

ORIGINAL RESEARCH ARTICLE

Genome-wide scans of three independent sets of 90 Irish multiplex schizophrenia families and follow-up of selected regions in all families provides evidence for multiple susceptibility genes

RE Straub^{1,3}, CJ MacLean^{1,3}, Y Ma^{1,3}, BT Webb^{1,3}, MV Myakishev⁷, C Harris-Kerr⁶, B Wormley^{1,3}, H Sadek⁸, B Kadambi^{1,3}, FA O'Neill⁴, D Walsh⁵ and KS Kendler^{1,2,3}

¹Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA; ²Department of Human Genetics, Virginia Commonwealth University, Richmond, VA, USA; ³Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA; ⁴The Department of Psychiatry, The Queens University, Belfast, Northern Ireland; ⁵The Health Research Board, Dublin, Ireland; ⁶Genaisance Pharmaceuticals, Inc, New Haven, CT, USA; ⁷Laboratory of Biochemistry, NIH/NCI, Bethesda, MD, USA; ⁸Commonwealth Biotechnologies, Inc, Richmond, VA, USA

From our linkage study of Irish families with a high density of schizophrenia, we have previously reported evidence for susceptibility genes in regions 5q21–31, 6p24–21, 8p22–21, and 10p15–p11. In this report, we describe the cumulative results from independent genome scans of three *a priori* random subsets of 90 families each, and from multipoint analysis of all 270 families in ten regions. Of these ten regions, three (13q32, 18p11–q11, and 18q22–23) did not generate scores above the empirical baseline pairwise scan results, and one (6q13–26) generated a weak signal. Six other regions produced more positive pairwise and multipoint results. They showed the following maximum multipoint H-LOD (heterogeneity LOD) and NPL scores: 2p14–13: 0.89 ($P = 0.06$) and 2.08 ($P = 0.02$), 4q24–32: 1.84 ($P = 0.007$) and 1.67 ($P = 0.03$), 5q21–31: 2.88 ($P = 0.0007$), and 2.65 ($P = 0.002$), 6p25–24: 2.13 ($P = 0.005$) and 3.59 ($P = 0.0005$), 6p23: 2.42 ($P = 0.001$) and 3.07 ($P = 0.001$), 8p22–21: 1.57 ($P = 0.01$) and 2.56 ($P = 0.005$), 10p15–11: 2.04 ($P = 0.005$) and 1.78 ($P = 0.03$). The degree of 'internal replication' across subsets differed, with 5q, 6p, and 8p being most consistent and 2p and 10p being least consistent. On 6p, the data suggested the presence of two susceptibility genes, in 6p25–24 and 6p23–22. Very few families were positive on more than one region, and little correlation between regions was evident, suggesting substantial locus heterogeneity. The levels of statistical significance were modest, as expected from loci contributing to complex traits. However, our internal replications, when considered along with the positive results obtained in multiple other samples, suggests that most of these six regions are likely to contain genes that influence liability to schizophrenia.

Molecular Psychiatry (2002) 7, 542–559. doi:10.1038/sj.mp.4001051

Keywords: schizophrenia; linkage; genome scan; Irish

Introduction

Schizophrenia is a relatively common, often chronic and debilitating mental illness. Evidence from a variety of types of studies has clearly demonstrated that genetic factors contribute substantially to the etiology.^{1,2} Several lines of evidence show that the mode of transmission is complex, including multiple, potentially interacting genes, along with locus heterogeneity and

incomplete penetrance.^{2–5} As with most complex diseases, despite tremendous efforts devoted to segregation, epidemiological, family, adoption, and twin studies, little detail is known about the fundamental features of the genetic architecture. Without a good estimate of the number of important genes, plausible hypotheses about their mechanism of action, or some idea of the relative roles of epistasis and heterogeneity, debates over 'the' optimal set of ascertainment and analytic strategies will inevitably continue.^{6,7} Furthermore, many power analyses^{8–11} have shown that the sample sizes needed to detect susceptibility loci for complex diseases by linkage methods alone are far greater than for Mendelian disorders. Since collection of large samples is difficult and expensive, false-negative linkage results may remain a major problem. These

Correspondence: RE Straub at his current address: Clinical Brain Disorders Branch, IRP, NIMH, NIH, Bldg 10, Room 4N-311, MSC 1385, Bethesda, MD 20892–1385, USA. E-mail: straubr@intr.nimh.nih.gov

Received 16 January 2001; revised 9 October 2001; accepted 22 October 2001

difficulties are compounded by the heterogeneity and variability of the symptoms that comprise the clinical phenotype and by the limited flexibility and functionality of current genetics software. It is no surprise, therefore, that susceptibility genes have not yet been identified by either the positional or functional candidate approaches.

However, the availability of the complete human genomic sequence, along with technology advances and cost reductions in single nucleotide polymorphism (SNP) identification and genotyping, are compelling reasons to be cautiously optimistic. In parallel to these two major developments, incremental but important progress has been made due to the coordination of genome-wide scans with more targeted replication studies of promising regions by worldwide consortia.^{12–15} Many regions now have support from multiple independent studies, although due to major methodological differences between studies, the cumulative strength of the evidence for each locus remains difficult to judge. Recent reviews,^{16–20} when taken together, provide a fairly comprehensive catalog of the most positive studies, but it is likely that negative results are substantially under-represented in the literature. At the moment, the following ten regions are, arguably, those most deserving of higher density mapping and follow-up studies: 1q21–22,^{19,21–23} 1q32–42,^{19,24–28} 5q21–35,^{29–33} 6p24–21,^{14,34–38} 6q13–26,^{39,40} 8p22–21,^{14,32,41–44} 10p15–11,^{45–47} 13q14–32,^{43,48–51} 15q13–15,^{40,52–54} and 22q11–13.^{12,13,55–61} A second tier of regions of interest might include 2p14–13,^{36,46,56,59,62} 2q12–13,^{36,46,63} 4q24–32,^{32,40,56,63,64} 9q21–22,^{36,64} 10q11–23,^{46,59,63,65,66} 18p11–q11⁶⁷ and Xp11.^{64,68–71}

The Irish Study of High Density Schizophrenia Families (ISHDSF) contains 1425 individuals with DNA available in 270 families systematically ascertained from psychiatric facilities in Ireland and Northern Ireland. In previous studies with these families we have, in conjunction with other research groups, generated support for the presence of four loci influencing the vulnerability to schizophrenia, on chromosomes 5q21–31,²⁹ 6p24–p21,³⁴ 8p22–21,⁴² and 10p15–11.⁴⁵ In these previous reports, we analyzed all families together, and the 6p report did not contain multipoint results. In this report, we describe the results from additional genotyping on 5q and 6p, and of pairwise linkage analysis of three independent genome scans, performed on subsamples of 90 families each. We also report multipoint results from each subsample and from all 270 families after follow-up of ten regions of current interest.

Materials and methods

Pedigree ascertainment and diagnostic assessment

A detailed description of the Irish Study of High Density Schizophrenia Families (ISHDSF) has been published.⁷² The linkage sample that was used for our previously reported studies of chromosomes 5q, 6p, 8p, and 10p contained 1408 individuals from 265 systematically ascertained multiplex schizophrenia families collected in Ireland and Northern Ireland. The linkage

sample that was used for the majority of the scan and for most of the follow-up analyses reported on here contained 1425 individuals from 270 families. Most of the 17 additional individuals were genotyped from the start, but their genotypes were included in the analyses only after we resolved familial inconsistencies by identity testing, and substitution of the correct DNA aliquot. Their inclusion had a negligible effect on the results.

The diagnostic instruments included the Structured Interview for DSM-III-R Diagnosis (SCID)⁷³ and the Structured Interview for Schizotypy (SIS).⁷⁴ Diagnosis was made, using DSM-III-R criteria, based on all available information (personal history, hospital record, family history report) by individuals blind to knowledge of genotypes and to the psychopathology of relatives. The diagnostic categories are as follows. NARROW (categories D1 through D2; 625 affecteds, 565 genotyped): ‘core schizophrenic phenotypes’; schizophrenia, poor outcome schizoaffective disorder (PO-SAD) and simple schizophrenia.⁷⁵ INTERMEDIATE (categories D1 through D5; 804 affecteds, 691 genotyped): a Narrow definition of the schizophrenia spectrum plus disorders which have been repeatedly shown to co-aggregate in families with narrowly defined schizophrenia.⁷⁶ This category adds to the Narrow definition schizotypal personality disorder, and all other nonaffective psychotic disorders (ie schizophreniform disorder, delusional disorder, atypical psychosis and good-outcome SAD). BROAD (categories D1 through D8; 888 affecteds, 744 genotyped): a broad definition of the schizophrenia spectrum and includes all disorders which significantly aggregated in relatives of schizophrenic probands in the Roscommon Family Study,^{77–79} an epidemiologic, controlled family study conducted in parallel in the west of Ireland. This category adds to the Intermediate definition mood incongruent and mood congruent psychotic affective illness, and paranoid, avoidant and schizoid personality disorder. VERY BROAD (categories D1 through D9; 1172 affecteds, 952 genotyped). This category adds to the Broad definition all other psychiatric disorders (eg nonpsychotic affective disorders, anxiety disorders, alcoholism and other non-schizophrenia spectrum personality disorders).

Genome scan—split sample design

The most severe problem encountered during a genome scan is that of type 2 errors (ie false negatives), which will, by definition, remain undetected. Therefore we sought study design features that might reduce the impact of those variables that hinder the ability to detect linkage. Both simulations and empirical results have consistently shown that for traits in which both the phenotype and the genetic architecture are complex, the relationship between the statistical significance of the linkage results for a region and the true likelihood of a susceptibility gene being present is far from straightforward.^{80–90} To further assume that it is a direct one and rely only on the absolute values of the linkage statistics⁹¹ is unfounded and unwise. Instead,

independent replication is the most useful statistical approach to distinguishing true from false positives. Prior to the start of the genome scan, we divided the 270 families at random into three subsets of 90 families each (Family sets A, B, and C). We have analyzed a number of important variables (family size distribution, number of affecteds, diagnostic classifications, geographical location, etc), and as expected, the characteristics of the three sets were similar. One notable difference is that while the number of affecteds genotyped is similar due to chance, the total number of affected sib pairs is greatest for Family set A and least for C.

Markers and maps The markers were predominantly tri- and tetra-nucleotide repeat microsatellites generated by the Cooperative Human Linkage Center (CHLC), and many of these are in the Weber screening set, version 8.0. For the scan, we used the marker order, estimated distances to the telomeres, and sex-averaged intermarker distances⁹² from maps ('Mfd98') provided by the Marshfield Medical Research Center. For multipoint analysis of the followup regions, the maps were constructed directly from the ISHDSF genotyping data using the program CRI-MAP.⁹³ A total of 166 markers were used; the total Mfd98 map distance was 472 cM and IHDS map distance was 396 cM. These regions were as follows (see bottom of page):

Genotyping and error checking

The genotyping methods have been described.^{94,95} Mendelian inheritance was checked using the program Geno (version 1.1 from the Marshfield Center for Medical Genetics), and marker data used for multipoint analysis was further checked for double recombinants using the program CRI-MAP.⁹³

Linkage analysis Pairwise LOD scores (H-LODs) were calculated under the assumption of heterogeneity using the program MENDEL.⁹⁶ The parameters used have been published,³⁴ and we have kept these parameters constant for all work to date except where noted. We have made no attempts to increase the significance of the findings by using additional programs or by varying the parameters or conditions of the analy-

sis. Pairwise H-LODs calculated using GENEHUNTER (v1.3)⁹⁷ gave very similar results for a test set of 20 markers over a large range of scores (>3.0 to <-3.0, data not shown). Marker allele frequencies were calculated from all available unrelated individuals ($n > 250$). We used four genetic models: dominant (Dom), additive on the penetrance scale (intermediate heterozygotic; Pen), additive on the liability scale (Lia) and recessive (Rec). We present only the maximum H-LOD score for each marker. Using the program GENEHUNTER (v1.3),⁹⁷ multipoint H-LODs and nonparametric linkage (NPL) scores were generated over the follow-up regions (where results are expressed in LOD and normal deviate (z) units, respectively). We report only the NPL_{all} statistic, which is calculated based on group-wise S(all) allele sharing among affecteds in sibships. For all marker-Family Set combinations, we ran 16 (4 genetic models \times 4 diagnostic categories) pairwise H-LOD calculations. In addition, for about one-third of the markers, we also ran pairwise NPL calculations, but under diagnostic categories Narrow and Broad only. The present report is both explicit and complete in terms of the number and types of tests performed. The four genetic models used are not independent and neither are the four diagnostic models, thus a correction based on 16 tests would be unduly conservative—no corrections were made.

Multiple testing The debate over the relative utility of parametric and non-parametric tests is unresolved, so we used both. Theoretically, non-parametric analyses are inherently less powerful than parametric analyses run using a suitably correct model,⁹⁸ and the LOD score will tend to maximize under the 'correct' model.^{98,99} However, not only is this model unknown, it is certainly not the same for all loci, families, and populations. The fieldwork to ascertain the families, plus the clinical diagnostic work thereafter, represents a tremendous investment of time and resources. To increase the chances of having included at least one analysis that is correct enough to generate a signal that is readily identifiable above background, we routinely ran many tests. The inflation of the type 1 error (ie false positive) rate we have incurred due to multiple testing is probably significant, but we think that this is far less detrimental than missing a true linkage. We expected

Region	No.	Marker	cM	Marker	cM	Mfd98	IHDS
2p14-13	23	D2S405	48.0	GATA113F01	161.3	113.3	83.7
4q24-32	22	D4S3250	126.2	D4S3335	195.1	68.9	64.1
5q21-31	35	D5S815	101.0	D5S658	142.9	41.9	36.1
6p24-21	34	D6S477	9.2	D6S291	49.5	40.3	33.0
6q13-26	6	D6S445	91.3	D6S262	130.0	38.7	34.3
8p21-11	15	D8S552	26.4	D8S283	60.9	34.4	28.2
10p15-11	12	D10S189	19.0	D10S183	70.2	51.2	36.2
13q32	7	D13S170	63.9	D13S173	93.5	29.6	31.8
18p11-q11	5	D18S53	41.2	D18S847	56.7	15.5	14.2
18q21-23	7	D18S41	80.4	D18S554	119.4	39.0	35.2

the split sample design to effectively control type 1 errors, and in practice we found that the use of the two remaining subsamples for 'internal replication' served to rapidly identify many signals (data not shown) that were likely to be false. With three subsamples, we had the opportunity to replicate not once, but twice, before reporting our results to other groups, and we did so before preparing our four previous chromosome specific reports.

Results

Sample characteristics

Some of the characteristics of the samples that are most relevant to the linkage results are shown in Table 1. Most (88%) of the genome was scanned using individual Family Sets, but it should be noted that each set contains more families (90) and, under the Broad diagnostic category, contains more affected individuals (287–309) and more total affected sib pairs (188–242) than most of the samples used in published genome scans. Since power considerations are central, the most germane comparisons between our results and those from other studies are those that use the results from *each* of the three Family Sets *separately*, rather than simply using the results from the total sample. Eventu-

ally, all 270 families were tested in ten regions (totaling about 12% of the genome), for which the prior probability of linkage was higher.

The actual course of the scan

We carried out three inter-digitating genome scans in parallel, testing three random subsets of families (sets A, B, and C), each initially at low resolution (>30 cM). For each Family Set, the majority of the genome was covered only at medium density, with sizable gaps, and so multipoint analysis was not possible for most of the genome. After each round of marker testing, we relied therefore on the pairwise results to prioritize further genotyping. Those regions with *multiple adjacent* markers that were positive (H-LOD greater than 0.5) in *more than one* Family Set were followed up. This was done by testing the other two subsets with the same markers, testing additional markers, and finally, for some regions, by multipoint analysis of all 270 families (ie Family Set ABC). During the course of these three genome scans, we also tested internal follow-up markers and many markers in other regions that were either positive in other studies or that contained 'functional' candidate genes. As a result of this mixture of scan and extensive follow-up work, marker spacing and progress on the individual scans was not systematic, and in the

Table 1 Selected characteristics of the samples

	Family Set			
	A	B	C	ABC
No. families	90	90	90	270
No. individuals (genotyped)	873 (478)	901 (488)	927 (459)	2701 (1425)
No. individuals per family (genotyped)	9.7 (5.3)	10.0 (5.4)	10.3 (5.1)	10.0 (5.3)
No. males (genotyped)	466 (259)	488 (267)	514 (257)	1468 (783)
No. females (genotyped)	407 (219)	413 (221)	413 (202)	1233 (642)
Diagnostic Categories				
Narrow (D1–D2)				
No. affected (genotyped)	215 (193)	206 (190)	204 (182)	625 (565)
No. affected sib-pairs: total (unique)	108 (70)	108 (61)	103 (65)	319 (196)
No. affected sib-pairs genotyped: total (unique)	92 (66)	88 (55)	81 (56)	261 (177)
Intermediate (D1–D5)				
No. affected (genotyped)	276 (239)	261 (229)	267 (223)	804 (691)
No. affected sib-pairs: total (unique)	184 (88)	168 (82)	154 (83)	506 (253)
No. affected sib-pairs genotyped: total (unique)	141 (78)	132 (70)	109 (70)	382 (218)
Broad (D1–D8)				
No. affected (genotyped)	309 (264)	287 (246)	292 (234)	888 (744)
No. affected sib-pairs: total (unique)	242 (92)	203 (84)	188 (89)	633 (265)
No. affected sib-pairs genotyped: total (unique)	181 (83)	159 (73)	125 (72)	465 (228)
Very Broad (D1–D9)				
No. affected (genotyped)	393 (325)	385 (319)	394 (308)	1172 (952)
No. affected sib-pairs: total (unique)	419 (104)	389 (99)	350 (103)	1158 (306)
No. affected sib-pairs genotyped: total (unique)	300 (94)	281 (87)	224 (81)	805 (262)

The diagnostic categories are described in the Subjects and Methods section. 'Total' affected sib pairs is a count of all possible pairs of affecteds in a sibship, and 'unique' pairs is a count of one pair per sibship.

final analysis, inefficient. This is reflected in the uneven final map distribution of markers, the number of sizeable gaps that remain, and a number of negative regions that contain a relatively high density of markers from follow-up of external positives.

Maps and final intermarker distances

For the scan, we used the sex-averaged, 3672 cM total, 'Mfd98' maps⁹² from the Marshfield Medical Research Center. For the follow-up regions, the genotypes from the families were used to construct new maps, which were used in the multipoint analysis. A total of 684 microsatellite markers were used (mean heterozygosity 75%), 85% of which contained tri- or tetra-nucleotide repeats. Table 2 shows the numbers of markers used and the final Mfd98 map distances. Overall, there were a total of 1048 marker/Family Set combinations tested: 353 (164 unique) on Family Set A, 341 (155 unique) on Family Set B, and 354 (169 unique) on Family Set C. Including the (estimated) distances to the telomeres, the final mean intermarker distance was 9.3, 9.8, and 9.6 cM for Family Sets A, B, and C respectively. If the Family Sets are assumed to be equivalent, the final intermarker distance is 5.3 cM, with no gaps >20 cM.

Genome-wide pairwise H-LOD results

The complete database of pairwise H-LOD and NPL results, including those generated under the Dom and Lia genetic models is entitled 'VCU-ISHDSF-LODS' and is available from the Virginia Institute for Psychiatric and Behavioral Genetics (VIPBG) web site. The

Table 2 Markers and map distances

	Family Set				
	A	B	C	Any	ABC
No. markers					
Total	353	341	354	684	166
Unique to Set	164	155	169	na	0
Intermarker distance (cM)					
Average	9.3	9.8	9.6	5.3	na
SD	10.2	9.6	9.6	3.6	na
Maximum	51.1	61.7	41.1	18.2	na
No. gaps					
>40 cM	7	4	4	0	na
30–40 cM	12	2	2	0	na
20–30 cM	39	8	16	0	na
15–20 cM	22	30	28	6	na
12.5–15 cM	13	15	20	12	na
10–12.5 cM	21	24	26	53	na
<10 cM	225	200	209	590	na

After completion of the scan and follow-up work, the final inter-marker distances were calculated from the marker locations and estimated distances to the telomeres on the maps provided by the Marshfield Medical Research Center (Mfd98; sex-averaged; genome = 3762 Kosambi cM).⁹² Values in the column labeled 'Any' were calculated by treating all three Family sets as equivalent.

genome-wide, maximum pairwise H-LODs are plotted for two of the four diagnostic categories in Figures 1 (Narrow) and 2 (Broad). For clarity, only the results under the Pen (filled diamonds) and Rec (open diamonds) genetic models are shown, since results Dom are usually similar to Pen, and results under Lia are similar to Rec. The vast majority of markers must be unlinked, but each will still produce a maximum H-LOD near or at zero at large recombination fractions. Overall, about 12% of all combinations of markers and Family Sets tested gave H-LODs greater than 0.5, regardless of the combinations of Diagnostic Category and Genetic Model used. Figures 1 and 2 show that the baseline of H-LOD scores is in the range of 0.0–0.5, which is the empirical baseline we used to define the linkage signals.

Pairwise results: effect of the genetic model

There did not seem to be any obvious, systematic effect of the genetic model (here Pen vs Rec) on the H-LODs, either for markers across the genome as a whole, or in the ten follow-up regions.

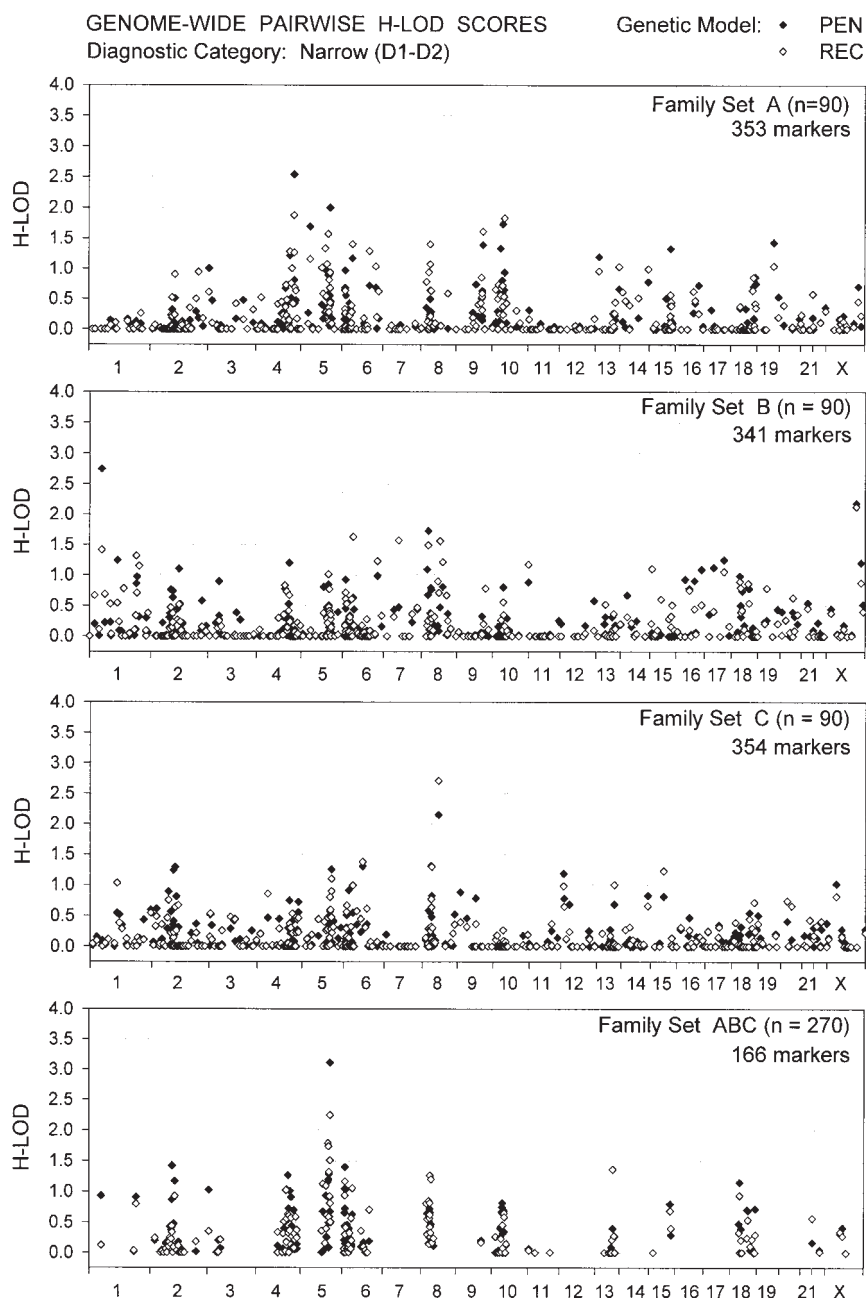
Pairwise results: effect of the diagnostic category

Some markers tested on Family Set ABC in some follow-up regions showed a marked dependence on the diagnostic category used. This can be seen in the pairwise results presented in Figures 1 and 2 and also in the multipoint results summarized in Table 4 and shown in more detail in Figures 3–6. For example, on 5q under the Rec model, the Narrow category produces by far the greatest multipoint H-LOD scores (see Figure 5b), while on 6p (Figure 6b), the Intermediate and Broad categories are both greater than either Narrow or Very Broad. On 10p (not shown), the Intermediate category is stronger than the other categories. Chromosomes 2p and 4q are less positive in general, and show few differences across diagnostic categories.

Our prior Irish family study¹⁰⁰ had suggested that the susceptibility genes for schizophrenia were perhaps largely distinct from those conferring risk to other non-schizophrenia spectrum psychiatric syndromes such as alcoholism or anxiety disorders. Therefore it was surprising that, for many markers, the scores under the Very Broad category were either comparable to, or greater than, those obtained under the Broad category. For example, three 8p markers tested on Family Set ABC gave the following results for Broad and Very Broad categories respectively: D8S1731 (0.17 and 2.40), D8S1715 (0.20 and 2.01), and D8S133 (0.46 and 1.79). Likewise, both H-LOD and NPL multipoint analysis of 8p gave similar results for the Broad and Very Broad diagnostic categories (not shown). Similar results (see also Table 3) were obtained with Family Set B and pairwise analysis of 6p markers D6S259, D6S285, D6S291, and 8p markers D8S1731 and D8S1752.

Pairwise results: effect of testing different Family Sets

Even though families were randomly assigned, for a given region the magnitude of the scores varied substantially across the three Family Sets. The most striking



Figures 1 and 2 The program MENDEL⁹⁶ was used to calculate pairwise LOD scores under the assumption of locus heterogeneity (H-LODs). The maximum score, attained at any recombination fraction and proportion linked, is plotted at the marker location. Results from three independent subsets of 90 families each (Family Sets A, B, and C) and the entire 270 families (Family Set ABC) are shown. For each Family Set, calculations were performed under four hierarchical diagnostic categories: Narrow (D1–D2, Figure 1), Intermediate (D1–D5, not shown), Broad (D1–D8, Figure 2), and Very Broad (D1–D9, not shown), and four genetic models: Dom (not shown), Pen (penetrance, filled diamonds), Lia (not shown), and Rec (recessive, open diamonds). Marker locations are as described in the legend to Table 2.

ing asymmetry is on 10p, where under the Intermediate category, Family Set A has two markers that produce H-LODs >3.0, and a number of others between 1.5 and 3.0 (data not shown). Family Set B is much weaker, and Family Set C produces no scores above baseline. Family Set A is stronger under the other three diagnostic categories also (eg Broad in Figure 2). The multi-point H-LOD maxima for each of the three Family Sets

were in almost the exact locations, but the values were very different: 2.86, 0.87 and 0.0 respectively (see Table 4 below). Similar results were found with multi-point NPL scores.

In contrast to 10p, where Family Set A was by far the most positive, for a number of markers on 2p, the pairwise results from Family Set C were greatest, especially under the Very Broad category. This was

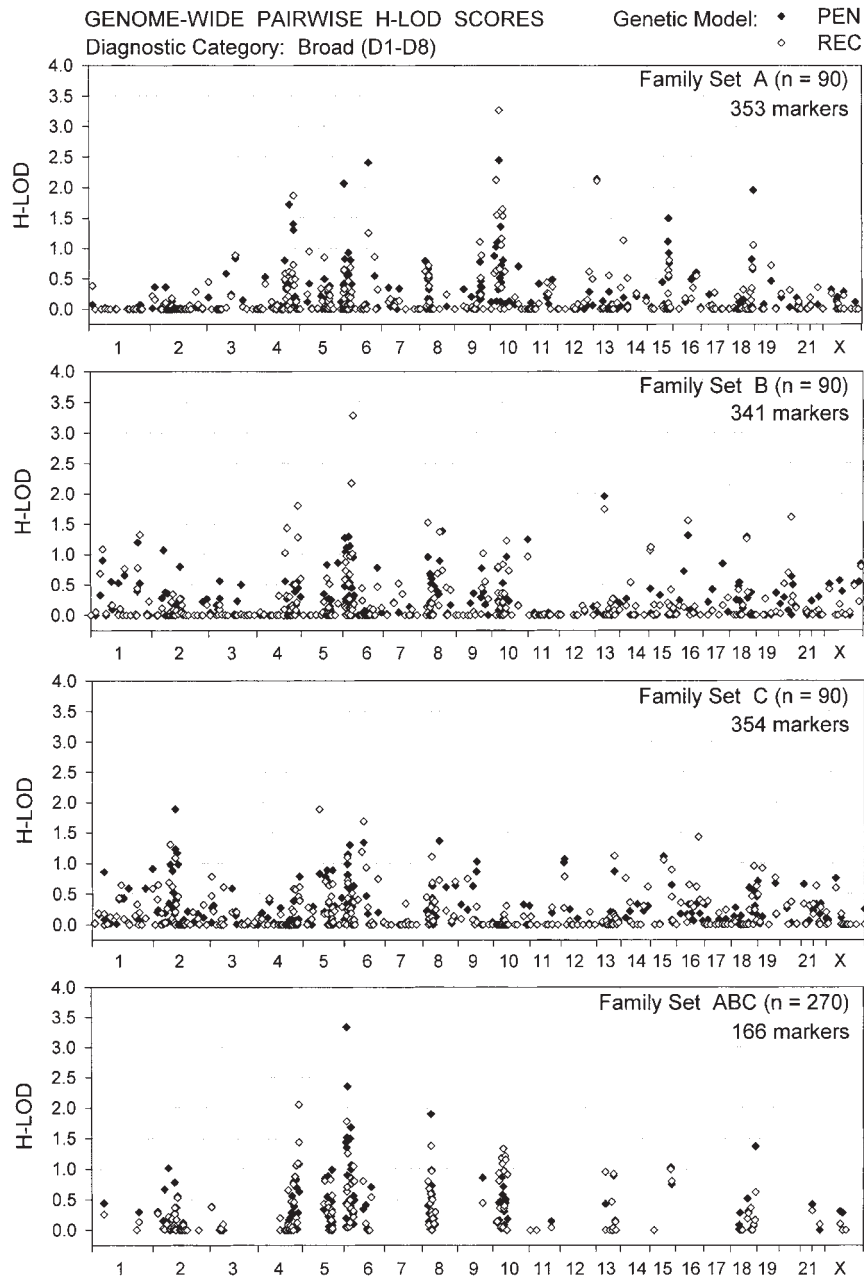


Figure 2 Caption on previous page.

also observed in the multipoint H-LOD scores (0.13, 0.07, and 1.20 for Family Sets A, B, and C respectively) and the NPL scores (0.61, 0.75, and 2.93 respectively).

Pairwise results: regions with primarily negative scores

The most noteworthy of the negative regions are 13q (see Table 4), and 22q, both of which have gathered substantial support in multiple studies. On chromosome 22, the maximum result score with marker D22S278 on all families was 0.51, and the vast majority of scores from individual Family Sets were less than 0.5. Since multiple flanking markers were also negative (data not shown), we did not pursue this region further.

The same was true of the majority of the markers in region 1q21–q32. Some of the markers on 6q did appear to be above the baseline of pairwise results, but the maximum multipoint H-LOD was only 0.71.

Pairwise results: markers with the greatest H-LODs

We identified those markers that, under any of the analyses, yielded pairwise H-LODs >2.0 on at least one Family Set. All 28 such markers are shown in Table 3, arranged in chromosomal order. The largest score observed in the scan and follow-up was 3.88, from D2S428 on Family Set C. This 2p result is noteworthy, but unfortunately Family Set C produces virtually all of the positive results. While the three best 4q results

Table 3 Markers with pairwise H-LOD scores greater than 2.0 on one or more Family Sets

CH	Marker	Locus	Set	cM	Narrow (D1-D2)			Intermediate (D1-D5)			Broad (D1-D8)			Very Broad (D1-D9)						
					DOM	PEN	LIA	REC	DOM	PEN	LIA	REC	DOM	PEN	LIA	REC	DOM	PEN	LIA	REC
1	ATA20F08	D1S1622	B	56.7	2.59	2.75	1.97	1.41	1.21	1.50	1.72	1.52	0.67	0.90	1.08	1.09	1.23	1.04	1.19	1.33
2	GATA14B12	D2S428	C	103.2	0.87	1.24	0.68	0.23	1.01	1.89	1.58	0.20	1.02	1.89	1.72	0.21	3.88	3.16	2.04	0.33
4	GATA5B09	D4S1653	A	161.9	1.34	1.22	1.39	1.28	0.87	0.91	0.95	0.72	2.23	1.72	1.53	0.61	0.67	0.53	0.48	0.18
4	AFM197xa11	D4S415	A	181.4	2.18	2.54	2.65	1.88	1.20	1.33	1.33	0.89	0.66	1.40	1.62	0.73	2.17	2.57	2.20	0.81
4	GATA42H02	D4S2417	A	181.9	0.88	0.81	1.19	1.27	0.66	0.92	1.07	1.41	1.18	1.31	1.42	1.87	2.17	1.91	1.76	1.35
5	GATA7C06	D5S1470	A	45.3	2.18	1.69	1.42	1.16	0.98	0.95	0.95	1.19	0.48	0.42	0.34	0.95	1.03	0.77	0.60	1.01
5	AFM042xd12	D5S393	A	140.7	1.96	2.00	1.65	0.93	0.37	0.78	1.12	0.78	0.00	0.11	0.13	0.17	0.24	0.15	0.17	0.05
6	AFM207xb2	D6S296	A	14.1	0.56	0.65	0.41	0.33	1.67	1.30	0.72	0.22	2.36	2.07	1.70	0.65	0.10	0.19	0.23	0.10
6	AFM035WC1	D6S259	B	27.8	0.27	0.57	0.33	0.27	0.55	0.98	0.91	0.50	0.89	1.29	1.17	0.51	3.30	2.94	2.21	1.13
6	AFM192yf2	D6S285	B	34.2	0.26	0.34	0.35	0.31	1.84	1.78	1.75	1.27	0.86	1.14	1.15	0.97	2.26	2.00	2.59	2.32
6	AFM203yg7	D6S291	B	49.5	0.41	0.63	1.40	1.63	0.50	0.88	1.41	2.24	0.29	0.95	1.61	3.29	1.73	2.90	3.43	3.46
6	AFM059yd6	D6S262	A	130.0	0.42	0.72	0.93	1.29	0.75	1.05	0.83	0.85	1.95	2.41	2.05	1.26	2.09	2.10	1.86	0.65
8	AFMa311wd1	D8S1731	B	31.7	1.72	1.73	1.20	1.49	0.97	0.97	0.75	1.28	1.04	0.96	0.66	1.52	2.20	2.24	2.05	2.03
8	AFMb055za9	D8S1752	B	46.3	0.58	0.76	0.60	0.36	0.21	0.42	0.67	0.22	0.30	0.52	0.85	0.43	1.00	1.71	2.23	2.00
8	GGAa8G07	D8S1113	C	77.9	1.86	2.15	2.43	2.71	1.21	1.38	1.17	0.51	1.24	1.36	1.33	0.73	0.91	0.74	0.67	0.64
8	ATA19G07	D8S1119	B	101.0	0.59	0.81	1.08	1.22	1.19	1.32	1.49	1.16	1.34	1.39	1.45	0.74	2.19	1.91	1.56	0.57
10	ATA31G11	D10S1412	A	28.3	0.32	0.29	0.41	0.50	0.75	0.77	0.74	0.65	0.84	1.03	1.21	2.12	1.07	1.30	1.23	2.12
10	GAAT5F06	D10S2325	A	32.8	0.47	0.64	0.79	0.73	1.95	2.10	2.31	1.94	0.67	1.09	1.21	1.55	1.49	1.88	1.98	1.81
10	GATA6E06	D10S674	A	41.8	1.12	1.33	0.91	0.56	3.04	3.18	3.19	3.10	1.95	2.45	2.60	3.27	1.01	1.35	1.64	2.23
10	GAAT13D02	D10S2443	A	50.0	0.75	0.81	0.93	0.72	2.30	2.49	2.84	3.10	1.25	1.36	1.49	1.60	1.12	1.19	1.10	0.49
10	GGAa7H02	D10S1215	A	52.9	1.50	1.73	0.94	0.37	2.19	2.29	1.95	1.42	0.71	0.72	0.59	0.35	0.85	0.98	0.88	0.55
10	GATA3G07	D10S611	A	54.2	0.35	0.37	0.26	0.24	1.67	1.67	1.42	2.03	0.53	0.36	0.30	1.16	0.56	0.45	0.44	0.51
10	MFD200L	D10S183	A	60.6	0.81	0.93	1.35	1.83	1.44	1.78	1.87	2.46	0.56	0.81	0.84	1.54	1.34	0.98	0.81	0.53
10	ATA20B07	D10S1220	B	70.2	0.25	0.29	0.33	0.09	1.65	1.67	2.09	1.40	0.98	0.96	1.13	1.23	0.15	0.41	0.80	1.74
13	UT1352	D13S252	A	16.2	1.19	1.20	1.04	0.96	2.00	2.53	2.45	1.83	1.34	2.14	2.47	2.11	2.27	3.07	2.91	1.24
18	AFM296wd5	D18S554	A	119.4	0.64	0.75	0.35	0.35	0.91	1.03	0.80	0.93	2.16	1.96	1.69	1.05	1.12	0.81	0.75	0.96
X	GATA186D06	DXS9907	C	45.0	0.94	1.02	0.93	0.81	0.67	0.74	0.70	0.31	0.66	0.75	0.79	0.60	2.58	2.70	2.70	1.75
X	HPRT 1	HPRT	B	142.0	2.02	2.18	2.25	2.12	0.50	0.67	0.82	0.71	0.40	0.52	0.57	0.48	0.35	0.55	0.74	1.14

For all markers tested, pairwise H-LOD scores were calculated for each Family set individually, as described in the legend to Figures 1 and 2. Markers with scores greater than 2.0 under any of the 16 combinations of diagnostic category and genetic model are shown. Scores between 1.50 and 1.99 are shown in italics, scores greater than 2 are shown in bold and those greater than 3 are also in italics and are underlined.

were with Family Set A, additional evidence comes from B. Overall, the two strongest *clusters* of pairwise H-LODs were on 6p and 10p—each cluster had two markers with scores >3.0 . On 10p, eight markers in the interval including D10S1412 and D10S1220, were positive under the Intermediate category, and across all genetic models, all on Family Set A. On 6p, D6S259, D6S285, and D6S291 were most positive, under the Very Broad diagnostic category. Note that D6S291 is actually distal to HLA and is over 15 cM proximal to the 6p23 multipoint peak (see Figure 6). As mentioned, the fact that the broadest diagnostic category would produce the greatest overall number of large pairwise H-LODs was quite unexpected.

Locations of the ten follow-up regions

We,¹⁰¹ and others^{102,103} well before us, have shown that with small samples, the location of maximal evidence from a gene of small effect size *can* vary considerably. For each region, therefore, we chose markers that included those regions containing the maximal evidence in the most convincing of the previous reports. The number of markers tested on all 270 families, the locations of the markers that bound the regions, and the length of each region are all given in the Methods section. The combined length according to the Mfd98 map is 473 cM, or about 12% of the genome. When calculated from genotypes of the family members, the length is 397 cM (84%). This could be due to a lower error rate in the genotypes we used to make our map, and/or to the considerably higher Mfd98 map density.

Multipoint analysis of ten regions: results from all families

During the course of the genome scan and follow-up work, we tested markers as dictated by our own independent scan results (6p, 5q, 10p) as well as by the findings of other studies of schizophrenia (2p, 4q, 5q, 6q, 8p, 13q, 18pq and many others) and bipolar disorder (13q and 18q). The markers on 8p and 10p used in the multipoint analysis presented here are identical to those used in our two previous publications.^{42,45} The number of markers presented here is greater on both 5q (35 here vs 14) and 6p (34 vs 11) than used previously.^{29,34} In addition, our report on 6p lacked multipoint results, and none of the four previous reports showed results from individual Family sets.

For each region, based on inspection of the patterns of the pairwise scores, we chose the 'best' combination of genetic model and diagnostic category, ie the one we thought most likely to produce the greatest multipoint scores. For some regions (eg 5q) the choice was clear, but for most it was not. In order to convey the substantial variability in the strength of the evidence, as well as its complexity, we present three graphs for each region. Figures 3–6 each contain three panels: (a) H-LODs calculated under the best diagnostic category, and all four genetic models; and H-LODs (b) and NPL scores (c), each calculated under different diagnostic categories. These multipoint curves are usually quite complex and highly dependent on the diagnostic cate-

gory. In some cases there are multiple peaks of similar magnitude, located at very different locations (eg 2p, Figure 3), and in other cases the evidence is relatively constant across very large genomic regions (eg 4q, Figure 4). In contrast, some results are reasonably simple, and thus more useful for fine mapping—for example the H-LOD peaks on 5q (Figure 5), 6p (Figure 6), and 10p (not shown) are both more distinct and more consistent with the pairwise results than those from most other regions. When 34 markers were analyzed, 6p showed two distinct peaks, a distal peak near D6S470 (6p25–24) and a proximal peak in the region between D6S260 and D6S1676 (6p23). Interestingly, at the 24 marker stage, these two peaks were not well defined.

Multipoint analysis of ten regions: results from family sets

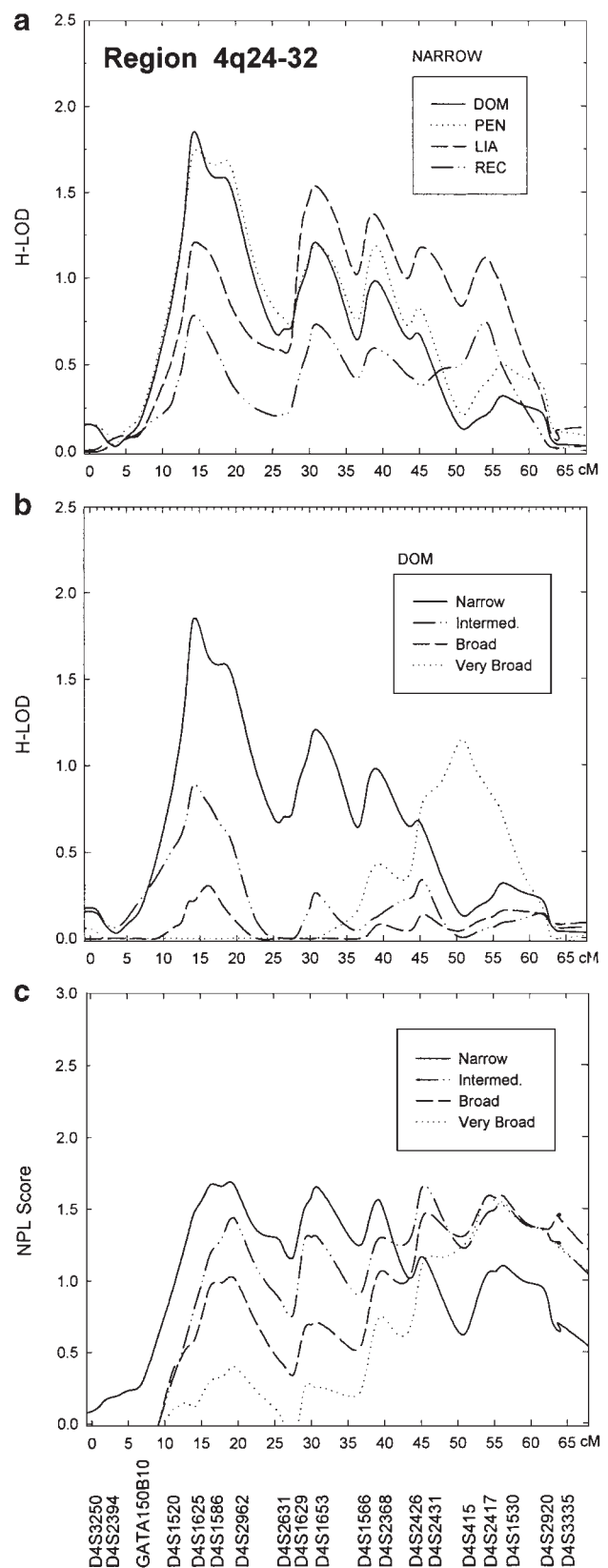
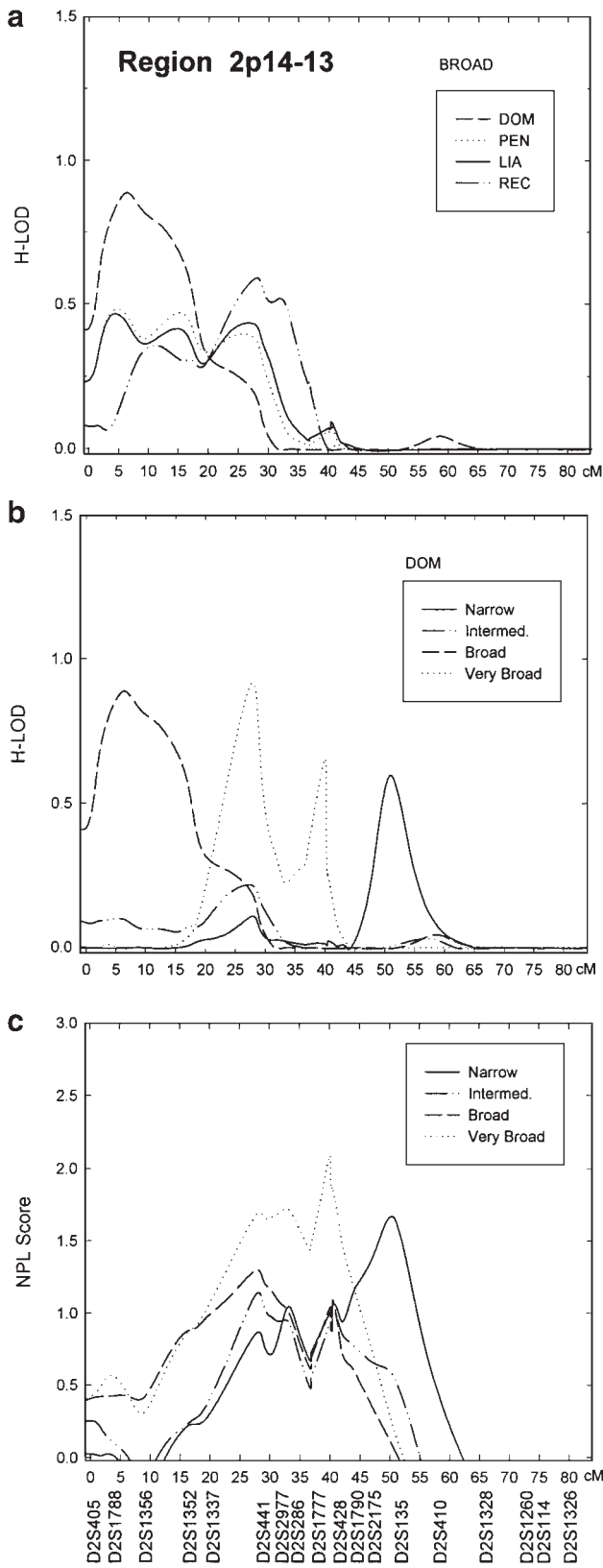
To provide a more convenient summary of the many multipoint results from Family Set ABC shown in Figures 3–6, as well as for six other regions for which the multipoint graphs are not shown, in Table 4 we show the maximum H-LOD and NPL scores and their locations for each region. Results from the individual Family Sets are also included. The curves are usually complex, and a great deal of contextual information is lost when only a small amount is extracted. For example, Family Set B produced positive scores throughout much of the region (ie the peaks are not well defined), whereas Family sets A and C resemble the results shown in Figure 6.

The multipoint results are in general agreement with the pairwise results. Regions 5q, 6p, and 8p produce the most consistent scores across Family Sets, and the largest scores on Family Set ABC. On 10p, Family Set A is by far the strongest, producing a maximum multipoint H-LOD of 2.86 and maximum NPL of 3.24. Conversely, on 2p, Family set C produces an H-LOD of 1.20 and an NPL of 2.93. There appears to be some degree of support for regions 4q, 6q, and 18q. However, inspection of the pairwise scores contained in the full database shows that across a number of additional markers tested only on individual family sets, global support for 18q is actually fairly weak. D18S554 is 7 cM from the telomere, and virtually all of the pairwise evidence from Family Sets A and C is from that marker alone. There is no support for regions 13q and 18pq.

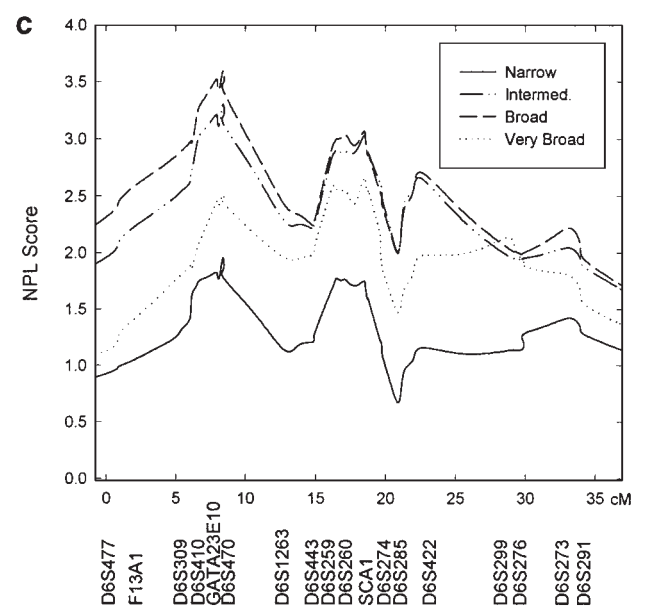
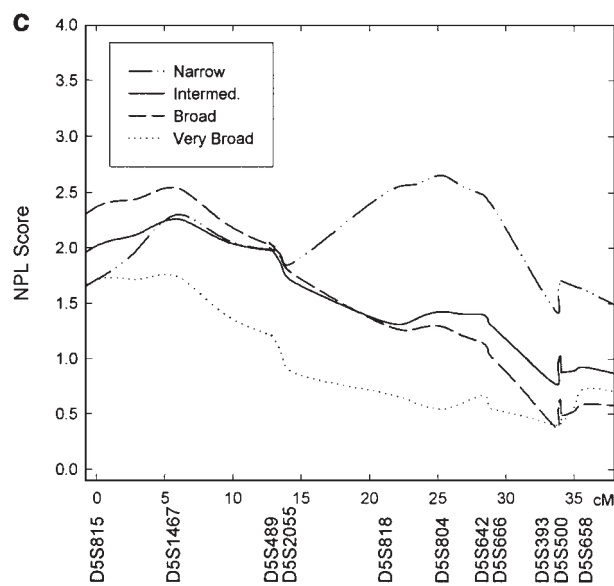
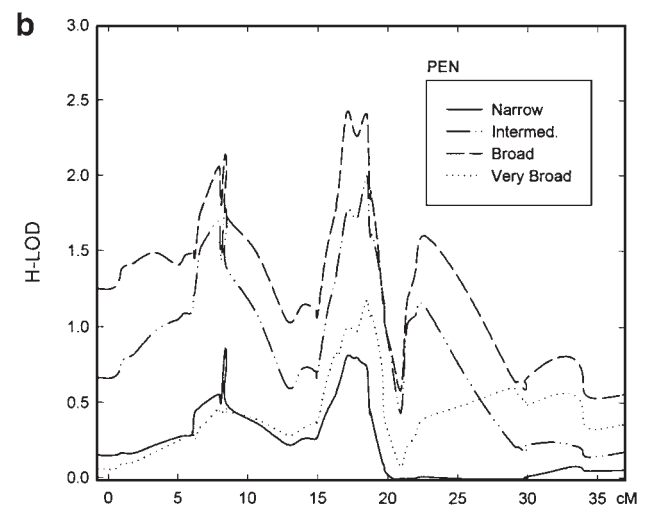
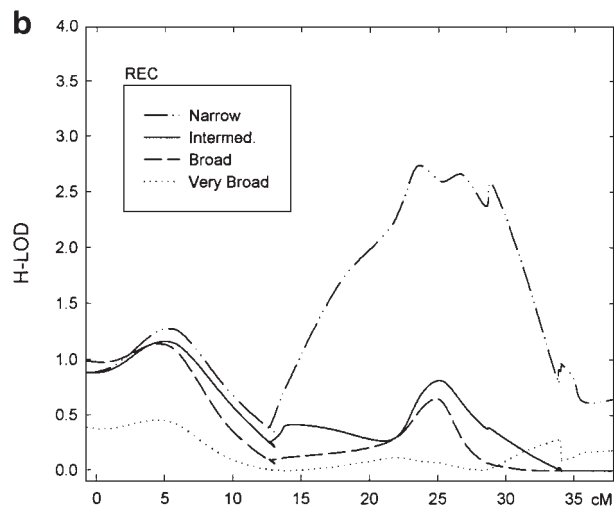
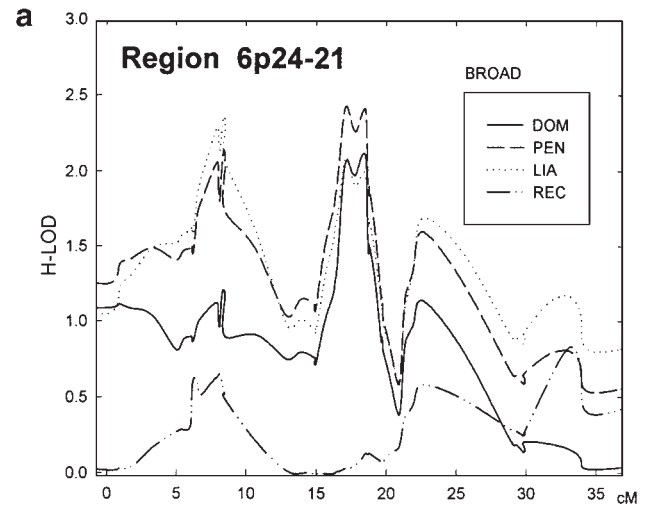
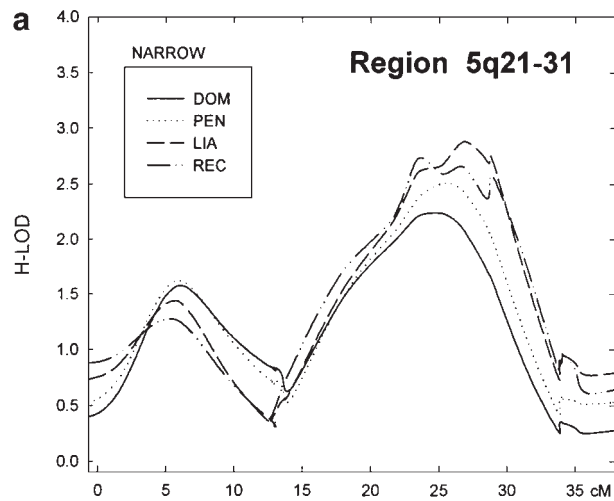
Multipoint results from individual families in the ten follow-up regions

From the analyses that produced the results shown in Table 4, we recorded the maximum scores achieved by each *individual* family, anywhere in each of the ten regions. In Table 5, we show the distribution of these maximum values, for each region individually. Below these results, for regions 5q, 6p, 8p, and 10p we show the numbers of families that are positive on combinations of two, three or all four chromosomes.

We have calculated the proportion of linked families, but given the potential problems associated with this estimation,¹⁰⁴ we can only tentatively posit a wide range of 5–30% for each of these four regions. We



Figures 3 and 4 Multipoint H-LOD and NPL results from all 270 families in four of the ten follow-up regions: 2p14–13 (Figure 3), 4q24–32 (Figure 4), 5q21–31 (Figure 5), and 6p24–21 (Figure 6). For each region, based on the pairwise results, we chose the diagnostic category which in general yielded the highest H-LODs and varied the genetic models (shown in Panel a). Caption continues on next page.



Figures 5 and 6 Caption continued from previous page. Likewise, we chose the 'best' genetic model and varied diagnostic category (shown in Panel b). Panel (c) shows the NPL results for the four diagnostic categories. Calculations were done using the program GENEHUNTER (v 1.0).⁹⁷ Maps were constructed from the ISHDSF genotypes using the program CRI-MAP,⁹³ which was also used for error checking.

Table 4 Maximum multipoint H-LOD and NPL scores from different Family Sets

		ABC (270 families)			A (90 families)			B (90 families)			C (90 families)			
Region	Diagnostic category	Model	H-LOD	cM	Near	H-LOD	cM	Near	H-LOD	cM	Near	H-LOD	cM	Near
2p	Very Broad (D1–D9)	Dom	0.92	27.7	S441	0.13	27.7	S441	0.07	50.8	S135	1.20	23.6	S441
4q	Narrow (D1–D2)	Dom	1.84	14.1	S1625	1.17	56.4	S1530	0.73	16.4	S1586	0.44	19.6	S2962
5q	Narrow (D1–D2)	Rec	2.73	23.5	S804	1.29	6.0	S1467	0.80	23.5	S804	1.31	29.2	S666
6p25-p24	Broad (D1–D8)	Pen	2.13	8.4	S470	0.69	8.4	S470	1.02	8.4	S470	0.75	13.4	S2434
6p23	Broad (D1–D8)	Pen	2.42	17.1	S260	0.70	18.5	S1605	**See legend.			0.83	16.2	S259
6q	Very Broad (D1–D9)	Lia	0.71	0.0	S445	1.01	34.3	S262	0.14	0.0	S445	0.70	0.0	S445
8p	Very Broad (D1–D9)	Dom	1.52	5.6	S1731	0.88	15.9	S136	0.91	3.5	S1731	0.62	28.4	S283
10p	Intermediate (D1–D5)	Rec	2.04	25.7	S1423	2.86	28.5	S245	0.87	35.9	S1426	0.00	11.4	S2325
13q	Narrow (D1–D2)	Dom	0.44	26.4	S797	0.06	31.8	S173	0.24	26.4	S797	0.23	17.5	S779
18pq	Narrow (D1–D2)	Pen	0.34	7.2	S975	0.12	7.2	S975	0.82	1.4	S37	0.00	0.0	S53
18q	Broad (D1–D8)	Dom	1.37	35.2	S554	0.96	35.2	S554	0.91	0.0	S41	0.98	29.5	S844
Region	Diagnosis		NPL	cM	Near	NPL	cM	Near	NPL	cM	Near	NPL	cM	Near
2p	Very Broad (D1–D9)		2.08	40.0	S428	0.61	27.7	S441	0.75	19.5	S1337	2.93	40	S428
4q	Narrow (D1–D2)		1.67	16.4	S1586	2.36	54.0	S2417	1.20	25.3	S2631	0.96	64.1	S3335
5q	Narrow (D1–D2)		2.66	25.2	S804	2.20	25.2	S804	1.30	28.7	S666	1.23	25.2	S804
6p25-p24	Broad (D1–D8)		3.59	8.4	S470	1.63	8.4	S470	2.72	7.7	S940	1.82	8.4	S470
6p23	Broad (D1–D8)		3.07	18.5	S1605	1.30	17.1	S260	**See legend.			1.76	18.5	S1605
6q	Very Broad (D1–D9)		1.08	0.0	S445	1.77	34.3	S262	0.91	0.0	S445	1.38	0.0	S445
8p	Very Broad (D1–D9)		2.33	3.5	S1731	1.66	12.8	S282	2.54	3.5	S1731	0.88	14.0	S298
10p	Intermediate (D1–D5)		1.78	35.8	S1426	3.24	27.0	S245	0.71	12.8	S2325	–0.08	36.2	S183
13q	Narrow (D1–D2)		–0.04	17.5	S779	0.65	31.8	S173	–0.14	12.9	S128	0.43	17.5	S779
18pq	Narrow (D1–D2)		0.42	1.4	S37	0.23	14.2	S847	1.04	0.7	S53	0.09	0.7	S37
18q	Broad (D1–D8)		1.92	35.2	S554	2.44	35.2	S554	1.53	0.0	S41	1.20	35.2	S554

From the multipoint graphs shown in Figures 3–6, and from other multipoint graphs not shown (regions 6q, 8p, 10p, 13q, 18p, and 18q), the magnitude and location ('cM' indicates the position on the IHDS map and 'Near' indicates the nearest marker tested) of the maxima are shown.

**For Family set B, positive H-LODs (>0.75) and NPL scores (>2.5) were observed over the majority of the length of the map. This is unlike Family sets A and C, which closely resemble the results found with all families combined, shown in Figure 6.

might expect, therefore, between 13 and 81 families to be linked to any one region. The (cumulative) cut-offs shown in Table 5 were chosen to reflect this estimate, even though with small families, linkage scores are inaccurate predictors of the presence of risk allele(s). Nevertheless, 13–81 families corresponds roughly to H-LODs of >0.5 to >0.7, and to NPL scores of >1.4 to >2.4. We found that a total of only 74 families had NPL scores greater than 2.0 on any of the ten regions (data not shown). Very few of the 270 families show large H-LOD or NPL scores on more than one region. Only one family (01226) shows NPL scores greater than 3.0 on all four of the regions reported previously.

In Table 6, we show data from the 12 families that gave scores >2.0 on three or more of the ten regions. An additional 11 families showed scores >2.0 on two or more regions (data not shown). Note that family 1226 yields NPL scores >3.0 on five of the ten regions, and is positive on two more (18q, NPL = 2.63 and 4q, NPL = 1.63) as well. Also shown in Table 6 are the rank of each of the NPL scores from family 01226 relative to all 270 families. It is extraordinary that the six best NPL scores produced are all within the top 1.5% (four of 270) produced in the entire study for those six regions. Family 01226 contains a sibship with four affected brothers, but is otherwise unremarkable—a number of other families are similar in structure.

Discussion

Effect of varying the genetic model

For all six regions that appear most promising in our sample, scores under the genetic models Dom and Pen are strongly correlated, as are those under Lia and Rec. Consistent with the work of Greenberg *et al*,^{98,99} we would have obtained nearly the same results with the use of two rather than four genetic models. For most positive regions, Pen and Rec are about equal in generating the best pairwise results. In contrast, the multipoint results are generally stronger under Dom or Pen models, with the exception of region 10p, where Rec and Lia models are somewhat better. No strong inferences can yet be made regarding the likely mode of action of any of these putative loci. These findings are consistent, though, with the assumption that different susceptibility loci can have different biological effects.

Effect of varying the diagnostic category

We used four diagnostic categories to minimize type 2 errors, and to try to gain insight into the nature of the risk conferred by individual loci. If for a given region a particular diagnostic category consistently produced much larger linkage signals, it may mean that the liability towards particular clinical symptoms is due to at least in part to variation at that locus. The magnitude of the H-LODs varied more, and varied on more

Table 5 Distribution of individual family ($n = 270$) maximum multipoint H-LOD and NPL scores from the follow-up regions

Greater than or equal to:	Number of H-LOD scores						Number of NPL scores						
	1.4	1.0	0.8	0.7	0.5	0.3	5.3	3.3	3.0	2.6	2.4	1.8	1.4
2p	0	5	21	32	62	96	0	4	4	8	13	30	83
4q	0	3	8	17	39	103	0	2	2	4	10	22	89
5q	0	7	12	20	75	98	0	0	1	5	11	17	78
6p	0	0	5	9	26	66	0	1	2	6	12	19	82
6q	0	2	4	8	15	40	0	3	3	6	9	15	33
8p	0	6	17	26	46	65	1	4	6	7	7	17	55
10p	1	4	9	14	46	96	1	2	3	4	10	15	68
13q	1	2	2	5	18	50	0	0	1	4	6	9	47
18q	0	4	13	20	38	65	0	1	1	3	9	22	44
18p	0	1	2	4	18	37	0	0	1	1	4	11	36
5q, 6p	0	0	0	1	6	24	0	0	1	2	3	3	25
5q, 8p	0	0	3	5	14	29	0	0	1	1	2	4	20
5q, 10p	0	0	0	2	12	31	0	0	1	2	2	3	19
6p, 8p	0	0	1	1	8	23	0	0	1	1	1	4	23
6p, 10p	0	0	1	3	6	24	0	0	1	2	2	3	26
8p, 10p	0	0	2	3	9	20	0	0	1	1	2	2	15
5q, 6p, 8p	0	0	0	0	1	11	0	0	1	1	1	2	10
5q, 6p, 10p	0	0	0	0	2	6	0	0	1	2	2	2	9
5q, 8p, 10p	0	0	0	0	2	5	0	0	1	1	1	1	6
6p, 8p, 10p	0	0	0	0	1	7	0	0	1	1	1	1	8
5q, 6p, 8p, 10p	0	0	0	0	0	1	0	0	1	1	1	1	5

Multipoint H-LOD and NPL scores were calculated for each family individually, as described in the legend to Figures 3–6. The diagnostic categories and genetic models used are those shown in Table 4. For all ten follow-up regions, the (cumulative) numbers of families producing scores greater than or equal to various cut-offs are shown. Below that, for regions 5q, 6p, 8p, and 10p, the number of scores greater than each cut-off on *each* of the different combinations of regions is shown. For example, there was only one family that had NPL >3.0 on *both* regions 5q and 6p.

Table 6 Individual family multipoint NPL scores in the follow-up regions

Family	Region:	Maximum multipoint NPL score										No. NPL scores	
		5q	6p	8p	10p	2p	4q	6q	13q	18pq	18q	>2.0	>3.0
1226		3.11	3.17	3.55	3.17	-0.09	1.63	-0.24	-0.47	3.08	2.64	6	5
	Rank:	1	2	4	3	245	24	219	239	1	3		
1540		2.45	2.77	2.05	0.51	4.06	1.07	0.95	2.45	2.45	1.51	6	1
1415		0.82	-0.81	0.82	2.45	2.45	2.45	2.45	-0.49	0.82	2.23	5	0
1347		2.65	2.86	0.01	5.79	2.48	1.41	0.19	-0.30	-0.35	0.40	4	1
1328		1.56	3.41	1.07	0.28	0.05	2.42	-0.24	2.98	0.45	2.03	4	1
1323		0.41	2.45	2.16	0.00	2.25	0.53	2.18	0.32	-1.09	-0.23	4	0
1348		0.11	0.82	3.14	2.45	0.49	1.41	1.20	1.41	1.41	2.45	3	1
1016		0.82	2.45	0.95	0.92	-0.41	1.32	2.45	0.78	1.10	2.45	3	0
1311		1.40	0.00	2.06	0.83	2.06	1.41	2.81	-0.09	0.04	0.93	3	0
1058		2.39	-0.29	-0.23	-0.51	2.66	-0.07	-0.38	1.57	2.48	0.38	3	0
1410		0.82	-0.81	-0.81	2.45	2.45	0.96	1.28	-0.28	-0.67	2.35	3	0
1353		2.66	1.69	0.82	0.73	0.69	2.65	0.08	0.55	2.07	-0.36	3	0

Multipoint NPL scores for those families yielding NPL scores >2.0 on three or more regions. For family 1226, the rank of the score amongst the 270 families is also shown. The diagnostic categories and genetic models used are those shown in Table 4.

regions, with changes in the diagnostic category than with changes in the genetic model. The region showing the greatest dependence on diagnostic category is 5q, where the Narrow category produced by far the highest

H-LODs and NPL scores. On 6p, the Intermediate and Broad categories produced similar and stronger signals than either the Narrow or Very Broad categories. Region 8p is the one where the Very Broad category is

the most positive. On 10p, the Intermediate category yielded better scores on average. One plausible interpretation of these results is that the 'spectrum of action' of individual susceptibility loci differs. That is, perhaps some of the 'schizophrenia' loci that we are observing may predispose individuals only to the narrowly defined psychotic illness of schizophrenia (eg 4q, 5q), while others may also increase risk to a broad range of schizophrenia spectrum disorders (6p23, 6p25–24, 10p), and others (8p) may influence vulnerability to an even wider range of psychiatric conditions, including anxiety and affective disorders.

Replication of linkage

The degree to which negative reports should detract from positive evidence for linkage in complex traits is controversial. Simulation studies predict that 'non-replications' of true linkages will be common when small, heterogeneous samples are used to detect genes of small effect.⁸ Furthermore, to date, because of cross-study heterogeneity of sample sizes, populations studied, ascertainment strategies, diagnostic methods, and statistical techniques,¹⁰⁵ all follow-up studies in schizophrenia are not truly 'replication' attempts in any strict sense. When follow-up results differ from the original ones, usually no firm conclusions as to the cause can be drawn. Replication in *comparable* independent samples is critical. We randomly assigned the families to the three Family Sets after ascertainment and diagnosis was completed, and before the systematic scan started. Each set had greater power to detect linkage than most other schizophrenia linkage samples. Therefore, by design, we had a good opportunity to perform a more valid and interpretable 'replication study' than had been achieved previously.

Distribution of positive evidence across the three Family Sets

Two of the six positive regions we examined carefully (2p and 10p) failed to 'replicate' in either of the two other subsamples. This may indicate that these findings are false positives. Alternatively, it is possible that these regions do contain susceptibility genes and that non-replication of true linkages may actually be much more prevalent than commonly thought. There was much less asymmetry in the distribution of positive evidence in regions 5q, 6p, and 8p, where the evidence for second best sample was usually similar to the best. Given our current state of ignorance about the stochastic properties of linkage in complex diseases, we can at most conclude from this study that the more consistently replicated regions are more likely to contain true susceptibility genes than are regions where positive evidence for linkage cannot be replicated.

Comparison with findings from other studies

A detailed evaluation of what has become a tremendous number of positive linkage findings in schizophrenia is not possible here. In brief, we found no evidence to support the putative schizophrenia loci in regions 13q32 or 18p11–q11, or the bipolar disorder

locus in 18q22–23.¹⁰⁶ Pairwise results were negative for regions 3p25–24,⁴¹ 5p14–p13,¹⁰⁷ and 5q11–13,^{108,109} and 15q.^{52,54} We have already published our primary positive evidence for regions 5q21–31, 6p24–21, 8p22–21, and 10p15–11. In this report, we provided additional detail on all four of these regions, including the evidence from each of the three Family Sets, and showed that for 10p, Sets B and C failed to corroborate the very strong findings from Set A. We also showed additional marker data on 6p24–21 that suggests the possibility of two genes—one in 6p25–24 and one in 6p23. Two additional regions also produced evidence in support of linkage, 2p14–13, and 4q24–32, and we noted that 2p failed to produce consistently positive scores on two of the three Family Sets. Region 6q13–26^{39,40} yielded only small positive scores, but the pattern is interesting and the region deserves further study.

The following seven candidate regions have not yet been subjected to a thorough investigation, (ie multipoint analysis of all families): 1q32.2–42,^{64,110} 2q12–13,^{36,46,63} 9q21–22,^{36,40,63} 10q11.2–23,^{46,59,63,65} 15q14,^{40,52,53} 22q12–13,^{12,13,55–61} and Xp11.4–11.3.^{64,68–71} While they are positive in some other studies, the pattern of our pairwise scores was sufficiently negative that we did not pursue them further. We have tested many markers in these regions (on either one or two Family Sets), in addition to those used for the multipoint analysis shown here, and there is important information contained in the pairwise results that are not shown. Since in scan data consideration of the context of the score from any individual marker is crucial, it is necessary to consider *all* of the data together when evaluating the overall support for these, and other, candidate regions. To this end, we have made all of the pairwise results available from the VIPBG web site.

Conclusions

In light of the relatively weak linkage statistics obtained, and the complexity of the results across the independent family sets and across analyses, only a few tentative, quite general conclusions are supported. First, three regions (13q, 18pq and 18q) clearly differ from the others in not having consistently generated pairwise scores greater than those typical of the scan baseline. Of the remaining seven regions, 2p and 6q show the weakest pairwise H-LOD scores and the least consistent multipoint peaks, although the NPL multipoint results look somewhat more promising. Regions 4q, 5q, 6p, 8p, and 10p all produce evidence of the sort that now appears to be characteristic of genes of moderate effect size in complex traits. Even in the absence of highly 'significant' results, overall the strength of the data for these regions is sufficient to justify high density mapping to detect linkage disequilibrium (LD). We have pursued such a strategy on 6p and have recently generated evidence of LD very close to each of the two multipoint peaks (manuscript in preparation). Second, the pattern of internal replication differed substantially between these positive

regions. While the meaning of these differences is unclear, were we restricted to our own data, we would have to conclude that the regions with clearer internal replications (5q, 6p, 8p) are more likely to be true findings than are the regions that replicate poorly, especially 2p and 10p. However, the evidence from other groups in support of the 10p locus is now quite strong. Third, it appears from the scores from individual families in the candidate regions, that locus heterogeneity, as opposed to a simple additive model for example, is likely to be a predominant feature of this Irish high-density schizophrenia sample, and perhaps of familial schizophrenia overall.

Thus far we have applied only the most basic of analytic approaches on the scan data. We have not attempted to maximize the statistical significance by varying the parameters used or by applying additional tests. We have just begun more exploratory types of analyses—for example our first attempt at detecting epistatic interactions between the loci in the six most positive regions was negative,¹¹¹ but more work is warranted. Approaches that are more synthetic and powerful, for example haplotype analysis that incorporates multivariate and quantitative approaches, and that allows for epistatic interactions, may serve to produce more definitive results. It is anticipated also that the application of newer methods, eg neural networks¹¹² or classification and regression trees¹¹³ and related algorithms, when applied to data from much larger samples probed with many more markers will also be productive. Even if such analytic advances are made, however, the use of intermediate phenotypes¹¹⁴ as well as a battery of assays of gene function (eg differential expression¹¹⁵) will be required to elucidate the genetic component of schizophrenia.

Acknowledgments

This project was largely supported by grants MH41953, MH52537, and MH45390 from the United States National Institute of Mental Health (NIMH). Data collection was conducted under the supervision of S Humphries, M Healy, and A Finnerty. Additional interviews were conducted by J Burke, B Murphy, F Duke, R Shinkwin, M Ni Nuallain, F McMahan, J Downing, T Hebron, B Hanratty, E Crowe, M Doherty, J Bray, and L Lowry. This project would not have been possible without the cooperation of the families themselves and the staffs of the many psychiatric hospitals and units in Ireland and Northern Ireland and their efforts are gratefully acknowledged. We also wish to acknowledge the assistance of R McClelland.

Electronic-database information

URLs for data in this article are as follows:

- (1) Cooperative Human Linkage Center (CHLC); <http://www.chlc.org/>
- (2) Marshfield Medical Research Center: <http://www.marshmed.org/genetics/>

- (3) The complete database, entitled VCU-ISHDSF-LODS, contains pairwise H-LODs for all markers and pairwise NPL scores for selected markers and will be available upon publication from VIPBG (Virginia Institute for Psychiatric and Behavioral Genetics) web site: (<http://electro.psi.vcu.edu/vipbg/vipbghome.html>).

References

- 1 Gottesman II, Shields J. *Schizophrenia: The Epigenetic Puzzle*. Cambridge University Press: New York, 1982.
- 2 Kendler KS. The genetics of schizophrenia. In: Kaplan HI and Sadock BJ (eds). *Comprehensive Textbook of Psychiatry*. Williams & Wilkins: New York, 1999.
- 3 O'Rourke DH, Gottesman II, Suarez BK, Rice J, Reich T. Refutation of the general single-locus model for the etiology of schizophrenia. *Am J Hum Genet* 1982; **34**: 630–649.
- 4 Prescott CA, Gottesman II. Genetically mediated vulnerability to schizophrenia [Review]. *Psych Clin North Am* 1993; **16**: 245–267.
- 5 Cloninger CR. Turning point in the design of linkage studies of schizophrenia. *Am J Med Genet* 1994; **54**: 83–92.
- 6 Badner JA, Gershon ES, Goldin LR. Optimal ascertainment strategies to detect linkage to common disease alleles. *Am J Hum Genet* 1998; **63**: 880–888.
- 7 Terwilliger JD. On the resolution and feasibility of genome scanning approaches. In: Rao DC, Province MA (eds). *Genetic Dissection of Complex Traits*. Academic Press: New York, 2001, pp 351–389.
- 8 Suarez BK, Hampe CL, Van Eerdewegh P. Problems of replicating linkage claims in psychiatry. In: Gershon ES, Cloninger CR (eds). *Genetic Approaches to Mental Disorders*. American Psychiatric Press: Washington, DC, 1994, pp 23–46.
- 9 Martinez M, Goldin LR. Power of the linkage test for a heterogeneous disorder due to two independent inherited causes: a simulation study. *Genet Epidemiol* 1990; **7**: 219–230.
- 10 Chen WJ, Faraone SV, Tsuang MT. Linkage studies of schizophrenia: a simulation study of statistical power. *Genet Epidemiol* 1992; **9**: 123–139.
- 11 Levinson DF. Power to detect linkage with heterogeneity in samples of small nuclear families. *Am J Med Genet* 1993; **48**: 94–102.
- 12 Gill M, Vallada H, Collier D, Sham P, Holmans P, Murray R *et al*. A combined analysis of D22S278 marker alleles in affected sibpairs: support for a susceptibility locus for schizophrenia at chromosome 22q12. *Am J Med Genet* 1996; **67**: 40–45.
- 13 Vallada H, Curtis D, Sham P, Kunugi H, Zhao JH, Murray R *et al*. A transmission disequilibrium and linkage analysis of D22S278 marker alleles in 574 families: further support for a susceptibility locus for schizophrenia at 22q12. *Schiz Res* 1998; **32**: 115–121.
- 14 Schizophrenia Linkage Collaborative Group for Chromosomes 3 and 6 and 8—coordinated by DF Levinson. Additional support for schizophrenia linkage on chromosomes 6 and 8: A multicenter study. *Am J Med Genet* 1996; **67**: 580–594.
- 15 Levinson DF, Holmans P, Straub RE, Owen MJ, Wildenauer DB, Gejman PV *et al*. Multicenter linkage study of schizophrenia candidate regions on chromosomes 5q, 6q, 10p, and 13q: schizophrenia linkage collaborative group III. *Am J Hum Genet* 2000; **67**: 652–663.
- 16 Karayiorgou M, Gogos JA. Dissecting the genetic complexity of schizophrenia. *Mol Psychiatry* 1997; **2**: 211–223.
- 17 Riley BP, McGuffin P. Linkage and associated studies of schizophrenia. *Am J Med Genet* 2000; **97**: 23–44.
- 18 Baron M. Genetics of schizophrenia and the new millennium: progress and pitfalls. *Am J Hum Genet* 2001; **68**: 299–312.
- 19 Gurling HMD, Curtis D, Kalsi G, McQuillin A, Lawrence J, Murphy P *et al*. Genome wide linkage analysis supports the presence of two separate susceptibility loci on chromosome 1q32 for bipolar disorder and 1q21–22.3 for schizophrenia [Abstract]. *Am J Med Genet* 2000; **96**: 400.
- 20 Bray NJ, Owen MJ. Searching for schizophrenia genes. *Trends Mol Med* 2001; **7**: 169–174.

- 21 Hovatta I, Varilo T, Suvisaari J, Terwilliger JD, Ollikainen V, Arajarvi R *et al*. A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci. *Am J Hum Genet* 1999; **65**: 1114–1124.
- 22 Brzustowicz LM, Hodgkinson KA, Chow EW, Honer WG, Bassett AS. Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21–q22. *Science* 2000; **288**: 678–682.
- 23 Ekelund J, Hovatta I, Parker A, Varilo T, Pauron T, Martin R *et al*. Fine-mapping of a 40 CM region on chromosome 1Q in schizophrenia [Abstract]. *Am J Med Genet* 2000; **96**: 461.
- 24 Ye FQ, Berman KF, Ellmore T, Esposito G, van Horn JD, Yang Y *et al*. H(2)(15)O PET validation of steady-state arterial spin tagging cerebral blood flow measurements in humans. *Magn Reson Med* 2000; **44**: 450–456.
- 25 Ekelund J, Lichtermann D, Hovatta I, Ellonen P, Suvisaari J, Terwilliger JD *et al*. Genome-wide scan for schizophrenia in the Finnish population: evidence for a locus on chromosome 7q22. *Hum Mol Genet* 2000; **9**: 1049–1057.
- 26 Ekelund J, Hovatta I, Parker A, Paunio T, Varilo T, Martin R *et al*. Chromosome 1 loci in Finnish schizophrenia families. *Hum Mol Genet* 2001; **10**: 1611–1617.
- 27 St Clair D, Blackwood D, Muir W, Carothers A, Walker M, Spowart G *et al*. Association within a family of a balanced autosomal translocation with major mental illness. *Lancet* 1990; **336**: 13–16.
- 28 Millar JK, Christie S, Anderson S, Lawson D, Loh DH, Devon RS *et al*. Genomic structure and localisation within a linkage hotspot of Disrupted In Schizophrenia 1, a gene disrupted by a translocation segregating with schizophrenia. *Mol Psychiatry* 2001; **6**: 173–178.
- 29 Straub RE, MacLean CJ, O'Neill FA, Walsh D, Kendler KS. Support for a possible schizophrenia vulnerability locus in region 5q21–31 in Irish families. *Mol Psychiatry* 1997; **2**: 148–155.
- 30 Schwab SG, Eckstein GN, Hallmayer J, Lerer B, Albus M, Borrmann M *et al*. 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 Psychiatry* 1997; **2**: 156–160.
- 31 Byerley W, Tiobech S, Blakis A, Zuo J, Zhao M, Hoff M *et al*. Evidence for a 5q31 schizophrenia locus in a large multiplex kindred from Palau, Micronesia [Abstract]. *Mol Psychiatry* 1999; **4** (Suppl 1): S4–S5.
- 32 Gurling HMD, Kalsi G, Blaveri E, McQuillin A, Read T, Murphy P *et al*. Initial genome wide parametric genetic linkage analysis of schizophrenia and schizophrenia spectrum disorders finds LOD scores above 3.00 on four chromosomes at 1q22–23, 5q22–35, 8p21–23 and 11q14–24. A further LOD above 3.00 at 4q21–31 was found within a single family [Abstract]. *Mol Psychiatry* 1999; **4** (Suppl 1): S4.
- 33 Blouin J-L, Lasseter VK, Nestadt G, Wolyniec PS, Gehrig C, Antonarakis SE *et al*. Linkage analysis for schizophrenia susceptibility loci in candidate regions on chromosomes 5q, 6q, 10q, and 13q: a collaborative follow-up effort in one sample [Abstract]. *Mol Psychiatry* 1999; **4** (Suppl 1): S32.
- 34 Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F *et al*. A potential vulnerability locus for schizophrenia on chromosome 6p24–22: evidence for genetic heterogeneity. *Nature Genet* 1995; **11**: 287–293.
- 35 Schwab SG, Albus M, Hallmayer J, Höning S, Borrmann M, Lichtermann D *et al*. Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nature Genet* 1995; **11**: 325–327.
- 36 Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F *et al*. An international two-stage genome-wide search for schizophrenia susceptibility genes. *Nature Genet* 1995; **11**: 321–324.
- 37 Maziade M, Bissonnette L, Rouillard E, Martinez M, Turgeon M, Charron L *et al*. 6p24–22 region and major psychoses in the eastern Quebec population. *Am J Med Genet* 1997; **74**: 311–318.
- 38 Lindholm E, Ekholm B, Balciumiene J, Johnsson G, Castensson A, Koisti M *et al*. Linkage analysis of a large Swedish kindred provides further support for susceptibility locus for schizophrenia on chromosome 6p23. *Am J Med Genet (Neuropsych Genet)* 1999; **88**: 369–371.
- 39 Cao Q, Martinez M, Zhang J, Sanders AR, Badner JA, Cravchik A *et al*. Suggestive evidence for a schizophrenia susceptibility locus on chromosome 6q and a confirmation in an independent series of pedigrees. *Genomics* 1997; **43**: 1–8.
- 40 Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD *et al*. NIMH Genetics Initiative Millennium Schizophrenia Consortium: Linkage analysis of African–American pedigrees. *Am J Med Genet* 1998; **81**: 282–289.
- 41 Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt G, Blouin J-L *et al*. Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. *Am J Med Genet* 1995; **60**: 252–260.
- 42 Kendler KS, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F *et al*. Evidence for a schizophrenia vulnerability locus on chromosome 8p in the Irish study of high-density schizophrenia families. *Am J Psychiatry* 1996; **153**: 1534–1540.
- 43 Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G *et al*. Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nature Genet* 1998; **20**: 70–73.
- 44 Blaveri K, Kalsi G, McQuillin A, Quedsted D, Read T, Murphy P *et al*. Linkage and allelic association studies of the 8p schizophrenia locus [Abstract]. *Mol Psychiatry* 1999; **4** (Suppl 1): S31.
- 45 Straub RE, MacLean CJ, Martin RB, Myakishev MV, Harris-Kerr C, O'Neill FA *et al*. A schizophrenia locus may be located in region 10p15–p11. *Am J Med Genet* 1998; **81**: 296–301.
- 46 Faraone SV, Matisse T, Svrakic D, Pepple J, Malaspina D, Suarez B *et al*. Genome scan of European-American schizophrenia pedigrees: results of the NIMH genetics initiative and millennium consortium. *Am J Med Genet* 1998; **81**: 290–295.
- 47 Schwab SG, Hallmayer J, Albus M, Lerer B, Hanses C, Kanyas K *et al*. Further evidence for a susceptibility locus on chromosome 10p14–p11 in 72 families with schizophrenia by nonparametric linkage analysis. *Am J Med Genet* 1998; **81**: 302–307.
- 48 Lin M-W., Curtis D, Williams N, Arranz M, Nanko S, Collier D *et al*. Suggestive evidence for linkage of schizophrenia to markers on chromosome 13q14.1–q32 [Abstract]. *Psychiatr Genet* 1995; **5**: 117–126.
- 49 Lin MW, Sham P, Hwu HG, Collier D, Murray R, Powell JF. Suggestive evidence for linkage of schizophrenia to markers on chromosome 13 in Caucasian but not Oriental populations. *Hum Genet* 1997; **99**: 417–420.
- 50 Kalsi G, Chen C-H, Smyth C, Brynjolfsson J, Sigmudsson Th, Curtis D *et al*. Genetic linkage analysis in Icelandic/British family sample fails to exclude the putative chromosome 13q14.1–q32 schizophrenia susceptibility locus [Abstract]. *Am J Hum Genet* 1996; **59**: A388.
- 51 Brzustowicz LM, Honer WG, Chow EWC, Little D, Hogan J, Hodgkinson K *et al*. Linkage of familial schizophrenia to chromosome 13q32. *Am J Hum Genet* 1999; **65**: 1096–1103.
- 52 Freedman R, Coon H, Myles-Worsley M, Orr-Urtreger A, Olincy A, Davis A *et al*. Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. *Proc Natl Acad Sci USA* 1997; **94**: 587–592.
- 53 Leonard S, Gault J, Moore T, Hopkins J, Robinson M, Olincy A *et al*. Further investigation of a chromosome 15 locus in schizophrenia: analysis of affected sibpairs from the NIMH Genetics Initiative. *Am J Med Genet* 1998; **81**: 308–312.
- 54 Freedman R, Leonard S, Gault JM, Hopkins J, Cloninger CR, Kaufmann CA *et al*. Linkage disequilibrium for schizophrenia at the chromosome 15q13–14 locus of the alpha7-nicotinic acetylcholine receptor subunit gene (CHRNA7). *Am J Med Genet* 2001; **105**: 20–22.
- 55 Coon H, Holik J, Hoff M, Reimherr F, Wender P, Myles-Worsley M *et al*. Analysis of chromosome 22 markers in nine schizophrenia pedigrees. *Am J Med Genet* 1994; **72**: 79.
- 56 Coon H, Jensen S, Holik J, Hoff M, Myles-Worsley M, Reimherr F *et al*. Genomic scan for genes predisposing to schizophrenia. *Am J Med Genet* 1994; **54**: 59–71.
- 57 Pulver AE, Karayiorgou M, Wolyniec PS, Lasseter VK, Kasch L, Nestadt G *et al*. Sequential strategy to identify a susceptibility gene for schizophrenia: report of potential linkage on chromosome 22q12–q13.1: Part 1. *Am J Med Genet* 1994; **54**: 36–43.
- 58 Vallada HP, Gill M, Sham P, Lim LCC, Nanko S, Asherson P *et*

- al. Linkage studies on chromosome 22 in familial schizophrenia. *Am J Med Genet* 1995; **60**: 139–146.
- 59 Shaw SH, Kelly M, Smith AB, Shields G, Hopkins PJ, Loftus J *et al*. A genome-wide search for schizophrenia susceptibility genes. *Am J Med Genet* 1998; **81**: 364–376.
- 60 Li T, Ball D, Zhao J, Murray RM, Liu X, Sham PC *et al*. Haplotypic transmission disequilibrium analysis: application to a potential locus for schizophrenia at chromosome 22q11 [Abstract]. *Mol Psychiatry* 1999; **4** (Suppl. 1): S37–S38.
- 61 Lin MW, Liu CM, Ou Yang WC, Lee SFC, Hwu HG. Weak evidence for a susceptibility locus on chromosome 22q12 in Taiwanese schizophrenic pedigrees [Abstract]. *Mol Psychiatry* 1999; **4** (Suppl. 1): S38.
- 62 Coon H, Myles-Worsley M, Tiobech J, Hoff M, Rosenthal J, Bennett P *et al*. Evidence for a chromosome 2p13–14 schizophrenia susceptibility locus in families from Palau, Micronesia. *Mol Psychiatry* 1998; **3**: 521–527.
- 63 Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A *et al*. *Am J Psychiatry* 1998; **155**: 741–750.
- 64 Pajukanta P, Terwilliger JD, Perola M, Hiekkalinna T, Nuotio I, Ellonen P *et al*. Genomewide scan for familial combined hyperlipidemia genes in Finnish families, suggesting multiple susceptibility loci influencing triglyceride, cholesterol, and apolipoprotein B levels. *Am J Hum Genet* 1999; **64**: 1453–1463.
- 65 Levinson DF, Mowry BJ, Ewen K, Derek J, Nancarrow DJ, Hayward NK *et al*. Genome scan of schizophrenia: results of genotyping of positive regions [Abstract]. *Mol Psychiatry* 1999; **4** (Suppl. 1): S37.
- 66 Mowry BJ, Ewen KR, Nancarrow DJ, Lennon DP, Nertney DA, Jones HL *et al*. Second stage of a genome scan of schizophrenia: study of five positive regions in an expanded sample. *Am J Med Genet* 2000; **96**: 864–869.
- 67 Schwab SG, Hallmayer J, Lerer B, Albus M, Borrmann M, Honig S *et al*. Support for a chromosome 18p locus conferring susceptibility to functional psychoses in families with schizophrenia, by association and linkage analysis. *Am J Hum Genet* 1998; **63**: 1139–1152.
- 68 DeLisi LE, Devoto M, Lofthouse R, Poulter M, Smith A, Shields G *et al*. Search for linkage to schizophrenia on the X and Y chromosomes. *Am J Med Genet* 1994; **54**: 113–121.
- 69 Dann J, DeLisi LE, Devoto M, Laval S, Nancarrow DJ, Shields G *et al*. Linkage study of schizophrenia to markers within Xp11 near the MAOB gene. *Psychiatry Res* 1997; **70**: 131–143.
- 70 Paterson AD. X-Chromosome workshop. *Psych Genet* 1998; **8**: 121–126.
- 71 Wei J, Hemmings GP. Linkage disequilibrium mapping of chromosome Xp11 for a schizophrenia susceptibility locus [Abstract]. *Mol Psychiatry* 1999; **4**: 416–417.
- 72 Kendler KS, O'Neill FA, Burke J, Murphy B, Duke F, Straub RE *et al*. Irish study of high-density schizophrenia families: field methods and power to detect linkage. *Am J Med Genet* 1996; **67**: 179–190.
- 73 Spitzer RL, Williams JBW, Gibbon M, First MB. The structured clinical interview for DSM-III-R (SCID): I: History, rationale, and description. *Arch Gen Psychiatry* 1992; **49**: 624–629.
- 74 Kendler KS, Lieberman JA, Walsh D. The Structured Interview for Schizotypy (SIS): a preliminary report. *Schizophr Bull* 1989; **15**: 559–571.
- 75 Kendler KS, McGuire M, Gruenberg AM, Walsh D. An epidemiologic, clinical, and family study of simple schizophrenia in County Roscommon, Ireland. *Am J Psychiatry* 1994; **151**: 27–34.
- 76 Kendler KS, Diehl SR. The genetics of schizophrenia: a current, genetic-epidemiologic perspective. *Schizophr Bull* 1993; **19**: 261–285.
- 77 Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The Roscommon family study: I. Methods, diagnosis of probands, and risk of schizophrenia in relatives. *Arch Gen Psychiatry* 1993; **50**: 527–540.
- 78 Kendler KS, McGuire M, Gruenberg AM, Spellman M, O'Hare A, Walsh D. The Roscommon Family Study: II. The risk of non-schizophrenic nonaffective psychoses in relatives. *Arch Gen Psychiatry* 1993; **50**: 645–652.
- 79 Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The Roscommon family study: III. Schizophrenia-related personality disorders in relatives. *Arch Gen Psychiatry* 1993; **50**: 781–788.
- 80 Daniels JK, Williams NM, Williams J, Jones LA, Cardno AG, Murphy KC *et al*. No evidence for allelic association between schizophrenia and a polymorphism determining high or low catechol O-methyltransferase activity. *Am J Psychiatry* 1996; **153**: 268–270.
- 81 Eaves LJ. Effect of genetic architecture on the power of human linkage studies to resolve the contribution of quantitative trait loci. *Heredity (Edinburgh)* 1994; **72**: 175–192.
- 82 Levinson DF. Pragmatics and statistics in psychiatric genetics. *Am J Med Genet* 1997; **74**: 220–222.
- 83 Morton NE. Significance levels in complex inheritance. *Am J Hum Genet* 1998; **62**: 690–697.
- 84 Vieland VJ. Statistical Genetics '98: Bayesian Linkage Analysis, or: How I Learned to Stop Worrying and Love the Posterior Probability of Linkage. *Am J Hum Genet* 1998; **63**: 947–954.
- 85 Vieland VJ, Hodge SE. Statistical evidence: a likelihood paradigm (book review). *Am J Hum Genet* 1998; **63**: 283–289.
- 86 Greenberg DA, Abreu P, Hodge SE. The power to detect linkage in complex disease by means of simple LOD-score analyses. *Am J Hum Genet* 1998; **63**: 870–879.
- 87 Elston RC. 1996 William Allan Award Address. Algorithms and inferences: The challenge of multifactorial diseases. *Am J Hum Genet* 1997; **60**: 255–262.
- 88 Elston RC. Methods of linkage analysis—and the assumptions underlying them. *Am J Hum Genet* 1998; **63**: 931–934.
- 89 Pontieri FE, Passarelli F, Calò L, Caronti B. Functional correlates of nicotine administration: similarity with drugs of abuse. *J Mol Med* 1998; **76**: 193–201.
- 90 Segman R, Neeman T, Heresco-Levy U, Finkel B, Karagichev L, Schlafman M *et al*. Genotypic association between the dopamine D3 receptor and tardive dyskinesia in chronic schizophrenia. *Mol Psychiatry* 1999; **4**: 247–253.
- 91 Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet* 1995; **11**: 241–247.
- 92 Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 1998; **63**: 861–869.
- 93 Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K *et al*. A genetic linkage map of the human genome. *Cell* 1987; **51**: 319–337.
- 94 Straub RE, Speer MC, Luo Y, Rojas K, Overhauser J, Ott J *et al*. A microsatellite genetic linkage map of human chromosome 18. *Genomics* 1993; **15**: 48–56.
- 95 Straub RE, Sullivan PF, Ma Y, Myakishev MV, Harris-Kerr C, Wormley B *et al*. Susceptibility genes for nicotine dependence: a genome scan and followup in an independent sample suggest that chromosomes 2, 4, 10, 16, 17 and 18 merit further study. *Mol Psychiatry* 1999; **4**: 129–144.
- 96 Lange K, Weeks D, Boehnke M. Programs for pedigree analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol* 1988; **5**: 471–472.
- 97 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
- 98 Abreu P, Greenberg DA, Hodge SE. Direct power comparisons between simple LOD scores and NPL scores for linkage analysis in complex diseases. *Am J Hum Genet* 1999; **65**: 847–857.
- 99 Hodge SE, Abreu PC, Greenberg DA. Magnitude of type I error when single-locus linkage analysis is maximized over models: a simulation study. *Am J Hum Genet* 1997; **60**: 217–227.
- 100 Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D. The Roscommon family study: IV. Affective illness, anxiety disorders, and alcoholism in relatives. *Arch Gen Psychiatry* 1993; **50**: 952–960.
- 101 Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS. Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Genet* 1999; **65**: 876–884.
- 102 Darvasi A, Weinreb A, Minke V, Weller JL, Soller M. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 1993; **134**: 943–951.
- 103 Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of

- qualitative and quantitative traits. *Am J Hum Genet* 1995; **57**: 439–454.
- 104 Whittemore AS, Halpern J. Problems in the definition, interpretation, and evaluation of genetic heterogeneity. *Am J Hum Genet* 2001; **68**: 457–465.
- 105 Straub RE, MacLean CJ, Kendler KS. The putative schizophrenia locus on chromosome 6p: a brief overview of the linkage studies. *Mol Psychiatry* 1996; **1**: 89–92.
- 106 Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P *et al*. Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22–q23. *Nature Genet* 1996; **12**: 436–441.
- 107 Silverman JM, Greenberg DA, Altstiel LD, Siever LJ, Mohs RC, Smith CJ *et al*. Evidence of a locus for schizophrenia and related disorders on the short arm of chromosome 5 in a large pedigree. *Am J Med Genet* 1996; **67**: 162–171.
- 108 Sherrington R, Brynjolfsson J, Petursson H, Potter M, Dudleston K, Barraclough B *et al*. Localization of a susceptibility locus for schizophrenia on chromosome 5. *Nature* 1998; **336**: 164–167.
- 109 McGuffin P, Sargeant M, Hetti G, Tidmarsh S, Whatley S, Marchbanks RM. Exclusion of a schizophrenia susceptibility gene from the chromosome 5q11–q13 region: new data and a reanalysis of previous reports. *Am J Hum Genet* 1990; **47**: 524–535.
- 110 Ekelund J, Lichtermann D, Hovatta D, Ellonen P, Suvisaari J, Terwilliger JD *et al*. Genome-wide scan for schizophrenia in the Finnish population: evidence for a locus on chromosome 7q22 [Abstract]. *Mol Psychiatry* 1999; **4** (Suppl. 1): S3.
- 111 Sullivan PF, O'Neill FA, Walsh D, Ma Y, Kendler KS, Straub RE. Analysis of epistasis in linked regions in the Irish study of high-density schizophrenia families. *Am J Med Genet* 2001; **105**: 266–270.
- 112 Lucek P, Hanke J, Reich J, Solla SA, Ott J. Multi-locus nonparametric linkage analysis of complex trait loci with neural networks. *Hum Hered* 1998; **48**: 275–284.
- 113 Rao DC. CAT scans, PET scans, and genomic scans. *Genet Epidemiol* 1998; **15**: 1–18.
- 114 Egan MF, Weinberger DR. Neurobiology of schizophrenia. *Curr Opin Neurobiol* 1997; **7**: 701–707.
- 115 Mirmics K, Middleton FA, Marquez A, Lewis DA, Levitt P. Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron* 2000; **28**: 53–67.