View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Am. J. Hum. Genet. 65:1114–1124, 1999

A Genomewide Screen for Schizophrenia Genes in an Isolated Finnish Subpopulation, Suggesting Multiple Susceptibility Loci

Iiris Hovatta,[1,2,3] Teppo Varilo,[1,3] Jaana Suvisaari,[2] Joseph D. Terwilliger,[6] Vesa Ollikainen,[4] Ritva Arajärvi,[2] Hannu Juvonen,[2] Marja-Liisa Kokko-Sahin,[2] Leena Väisänen,[5] Heikki Mannila,[4] Jouko Lönnqvist,[2] and Leena Peltonen[1,3]

Departments of [1]Human Molecular Genetics and [2]Mental Health and Alcohol Research, National Public Health Institute, Departments of [3]Medical Genetics and [4]Computer Science, University of Helsinki, Helsinki; [5]Department of Psychiatry, University of Oulu, Oulu, Finland; and [6]Department of Psychiatry and Columbia Genome Center, Columbia University, New York

Summary

Schizophrenia is a severe mental disorder affecting ~1% of the world's population. Here, we report the results from a three-stage genomewide screen performed in a study sample from an internal isolate of Finland. An effort was made to identify genes predisposing for schizophrenia that are potentially enriched in this isolate, which has an exceptionally high lifetime risk for this trait. Ancestors of the local families with schizophrenia were traced back to the foundation of the population in the 17th century. This genealogical information was used as the basis for the study strategy, which involved screening for alleles shared among affected individuals originating from common ancestors. We found four chromosomal regions with markers revealing pairwise LOD scores >1.0: 1q32.2-q41 (Z_{max} = 3.82, dominant affecteds-only model), 4q31 (Z_{max} = 2.74, dominant 90%-penetrance model), 9q21 (Z_{max} = 1.95, dominant 90%-penetrance model), and Xp11.4-p11.3 (Z_{max} = 2.01, recessive 90%-penetrance model). This finding suggests that there are several putative loci predisposing to schizophrenia, even in this isolate.

Introduction

Schizophrenia (MIM 181500) is a mental disorder characterized by delusions, hallucinations, disturbed thinking, and bizarre behavior. The prevalence of schizophrenia varies; the disorder occurs in ~1% of most populations that have been studied, although some local subpopulations with a notably higher prevalence have been described (Böök et al. 1978; Hovatta et al. 1997). On the basis of the results of twin and adoption studies, genetic components, in combination with environmental factors, seem to play a role in the etiology of the disease (Tienari et al. 1987; Cannon et al. 1998).

To date, the results of several genomewide scans have been published; these findings are from family studies done in a number of different, mostly heterogeneous white populations (Barr et al. 1994; Coon et al. 1994; Moises et al. 1995; Levinson et al. 1998). The results of these studies and other, not fully published genome screens have shown some evidence of linkage to several chromosomal regions, including 2q (Levinson et al. 1998), 3p (Pulver et al. 1995), 4q (Levinson et al. 1998), 5q (Straub et al. 1997), 6p (Moises et al. 1995; Wang et al. 1995), 8p (Pulver et al. 1995), 9 (Moises et al. 1995; Levinson et al. 1998), 10q (Levinson et al. 1998), 11q (Levinson et al. 1998), 20 (Moises et al. 1995), and 22q (Coon et al. 1994; Pulver et al. 1994), that might harbor genes predisposing to schizophrenia. To date, the statistical significance of the findings and the subsequent replication attempts has not been conclusive, and, despite the intensive work being done, positional cloning efforts have not yet revealed any specific susceptibility genes for schizophrenia.

To increase the power to detect correlations between genetic markers and the disease gene, studies of isolated human populations have been undertaken. In small homogeneous populations, the genetic variability can be substantially reduced, and the environmental and cultural variabilities are, likewise, typically much lower than those in larger, more-cosmopolitan populations. Furthermore, in populations with a very small number of founders, the correlation between marker loci and the disease gene may also be increased as a result of the greater potential for linkage disequilibrium (LD) to be created and maintained (Thompson and Neel 1978; O'Brien et al. 1994; Terwilliger et al. 1998). All these factors should increase the chances of detecting at least

Received July 7, 1998; accepted July 27, 1999; electronically published September 3, 1999.

Address for correspondence and reprints: Dr. Leena Peltonen, UCLA School of Medicine, Department of Human Genetics, Gonda Center, Suite #6506, 695 Charles E. Young Drive South, Los Angeles, CA 9005-7088. E-mail: lpeltonen@mednet.ucla.edu

© 1999 by The American Society of Human Genetics. All rights reserved.
0002-9297/1999/6504-0022$02.00

some possible genetic loci, while decreasing the noise caused by other etiologic agents.

Elsewhere, we have described an isolate with 18,000 inhabitants in the northeastern part of Finland, where the age-corrected lifetime risk of schizophrenia is 3.2%, compared with the national average risk of 1.1% (Hovatta et al. 1997). This subpopulation was founded, by 40 families, at the end of the 17th century, and precise details of the subpopulation's births, deaths, marriages, and movements, kept by the Finnish Church, have been preserved. A well-known population history and high-quality health-care records, including hospitalizations and medications, make this isolate valuable for molecular genetic studies. Here, we describe the data obtained from a genome scan performed in this unique study sample with well-established genealogy.

Material and Methods

Genealogical Search

The genealogical study was performed in accordance with published criteria (Varilo et al. 1996). The names, dates, and places of birth of each patient's parents were used to trace ancestors from local church registries. Microfilm copies of records obtained from the Finnish National Archives were used for earlier periods.

Collection of the Families

All patients with schizophrenia who were born between 1940 and 1969 and who have either been hospitalized for schizophrenia or receive, from the Social Insurance Institution of Finland, a free-medication or disability pension for schizophrenia were identified in Finland (n = 29,124). These data were linked with data from the National Population Register to find first-degree relatives of the patients and to construct families with multiple affected individuals (Hovatta et al. 1997). We could identify a total of 365 families with at least one patient with schizophrenia and with at least one of the parents born in the isolate. The number of families with at least two affected siblings was 69 (Hovatta et al. 1997). A 20-ml EDTA blood sample was collected from all available family members who were willing to participate in the study and who came from families with at least two affected children (n = 20). In addition, blood samples from single affected patients with parents were collected (n = 25). This study has been approved by the ethical review board of the National Public Health Institute of Finland.

All available inpatient and outpatient case notes were collected for probands and relatives with any psychiatric diagnosis in any of the registers or for any psychiatric disturbance reported by key informants. Two pairs of psychiatrists, who were blind to the family structure, independently made a lifetime diagnosis of schizophrenia, according to the *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition* (DSM-IV). One of them also filled out the Operational Criteria (OPCRIT) checklist (McGuffin et al. 1991). In instances of diagnostic disagreement, a third psychiatrist reanalyzed the case to achieve consensus best-estimate DSM-IV lifetime diagnoses. The reliability of the diagnoses was tested using the κ statistic. Both J.S. and R.A. rated the diagnoses of 66 patients. The κ values (95% confidence interval) were excellent for schizophrenia, .907 (.805–1.010); schizoaffective psychosis, .816 (.612–1.019); liability class 3 (schizophrenia spectrum diagnoses), .817 (.568–1.067); and liability class 4 (bipolar disorder and major depressive disorder), .891 (.742–1.040). Both J.S. and M.-L.K.-S. rated the diagnoses of 25 patients without any disagreements. J.S. and H.J. agreed on the diagnoses of three patients, whereas one other patient was rated, by H.J., as having schizophrenia and, by J.S., as having schizoaffective disorder. We chose not to conduct direct interviews in this stage of the study, because the reliability of the diagnosis of schizophrenia in the Hospital Discharge Register of Finland has been assessed in several studies, and it has been shown to be good (Pakaslahti 1987; Isohanni et al. 1997; Cannon et al. 1998; Mäkikyrö et al. 1998). Therefore, we were able to identify every inpatient treatment for each patient and to contact the treating hospital. The outpatient clinic for treatment was identified through the treating psychiatrist.

DNA Analysis

DNA samples were extracted, from 20-ml EDTA blood, according to a standard procedure (Blin and Stafford 1976) modified to adapt Phase Lock Gel tubes (5′→3′). PCR amplification and gel electrophoresis performed with the use of either the A.L.F.express automated DNA sequencer or radioactivity have been described elsewhere (Pekkarinen et al. 1998).

Genome Scan Stage I

We typed 351 microsatellite markers from a modified Weber screening set, version 6.0, in 17 affected individuals from the large pedigree shown in figure 1 and in 18 controls (Sheffield et al. 1995). The average intermarker distance was 10.9 cM, with the largest gap being 28 cM on the X chromosome. Some of the markers were replaced with markers from the Généthon marker map (Dib et al. 1996). All affected individuals had a DSM-IV diagnosis of schizophrenia. Controls were either healthy married individuals from the large pedigree (n = 8) or healthy individuals from other families with schizophrenia that were collected from the isolate (n =

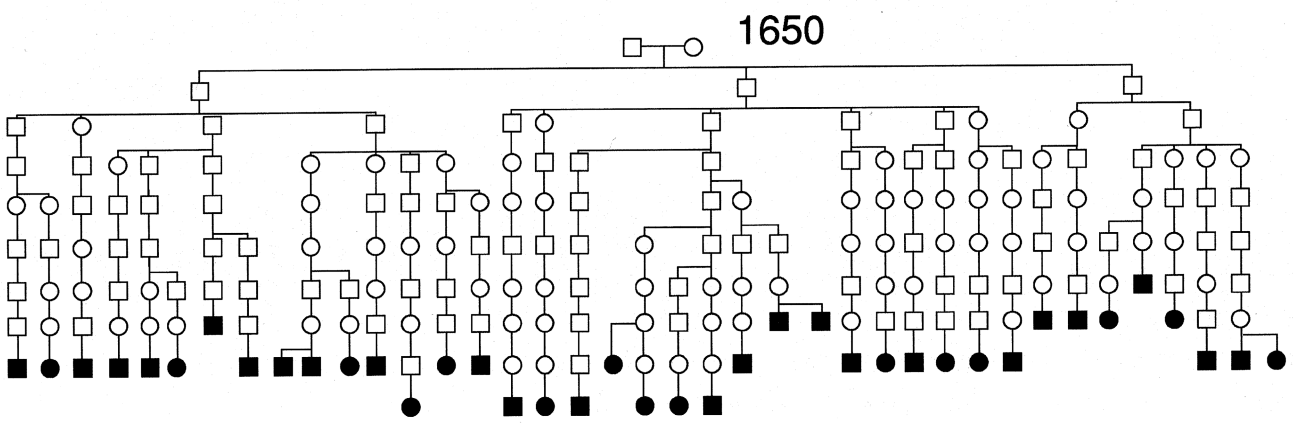

Figure 1 Large pedigree from the isolate with a founder couple born in approximately 1650. Circles represent females; squares represent males. Blackened circles and squares represent individuals with at least two children affected with schizophrenia.

10). Two case-control association tests were used to analyze the data from these 17 cases and 18 controls. Linkage analysis was performed, by use of a pseudomarker approach (Trembath et al. 1997) (see Statistical Analyses section, below), in nine affected individuals belonging to two larger reconstructed pedigrees (these individuals are denoted by an asterisk in fig. 2). Eight of these individuals were the same individuals who were part of the association analysis.

Genome Scan Stage II

In stage II, we analyzed 20 families with at least two affected individuals collected from the isolate (fig. 2). At this point, all possible connections between these nuclear families were traced back for three generations, and 13 families were found to be related to 1–4 other families. Not all of these families could be shown to be linked to the large pedigree shown in figure 1, even though such linkage is highly likely. A total of 27 markers meeting one of the following criteria from the stage I analyses were genotyped in this additional set of data: (1) markers yielding a P value $< .01$, or pairs of adjacent markers where both yield $P < .05$ in either of the association tests; (2) markers with a LOD score of $Z_1 > 1.0$ in the pseudomarker linkage analysis; and (3) pairs of adjacent markers that have one allele shared at each marker in $\geq 50\%$ of cases (i.e., those markers for which $\geq 50\%$ of cases possibly share a "two-locus haplotype"). These regions are denoted by red vertical lines in figure 3. The data were analyzed by use of both model-based linkage analysis and model-free affected sib-pair analysis. Two different diagnostic classes were considered: in class 1, only patients with schizophrenia were considered to be affected, and, in class 2, patients with schizoaffective disorder were also considered to be affected. Patients with schizophrenia-spectrum diagnosis, bipolar disorder, or major depressive disorder were considered to be of unknown status throughout.

Genome Scan Stage III

In stage III, a denser marker map was typed on the four chromosomal regions where markers showed a LOD score greater than 1.0 in stage II. The markers analyzed in stage III were chosen from the Généthon marker map and were radioactively labeled. The distances between the markers were derived from the genetic maps of both The Cooperative Human Linkage Center and Généthon and from the radiation hybrid map of the Whitehead Institute for Biomedical Research/MIT Center for Genome Research. The data were evaluated by linkage and affected sib-pair analyses.

Statistical Analyses

After the first screen, markers were analyzed, under the assumption of a founder effect, both by a likelihood-ratio test for LD, and by a $2 \times N$ contingency-table χ^2 analysis; both analyses were performed by means of the DISLAMB program (Terwilliger 1995). Pairwise linkage analysis was performed, by use of the MLINK program of the LINKAGE package (Lathrop and Lalouel 1984; Lathrop et al. 1984, 1986), in two subpedigrees containing nine of the typed affected individuals from the large pedigree (fig. 2). A pseudomarker approach (Trembath et al. 1997; Terwilliger 1998; J. Terwilliger, unpublished data), in which every meiosis was made informative at the trait locus to maximize the power of this study, was used to perform a traditional linkage analysis that has properties analogous to an affected relative-pair design, while also allowing for correlations to be made between multiple related individuals in an

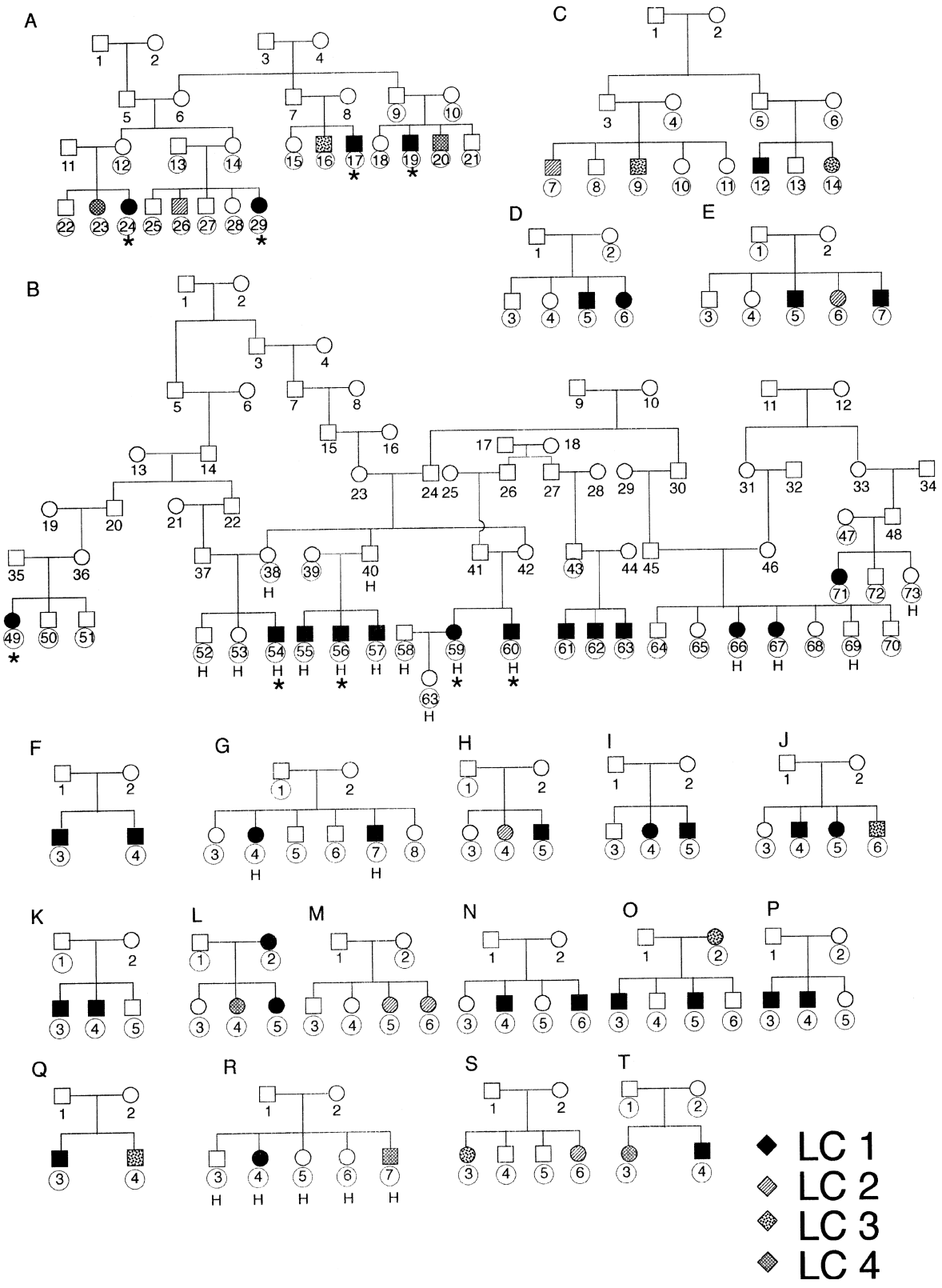

Figure 2 Families that underwent genotyping in stages II and III. Individuals with a circled number underwent genotyping, and those with an asterisk underwent linkage analysis in stage I. Individuals with an "H" below the circled number have a putative 6.6-cM haplotype on chromosome 1. Liability class (LC) 1 includes individuals with a DSM-IV diagnosis of schizophrenia; LC 2, those with schizoaffective disorder; LC 3, those with schizophreniform disorder, schizotypal personality disorder, schizoid personality disorder, delusional disorder, brief psychotic disorder, or psychotic disorder not otherwise specified; and LC 4, those with bipolar disorder and major depressive disorder.

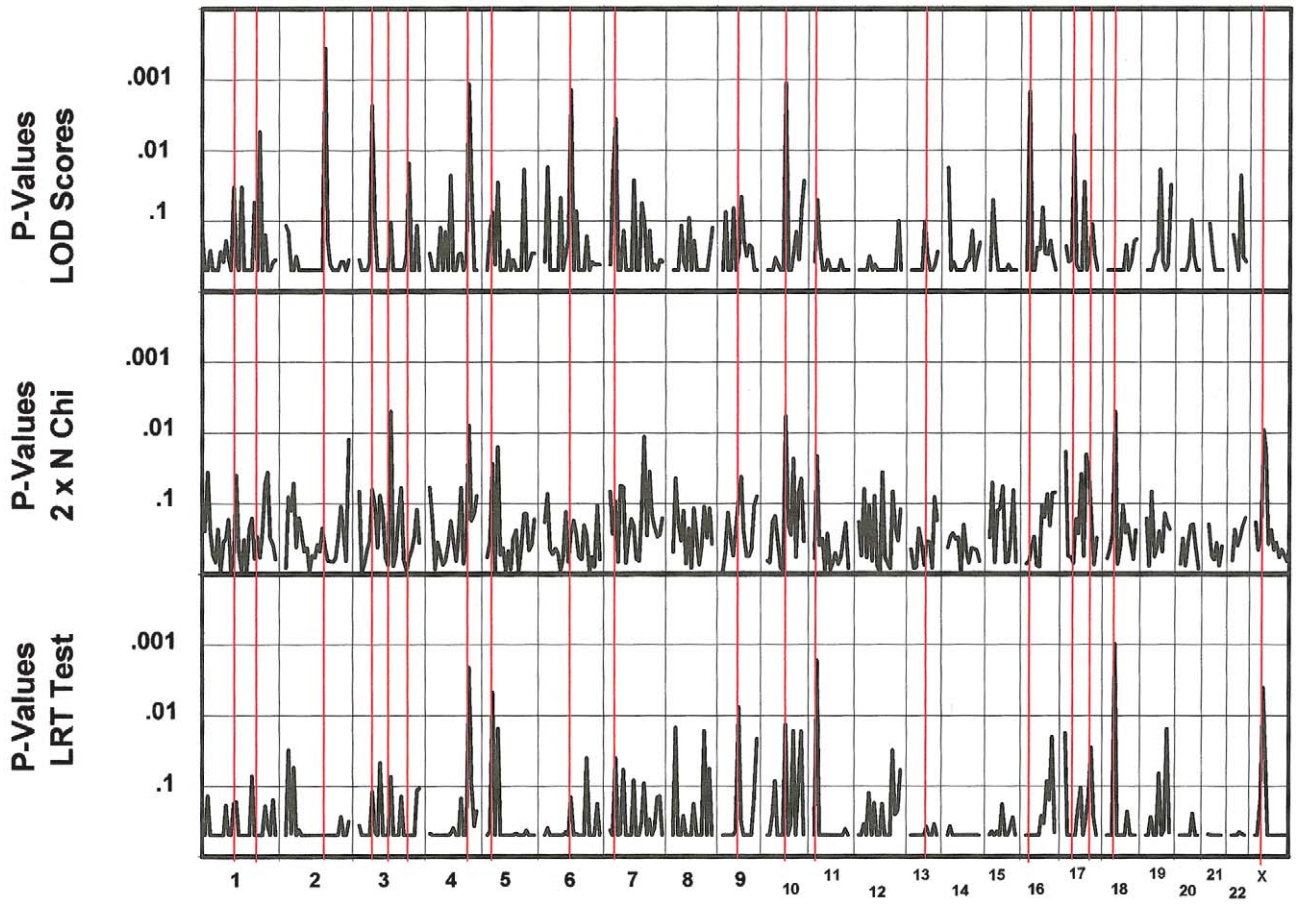

Figure 3 Results from stage I of the genome screen. The top panel shows results from the pseudomarker LOD-score analysis. The middle panel shows the results from the $2 \times N \chi^2$ test. The bottom panel shows the results from the likelihood ratio test (LRT). For each marker in the genome scan and for each statistic, the "theoretical *P* value" for each statistic is plotted on a log scale, such that higher peaks are more significant than lower peaks. The *P* values used on the *y*-axis compare the results of different statistics more easily, since they are expressed in a common currency. Red vertical lines indicate regions that were followed up in stage II.

efficient and unbiased manner, as described elsewhere (Trembath et al. 1997). This algorithm has been shown elsewhere to consistently be among the most powerful model-free methods, under a variety of true states of nature (Davis and Weeks 1997). LOD scores derived from this analysis are presented as Z_1.

In stages II and III, the genotype data were analyzed by pairwise linkage analysis, with the use of dominant and recessive inheritance models and two different penetrances, a 90%-penetrance model and an analogous affecteds-only model. No phenocopies were allowed, and a disease-gene frequency of .001 was assumed. Results from this linkage analysis are presented as Z_2. Pairwise linkage analysis was performed with the use of the LINKAGE package (Lathrop and Lalouel 1984; Lathrop et al. 1984, 1986), and the heterogeneity analysis was performed with the use of the HOMOG program (Ott 1986). Sib-pair analysis was performed by means of the SIBPAIR program (Kuokkanen et al. 1996).

Simulation of the Population Subisolate

To examine the effects of the inbred population structure on the power of association studies in this data, a population simulation study was done with the use of the algorithm described elsewhere (Terwilliger et al. 1998) (data available from the authors of the present study). The power to detect some evidence of LD in this study is good, although, as in the case of a complex disease, it is not expected to find a single disease-associated haplotype shared by all affected individuals—in contrast to the rare diseases of the Finnish disease heritage (Höglund et al. 1995; Nikali et al. 1995; Varilo et al. 1996).

Estimation of the Genomewide Significance

Since this population began, with a maximum of 160 chromosomes, in the 17th century and since today there are >36,000 chromosomes, there should be a high amount of identity by descent from these founders. Combined with the unavoidability of there being consanguinities over several generations in a finite closed system, the conventional assumptions of Hardy-Weinberg equilibrium and linkage equilibrium are not expected to hold true. For this reason, several statistical analyses were performed to address the genomewide significance of our findings. First, the inbreeding coefficients F_{is} and F_{st} were estimated. F_{st} is normally used to measure differentiation between allele-frequency distributions in two populations. We defined the cases and controls as "populations" and measured the degree of difference between these two populations genetically. Our hypothesis was that, near a disease gene, the differentiation would be larger, and, in other regions, it would fluctuate around some fixed value (which would be 0 if the cases and controls were randomly sampled from the whole population, but which would be some value >0 in this example, since the cases are more closely related to each other than they are to the controls). Although F_{st} is mathematically related to the $2 \times N$ contingency-table statistic we presented in figure 3 (Weir 1996), it has the property of being a quantifiable measurement of the strength of the association, which is not as dependent on the marker-allele frequencies as are the *P* values in that graph.

Second, we used a permutation test to determine whether our cases are more closely related to each other than they are to the controls. According to the null hypothesis of independence of the case and control samples, the marker-locus genotypes should be independent of the phenotype. Since it is possible that marker-locus genotypes may be correlated within individuals (i.e., LD may exist), the phenotypes were randomized for each individual, keeping the genotypes fixed across the entire genome-spanning set of markers. In each replicate, a set of 17 of the 35 genotyped individuals were randomly assigned to be "cases," and the remaining 18 individuals were designated as "controls"; association analyses were repeated for every marker in the genome. For each replicate, the number of observed *P* values <.05, .01, and .001 was determined, and this process was repeated 1,000 times to determine the expected genomewide rate of false positives for each statistic.

Simulations on the Significance of the Chromosome 1 Haplotype

To test the significance of the shared 6.6-cM haplotype in chromosome 1, which was found in the linkage analysis of extended pedigree B (fig. 2), we used the following method. We generated random chromosomes for the founders in pedigree B, by use of the marker maps from genome scan stage III. For allele frequencies, we used the frequencies from the whole study population. For each set of simulated chromosomes, we then simulated the meioses in the pedigree, with recombination probability depending on the genetic distance. For the resulting chromosomes of the 10 affected individuals in the pedigree, we computed the length of the longest shared haplotype (identical by state). To test the significance and frequency of the shared haplotype, further simulations were performed. Individuals 17 and 18, 25–28, 43–44, and 61–63 were omitted from the analysis, since they are related to the pedigree only through marriage. At least eight individuals were required to have the haplotype.

This simulation estimate is potentially vulnerable to undetected relationships between the founders of the pedigree. To test the possible effect of such relationships, we evaluated the findings when: (1) the two founders of the pedigree—founders 1 and 2—were assumed to be second cousins; (2) individuals 17–20 were assumed to be children of founders 1 and 2; and (3) founders 1 and 2 and individuals 17–20 were all assumed to be second cousins. These additions to the pedigree did not alter the results in any significant way.

Results

Genealogical Search

We first performed an extensive genealogical search to trace the ancestors of identified families with schizophrenia in the isolate. We identified a total of 365 nuclear families with at least one patient with schizophrenia and at least one of the parents born in the isolate. Our next step was to search the nationwide genealogical records to identify which of these families were related to each other within the past two generations; 151 families were found to have such a connection with at least one other family from the same set of 151 families. This subset of interrelated nuclear families was then traced back to the founding families of our northeastern subpopulation in the 17th century, and it was found to be related along multiple genealogical paths.

During the second phase of our study, all nuclear families with at least two children affected with schizophrenia were selected. Of the 151 nuclear families, 48 had at least two affected children, and 39 (81%) could be merged into a single large pedigree, with ~80 affected individuals in the last generation and a founder couple born, approximately, in 1650 (i.e., 7–10 generations from the patients' generation; fig. 1). The pedigree pre-

sented in figure 1 shows only one of the many possible genealogical connections between the individuals, and, in reality, the genealogy is much more complex. To identify all possible connections between individuals would be an enormous task, and we only wanted to confirm that patients in the last generation have a common ancestor, which would, hopefully, reduce genetic heterogeneity among affected individuals.

Genome Scan Stage I and Estimation of the Genomewide Significance

Results from the first stage of the genome scan are summarized in figure 3, and numerical details are available from table A on our web site. The most significant result obtained from the association tests was seen with the use of marker D18S542 (P = .00097), and the highest pairwise LOD score in the linkage analysis (Z_1 = 2.49) was obtained with marker D2S1399.

Over the entire genomewide association set of data, the inbreeding coefficient F_{is} was estimated to be .015, and, for individual marker loci, the point estimates ranged between −.15 and .22. To quantify the independence of case and control samples, a primitive analysis was performed, with the use of F_{st}, to measure the similarity between case and control populations (Terwilliger 1998). We found significant evidence ($P <$ .001) of a mild amount of subdivision (F_{st} = .025), with significance judged by permutation test. This would provide evidence that the affected individuals in our study data are more closely related to each other than they are to the controls. This is not too surprising, since controls were individuals who married into the pedigree and were not the members of the large family. Since the basis of our initial genome scan was the monitoring of shared alleles among affected individuals, the multiple connections between family members can be expected to result in an excessive number of false positive regions. However, this initial scan was not meant to be conclusive in itself but, rather, was designed to have maximum sensitivity while, admittedly, leading to the admission of a few too many false positives.

Similar results were obtained in another permutation test done on the basis of randomization of the phenotypes. As shown in table 1, several shared regions among affected individuals are expected to occur simply by chance, as a result of inbred pedigree structures. However, one still expects the cases to be even more similar to each other with regard to having marker loci located near a shared disease allele rather than elsewhere in the genome, and, therefore, the relative significance of the association tests for different markers, in the context of the entire genome, should still be useful for discriminating which regions of the genome are more likely to harbor a schizophrenia-predisposing locus.

Table 1

Results from the Permutation Test Showing Numbers of Markers with Corresponding *P* Values

P Value	No. in Genome Screen		
	LRT	2 × *N*	Both
	Observed Positives		
<.05	24	24	13
<.01	6	5	3
<.001	1	0	0
<.0001	0	0	0
	Expected False Positives		
<.05	12.21	15.86	5.929
<.01	2.857	2.429	.750
<.001	.393	.107	.036
<.0001	.036	.000	.000

F_{st} estimates can also be used as a quantifier of the strength of allelic association between a marker locus and a disease phenotype (Weir 1996). In this genome scan, the largest values of F_{st} were also found for markers on chromosomes 4 (F_{st} = .13), 10 (F_{st} = .11), and 11 (F_{st} = .09), each of which had a P value <.01. At least in this study, the F_{st} estimates correlated well with the P values obtained in the association test.

Genome Scan Stage II

In stage II, we typed, on 19 chromosomal regions in 20 families collected from the isolate, 27 markers that provided the most interesting results in stage I (see the Material and Methods section, above, for further details). Results of the linkage and sib-pair analyses are shown in table B*a* on our web site. We were left with four markers showing a LOD score of $Z_2 > 1.0$. These markers were D1S2141, on chromosome 1, which had a maximum LOD score of Z_2 = 3.73; D4S1629, on chromosome 4, which had a maximum LOD score of Z_2 = 2.36; D9S922, on chromosome 9, which had a maximum LOD score of Z_2 = 1.95; and DXS6810, on chromosome X, which had a maximum LOD of score of Z_2 = 1.19.

Genome Scan Stage III

In stage III, a denser marker map was typed on the four regions from stage II where markers showed a LOD score >1.0. On chromosome 1, 11 markers were genotyped to an 18-cM region (tables C*a* and C*b* on our web site). Three of the markers gave LOD scores $Z_2 > 3.0$, with the maximum LOD score being Z_2 = 3.82, which was attained with marker D1S2891, by use of a dominant affecteds-only model in diagnostic class 2. No significant evidence for locus heterogeneity was seen, although the HOMOG program might not optimally

detect locus heterogeneity, even though it may exist, in such a small set of family data as ours. In addition, we observed a 6.6-cM haplotype, consisting of markers D1S2891, D1S491, D1S205, and D1S425, that is segregating within three of the families with schizophrenia from this isolate (fig. 2). After initial recognition of the haplotype, we performed genotyping in 25 additional patients and their parents from the isolate, but only one additional patient carried this haplotype, and none of the 50 nontransmitted chromosomes from the parents of these individuals had the haplotype (not significant; $P > .5$). Overall, 12 patients with schizophrenia, 1 patient with major depressive disorder, 7 healthy siblings of patients with schizophrenia, and 1 healthy spouse and his daughter, all of whom were from four families, carried this particular haplotype. One individual in the data set (pedigree B, individual 63) had two copies of this haplotype in a homozygous state, yet she was phenotypically healthy. A further genealogical search revealed that her father (individual 58) is a grandson of individual 10's cousin.

Of the 11 markers typed on chromosome 4, 2 had a LOD score of $Z_2 > 2.0$, with the maximum LOD score being $Z_2 = 2.74$ with marker D4S1586, by use of a dominant model with 90% penetrance and a narrow diagnostic class (tables C*c* and C*d* on our web site). In addition, an affected sib-pair LOD score of 2.09 ($P = .00097$) was observed with this marker. Positive LOD scores were observed for a fairly large 22-cM region, but no common haplotype was seen among affected individuals.

Only one of the six markers typed on chromosome 9, D9S922, yielded a LOD score of $Z_2 > 1$ (tables C*e* and C*f* on our web site). This is the same marker that was positive in stages I and II, and the obtained maximum LOD score was $Z_2 = 1.95$, by use of a dominant model with 90% penetrance under diagnostic class 2.

Several of the nine genotyped chromosome-X markers showed LOD scores of $Z_2 > 1$, with the maximum being $Z_2 = 2.01$ with marker *MAOB*, under a recessive inheritance model with 90% penetrance and diagnostic class 1 (tables C*g* and C*h* on our web site). On this chromosomal region, we also saw positive LOD scores for a quite-wide region of 15 cM, but no evidence for a common haplotype emerged among affected individuals.

Simulations on the Significance of the Chromosome 1 Haplotype

We performed simulations to test both the frequency and the length of the haplotype, to evaluate the effect of the inbred population structure on the results of the potential 6.6-cM haplotype shared among affected individuals. A simulation test (see the Material and Methods section, above) indicated that the observed amount of haplotype sharing would occur by chance in our study sample, with a probability <.001. Further simulation showed that adding some potentially undetected relationships between a small number of the founders of the pedigree did not significantly increase the probability of the occurrence of such a haplotype. We also tested the significance of the frequency of the observed haplotype, by use of a similar simulation method. For each simulated set of chromosomes, we recorded the rarest haplotype that was shared by the eight affected individuals in the pedigree. The results are shown in table D on our web site. The frequency of the shared 6.6-cM haplotype was found to be .0008, on the basis of the allele frequencies of the study population. According to the simulation, this corresponds to a P value of .0002. These data would suggest that the putative haplotype does not represent a by-chance finding in this inbred pedigree.

Discussion

We have performed a three-stage genomewide screen to identify loci predisposing to schizophrenia in an isolated population of 18,000 inhabitants from northeastern Finland. In this community, the lifetime risk of schizophrenia is 3.2%, whereas the national average is 1.1% (Hovatta et al. 1997). We made a special effort to genealogically define our study sample and could confirm that affected individuals shared common ancestors; we based our genome-scan strategy on this information. We found four chromosomal regions showing LOD scores >1.0. Estimation of the genomewide significance level in isolated populations is difficult because of the high level of inbreeding, and this question has not been taken into account in many earlier studies dealing with data sets collected from isolated populations. We made a special effort to address this problem by performing several simulation analyses to evaluate the probability of seeing, in this inbred study sample, sharing of alleles by chance. In the first stage of the screen, the controls we selected were healthy individuals who married into families with schizophrenia, as opposed to family-based controls. This method requires genotyping only one control individual, as opposed to two control individuals, which is necessary when family-based controls are used. After the first stage of the screen, we observed that the cases are more closely related to each other than to these controls. That affected individuals are more closely related to each other than to the controls provides further evidence that the controls represent a more random sample of this isolated subpopulation than do members of the large pedigree; however, this fact should not increase the false negative rate of our study but, rather, should increase the false positive rate. Being aware that some identified regions are shared by affected individuals simply as a result of inbreeding, we still pursued further

analyses of the chromosomal regions showing the most significant results after the first and second stages.

Linkage analysis in complex diseases is problematic as a result of inheritance models that are incorporated into the analysis because the true model is typically unknown. Some assumptions about the mode of inheritance must be made, and such linkage analysis done with the use of extended pedigrees has still been, to date, the most successful gene-mapping protocol of complex diseases (St. George-Hyslop et al. 1987; Hall et al. 1990; Pericak-Vance et al. 1991; Schellenberg et al. 1992). We used dominant and recessive inheritance models, initially with an assumption of 90% penetrance and, subsequently, with an analogous affecteds-only model. We allowed no phenocopies (individuals who are affected without having the particular susceptibility gene), and the gene frequency was set to .001. It is obvious that such models are incorrect, but we wanted to use robust models to maximize power (Davis and Weeks 1997; Trembath et al. 1997; Terwilliger 1998). We actually tested the effect of changing the linkage-analysis parameters in a subset of the data. The use of a lower penetrance value and a higher disease-gene frequency, in addition to allowing for phenocopies, did not have considerable effect on the results; the LOD scores neither increased nor decreased significantly (data not shown). In addition to undergoing linkage analysis, the data were also analyzed with analogous model-free affected sib-pair methods. Since deviations from Hardy-Weinberg equilibrium and linkage equilibrium can occur in an inbred population, the *P* values obtained in studies using isolated populations should be interpreted with caution. Since we used multiple different models, multiple testing should be taken into account when interpreting the results.

The highest pairwise LOD score in this genomewide scan was $Z_2 = 3.82$, which was obtained with marker D1S2891, by use of a dominant affecteds-only model in linkage analysis and by considering patients with schizophrenia or schizoaffective disorder to be affected. No significant evidence of locus heterogeneity was found. On the same chromosomal region, we also observed a 6.6-cM haplotype segregating in four core families. The haplotype was also seen in healthy family members, and, assuming a locus for schizophrenia on this region, these individuals would represent nonpenetrant cases, as is expected for predisposing loci for a complex disease. The haplotype was present in 8 of 13 affected individuals in the largest family, as well as in 8 of 18 healthy individuals. The putative haplotype would narrow, to a much more restricted 6.6-cM region, the otherwise wide 18-cM region in which the positive LOD scores are seen. However, further genotyping with the use of more markers is needed for additional individuals from the same community, to confirm the significance of the haplotype in practice. It will be interesting to know the frequency of the haplotype both in this community and in the general Finnish population. In this study, we used 90% penetrance and low-penetrance affecteds-only models in linkage analysis. If all haplotype carriers will have the putative disease gene, the true penetrance of the gene will be lower, ~50%. Determining the correct penetrance of the putative susceptibility gene should increase the power of linkage analysis.

With the use of the markers analyzed so far, no statistically significant evidence for LD could be observed, and it might be possible that extended LD is not detectable in this population for such complex diseases as schizophrenia (Weiss 1996). However, because of the young age of the population, the odds are better that, even with the use of a marker map as sparse as that used in this study, some evidence of LD will be found in this population rather than in the larger, more-cosmopolitan study populations. The statistical significance of the observed haplotype, on the basis of simulation studies, has approximately the same magnitude as the significance of the linkage finding. Taken together, these findings add to our confidence that there might be a true susceptibility locus for schizophrenia in this chromosomal region; however, extensive studies in this population and in other populations are needed to determine the frequency of this haplotype in both affected individuals and healthy individuals.

Another promising region was identified on chromosome 4, where the marker D4S1586 resulted in a maximum LOD score of $Z_2 = 2.74$, by use of a dominant model with 90% penetrance and a narrow diagnostic model. Even though we could not see any evidence for a shared haplotype among affected individuals, the region still looks promising, since positive LOD scores are seen on a fairly large 22-cM region, and the data are supported by affected sib-pair analysis that reveals a maximum LOD score of 2.09, with marker D4S1586 corresponding to a *P* value of .00097.

On chromosome 9, a maximum LOD score of $Z_2 = 1.95$ was obtained with the marker D9S922, but all surrounding markers showed LOD scores <1.0. Interestingly, positive findings have been reported in genome screens ~25 cM from D9S922 (Moises et al. 1995) and 4 cM from the same marker (Levinson et al. 1998). Clearly, this region needs to be studied by other groups to determine whether it would harbor a locus predisposing to schizophrenia in some families.

In addition, our chromosome-X finding looks interesting because several other groups have also reported positive results around our best marker, *MAOB* (an intragenic marker for monoamine oxidase B gene), which yielded a LOD score of $Z_2 = 2.01$, with a recessive 90%-penetrance model and narrow diagnostic classification. The use of two other markers also resulted in LOD

scores >1.0. Originally, with the use of marker DXS7, a maximum LOD score of $Z_2 = 1.83$ was reported in the close vicinity of *MAOB* (DeLisi et al. 1994). The highest LOD score reported on DXS7 is $Z_2 = 2.16$, under a dominant inheritance model in a multicenter study (Dann et al. 1997). Some other groups have not been able to see evidence for linkage on this region (Okoro et al. 1995), and no sequence variation associated with schizophrenia could be seen when the *MAOB* gene was screened, by dideoxy fingerprinting, in 100 male patients (Sobell et al. 1997). In an association study (Coron et al. 1996), no significant difference in allele frequencies was observed between patients and controls, but a trend toward an association between allele 1 of the *MAOB* gene and paranoid schizophrenia was found.

In sum, in this study, in which we aimed to maximize the benefits of genetically isolated populations in the search of complex disease loci, positive results for the schizophrenia trait were obtained on four chromosomal regions: 1q32.2-q41, 4q31, 9q21, and Xp11.4-p11.3. That they were found warrants further studies in this family set and in independent family sets from other populations, to determine the significance of these findings in other parts of Finland and in other ethnic groups. Although we are cautious about the significance of these loci until further examination has been done, our results from this genome screen may eventually help to identify specific genes contributing to susceptibility to schizophrenia.

Acknowledgments

We thank Eero Vierimaa for his invaluable help in the isolate. We thank Päivi Kortelainen, Liisa Moilanen, and Merja Nissi for helping in the sample collection, and we thank Jaana Hartiala, Soili Johansson, Anne Jokiaho, Elli Kempas, Mari Sipilä, and Anne Vikman for their excellent technical assistance. We thank Tero Juntunen and Aki Suomalainen for their help in data management, and we thank Jesper Ekelund and Petra Pekkarinen for their helpful comments during this study. Above all, we wish to thank the participating families with schizophrenia. This work was financially supported by the Jalmari and Rauha Ahokas Foundation, the Oskar Öflund Foundation, The Foundation for Pediatric Research (Ulla Hjelt Fund), The Finnish Cultural Foundation, The Juselius Foundation, The Paulo Foundation, and the Hitchings-Elion Fellowship from the Burroughs-Wellcome Fund (to J.D.T.).

Electronic-Database Information

Accession number and URLs for the data in this article are as follows:

Authors' Web Site for Additional Article Data, http://www.ktl.fi/molbio/sch

Généthon, http://www.genethon.fr/genethon_en.html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for schizophrenia [MIM 181500])

The Cooperative Human Linkage Center, http://lpg.nci.nih.gov/CHLC/

Whitehead Institute for Biomedical Research/MIT Center for Genome Research, http://www.genome.wi.mit.edu/

References

Barr CL, Kennedy JL, Pakstis AJ, Wetterberg L, Sjögren B, Bierut L, Wadelius C, et al (1994) Progress in a genome scan for linkage in schizophrenia in a large Swedish kindred. Am J Med Genet 54:51–58

Blin N, Stafford DW (1976) A general method for isolation of high molecular weight DNA from eukaryotes. Nucleic Acids Res 3:2303–2308

Böök JA, Wetterberg L, Modrzewska K (1978) Schizophrenia in a North Swedish geographical isolate, 1900–1977: epidemiology, genetics and biochemistry. Clin Genet 14:373–394

Cannon TD, Kaprio J, Lönnqvist J, Huttunen M, Koskenvuo M (1998) The genetic epidemiology of schizophrenia in a Finnish twin cohort: a population-based modeling study. Arch Gen Psychiatry 55:67–74

Coon H, Jensen S, Holik J, Hoff M, Myles-Worsley M, Reimherr F, Wender P, et al (1994) Genomic scan for genes predisposing to schizophrenia. Am J Med Genet 54:59–71

Coron B, Campion D, Thibaut F, Dollfus S, Preterre P, Langlois S, Vasse T, et al (1996) Association study between schizophrenia and monoamine oxidase A and B DNA polymorphism. Psychiatry Res 62:221–226

Dann J, DeLisi LE, Devoto M, Laval S, Nancarrow DJ, Shields G, Smith A, et al (1997) Linkage study of schizophrenia to markers within Xp11 near the MAOB gene. Psychiatry Res 70:131–143

Davis S, Weeks DE (1997) Comparison of nonparametric statistics for detection of linkage in nuclear families: single-marker evaluation. Am J Hum Genet 61:1431–1444

DeLisi LE, Devoto M, Lofthouse R, Poulter M, Smith A, Shields G, Bass N, et al (1994) Search for linkage to schizophrenia on the X and Y chromosomes. Am J Med Genet 54:113–121

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, et al (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380:152–154

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684–1689

Höglund P, Sistonen P, Norio R, Holmberg C, Dimberg A, Gustavson K-H, de la Chapelle A, et al (1995) Fine mapping of the congenital chloride diarrhea gene by linkage disequilibrium. Am J Hum Genet 57:95–102

Hovatta I, Terwilliger JD, Lichtermann D, Mäkikyrö T, Suvisaari J, Peltonen L, Lonnqvist J (1997) Schizophrenia in the genetic isolate of Finland. Am J Med Genet 74:353–360

Isohanni M, Mäkikyrö T, Moring J, Räsänen P, Hakko H, Partanen U, Koiranen M, et al (1997) A comparison of clinical and research DSM-III-R diagnoses of schizophrenia in

a Finnish national birth cohort. Soc Psychiatry Psychiatr Epidemiol 32:303–308

Kuokkanen S, Sundvall M, Terwilliger JD, Tienari PJ, Wikström J, Holmdahl R, Pettersson U, et al (1996) A putative vulnerability locus to multiple sclerosis maps to 5p14-p12 in a region syntemic to the murine locus Eae2. Nat Genet 13:477–480

Lathrop GM, Lalouel JM (1984) Easy calculations of LOD scores and genetic risks on small computers. Am J Hum Genet 36:460–465

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Lathrop GM, Lalouel JM, White RL (1986) Construction of human linkage maps: likelihood calculations for multilocus linkage analysis. Genet Epidemiol 3:39–52

Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A, Hayward NK, et al (1998) Genome scan of schizophrenia. Am J Psychiatry 155:741–750

Mäkikyrö T, Isohanni M, Moring J, Hakko H, Hovatta I, Lönnqvist J (1998) Accuracy of register-based diagnoses in a genetic study. European Psychiatry 13:57–62

McGuffin P, Farmer AE, Harvey I (1991) A polydiagnostic application of operational criteria in psychotic illness: development and reliability of the OPCRIT system. Arch Gen Psychiatry 48:764–770

Moises HW, Yang L, Kristbjarnarson H, Wiese C, Byerley W, Macciardi F, Arolt V, et al (1995) An international two-stage genome-wide search for schizophrenia susceptibility genes. Nat Genet 11:321–324

Nikali K, Suomalainen A, Terwilliger J, Koskinen T, Weissenbach J, Peltonen L (1995) Random search for shared chromosomal regions in four affected individuals: the assignment of a new hereditary ataxia locus. Am J Hum Genet 56: 1088–1095

O'Brien E, Kerber RA, Jorde LB, Rogers AR (1994) Founder effect: assessment of variation in genetic contributions among founders. Hum Biol 66:185–204

Okoro C, Bell R, Sham P, Nanko S, Asherson P, Owen M, Gill M, et al (1995) No evidence for linkage between the X-chromosome marker DXS7 and schizophrenia. Am J Med Genet 60:461–464

Ott J (1986) Linkage probability and its approximate confidence interval under possible heterogeneity. Genet Epidemiol Suppl 1:251–257

Pakaslahti A (1987) On the diagnosis of schizophrenic psychosis in clinical practice. Psychiatrica Fennica 18:63–72

Pekkarinen P, Hovatta I, Hakola P, Järvi O, Kestilä M, Lenkkeri U, Adolfsson R, et al (1998) Assignment of the locus for PLO-SL, a frontal-lobe dementia with bone cysts, to 19q13. Am J Hum Genet 62:362–372

Pericak-Vance MA, Bebout JL, Gaskell PC Jr, Yamaoka LH, Hung WY, Alberts MJ, Walker AP, et al (1991) Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. Am J Hum Genet 48:1034–1050

Pulver AE, Karayiorgou M, Wolyniec PS, Lasseter VK, Kasch L, Nestadt G, Antonarakis S, et al (1994) Sequential strategy to identify a susceptibility gene for schizophrenia: report of potential linkage on chromosome 22q12-q13.1: part 1. Am J Med Genet 54:36–43

Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt G, Blouin J-L, Kimberland M, et al (1995) Schizophrenia: a genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet 60:252–260

Schellenberg GD, Bird TD, Wijsman EM, Orr HT, Anderson L, Nemens E, White JA, et al (1992) Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. Science 258:668–671

Sheffield VC, Weber JL, Buetow KH, Murray JC, Even DA, Wiles K, Gastier JM, et al (1995) A collection of tri- and tetranucleotide repeat markers used to generate high quality, high resolution human genome-wide linkage maps. Hum Mol Genet 4:1837–1844

Sobell JL, Lind TJ, Hebrink DD, Heston LL, Sommer SS (1997) Screening the monoamine oxidase B gene in 100 male patients with schizophrenia: a cluster of polymorphisms in African-Americans but lack of functionally significant sequence changes. Am J Med Genet 74:44–49

St. George-Hyslop PH, Tanzi RE, Polinsky RJ, Haines JL, Nee L, Watkins PC, Myers RH, et al (1987) The genetic defect causing familial Alzheimer's disease maps on chromosome 21. Science 235:885–890

Straub RE, MacLean CJ, O'Neill FA, Walsh D, Kendler KS (1997) Support for a possible schizophrenia vulnerability locus in region 5q22-31 in Irish families. Mol Psychiatry 2:148–155

Terwilliger JD (1995) A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. Am J Hum Genet 56: 777–787

——— (1998) Linkage analysis—model-based. In: Encyclopedia of biostatistics. John Wiley and Sons, Chichester, New York, pp 2279–2291

Terwilliger JD, Zöllner S, Laan M, Pääbo S (1998) Mapping genes through the use of linkage disequilibrium generated by genetic drift: "drift mapping" in small populations with no demographic expansion. Hum Hered 48:138–154

Thompson EA, Neel JV (1978) Probability of founder effect in a tribal population. Proc Natl Acad Sci USA 75:1442–1445

Tienari P, Lahti I, Sorri A, Naarala M, Moring J, Wahlberg KE, Wynne LC (1987) The Finnish adoptive family study of schizophrenia. J Psychiatr Res 21:437–445

Trembath RC, Clough RL, Rosbotham JL, Jones AB, Camp RDR, Frodsham A, Browne J, et al (1997) Identification of a major susceptibility locus on chromosome 6p and evidence for further disease loci revealed by a two stage genome-wide search in psoriasis. Hum Mol Genet 6:813–820

Varilo T, Savukoski M, Norio R, Santavuori P, Peltonen L, Järvelä I (1996) The age of human mutation: genealogical and linkage disequilibrium analysis of the CLN5 mutation in the Finnish population. Am J Hum Genet 58:506–512

Wang S, Sun CE, Walczak CA, Ziegle JS, Kipps BR, Goldin LR, Diehl SR (1995) Evidence for a susceptibility locus for schizophrenia on chromosome 6pter-p22. Nat Genet 10: 41–46

Weir BS (1996) Genetic data analysis II. Sinauer, Sunderland

Weiss KM (1996) Is there a paradigm shift in genetics? lessons from the study of human diseases. Mol Phylogenet Evol 5: 259–265