

Report

Linkage Analysis of a Complex Pedigree with Severe Bipolar Disorder, Using a Markov Chain Monte Carlo Method

Chad Garner,^{1,2} L. Alison McInnes,^{2,3,4} Susan K. Service,^{2,*} Mitzi Spesny,⁵ Eduardo Fournier,⁵ Pedro Leon,⁵ and Nelson B. Freimer^{2,3,4,*}

¹Department of Integrative Biology, University of California Berkeley, Berkeley; ²Neurogenetics Laboratory, ³Center for Neurobiology and Psychiatry, and ⁴Department of Psychiatry, University of California San Francisco, San Francisco; and ⁵School of Medicine and Cell and Molecular Biology Research Center, University of Costa Rica, San Jose, Costa Rica

Recently developed algorithms permit nonparametric linkage analysis of large, complex pedigrees with multiple inbreeding loops. We have used one such algorithm, implemented in the package SimWalk2, to reanalyze previously published genome-screen data from a Costa Rican kindred segregating for severe bipolar disorder. Our results are consistent with previous linkage findings on chromosome 18 and suggest a new locus on chromosome 5 that was not identified using traditional linkage analysis.

A single large pedigree can provide a powerful sample for mapping complex traits; compared with a collection of independent nuclear families, a single pedigree may contain more linkage information and less etiologic heterogeneity and yields a greater possibility of identifying genotyping errors. Large pedigrees from recently founded population isolates may be particularly valuable, as affected individuals in such populations are more likely to share common ancestry than in admixed populations.

The increase in power associated with the large-pedigree study design comes at the cost of computational feasibility, and, until recently, pedigree size and consanguinity were limiting factors for both model-based and model-free linkage analysis. Investigators who collected large complex pedigrees traditionally had to break up their samples into smaller family units that available algorithms could handle. An example is the extended Old Order Amish pedigree that has been investigated in a series of linkage studies of bipolar disorder (BP) (Ege-

land et al. 1987; Ginns et al. 1996). Although genealogical information has shown that this sample of 207 individuals (including 81 affected with BP) could be represented as a single, highly consanguineous 10-generation kindred, for linkage analyses, the family has been broken into smaller pedigrees, each covering, at most, five generations. None of these analyses has produced unequivocal localization of BP genes.

As a result of the perceived failures in identification of linkage using large pedigrees, mapping studies of complex traits now mainly use less-powerful nuclear-family study designs. Software packages have recently become available, however, that use new algorithms that can compute linkage statistics on highly complex pedigrees. We used one such package, SimWalk2 (Sobel and Lange 1996), to reanalyze data from a previously published genome screen (Freimer et al. 1996a; McInnes et al. 1996) for severe BP (BP-I) in a kindred from the genetically isolated Costa Rican population. The previous analyses of the kindred used a model-dependent method, assuming a nearly dominant mode of inheritance. With the algorithms available at the time of the previous analysis, it was necessary to analyze the kindred as two families without including inbreeding loops. In contrast, the SimWalk2 analysis reported here takes advantage of the power provided by the full-pedigree structure. In addition, SimWalk2, which uses Markov chain Monte Carlo (MCMC) methods to compute allele-sharing statistics, provides a model-free (or nonparametric) analysis; this type of analysis is more robust than model-dependent

Received December 21, 2000; accepted for publication January 29, 2001; electronically published February 14, 2001.

Address for correspondence and reprints: Dr. Nelson B. Freimer, Center for Neurobehavioral Genetics, University of California Los Angeles, Gonda Center, Room 3506, 695 Charles E. Young Drive South, Box 951761, Los Angeles, CA 90095-1761. E-mail: NFreimer@mednet.ucla.edu

* Present affiliation: Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles.

© 2001 by The American Society of Human Genetics. All rights reserved. 0002-9297/2001/6804-0028\$02.00

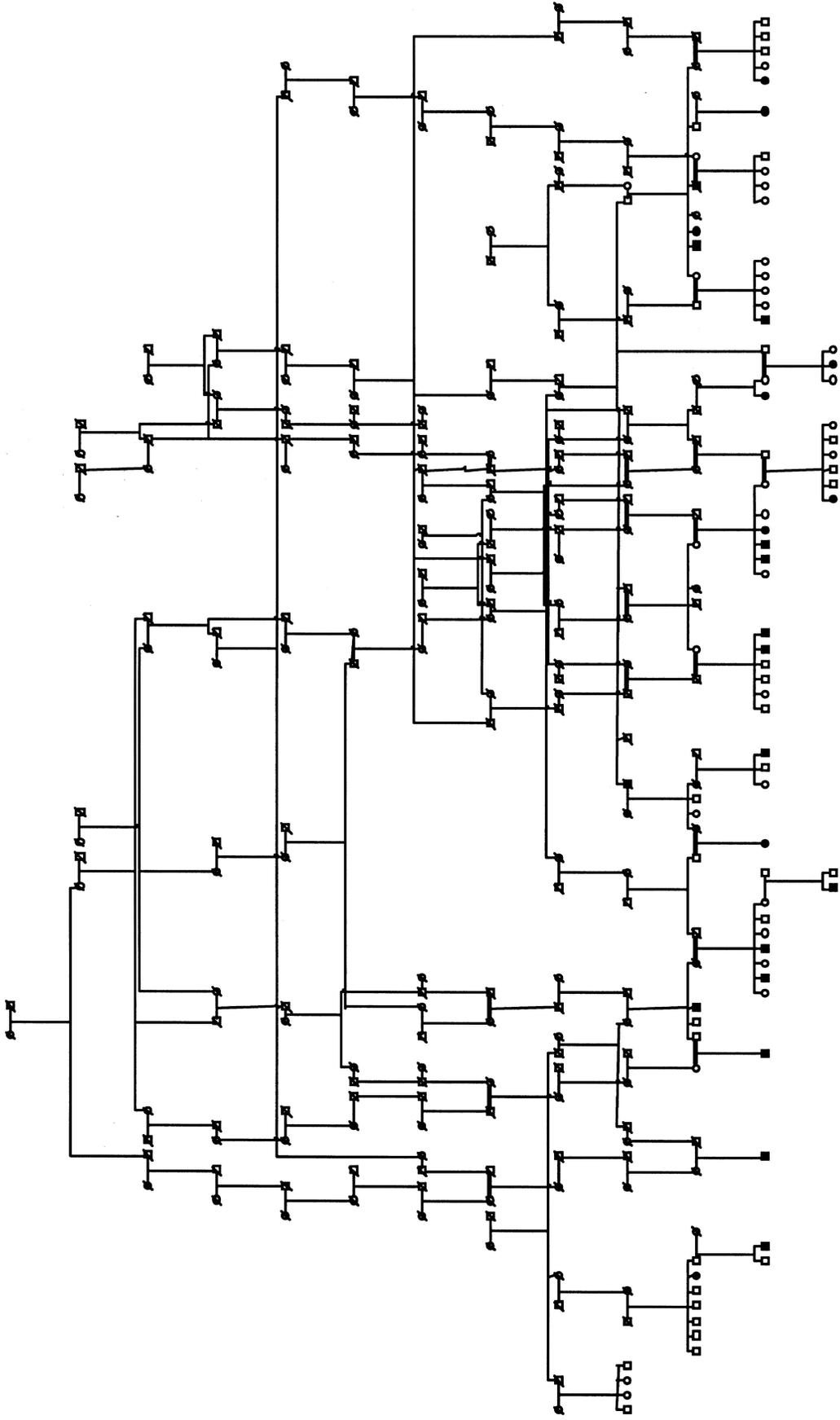


Figure 1 Full Costa Rican kindred. Affected individuals are shown with blackened symbols. Genealogical information is represented for 13 generations. All individuals in the first seven generations were considered phenotype unknown. The consanguineous marriages (thick marriage lines) include seven second-cousin marriages (once or twice removed), two first-cousin marriages (once or twice removed), and one third-cousin marriage.

Table 1

Locations, Allele-Sharing Statistics, and *P* Values for All 25 Markers Tested on Chromosome 18q

Marker	Location ^a	Allele-Sharing Statistic	<i>P</i>
D18S56	0.0	1.122	.0784
D18S57	2.5	1.109	.0917
D18S67	5.0	1.205	.0811
D18S450	7.5	1.330	.0445
D18S69	14.5	.920	.1853
D18S64	16.5	.946	.1552
D18S38	18.5	.943	.1558
D18S60	20.5	1.034	.0976
D18S68	26.0	1.025	.0829
D18S55	27.5	1.021	.0844
D18S483	29.5	1.061	.0724
D18S477	31.5	1.204	.0246
D18S61	35.5	1.223	.0328
D18S488	37.5	1.161	.0372
D18S485	40.5	.951	.1358
D18S870	42.5	.992	.1168
D18S469	43.5	1.014	.1093
D18S1161	46.5	1.093	.0652
D18S1009	48.5	1.010	.1188
D18S1121	50.5	1.565	.0168
D18S380	54.0	1.599	.0110
D18S554	57.0	1.477	.0157
D18S462	60.0	1.474	.0144
D18S461	64.0	1.487	.0127
D18S70	68.0	1.819	.0038

^a Locations are taken from the most centromeric marker, D18S56.

analysis when the mode of inheritance is unknown, as is the case with BP. Although there are powerful methods for computing exact nonparametric linkage statistics (Lander and Green 1984; Kruglyak and Lander 1995), these methods could not accommodate the size and complexity of the Costa Rican BP kindred, thus necessitating the application of a stochastic method such as SimWalk2.

Figure 1 shows the pedigree as analyzed in the present study, with all known connections specified. We identified the eight great-grandparents for each affected individual to verify that there were no connections between these individuals closer than those depicted in the figure. Given the demographic history of the Costa Rican population, it is likely that there are still unknown remote connections between these individuals; however, such distant connections would not likely substantially affect the linkage analysis (L. A. McInnes and N. B. Freimer, unpublished data).

We reanalyzed genotypes from the 459 markers in the genomewide linkage analysis of the kindred. The marker selection and genotyping procedures for the genomewide data have been described elsewhere (Freimer et al. 1996a; McInnes et al. 1996). Marker allele frequencies were estimated from the families using known relation-

ships among the individuals but without linkage to the disease phenotype (Boehnke 1991), by means of the program ILINK (Lathrop and Lalouel 1984) and using the simplified pedigree structure from Freimer et al. (1996b).

Nonparametric linkage analysis was performed using SimWalk2. SimWalk2 uses MCMC methods to sample from the complete distribution of underlying inheritance patterns proportional to their likelihood, which is calculated from the observed genotype data. Statistic D, calculated by SimWalk2, measures the extent of allele sharing among affected relative pairs as the average across the sampled inheritance patterns. A large value of the statistic indicates a high degree of identity-by-descent allele sharing among the affected relatives. We chose statistic D over other nonparametric statistics calculated by SimWalk2 because it is generally powerful when the model of inheritance is unknown and because similar statistics have been studied by others (Weeks and Lange 1988; Whittemore and Halpern 1994). All marker information was used in this multipoint computation of allele-sharing statistics. Empirical *P* values are obtained by comparing the observed value of the statistic to that found under the null hypothesis, which is generated by repeated sampling of marker data simulated with a gene-dropping algorithm, without linkage to the phenotype. Sobel and Lange (1996) suggest that *P* values from this procedure will be slightly conservative; thus, statistical significance will be potentially understated.

All the markers showing nominal *P* values <.05 in the current analysis were on chromosomes 18q and 5q. Markers on 18q had provided, by far, the strongest evidence of any portion of the genome for linkage in the prior analysis of these data (McInnes et al. 1996), and the majority of affected individuals shared a marker haplotype in this region (Freimer et al. 1996a). Table 1 shows the relative locations, allele-sharing statistics, and significance levels for all 25 markers tested on chromosome 18q in the current analysis; each of these markers showed allele-sharing statistics that were >1 SD above the genomewide mean, with a range of 0.920–1.819 SD. Two regions within 18q contained clusters of markers for which allele-sharing statistics re-

Table 2

Locations, Allele-Sharing Statistics, and *P* Values for Five Markers on Chromosome 5q

Marker	Location ^a	Allele-Sharing Statistic	<i>P</i>
D5S658	0	1.325	.0125
D5S436	5.0	1.430	.0057
D5S636	10.0	1.392	.0054
D5S673	12.0	1.397	.0059
D5S410	15.0	1.355	.0135

^a Locations are taken from the most centromeric marker, D5S658.

sulted in P values $<.05$. These two clusters of markers (from D18S477 to D18S488 and from D18S1121 to D18S70) correspond to the 18q segments highlighted in the prior analyses of these data (Freimer et al. 1996a).

Five consecutive markers, spanning ~ 15 cM of chromosome 5q, showed evidence for linkage in the current nonparametric analysis. The allele-sharing statistics for markers D5S658, D5S436, D5S636, D5S673, and D5S410 had P values of .015, .0057, .0054, .0059, and .0135, respectively (table 2). Our prior parametric analyses of the genome-screen data (McInnes et al. 1996) found no evidence for linkage to 5q. The six markers now providing such evidence only did so when analyzed with the data from neighboring markers. Visual examination of the genotypes of the individual markers showed that there is not a clear association between their alleles and BP, suggesting that the evidence in 5q derived from an informative haplotype rather than from information at individual markers. Visual inspection also subsequently confirmed that the majority of affected individuals in the kindred shared a single haplotype over this region of 5q (data not shown). We carried out tests to assess the sensitivity of the results observed for the five consecutive markers showing P values $<.05$ on chromosome 5q, to the prespecified marker allele frequencies (data not shown). These additional tests showed that the results in 5q were not sensitive to the allele frequency used.

In the prior analysis, markers on chromosomes 11 and 16 provided linkage evidence that surpassed a predefined threshold (LOD 1.6 in the combined pedigrees) (McInnes et al. 1996). In the current analysis, neither of these locations showed linkage evidence at a nominal significance of $P < .05$. The variability in these results between the two analyses is difficult to evaluate, given the differences in the methods of analysis and pedigree structure used.

By reanalyzing the Costa Rican pedigrees as a single kindred using SimWalk2, we continue to detect the most suggestive linkage evidence identified in the original analyses, that for 18q22–q23. The fact that a previously undetected region on 5q was identified with the new methods demonstrates the utility of haplotype information in linkage analysis of genome-scan data from large complex pedigrees. We suggest that similar analyses should be applied to genotype data from other such pedigrees—for example, the Old Order Amish BP kindred.

Acknowledgments

This work was supported by the National Institutes of Health (NIH) grants MH-01748, to L.A.M., and MH-00916 and MH-49499, to N.B.F.; by Fundacion de la Universidad de Costa Rica para la Investigacion (FUNDEVI); and by the vice

rectory of research of the University of Costa Rica. C.G. is supported by NIH grant GM-40282. We thank the Wellcome Trust Centre for Human Genetics, for the use of computer resources, and Eric Sobel and Lodewijk Sandkuijl, for helpful comments. We thank the families who participated in this project and Costa Rican institutions that made this work possible: Hospital Nacional Psiquiatrico, Hospital Calderon Guardia, Caja Costarricense de Seguro Social, Archivo Nacional de Costa Rica, and Iglesia Catolica de Costa Rica. A complete list of genomewide results can be obtained from N.B.F.

References

- Boehnke M (1991) Allele frequency estimation from data on relatives. *Am J Hum Genet* 48:22–25
- Egeland JA, Gerhard DS, Pauls DL, Sussex JN, Kidd KK, Allen CR, Hostetter AM, and Housman DE (1987) Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature* 325:783–787
- Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P, Service SK, Smith LB, Silva S, Rojas E, Gallegos A, Meza L, Fournier E, Baharloo S, Blankenship K, Tyler D, Batki S, Vinogradov S, Weissenbach J, Barondes S, Sandkuijl L (1996a) 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, Smith L, Service S, Gallegos A, Meza L, Batki S, Vinogradov S, Leon P, Sandkuijl L (1996b) An approach to investigating linkage for bipolar disorder using large Costa Rican pedigrees. *Am J Med Genet* 67:254–263
- Ginns EI, Ott J, Egeland JA, Allen CR, Fann CS, Pauls DL, Weissenbachoff J, Carulli JP, Falls KM, Keith TP, Paul SM (1996) A genomewide search for chromosomal loci linked to bipolar affective disorder in the Old Order Amish. *Nat Genet* 12:431–435
- Kruglyak L, Lander ES (1995) Complete multipoint sib pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 57:439–454
- Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci USA* 84:2363–2367
- Lathrop G, Lalouel J (1984) Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet* 36:460–465
- McInnes LA, Escamilla MA, Service SK, Reus VI, Leon P, Silva S, Rojas E, Spesny M, Baharloo S, Blankenship K, Peterson A, Tyler D, Shimayoshi N, Tobey C, Batki S, Vinogradov S, Meza L, Gallegos A, Fournier E, Smith L, Barondes S, Sandkuijl L, Freimer N (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
- Sobel E, Lange K (1996) Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet* 58:1323–1337
- Weeks DE, Lange K (1988) The affected-pedigree-member method of linkage analysis. *Am J Hum Genet* 42:315–326
- Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. *Biometrics* 50:118–127