.L., Kimberland, M., Babb, R., Vourlis, S. and Chen, H. (1995) Schizophrenia—a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. *Am. J. Med. Genet.*, **60**, 252–260.
42. Pulver, A.E., Karayiorgou, M., Wolyniec, P.S., Lasseter, V.K., Kasch, L., Nestadt, G., Antonarakis, S., Housman, D., Kazazian, H.H. and Meyers, D. (1994) Sequential strategy to identify a susceptibility gene for schizophrenia—report of potential linkage on chromosome 22q12–q13.1. *Am. J. Med. Genet.*, **54**, 36–43.
43. Pulver, A.E., Karayiorgou, M., Lasseter, V.K., Wolyniec, P., Kasch, L., Antonarakis, S., Housman, D., Kazazian, H.H., Meyers, D. and Nestadt, G. (1994) Follow-up of a report of a potential linkage for schizophrenia on chromosome 22q12–q13.1.2. *Am. J. Med. Genet.*, **54**, 44–50.
44. Levinson, D.F., Wildenauer, D.B., Schwab, S.G., Albus, M., Hallmayer, J., Lerer, B., Maier, W., Blackwood, D., Muir, W., St Clair, D., Morris, S., Moises, H.W., Yang, L., Kristbjarnarson, H., Helgason, T., Wiese, C., Collier, D.A., Holmans, P., Daniels, J., Rees, M., Asherson, P., Roberts, Q., Cardno, A., Arranz, M.J., Vallada, H. *et al*. (1996) Additional support for schizophrenia linkage on chromosome-6 and chromosome-8—A multicenter study. *Am. J. Med. Genet.*, **67**, 580–594.
45. Lin, M.W., Curtis, D., Williams, N., Arranz, M., Nanko, S., Collier, D., McGuffin, P., Murray, R., Owen, M. and Gill, M. (1995) Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1–q32. *Psychiat. Genet.*, **5**, 117–126.
46. Blouin, J.L., Dombroski, B.A., Nath, S.K., Lasseter, V.K., Wolyniec, P.S., Nestadt, G., Thornquist, M., Ullrich, G., McGrath, J., Kasch, L., Lamacz, M., Thomas, M.G., Gehrig, C., Radhakrishna, U., Snyder, S.E., Balk, K.G., Neufeld, K., Swartz, K.L., DeMarchi, N., Papadimitriou, G.N., Dikeos, D.G., Stefanis, C.N., Chakravarti, A., Childs, B. and Pulver, A.E. (1998) Schizophrenia susceptibility loci on chromosomes 13q32 and 8p2. *Nature Genet.*, **20**, 70–73.
47. Schwab, S.G., Eckstein, G.N., Hallmayer, J., Lerer, B., Albus, M., Borrmann, M., Lichtermann, D., Ertl, M.A., Maier, W. and Wildenauer, D.B. (1997) Evidence suggestive of a locus on chromosome 5q31 contributing to susceptibility for schizophrenia in German and Israeli families by multipoint affected sib-pair linkage analysis. *Mol. Psychiat.*, **2**, 156–160.
48. Straub, R.E., Maclean, C.J., O'Neill, F.A., Walsh, D. and Kendler, K.S. (1997) Support for a possible schizophrenia vulnerability locus in region 5q22–31 in Irish families. *Mol. Psychiat.*, **2**, 148–155.
49. Straub, R.E., MacLean, C.J., Martin, R.B., Ma, Y., Myakishev, M.V., Harris-Kerr, C., Webb, B.T., O'Neill, F.A., Walsh, D. and Kendler, K.S. (1998) A schizophrenia locus may be located in region 10p15–p11. *Am. J. Med. Genet.*, **81**, 296–301.
50. Schwab, S.G., Hallmayer, J., Albus, M., Lerer, B., Hanses, C., Kanyas, K., Segman, R., Borrman, M., Dreikorn, B., Lichtermann, D., Rietschel, M., Trixler, M., Maier, W. and Wildenauer, D.B. (1998) Further evidence for a susceptibility locus on chromosome 10p14–p11 in 72 families with schizophrenia by nonparametric linkage analysis. *Am. J. Med. Genet.*, **81**, 302–307.
51. Faraone, S.V., Matise, T., Svrakic, D., Pepple, J., Malaspina, D., Suarez, B., Hampe, C., Zambuto, C.T., Schmitt, K., Meyer, J., Markel, P., Lee, H., Harkavy Friedman, J., Kaufmann, C., Cloninger, C.R. and Tsuang, M.T. (1998) Genome scan of European-American schizophrenia pedigrees: results of the nimh genetics initiative and millennium consortium. *Am. J. Med. Genet.*, **81**, 290–295.
52. Daniels, J.K., Spurlock, G., Williams, N.M., Cardno, A.G., Jones, L.A., Murphy, K.C., Asherson, P., Holmans, P., Fenton, I., McGuffin, P. and Owen, M.J. (1997) Linkage study of chromosome 6p in sib-pairs with schizophrenia. *Am. J. Med. Genet.*, **74**, 319–323.
53. Levinson, D.F., Mahtani, M.M., Nancarrow, D.J., Brown, D.M., Kruglyak, L., Kirby, A., Hayward, N.K., Crowe, R.R., Andreasen, N.C., Black, D.W., Silverman, J.M., Endicott, J., Sharpe, L., Mohs, R.C., Siever, L.J., Walters, M.K., Lennon, D.P., Jones, H.L., Nertney, D.A., Daly, M.J., Gladis, M. and Mowry, B.J. (1998) Genome scan of schizophrenia. *Am. J. Psychiat.*, **155**, 741–750.
54. Moises, H.W., Yang, L., Kristbjarnarson, H., Wiese, C., Byerley, W., Macciardi, F., Arolt, V., Blackwood, D., Liu, X. and Sjogren, B. (1995) An international 2-stage genome-wide search for schizophrenia susceptibility genes. *Nature Genet.*, **11**, 321–324.
55. Suarez, B.K., Hampe, C.L. and Van Eerdewegh, P. (1994) Problems of replicating linkage claims in psychiatry. In Gershon, E.S. and Cloninger, C.R. (eds), *Genetic Approaches to Mental Disorders*. American Psychiatric Press, Washington, DC, pp. 23–46.
56. Hauser, E.R., Boehnke, M., Guo, SW. and Risch, N.J. (1996) Affected sib pair interval mapping and exclusion for complex traits: sampling considerations. *Genet. Epidemiol.*, **13**, 117–137.
57. Scott, W.K., Pericak-Vance, M.A. and Haines, J.L. (1997) Genetic analysis of complex diseases. *Science*, **275**, 1327.
58. American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Press, Washington, DC.
59. Goring, H.H.H. and Ott, J. (1995) Verification of sib relationship without knowledge of parental genotypes. *Am. J. Hum. Genet.*, **57**, 1102.
60. Wing, J.K., Cooper, J.E. and Satorius, N. (1974) *The Measurement and Classification of Psychiatric Illness*. Cambridge University Press, Cambridge, UK.
61. Wing, J.K., Babor, T. and Brugha, T. (1990) SCAN: schedules for the clinical assessment in neuropsychiatry. *Arch. Gen. Psychiat.*, **47**, 589–593.
62. McGuffin, P., Farmer, A.E. and Harvey, I. (1992) A polydiagnostic application of operational criteria in studies of psychotic illness: development and reliability of the OPCRIT system. *Arch. Gen. Psychiat.*, **48**, 643–647.
63. Reed, P.W., Davies, J.L., Copeman, J.B., Bennett, S.T., Palmer, S.M., Pritchard, L.E., Gough, S.C., Kawaguchi, Y., Cordell, H.J. and Balfour, K.M. (1994) Chromosome-specific microsatellite sets for fluorescence-based, semiautomated genome mapping. *Nature Genet.*, **7**, 390–395.
64. Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. and Weissenbach, J. (1996) A comprehensive genetic-map of the human genome based on 5,264 microsatellites. *Nature*, **380**, 152–154.
65. Dubovsky, J., Sheffield, V.C., Duyk, G.M. and Weber, J.L. (1995) Sets of short tandem repeat polymorphisms for efficient linkage screening of the human genome. *Am. J. Hum. Genet.*, **4**, 449–452.
66. Kruglyak, L., Daly, M.J., ReeveDaly, M.P. and Lander, E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.*, **58**, 1347–1363.
67. Fenton, I. and Sandkuijl, L.A. (1992) MEGABASE-PKD: a genetic database for polycystic kidney disease. *Contrib. Nephrol.*, **97**, 118–127.
68. Holmans, P. and Clayton, D. (1995) Efficiency of typing unaffected relatives in an affected-sib-pair linkage study with single-locus and multiple tightly linked markers. *Am. J. Hum. Genet.*, **57**, 1221–1232.
69. Kruglyak, L. and Lander, E.S. (1995) Complete multipoint sib pair analysis of qualitative and quantitaive traits. *Am. J. Hum. Genet.*, **57**, 439–454.