Am. J. Hum. Genet. 64:1670–1678, 1999

Assessing the Feasibility of Linkage Disequilibrium Methods for Mapping Complex Traits: An Initial Screen for Bipolar Disorder Loci on Chromosome 18

Michael A. Escamilla,[1,2,*,†] L. Alison McInnes,[1,2,*] Mitzi Spesny,[8] Victor I. Reus,[2] Susan K. Service,[1] Norito Shimayoshi,[1,‡] David J. Tyler,[1] Sandra Silva,[8] Julio Molina,[8] Alvaro Gallegos,[9,§] Luis Meza,[9] Maria L Cruz,[8] Steven Batki,[3] Sophia Vinogradov,[2] Thomas Neylan,[4] Jasmine B. Nguyen,[1] Eduardo Fournier,[8] Carmen Araya,[8] Samuel H. Barondes,[2] Pedro Leon,[8] Lodewijk A. Sandkuijl,[5,6,7] and Nelson B. Freimer[1,2]

[1]Neurogenetics Laboratory, [2]Center for Neurobiology and Psychiatry, [3]Department of Psychiatry, San Francisco General Hospital, and [4]Department of Psychiatry, Veterans Administration Medical Center, University of California San Francisco, San Francisco; [5]Department of Clinical Genetics, Erasmus University, Rotterdam, [6]Department of Human Genetics, Leiden University, Leiden, and [7]Department of Medical Genetics, Groningen University, Groningen, the Netherlands; and [8]Cell and Molecular Biology Research Center and Escuela de Medicina, Universidad de Costa Rica, and [9]Hospital Calderon Guardia, San José, Costa Rica

Summary

Linkage disequilibrium (LD) analysis has been promoted as a method of mapping disease genes, particularly in isolated populations, but has not yet been used for genome-screening studies of complex disorders. We present results of a study to investigate the feasibility of LD methods for genome screening using a sample of individuals affected with severe bipolar mood disorder (BP-I), from an isolated population of the Costa Rican central valley. Forty-eight patients with BP-I were genotyped for markers spaced at ~6-cM intervals across chromosome 18. Chromosome 18 was chosen because a previous genome-screening linkage study of two Costa Rican families had suggested a BP-I locus on this chromosome. Results of the current study suggest that LD methods will be useful for mapping BP-I in a larger sample. The results also support previously reported possible localizations (obtained from a separate collection of patients) of BP-I–susceptibility genes at two distinct sites on this chromosome. Current limitations of LD screening for identifying loci for complex traits are discussed, and recommendations are made for future research with these methods.

Introduction

Identifying genes for disorders with complex inheritance patterns is one of the greatest challenges in biomedical research (Lander and Schork 1994). Such disorders, which include many of the most prevalent human diseases, are difficult to map with standard linkage methods. It has been suggested that the availability of dense marker maps covering the genome will make linkage disequilibrium (LD) analysis a feasible approach for screening the genome to map complex disorders (Risch and Merikangas 1996). Current marker maps are not sufficiently dense to enable such studies to be performed in heterogeneous populations or in populations that were founded in the distant past. However, the success of genome-screening LD-mapping studies of genetically simple and/or rare diseases in recently founded isolated populations (Houwen et al. 1994; Puffenberger et al. 1994; Friedman et al. 1995; Newport et al. 1996) provide the impetus for testing the utility of LD methods for mapping complex diseases in such populations (Escamilla et al. 1996). In populations where randomly sampled patients are on average <20 generations removed from their last common ancestor, LD may be maintained for sizable regions around disease genes. Such LD should be manifested by affected individuals sharing alleles, identical by descent (IBD), at markers spaced at intervals of several centimorgans surrounding

Received February 5, 1999; accepted for publication March 16, 1999; electronically published May 6, 1999.

Address for correspondence and reprints: Dr. Nelson Freimer, University of California San Francisco, Box F-0984 (regular mail), 401 Parnassus Avenue (courier mail), San Francisco, CA 94143. E-mail: nelson@ngl.ucsf.edu

0002-9297/99/6406-0020$2.00

*These authors contributed equally to this work.

†Present affiliation: Department of Psychiatry, University of Texas Health Science Center at San Antonio, San Antonio, Texas.

‡Present affiliation: Department of Psychiatry, Nara Medical School, Nara, Japan.

§Dr. Gallegos died after this paper was accepted for publication. This paper is dedicated to his memory.

a disease gene. We now present the results from the first stage of a study in which LD methods were used to screen for loci that predispose to severe bipolar mood disorder (BP-I), which is common and is almost certainly characterized by a complex mode of inheritance. The study was done in a relatively recently founded isolated population, that of the central valley of Costa Rica (CVCR) (Escamilla et al. 1996), where founder effects have already been observed for several inherited diseases (Saborio 1992; Uhrhammer et al. 1995; Shah et al. 1997).

Despite long-standing evidence that BP-I has a genetic basis (Escamilla et al. 1997), genome scans for linkage have provided equivocal results (Risch and Botstein 1996; Nurnberger et al. 1997) that fail to satisfy the levels of significance suggested for genomewide screens by Lander and Kruglyak (1995). The failure to identify BP-I loci definitively, by standard linkage approaches, probably reflects uncertainty regarding mode of inheritance, high phenocopy rates, difficulty in demarcation of distinct phenotypes, and presumed genetic heterogeneity. LD-based mapping approaches within population isolates may offer a means of diminishing several of these obstacles. An approach (such as LD mapping) that samples individuals from an entire population can more easily ascertain a large set of patients with a narrowly defined, reliably diagnosed phenotype (in this case, BP-I) than linkage-based approaches that require ascertainment of family units with multiple affected cases. Within a population isolate, genetic heterogeneity of BP-I may also be less than in larger, genetically mixed populations, as there is a high probability that individuals with such a phenotype share descent from a few common ancestors.

We collected a sample of patients with BP-I, for LD analysis, by identifying individuals currently living in the CVCR who had known CVCR ancestry. This sample was collected independently of our previous pedigree-based studies of BP-I in Costa Rica. Our aim in the current study was to evaluate the feasibility of identifying BP-I loci by LD screening in this population, as proposed in Escamilla et al. (1996). To do this, we conducted an LD screen of an entire single chromosome (chromosome 18). This chromosome was chosen because previous linkage studies in Costa Rica and in other populations suggested that it possibly contained bipolar disorder loci (Berretini et al. 1994; Stine et al. 1995; Freimer et al. 1996*a*). Genealogical studies indicated that the individuals in our current study did not share common ancestry over the past several generations (Escamilla et al. 1996). We therefore anticipated that we would not detect random genome regions shared IBD by more than a few individuals and that regions of high IBD sharing would thus be areas containing possible BP-I–susceptibility genes inherited from a common founder.

Samples and Methods

Sample Collection

To diminish the likelihood of investigating phenocopies, we limited the sample to individuals with a definite diagnosis of BP-I, with onset by age 50 years and a history of at least two psychiatric hospitalizations. The 48 patients with BP-I (25 female patients and 23 male patients) in the current study were recruited independently from psychiatric hospitals and clinics in the CVCR. First-degree relatives of patients were also recruited, to determine genetic phase. The study was approved by institutional review boards at the Costa Rican Ministry of Health, the University of Costa Rica, and the University of California at San Francisco, and informed consent was obtained from all participating subjects. Of the 48 BP-I subjects, 8 individuals had both parents available for genotyping, 20 individuals had one parent available, 10 individuals had one or more children available, 1 individual had two siblings available, and 9 individuals had no relatives available. In nuclear families, only one individual (the proband) was designated as affected, and all others were considered to have unknown phenotype. Details of ascertainment and diagnostic procedures, and the clinical and genealogic profiles of the study sample, can be found in Escamilla et al. (1996).

Genotyping

We used 26 markers, spanning chromosome 18, to genotype all 48 affected individuals (as well as 53 relatives, to establish phase). Of the 25 regions, 21 were $\leq$6 cM, and 4 were 6–7 cM. The average distance between markers was 4.8 cM. When choices were available, we chose the most polymorphic marker (Gyapay et al. 1994). The average heterozygosity of the markers used in this screen (in the CEPH pedigree collection) was 0.75. (The only screening markers with heterozygosity values <0.70 were D18S464, D18S60, D18S378, and D18S469.) We screened chromosome 18 at a marker density of 6 cM because available marker maps had gaps $\leq$6 cM, and our goal was to have an equal density of coverage across the chromosome (Gyapay et al. 1994). We chose markers from the maps available, at the time of the study, from Généthon (Gyapay et al. 1994), the Cooperative Human Linkage Center (Murray et al. 1994), and the public database of the Utah Center for Genome Research. Genotyping procedures used for all experiments were as previously described by Di Rienzo et al. (1994). In brief, one of the two primers was labeled radioactively with a polynucleotide kinase, and PCR products were separated, by electrophoresis, onto polyacrylamide gels. Autoradiographs were scored independently by two raters. Data for each marker were entered

into the computer database twice, and the resultant files were compared for discrepancies. Scoring was done without knowledge of affected status.

Simulations

We conducted simulations to evaluate the power of a likelihood-based test of LD (Terwilliger 1995), to detect a result significant at the .05 level, with these assumptions: a 6-cM marker map; a disease gene in the middle of the 6-cM segment; affected subjects, with one copy of the disease gene, separated by 10 generations from a common ancestor; and four equally frequent marker alleles at each marker site. (The disease gene was associated with the "1" allele at the marker locus.) Under these assumptions, and with a phenocopy rate of 0%, normal chromosomes carried each marker allele with a probability of 25% (normal-chromosome distribution), and disease chromosomes carried the "1" allele with a probability of 80%. The probability of disease-chromosome distribution was calculated with the formula $(1-\theta)^G + [1-(1-\theta)^G \times f]$, where θ = recombination fraction, G = number of generations from a common ancestor, and f= the frequency of the allele in the population. Thus, the disease chromosomes carry the "1" allele with a probability of 80% and each of the remaining three alleles with a probability of 6.7%. Because the true genetic structure of bipolar disorder is unknown, we examined several different conditions of etiologic heterogeneity (which would include locus and allelic heterogeneity, as well as phenocopies). We investigated phenocopy rates of 0%, 33%, and 67% (with phenocopy rates of 33% and 67%, the percentages of chromosomes from affecteds with the "1" allele are 62% and 43%, respectively). If an affected individual was randomly selected as a phenocopy (with a probability equal to the phenocopy rate), then the marker allele on all four parental chromosomes was randomly chosen from the normal chromosome distribution. If the affected individual was randomly chosen as a true case, (with a probability of 1 minus the phenocopy rate) the marker allele for one chromosome of that individual was randomly chosen from the normal chromosome distribution, and the other chromosome's marker alleles were randomly chosen from the disease-chromosome distribution. Recombination occurred on parental chromosomes in proportion to the marker map. Marker alleles for nontransmitted chromosomes of the parents were randomly chosen from the normal chromosome distribution. We performed these analyses by using the 48 patients with BP-I plus their available relatives. One hundred replications were performed for each simulation. Available relatives were considered to have unknown disease phenotype. For the 10 affected individuals with at least one child available for genotyping, one chromosome from the affected parent was randomly simulated to be transmitted to available children, and the other chromosome was randomly selected from the normal chromosome distribution. Although data were simulated for parents of all affected individuals, if parents were not available for genotyping, their simulated genotypes were not used in these analyses.

We also did power simulations (100 replications for each model) of larger sample sizes, using an ideal situation in which both parents are available for genotyping, to aid in planning future studies. In these simulations we used sample sizes of 90, 200, 300, and 400 affected individuals; phenocopy rates of 50% and 75%; and a marker map of 2.5 cM, with all other assumptions as described above. With this denser marker map, at a phenocopy rate of 0%, disease chromosomes carried the "1" allele with a probability of 90%, calculated by the formula $(1-\theta)^G + [1-(1-\theta)^G \times f]$, and each of the remaining three alleles with a probability of 3.3%. Details of the likelihood-ratio test used in analyzing simulation results are described inAnalysis.

Analysis

We used two different procedures to identify regions potentially shared IBD by patients with BP-I. The first approach, a search for shared segments, has the advantage of being nonparametric. The second approach, although requiring parameters of the illness to be specified, has the advantage of providing a formal test statistic, allowing for the calculation of P values. These two tests thus offer compensatory strengths and weaknesses when used in the search for genes in a complex disease.

We first searched for shared segments (Houwen et al. 1994). For each individual, we evaluated two marker haplotypes in each of the 25 intermarker intervals, by using a preselected threshold (the possible sharing of a haplotype by $\geq$50% of patients) to select segments for further investigation. Since this screen does not differentiate between sharing that is IBD and sharing that is identical by state (IBS), use of lower thresholds would lead to too many segments passing the screen.

We also applied a likelihood-ratio test for LD to each of the 26 initially tested markers. This test was done independently of the results of the shared-segment evaluation. We applied a modified version of the procedure of Terwilliger (1995), which only includes case and control chromosomes or chromosomes transmitted and not transmitted to patients. In our sample there were several affected individuals whose parents were not available but whose children were available. DNA from these latter individuals could not be analyzed with the original Terwilliger program but could be analyzed with our implementation of the same procedure, as described by Freimer et al. (1996*a*). This procedure examines the likelihood that a particular allele (or alleles) is (are) overrepresented on disease chromosomes compared with

Table 1

Heterozygosity of Markers Used in the Genome Screen of Chromosome 18

Marker Name	Heterozygosity in Généthon Database	Heterozygosity in Costa Rican Sample[a]
D18S1140	.49	.39
D18S59[b]	.81	.81
D18S476[b]	.76	.62
D18S481	.76	.74
D18S391	.75	.69
D18S452	.83	.85
D18S843	NA	.73
D18S464	.65	.51
D18S1153	.78	.69
D18S378	NA	.54
D18S53	.79	.81
D18S453	.82	.81
D18S40	NA	.81
D18S66	.85	.81
D18S56	.73	.74
D18S57	.87	.85
D18S467[b]	.73	.64
D18S460	.62	.67
D18S450	.79	.74
D18S474	.82	.73
D18S69	.79	.78
D18S64	.74	.65
D18S1134	.73	.68
D18S1147	.85	.86
D18S60	.37	.58
D18S55	.77	.80
D18S68	.79	.79
D18S477	.62	.70
D18S61[b]	.87	.86
D18S488	.87	.82
D18S485[b]	.79	.79
D18S541	NA	.63
D18S870[b]	NA	.66
D18S469[b]	.65	.64
D18S874	NA	.64
D18S380	NA	.63
D18S1121[b]	.74	.77
D18S1009	.74	.66
D18S844	NA	.76
D18S554	.82	.79
D18S461	.77	.65
D18S70	.83	.86

NOTE.—NA = data not available.

[a] Allele frequencies were calculated from the entire sample, accounting for known relationships among individuals.

[b] Markers with -2l*n*(LR) >1.0.

nondisease chromosomes (Terwilliger 1995; Freimer et al. 1996*a*). A single parameter, λ, is estimated, which quantifies such overrepresentation of marker alleles on disease chromosomes. Designation of chromosomes of probands as disease carrying or non–disease carrying was achieved by specification of a genetic model for the disease. The same model of transmission was used in this LD-likelihood test as was used in the initial genome screen of the Costa Rican families, described in McInnes et al. (1996). In brief, this model assumes that the disease is nearly dominant (assuming penetrance of .81 for heterozygous individuals and .90 for homozygous individuals with the disease mutation), that the population prevalence of BP-I in Costa Rica is .015, and that the frequency of the disease gene in the population is .003. In the likelihood calculations, all possible disease-genotype combinations of all relatives are considered. With the model that was used, in which the disease-gene frequency is very low, the LD-likelihood test, in most cases, treats the probands as effectively heterozygous at the disease locus, and chromosomes of other relatives not occurring in the probands are treated as non–disease-carrying chromosomes. We did not specify a phenocopy rate in the genetic model, because the effect of phenocopies will be absorbed by the parameter λ; the presence of phenocopies in our sample will serve to erode the association between marker alleles and disease and hence will reduce the estimate of λ. Because, in the present LD study, we were attempting to gather further evidence regarding the findings published in our initial genome screen, we limited ourselves to this one model in performing the likelihood analyses. However, both the BP-I family sample and the current LD sample will ultimately be analyzed with use of other models. We considered as promising those markers that gave evidence of overrepresentation of an allele on affected chromosomes, with a $-2ln$(likelihood ratio [LR]) statistic >1.0.

Follow-up genotyping and LD-analysis studies were performed on markers that gave suggestive findings in the shared-segment evaluation. Within each segment that passed the threshold described above, 1–3 additional markers were typed to permit us to test for LD across regions of 1–2 cM. Markers that provided suggestive evidence of LD by the initial likelihood-ratio test, but had not been suggested as promising by the shared-segment screen, were also followed up, in this case by typing two additional nearby markers. In all, a total of 42 markers from chromosome 18 were used to genotype the study sample (table 1 and fig. 1). LD analysis of the additionally typed markers was conducted by use of the likelihood-ratio test.

Results

Simulations

Simulation results for the sample of 48 patients with BP-I and available relatives showed relatively high power to detect suggestions of association ($P \leq .05$) with low phenocopy rates (94% for a phenocopy rate of 0%, and 54% for a phenocopy rate of 33%) but a dramatically decreased power under high phenocopy rates (e.g., 9% for a phenocopy rate of 67%). Additional simulations showed that, under higher phenocopy rates, the power to detect LD can be improved by increasing the sample

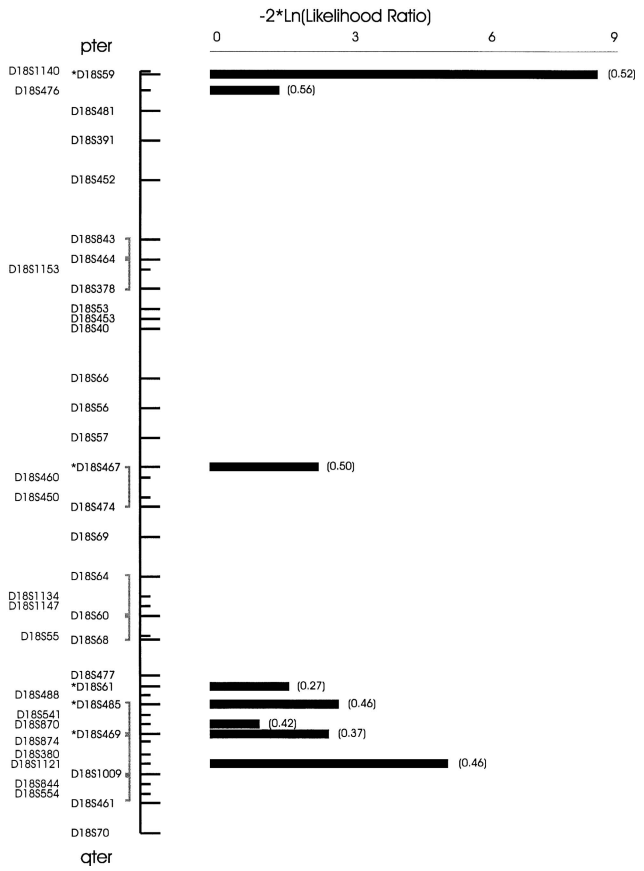

Figure 1 Results from the LD screen of chromosome 18. The 26 markers used in the first stage of the screen are listed in the right column. Sixteen markers used to follow up interesting regions are listed in the left column. Approximate chromosomal locations of the 26 initial markers and the 16 follow-up markers are indicated by long and short tick marks, respectively. The eight segments that passed the initial screen threshold for segment sharing (50% of individuals or 25% of chromosomes sharing a two-marker haplotype) and the five markers that passed the initial threshold for the Terwilliger LR test ($-2ln$[LR] > 1.0) are indicated by blackened bars and asterisks, respectively. Two marker segments that passed the initial threshold were followed up by at least one marker within the segment, if possible (at the time of the study no markers were available between D18S843 and D18S464, and only one marker was available between D18S464 and D18S378). Markers that passed the initial threshold for the Terwilliger LR test were followed up with two additional markers. These additional markers flanked the original finding. The value of the $-2ln$(LR) statistic, from the Terwilliger test, is plotted as a solid bar. This statistic is distributed as a one-sided χ^2 random variable with one degree of freedom. The estimate of the λ value, for the eight markers with positive results, is indicated in parentheses after the $-2ln$(LR) statistic. Markers without a $-2ln$(LR) statistic plotted had estimates of $\lambda = 0$, with the exception of three markers that had estimates of $0 < \lambda < 0.62$.

size and/or the marker density of screening (table 2). For instance, with a phenocopy rate of 75%, the power increases to 82% with a sample of 300 affected individuals and a 2.5-cM marker map.

Shared-Segment Screen

We evaluated 25 possible shared segments (defined by the 26 markers genotyped in the sample). Eight regions passed the threshold of possible IBD sharing by ⩾50% of patients. These regions were bounded by the following markers: D18S843-D18S464, D18S464-D18S378, D18S467-D18S474, D18S64-D18S60, D18S60-D18S68, D18S485-D18S469, D18S469-D18S1009, and D18S1009-D18S461 (fig. 1).

Linkage-Disequilibrium Testing

Five (D18S59, D18S467, D18S61, D18S485, and D18S469) of the original 26 markers displayed evidence of possible LD, by means of a likelihood procedure ($-2ln$[LR] statistic >1.0; table 3). Two (D18S59 and D18S61) of these five markers had not been identified as markers of interest by the shared-segment evaluation. D18S59, located near 18pter, displayed the strongest pointwise evidence for LD ($-2ln$[LR] statistic of 8.3, $P = .002$) of all the markers tested in this sample.

Follow-up of Initial Results

Using the protocol discussed in Samples and Methods, we genotyped additional markers within the segments that passed the shared-segment screen as well as follow-up markers surrounding one (D18S59) of the two markers that had passed only the LD screen. We were unable to follow up one shared-segment region (D18S843-D18S464), because additional polymorphic markers were not available within the segment. We were also unable to follow up the finding for D18S61, for the same reason. Three (D18S476, D18S870, and D18S1121) of the 16 follow-up markers typed displayed additional evidence of possible LD (fig. 1).

These additional results brought to eight the total number of markers with $-21n$(LR) statistics >1.0 (table 3). Five of these eight marker loci were clustered within a small region of 18q22-23. The most significant LD in 18q22-23 was observed at D18S1121, with $-2ln$(LR) of 5.03 and $P = .01$, and two were in 18pter.

For the two 18pter markers (D18S59 and D18S476),

Table 2

Power-of-Likelihood–Analysis Test of LD

	Power to detect LD for Sample Size (n) =			
Penocopy Rate	90	200	300	400
50%	82%	99%	100%	100%
75%	33%	62%	82%	90%

Note.—Assumptions included that subjects were removed from a common ancestor by 10 generations, that a marker map of 2.5 cM was used, and that each marker had four equally frequent alleles. Values are the percentage of replicates to have P values <.05.

Table 3

Frequencies of Marker Alleles Overrepresented in Disease Chromosomes, as Compared with Nondisease Chromosomes, for Markers Where -2ln(LR) >1.0

		Frequency on	
Marker	Allele	Nondisease Chromosomes	Disease Chromosomes
D18S59[a]	154	.121	.572
D18S476	271	.470	.771
D18S467[a]	172	.384	.693
D18S61[a]	177	.074	.326
D18S485[a]	182	.237	.586
D18S870	179	.405	.657
D18S469[a]	234	.128	.450
D18S1121	168	.171	.553

[a] Markers from the screening stage.

the alleles overrepresented on BP-I chromosomes (154 and 271 bp, respectively) form a haplotype that occurs in 48% of the patients with BP-I. Overall, this haplotype occurs on 26% of the chromosomes of individuals with BP-I and on 4% of the chromosomes not transmitted from parents to individuals with BP-I (definite phase for these two markers could be assigned in 25 patients with BP-I [50 chromosomes] and 25 nontransmitted parental chromosomes). Because the composite genetic and physical maps of the 18q22-23 region had not yet been completed at the time of this study, the relative order of the five markers in 18q22-23, for which evidence of LD was observed, was still too uncertain to permit construction of definitive marker haplotypes in our study sample.

Marker D18S467, in the 18q12.3 region, was the one marker outside 18q22-23 and 18pter to show a $-2\ln(\mathrm{LR}) > 1$ ($-2\ln[\mathrm{LR}] = 2.5$, $P = .06$). The additional markers used to follow up this result (D18S450, D18S460, and D18S57) displayed no evidence of association.

Marker Heterozygosity in the Costa Rican Sample

We calculated heterozygosity values for the markers used, on the basis of the allele frequencies, estimated from the entire sample, accounting for known relationships among individuals. These heterozygosities are shown in table 1, along with the corresponding heterozygosity values of these markers in the CEPH population, used by Généthon.

Discussion

Screening for Complex Disease Loci by LD Approaches

Our intention in this work was to explore the feasibility of using LD methods to screen the genome for susceptibility genes for a common, genetically complex disorder. The results obtained in our LD-based search for possible BP-I gene–loci on chromosome 18 were encouraging (specific susceptibility regions were suggested), but they highlight a number of issues that must be considered before LD screening is widely adopted.

Successful application of a shared-segment approach to any LD study depends on (1) a marker-map density that is appropriate to the age of the population isolate being studied and (2) a sharing threshold that will not be too high to allow true IBD areas to be identified and that will not be so low as to include many areas that are IBS false-positive signals. An appropriate marker map for an LD-screening study should have segments of a size expected to be shared IBD by many of the affected individuals. In addition to the density of the marker map used, the number of generations separating affected individuals from their common ancestor and the rate of etiologic heterogeneity in the population will also influence the choice of the sharing threshold, used to trigger further study. For example, if the common (disease–gene bearing) ancestor is removed from the current descendants by >10 generations, the length of true IBD haplotypes shared by ⩾50% of the descendants may be <5 cM (and certainly <6 cM, as is the screen used in this study) (Te Meerman et al. 1994; Durham and Feingold 1997). Our choice of a threshold of 50% of affected individuals sharing a possible haplotype therefore effectively meant that we were likely to identify only BP-I genes of a major effect in this population (phenocopy rate approaching zero), and even then, only if the distance from a common ancestor is not > ~10 generations. Although this was probably too stringent a screen threshold, given the complex etiology of bipolar disorder, the alternative we faced—reducing the threshold to a lower percentage of potential IBD sharing—would have drastically decreased the specificity (and hence the utility) of the screen. For instance, in this particular study, lowering the threshold to a possible IBD haplotype shared in ⩾25% of the patients would have resulted in 24 of the 25 regions tested being determined as regions of interest. If, in future studies, definite phase information can be set for a greater proportion of the probands (obtained from phasing information supplied by additional relatives) the "possible IBD" threshold will be more useful as a screening criterion at thresholds approaching 25% sharing (almost one in five of the patients with BP-I in the current study had no relative available for phase construction). Finally, regardless of the threshold chosen, there is no widely accepted statistical test available to evaluate the significance of the number of shared haplotypes observed, although several statistical approaches are under development (reviewed by Kruglyak 1997).

The use of markers with low heterozygosity will increase the number of false-positive results in a shared-

segment screen, as some regions may pass the threshold because of IBS sharing of a common allele. For example, the four regions that passed our shared-segment screen, but gave no evidence of LD in the likelihood-ratio tests (D18S843-D18S464, D18S464-D18S378, D18S64-D18S60, and D18S60-D18S68), included markers that had relatively low heterozygosities in the study population (D18S464, D18S60, and D18S378; table 1).

There are two ways to overcome the limitations of shared-segment analysis, as seen in this study. The first is to increase the density of markers in the initial screen (i.e., increase the proportion of BP-I individuals in whom a shared haplotype can be detected, thus decreasing the number of false-negative results). Second, future screening studies may focus on individuals who have available parents (i.e., increase the number of patients for whom we can set phase, thus allowing the threshold to be lowered in a meaningful way and decreasing the number of false-positive results).

For a formal statistical test of LD, we used the likelihood-ratio test rather than another frequently used method, the transmission disequilibrium test (TDT) (Spielman et al. 1993), because data from all 48 patients with BP-I could be used in the likelihood approach. Effective use of the original TDT requires parental genotypes, which were unavailable for 20 of the 48 patients with BP-I. One potential source of false-negative results, in our application of the likelihood-ratio test for LD, is that it is dependent on the specific genetic model for the disease used in the analysis. For instance, the results of the likelihood analysis presented here are applicable only to transmission of dominantly inherited BP-I genes in the CVCR population. The power of the likelihood test is also critically dependent on the polymorphism content of the markers tested and the density of the markers used for a screening analysis.

Evaluation of Potential BP-I Loci on Chromosome 18

Our previous linkage study of BP-I in two Costa Rican pedigrees had provided several possible localizations for BP-I, throughout the genome. Since the 48 Costa Rican patients in the present study (collected independently of the pedigree studies and with no known relation to the pedigree members) are descended from the same ancestral population as the patients in those pedigrees (CR001 and CR004; Freimer et al. 1996*b*), we had reasoned that LD could be present in the population sample at markers surrounding any true BP-I loci identified in the pedigree study. LD screening of patients with BP-I, in this population, might also yield important BP-I loci that were not identified in the pedigree study. Pedigree-based linkage studies involve selection of certain subsets (individual families) of the population in which there is a clustering of affected individuals. In a complex disease, such studies may be useful in finding genes of large effect in those particular subsets, but they might not identify loci that are important in understanding the basis of the disease in the general population. In this LD screen of chromosome 18 in Costa Rican patients with BP-I, two regions were highlighted as being of particular interest, and both regions correspond to segments highlighted in the previous pedigree studies from Costa Rica.

We previously highlighted the 18q22-23 chromosomal region (Freimer et al. 1996*a*) because this area showed the strongest evidence suggestive of linkage in the two pedigrees of any region tested in a genome screen conducted with ~500 microsatellite markers (McInnes et al. 1996). In the pedigree study, portions of a haplotype of >40 cM in this region were shared by 22 of 26 individuals with BP-I (Freimer et al. 1996*a*), although formal LOD scores for markers in this area were below the level of significance required for proof of linkage. In the current study, five markers in the 18q22-23 region provide possible evidence of LD in Costa Rican patients with BP-I. The marker that gave the strongest evidence of possible LD in the current study, D18S1121, is located within the 3-cM region of highest haplotype sharing observed in the individuals with BP-I from the pedigree study. The specific allele (of 168 bp), which is overrepresented on the disease chromosomes at this locus (D18S1121) in the sample of the population with BP-I, is also the allele that occurs on the putative high-risk haplotype within the pedigrees (Freimer et al. 1996*a*).

Our pedigree studies had also highlighted a region at 18pter deserving of further study (McInnes et al. 1996), albeit in only one of the two families, CR001. The second-highest LOD score in the genome observed for family CR001 was at D18S59, located near 18pter, and a nearby marker, D18S476, also gave a positive LOD score in this family. This current study of 48 patients with BP-I now provides additional evidence for a BP-I locus in this region, with the same two markers showing evidence of LD. Because genomewide significance levels have yet to be calculated for LD tests (Kruglyak 1997), we can at present only interpret the evidence for LD in the 18pter region (a pointwise P value of .002 for marker D18S59) as being roughly equivalent to Lander and Kruglyak's criteria for suggestive, but not significant, linkage in a genomewide screen (Lander and Kruglyak 1995). The alleles at D18S59 and D18S476 that are overrepresented among the patients with BP-I, from the population sample (154 and 271 bp, respectively), are also overrepresented in the patients with BP-I from pedigree CR001 (all patients with BP-I in family CR001 have at least one copy of the 154 allele at D18S59), possibly indicating that the patients with BP-I in the pedigree share this region IBD with those 48% of patients with BP-I from the population sample who also carry this haplotype.

The third region that showed possible evidence of LD in our population sample was identified through a single marker (D18S467), in the 18q12.3 region. Additional markers typed near this one did not support the initial suggestion of LD in this region. Evidence from a linkage test that yielded a significance level of $P = .06$ would be expected to occur, by chance, ~24 times (about once on most chromosomes) in a genomewide screen.

Our possible BP-I localizations at 18pter and 18q22-23, in the current sample, are distinct from regions on chromosome 18 suggested by other groups as being possibly linked to mood disorder (Berretini et al. 1994; Stine et al. 1995). We detected no evidence of association with these areas (near the centromere and in 18q21) in our BP-I population sample; nevertheless, the power of our current sample is not great enough to rule out these regions as potential BP-I loci in the Costa Rican population. McMahon et al. (1997) have recently reported excess allele sharing in sib pairs at 18S541, which is in the 18q22-23 region, although their affected status included not only BP-I, but also bipolar type II and schizoaffective patients.

Future Directions

The results of this study suggest that shared-segment–screening approaches will only be useful with the development of denser marker maps (Collins et al. 1997) and with the development of tests that permit statistical comparison of disease-chromosome haplotypes with control-chromosome haplotypes. Because the potential advantages of a shared-segment approach are substantial (this type of approach takes maximum advantage of the fact that haplotypes, not just single alleles, are inherited IBD in population isolates, and it is nonparametric), and because marker maps (Dib et al. 1996; Yuan et al. 1997) and statistical methods continue to improve, we remain optimistic about this method of mapping genes for complex disorders.

Both the 18pter and the 18q22-23 regions would have been identified as regions with possible LD at a significance of $P < .05$, even if we had not used the shared-segment approach but had instead screened for evidence by using only the likelihood-ratio test, with the original 26 markers. Our results indicate that, in population isolates, such as the CRCV, and with suitably dense marker coverage, tests similar to the likelihood-ratio test of LD (Terwilliger 1995) are promising tools for genome screening of complex diseases. It is not clear, however, whether currently available tests will be powerful enough to detect unequivocal proof of association in a genomewide scan for such diseases, given sample sizes that are easily obtained. More-powerful tests are needed and may emerge from efforts to develop measures that make use of haplotype information (Service et al. 1999 [in this issue]; Durham and Feingold 1997; Goldin and Chase 1997).

The test of LD screening conducted in the current study points out the need to do a more complete LD screening analysis. We thus intend to perform an LD screen of chromosome 18, using an expanded sample of patients with BP-I and a denser marker map. The addition of more-polymorphic markers to genome maps, and the application of haplotype-based statistical tests currently under development, should facilitate efforts to definitively identify BP-I susceptibility genes in Costa Rica.

Acknowledgments

This work was supported by National Institutes of Health grants MH00916, MH49499, MH48695, and MH47563; the National Alliance for Research on Schizophrenia and DepressionYoung Investigator Awards (to M.A.E. and N.B.F.); the American Psychiatric Association's Program for Minority Research Training in Psychiatry Award (to M.A.E.); and the Veterans Administration Research Psychiatrist Award (to L.A.M.). We thank Lauren Smith for assistance in collating the best estimate materials. M.A.E. thanks Dr. Carlos Martinez for his support and suggestions regarding this research work. Above all we would like to thank the patients and family members who participated in this project and the Costa Rica institutions that made this work possible, including the Hospital Nacional Psiquiatrico, The Hospital Calderon Guardia, The Ministry of Science, and the Caja Costarricense de Seguro Social.

Electronic-Database Information

URLs for data in this article are as follows:

Cooperative Human Linkage Center, http://www.chlc.org/ (for marker maps)
Généthon, http://www.genethon.fr/ (for marker maps)
Utah Center for Genome Research, http://www.genome.utah.edu/ (for marker maps)

References

Berretini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger JI Jr, Gershon E (1994) Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. Proc Natl Acad Sci USA 91: 5918–5921

Collins FS, Guyer MS, Chakravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278:1580–1581

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

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence

repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Durham LK, Feingold E (1997) Genome scanning for segments shared identical by descent among distant relatives in isolated populations. Am J Hum Genet 61:830–842

Escamilla M, Spesny M, Reus V, Gallegos A, Meza L, Molina J, Sandkuijl L, et al (1996) Use of linkage disequilibrium approaches to map genes for bipolar disorder in the Costa Rican population. Am J Med Genet 67:244–253

Escamilla MA, Freimer NB, Reus VI (1997) The genetics of bipolar disorder and schizophrenia. In: Rosenberg R, Prusiner S (eds) Molecular and genetic basis of neurological disease, 2d ed. Butterworth Heinemann, Newton, pp. 1343–1362

Friedman TB, Liang Y, Weber JL, Hinnant JT, Barber TD, Winata S, Arhya IN, et al (1995) A gene for congenital, recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. Nat Genet 9:86–91

Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P, Service SK, et al (1996*a*) Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22–q23. Nat Genet 12:436–441

Freimer NB, Reus VI, Escamilla M, Spesny M, Service S, Gallegos A, Meza L, et al (1996*b*) An approach to investigating linkage for bipolar disorder using large Costa Rican pedigrees. Am J Med Genet 67:254–263

Goldin LR, Chase GA (1997) Improvement of power to detect complex disease genes by regional inference procedures. Genet Epidemiol 14:785–789

Gyapay G, Morisette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, et al (1994) The 1993–1994 Généthon human genetic linkage map. Nat Genet 7:246–339

Houwen RHJ, Baharloo S, Blankenship K, Raeymakers P, Juyn J, Sandkuyl LA, Freimer NB (1994) Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet 8:380–386

Kruglyak L (1997) What is significant in whole-genome linkage disequilibrium studies? Am J Hum Genet 61:810–812

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247

Lander E, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2948

McInnes LA, Escamilla MA, Service SK, Reus VI, Leon PE, Silva S, Rojas E, et al (1996) A complete genome screen for genes predisposing to severe bipolar disorder in two Costa Rican pedigrees. Proc Natl Acad Sci USA 93:13060–13065

McMahon FJ, Hopkins PJ, Xu J, McInnes MG, Shaw S, Cardon L, Simpson SG, et al (1997) Linkage of bipolar affective disorder to chromosome 18 markers in a new pedigree series. Am J Hum Genet 61:1397–1404

Murray JC, Buetow KH, Weber JL, Ludwigsen S, Scherpbier-Heddema T, Manion F, Quillen J, et al (1994) A comprehensive human linkage map with centimorgan density. Science 265:2049–2054

Newport MJ, Huxley CM, Huston S, Hawrylowicz CM, Oostra BA, Williamson R, Levin M (1996) A mutation in the interferon-γ-receptor gene and susceptibility to mycobacterial infection. New Engl J Med 335:1941–1949

Nurnberger JI, DePaulo RJ, Gershon ES, Reich T, Blehar MC, Edenberg HJ, Foroud T, et al (1997) Genomic survey of bipolar illness in the NIMH genetics initiative pedigrees: a preliminary report. Am J Med Genet 74:227–237

Puffenberger EG, Kauffman ER, Bold S, Matise TC, Washington SS, Angrist M, Weissenbach J, et al (1994) Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. Hum Mol Genet 3:1217–1225

Risch N, Botstein D (1996) A manic depressive history. Nat Genet 12:351–353

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Saborio M (1992) Experience in providing genetic services in Costa Rica. In: Birth Defects Original Article Series 28: 96–102

Service SK, Temple Lang DW, Freimer NB, Sandkuijl LA (1999) Linkage-disequilibrium mapping of disease genes by reconstruction of ancestral haploytpes in founder populations. Am J Hum Genet 64:1729–1739 (in this issue)

Shah AB, Chernov I, Zhang HT, Ross BM, Das K, Lutsenko S, Parano E, et al (1997) Identification and analysis of mutations in the Wilson disease gene (ATP7B): population frequencies, genotype-phenotype correlation, and functional analysis. Am J Hum Genet 61:317–328

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52:506–516

Stine OC, Xu J, Koskela R, McMahon FJ, Gschwend M, Friddle C, Clark CD (1995) Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. Am J Hum Genet 57:1384–1394

Te Meerman GJ, van der Meulen MA, Sandkuijl LA (1994) Expected size of shared haplotypes surrounding a common disease gene. Am J Hum Genet Suppl 55:A205

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

Uhrhammer N, Lange E, Porras O, Naeim A, Chen X, Sheikhavandi S, Chiplunkar S (1995) Sublocalization of an ataxia-telangiectasia gene distal to D11S384 by ancestral haplotyping in Costa Rican families. Am J Hum Genet 57: 103–111

Yuan B, Vaske D, Weber JL, Beck J, Sheffield VC (1997) Improved set of short-tandem-repeat polymorphisms for screening the human genome. Am J Hum Genet 60:459–460