© 1996 Nature Publishing Group http://www.nature.com/naturegenetics

Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23

Nelson B. Freimer[1,2,3], Victor I. Reus[2], Michael A. Escamilla[1,2], L. Alison McInnes[1,2], Mitzi Spesny[4], Pedro Leon[4], Susan K. Service[1], Lauren B. Smith[1], Sandra Silva[4], Eugenia Rojas[4], Alvaro Gallegos[5], Luis Meza[5], Eduardo Fournier[4], Siamak Baharloo[1], Kathleen Blankenship[1], David J. Tyler[1], Steven Batki[6], Sophia Vinogradov[2], Jean Weissenbach[7], Samuel H. Barondes[2] & Lodewijk A. Sandkuijl[8]

[1]*Neurogenetics Laboratory, University of California San Francisco, San Francisco California 94143, USA*
[2]*Center for Neurobiology and Psychiatry, Department of Psychiatry, University of California San Francisco, San Francisco California 94143, USA*
[3]*Programs in Genetics and Biomedical Sciences, University of California, San Francisco, San Francisco, California 94143*
[4]*Cell and Molecular Biology Research Center and Escuela de Medicina (P.L) Universidad de Costa Rica, San Jose, Costa Rica*
[5]*Hospital Calderon Guardia, San Jose Costa Rica*
[6]*Department of Psychiatry, University of California San Francisco, San Francisco General Hospital, San Francisco California 94110, USA*
[7]*Généthon 1 rue de l'Internationale, 91000 Evry, France and Unité de Génétique Moléculaire Humaine, CNRS URA 1445, Institut Pasteur, 75724 Paris Cédex, France*
[8]*Department of Clinical Genetics, Erasmus University, Rotterdam, Department of Human Genetics, Leiden University, Department of Medical Genetics, Groningen University, The Netherlands*

Correspondence should be addressed to N. B. F. at U.C.S.F. email: nbfr@itsa.ucsf.edu.

Manic-depressive illness, or bipolar disorder (BP), is characterized by episodes of elevated mood (mania) and depression[1]. We designed a multistage study in the genetically isolated population of the Central Valley of Costa Rica[2,3] to identify genes that promote susceptibility to severe BP (termed BPI), and screened the genome of two Costa Rican BPI pedigrees (McInnes *et al.*, submitted). We considered only individuals who fullfilled very stringent diagnostic criteria for BPI to be affected. The strongest evidence for a BPI locus was observed in 18q22-q23. We tested 16 additional markers in this region and seven yielded peak lod scores over 1.0. These suggestive lod scores were obtained over a far greater chromosomal length (about 40 cM) than in any other genome region. This localization is supported by marker haplotypes shared by 23 of 26 BPI affected individuals studied. Additionally, marker allele frequencies over portions of this region are significantly different in the patient sample from those of the general Costa Rican population. Finally, we performed an analysis which made use of both the evidence for linkage and for association in 18q23, and we observed significant lod scores for two markers in this region.

Using two Costa Rican pedigrees, each heavily loaded for individuals affected with BPI, we screened the genome with 473 microsatellite markers (McInnes *et al.*, submitted). We used a stringent approach in analysing these marker data, considering only individuals with a narrowly defined phenotype for BPI to be affected and assuming a high probability that affected individuals were phenocopies[2]. We employed a nearly dominant transmission model[2]. Six markers over the entire genome surpassed our lod score threshold ($\geq$1.6 in the combined pedigrees) for deciding that a region warranted further study. Our lod score threshold was based on simulation analyses showing expected distributions of lod scores under linkage and non-linkage[2]. Three of the six markers were contiguous within 18q22-q23 (McInnes *et al.*, submitted).

The clustering in the genome screen of suggestive lod scores within 18q22-q23, indicated co-segregation between BPI and markers in this region, but did not prove linkage. Our study was designed with the premise that conventional linkage analysis, based on calculating lod scores in pedigrees, would probably not yield unequivocal evidence for localization of a BP gene (in contrast to expectations for mendelian disorders). This form of analysis requires specifying a genetic model and, as with any model, one makes simplifying assumptions that may be wrong, particularly for traits as complex as BP. These assumptions are acceptable, however, in screening the genome for suggestive localizations, because lod score analysis using extended pedigrees is still a very good method for this purpose; since relatively few recombination events are likely to be observed between individuals as closely related to each other as are the members of even large pedigrees, this form of analysis can detect evidence hinting at linkage over extensive regions[4]. One simplifying assumption usually made in standard lod score analysis is that markers are in linkage equilibrium with each other (that is their alleles are not associated). Even in isolated populations this assumption is reasonable for the purposes of rough genome screening; linkage disequilibrium (allelic association) is not expected over genome regions of the size examined in lod score analysis[4,5]. In such populations, however, LD is predictably present in the immediate vicinity of a disease gene, as a high proportion of affected individuals are likely to share the chromosomal segment containing the gene identical by descent (IBD) from a remote common ancestor[4]; it is therefore necessary to explicitly account for LD, in follow up investigations of suggestive localizations using large numbers of closely spaced markers in a single region. In fact, studies of several genetically complex disorders have shown that evaluating allele association or haplotype sharing (which does not require specification of a genetic model) among affected individuals in an isolated population is a powerful means of mapping disease genes in situations where conventional lod score analysis cannot provide adequate proof of linkage[6-9]. Because of the opportunity offered for detecting LD between a disease phenotype and marker haplotypes surrounding a disease gene, genetically isolated populations are extremely valuable for mapping genes for complex traits. Our pedigrees derive from the Costa Rican Central Valley, which has remained isolated until relatively recently; the majority of the population is descended from a small number of founders in the 16th-18th Centuries[3]. We therefore followed up the suggested localization of BPI to 18q22-q23, using LD based methods: haplotype evaluation, association tests and a joint linkage/association analysis.

We constructed 18q22-q23 marker haplotypes for all genotyped individuals (and for deceased individuals whose genotypes could be reconstructed). BPI segregates with particular marker haplotypes in both families (Figs 1 & 2). In pedigree CR004, 16 of 17 affected individuals share portions of a marker haplotype from *D18S64* to *D18S70*, a distance of about 40 cM (Fig. 1*a*), with the majority of these individuals sharing at least a 30 cM segment in common. Most affected individuals in CR004 share the distal segment of this haplotype, consisting largely of marker alleles that are rare

© 1996 Nature Publishing Group http://www.nature.com/naturegenetics

Distance from pter (in cM)	84	95.5	104	106	108	110	113	115	115	116	118	119	119	125
Marker	64	55	61	485	870	469	1161	1121	1009	380	554	462	461	70
a														
V-17	188	138	175	176	179	236	106	173	162	154	218	193	166	124
V-16	188	138	175	176	179	236	106	168	150	150	218	193	166	124
V-8	188	138	175	176	179	236	106	168	150	150	222	187	168	124
IV-24	188	138	175	176	179	236	106	168	150	150	222	187	168	124
IV-30	188	138	175	176	179	236	106	168	150	150	222	187	168	124
V-6 #	188	138	175	176	179	236	106	168	150	150	222	187	168	124
V-15	188	138	175	176	179	236	106	168	150	150	222	187	168	124
V-19	188	138	175	176	179	236	106	168	150	150	222	187	168	124
V-9 *	188	146	175	176	179	236	106	168	150	150	222	187	168	124
V-13	188	146	175	176	179	236	106	168	150	150	222	187	168	124
V-14	188	146	175	176	179	236	106	168	150	150	222	187	168	124
IV-17 #	188	142	159	182	179	242	106	168	150	150	222	187	168	124
IV-18 @	188	142	159	182	179	242	106	168	150	150	222	187	168	124
V-11	188	142	159	182	179	242	106	168	150	150	222	187	168	124
VI-2	192	138	157	188	179	242	106	168	150	150	222	187	168	124
VI-1 #	200	142	173	182	179	236	94	168	150	150	222	187	167	122
VI-1 #	188	138	177	186	179	236	96	168	150	150	216	185	164	124
V-7 #	188	138	177	186	179	236	96	168	150	150	216	185	164	114
V-2 *						236	94	166	150	150				
IV-6	192	140	159	178	171	236	94	166	150	150	222	183	163	110
IV-9 @	192	140	159	178	171	236	94	166	150	150	222	183	163	110
V-3	192	140	159	178	171	236	94	166	150	150	222	183	163	110
V-6 #	192	140	159	178	171	236	94	166	150	150	222	183	163	110
V-7 #	192	140	159	178	171	236	94	166	150	150	222	183	163	110
V-10	192	142	159	186	171	236	100	173	150	150	222	183	164	122
IV-19	204	144	171	182	171	236	100	173	150	150	222	183	164	122
IV-17 #	202	146	179	182	171	236	100	173	150	150	222	183	164	122
b														
IV-6	188	146	173	178	175	236	96	175	150	148	216	189	164	112
IV-9	188	146	173	186	179	236	94	166	146	145	218	183	164	114
IV-18	204	144	171	182	179	236	108	166	154	146	216	187	164	111
IV-19	192	142	159	186	179	236	96	164	150	155	218	185	164	118
IV-24	190	140	177	186	179	236	90	173	162	154	218	193	166	124
IV-30	190	144	175	178	175	236	90	175	150	155	218	185	164	112
V-3	200	144	169	182	187	242	96	168	154	150	218	183	169	106
V-8	192	142	171	182	179	236	102	166	150	154	216	187	164	128
V-10	192	142	179	186	179	236	96	168	156	145	216	183	167	118
V-11	192	138	179	180	175	242	104	166	158	150	216	183	162	118
V-13	192	146	171	182	179	238	90	175	150	150	216	183	167	124
V-14	192	138	173	182	179	236	94	164	152	150	222	185	167	124
V-19	190	150	177	182	179	236	104	168	150	155	218	191	167	124
V-15,16,17	192	138	157	188	179	242	104	168	148	145	222	185	160	114
	206	144	173	186	179	236	94	166	146	145	218	183	164	114
	200	144	159	186	NT	242	NT	166	148	150	216	193	164	124
	192	142	171	182	179	236	90	175	150	150	216	183	164	111
	190	140	177	186	179	236	90	173	150	154	218	185	164	118
	188	138	179	182	183	240	96	168	150	150	216	187	164	120
	188	146	159	186	179	236	96	164	150	154	218	185	164	128
	188	142	157	178	175	236	98	181	148	150	218	185	163	112

Fig. 1 *a*, Affected members of CR001 and CR004 with depiction of the shared marker haplotypes observed. These haplotypes were defined by sharing over ≥ 3 markers. The unshaded area outlined in solid lines is the more extensive haplotype conserved in CR004, the unshaded area outlined in dashed lines is the more extensive haplotype conserved in CR001. The shaded area indicates a haplotype that extends from *D18S1009* to *D18S554* and is apparently common to both larger haplotypes. The ID numbers in the first column refer to the pedigree displayed in Fig. 2. In the remaining columns are the allele sizes at the indicated markers. "*" indicates an uncertain haplotype, "#" indicates that individual received two copies of the shaded haplotype (both haplotypes are displayed); "@" indicates an inferred haplotype. The markers used for haplotyping are shown at the top of the figure with inter-marker distances in cM. The marker order towards qter, is: *D18S64, D18S55, D18S61, D18S485, D18S870, D18S469, D18S1161, D18S1121, D18S1009, D18S380, D18S554, D18S462, D18S461, D18S70*. *b*, The other haplotypes that could be unambiguously reconstructed in the pedigree in Fig. 2 are shown. Those that are present in affected individuals are indicated by the ID numbers at the left of the diagram. "NT" indicates that an individual was not typed for a given marker.

in the general population of Costa Rica. This haplotype is far longer and shared by a much higher proportion of the BPI affected individuals than any observed in this pedigree in other genome regions (we constructed haplotypes around all markers that exceeded our lod score thresholds for either family or for the combined data set in the genome screening study, data not shown). A distinct, but similarly sized haplotype is shared by seven of the nine BPI individuals in pedigree CR001 (Fig. 1*a*). Three of these BPI individuals in CR001 also share portions of the high risk haplotype observed in CR004; this haplotype is transmitted by individual IV-10, a member of CR004 who married into CR001.

In both pedigrees maximal haplotype sharing is observed within the eight cM interval between *D18S469* and *D18S554*. The haplotypes in CR001 and CR004 appear identical in a five to six cM portion of this segment (including markers *D18S1009-D18S380-D18S554*). This 'core' haplotype may be shared IBD by the affected individuals in the two families; all individuals who share this portion of 18q23 are descended from a founding couple who are ancestral to the main branches of CR004 as well as to one branch of CR001. In total, of the BPI individuals whose genotypes can be fully reconstructed, 23 of 26 share some portion of the conserved haplotypes observed in these families, and three individuals (IV-3, V-1, V-12) do not display any portion of these haplotypes.

To evaluate statistical evidence for a BPI localization in 18q22-q23, we used a non-parametric approach. In isolated populations, genome regions where marker allele frequencies differ between affected individuals and the background population likely harbor disease loci; this fact has been exploited in previous mapping studies[9]. When such deviation is based on a substantially increased frequency of one or a few alleles in affected individuals, the region is almost certainly inherited IBD, with the disease gene, from common ancestors. In our genome screening experiments, we followed established procedures to directly estimate allele frequencies using genotypes from the family members in the study[10]. For several markers tested in 18q23, we noted that the alleles most commonly observed in the affected individuals are rare in reference pedigrees of European descent[11]. For example, an allele of 124 bp in length at *D18S70* is observed in 67% of the BPI individuals tested but has a frequency of only 3% in the reference families. To evaluate whether these allele frequency differences are simply the result of genetic drift in the isolated population of the Costa Rican Central Valley, we used several of the 18q23 markers to genotype a set of randomly collected, unrelated, unaffected

© 1996 Nature Publishing Group http://www.nature.com/naturegenetics

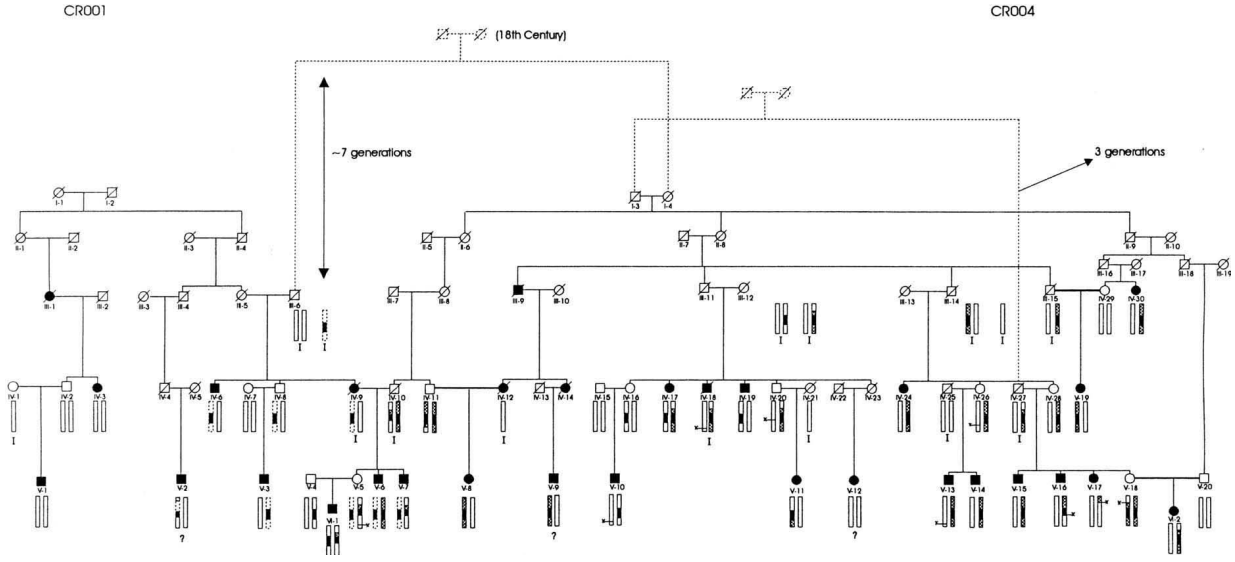

Fig. 2 Pedigree showing both CR001 and CR004. Affected individuals are denoted by black symbols, deceased individuals by a diagonal slash. A schematic of each individual's haplotype (where available) is shown below the ID number for the region *D18S870 - D18S70*. Apparent recombinations are denoted by "-x"; consanguineous marriages by a double bar and the conserved haplotypes as black shading within the haplotype bars. The larger conserved region for CR004 is stippled, the larger conserved region for CR001 is indicated by a dashed outline. "I" underneath the haplotype bars indicates inferred haplotype, a "?" indicates phase is uncertain. The connection between CR001 and CR004, dating to an 18th Century founding couple, is indicated by the dashed lines joining persons III-6 and I-4. Similarly, a dashed lined indicates the connection of person IV-27 to the rest of CR004. Three individuals who carry diagnoses of BPI are deceased but their haplotypes could be partially (IV-12) or nearly fully reconstructed (IV-9, IV-18). Other deceased individuals who are designated as affected in this figure received best estimate diagnoses of BPI (based on review of hospital records) but their haplotypes could not be reconstructed. Individuals whose diagnostic status is left blank did not fulfill 'best estimate' criteria for BPI or SAD-M (schizoaffective disorder, manic type) but some of them have well documented mood disorder. For example, individual V-18 was hospitalized three times for depressive and manic episodes, with psychosis, but as one episode was subsequent to treatment with thyroid hormone, she was assigned a diagnosis of organic mood disorder rather than BPI. We have not yet assigned final diagnoses to many members of CR001 and CR004 with tentative diagnoses other than BPI or SAD-M. However, of those individuals who have been assigned final diagnoses, the shared marker haplotype is observed in all six of the individuals with BPII, MDD or organic mood disorder. Six of eleven individuals with no apparent psychiatric diagnosis share this haplotype, however four of them (IV-8, IV-11, IV-26, IV-28) transmit the haplotype (and presumably the risk of BP) to affected offspring. This pedigree drawing is reduced from more complete ones[2] in that only BPI and SAD-M individuals and their direct ancestors are depicted.

individuals from this population. These individuals were originally sampled for population genetic studies in which allele frequencies were compared, in diverse populations, for microsatellites from several genome regions[12,13]. We compared marker allele frequencies in the samples of affected individual (taking into consideration the connection between the pedigrees and consanguinity loops) with those observed in the general Costa Rican population sample, without making any assumptions concerning linkage. For four of the 18q23 markers (*D18S469, D18S554, D18S461, and D18S70*), allele frequencies in the affected individuals differed significantly from those in the Costa Rican population (Table 1); no such differences were observed for markers tested from other genome regions (data not shown).

In the above analyses, the evidence based on population association was separated from the evidence based on linkage within the pedigrees. We subsequently carried out an analysis to evaluate the joint evidence for linkage and association in 18q23; we compared the likelihood obtained when parameters relating to both association and linkage are estimated, to the likelihood obtained when both of these parameters are set to their null values (no association and no linkage). Significant linkage/association-based lod scores were obtained for *D18S554* and *D18S70*, with lod scores respectively of 3.70 and 4.06 (Table 1).

The strongest evidence for the suggested localization of a BPI gene to 18q23 is provided by the sharing of marker haplotypes by affected individuals in the study pedigrees. These haplotypes must have been introduced into these families through several ancestors, who were themselves distantly related to each other (for example, see individuals I-4, III-6, and IV-27 in Fig. 2). Although the pedigrees share common ancestry, they have been mainly separated for at least seven generations. Additionally, within each pedigree, haplotype sharing is displayed between individuals who are separated by several generations from any common ancestor (for example, V-2 and the descendants of III-5). Several studies have shown that such distantly related affected individuals should demonstrate IBD sharing of marker alleles over segments of several cM surrounding a disease susceptibility gene, and that it is extremely unlikely that they would share such segments in random genome regions[4-7]. Our studies of background LD within the Costa Rican population support this contention; intensive study using multiple, densely spaced markers of a different chromosomal region in a population sample, revealed virtually no evidence of LD between markers more than one to two cM apart (L. Bull *et al.*, unpublished data). Although methods have been proposed to evaluate the statistical significance of shared haplotypes among distantly related affected individuals[4], it is not currently possible to measure the significance of such haplotype sharing in large pedigrees, where many different degrees of relationship exist between such individuals.

© 1996 Nature Publishing Group http://www.nature.com/naturegenetics

Table 1 Results of association and lod score analyses for 18q23 markers

	Association analyses		Joint linkage and association analyses	
Marker	χ^2 value	*P* value	Lod score (Z_{max})	θ
D18S469 (4)	16.56	0.0009	1.33	0.0
D18S380 (4)	7.62	ns	1.16[a]	0.5
D18S1009 (5)	4.98	ns	0.44	0.16
D18S554 (6)	17.72	0.0034	3.70	0.0
D18S462 (5)	1.41	ns	0.12[a]	0.50
D18S461 (7)	13.97	0.03	1.70	0.0
D18S70 (8)	26.0	0.0005	4.06	0.0

Number of alleles for each marker shown in parentheses. The association analyses used likelihood ratio tests to examine the null hypothesis of no difference between allele frequencies of the Costa Rican population sample and the families. The *P* values associated with the given Chi-square values are also indicated. The method of joint linkage and association analysis is detailed in the text and estimates both the recombination fraction (θ) (assessing linkage) and the degree of association between BP and the markers. ns: not significant. [a]The (non-significant) evidence for these markers is exclusively based on association; the markers are almost completely uninformative for linkage within these families.

Standard methods of linkage analysis cannot be practically applied to evaluate the evidence provided by the shared multi-locus haplotypes in these pedigrees as their structures are too complicated to enable accurate evaluation of all possible routes of transmission of a phenotype and marker alleles[4,14,15]. While fast algorithms have recently been developed for multilocus linkage analysis, these cannot currently run if consanguinity loops are present[16]. Two non-parametric linkage approaches have been widely used in psychiatric genetic studies[17,18]. We did not use the affected pedigree method[19] (APM) because it is highly sensitive to the population frequencies of the alleles shared by affected relatives and is thus inappropriate in our data set as the BPI individuals in our study share extremely rare alleles for several markers. We also did not apply affected sib-pair tests to the pedigree data as there are too few independent sibs for such tests to attain significant results.

In the absence of appropriate statistical tests, we visually inspected the evidence from shared haplotypes. Such examination in our study pedigrees shows multiple entries of the conserved haplotypes, and indicates that the 18q23 region is shared IBD by most of the affected individuals. It is highly improbable that the observed haplotype sharing simply reflects random LD in the isolated population of Costa Rica; in addition to our results indicating that background LD in this population extends for very short distances (L. Bull *et al.*, unpublished data), the association tests indicate that allele sharing among the affected individuals, for several 18q23 markers, is non-random with respect to an unselected sample from the same population. Nor does the observed sharing reflect genetic isolation of these families within Costa Rica; the pedigrees have been dispersed throughout the Costa Rican Central Valley for several generations. As depicted in Fig.1*b*, a large number of different alleles and haplotypes for the 18q22-q23 markers are observed in the study pedigrees in addition to the ones shared by the BPI individuals. For example, at *D18S1161*, 13 different alleles are observed in the two pedigrees, but almost all of the affected individuals share one or the other of two alleles.

As maximal haplotype sharing is observed in the interval between *D18S469* and *D18S554* among most distantly and closely related BPI individuals, a putative gene for BPI susceptibility probably lies within this approximately eight cM segment. Identifying the most likely location of such a gene within this region may be facilitated by possible recombination events in three affected individuals (V-16, V-17, and VI-1 in Fig.2). This evidence should be interpreted cautiously, however, given the uncertainty regarding the relationship between genotype and phenotype for such a complex disorder, and the fact that the exact order of these closely spaced markers cannot be confirmed until genetic and physical maps are fully integrated. Interestingly, the revised placement on a recent whole genome physical map, of one of the 18q23 markers that we used, may further support the localization of a putative BPI gene in or near the interval from *D18S469* to *D18S554*. On this map, *D18S70*, which displayed the highest lod score in the joint linkage/association analysis is placed immediately telomeric to *D18S554* (ref. 20) (the other marker demonstrating a significant lod score). In contrast, these loci are separated by two markers and by seven cM on the genetic map[21] (Fig.1).

The candidate segment defined by haplotype sharing in these families does not appear to overlap with recently suggested localizations of susceptibility loci for (broadly defined) BP on chromosome 18 (near the centromere[17] and in 18q21 (ref. 18)). Overlap with the putative pericentromeric location can be excluded. Although the possible 18q21 localization is apparently at least 10 cM distant from the edge of the candidate region described here, its boundaries are not clearly defined, so that overlap cannot be excluded.

Although BPI is transmitted along with conserved 18q22-q23 haplotypes across multiple generations in the study pedigrees, the observed data are not adequately explained by simple dominant transmission of a single disease allele. For example, BPI illness in the three individuals who do not share the conserved haplotypes probably reflects either locus or allelic heterogeneity with respect to the rest of the study population or non-genetic causation of the disorder. Such aetiologic heterogeneity is suggested by the clustering of two of these individuals (IV-3, and V-1) in a CR001 branch that is not descended from the founding couple who connect this family to CR004. The conserved haplotypes observed in our pedigrees demonstrate evident incomplete penetrance (for example, individual IV-21 is not affected with BPI but apparently transmits illness to two sons). Such non-penetrance of BPI is concordant with the predictions of genetic epidemiological studies[1, 22, 23], and with usually observed segregation patterns of BPI. However, as indicated by the results of mutational analyses for other complex traits[14, 24], delineation of the mode of transmission of BPI in these families will be impossible until causative mutations are identified.

© 1996 Nature Publishing Group http://www.nature.com/naturegenetics

The approach used to obtain the localization suggested here differs from those employed in previous mapping studies of BP[17,18,25-27]. Most importantly, all evaluations focused exclusively on individuals with a severe, reliably diagnosed and narrowly defined phenotype, BPI. It is thus relatively unlikely that evidence for this localization could disappear due to new onsets of illness in individuals previously considered unaffected, as has occurred in earlier mapping studies of BP[28, 29]. Additionally, we completed a genome screen before following up the initial results for 18q22-q23 markers; this region was studied intensively because it showed more consistent evidence of linkage than any other genome segment.

The localization of a BPI gene to 18q23 is supported by three lines of evidence: (i) a haplotype, extending over several markers, is shared by 23 of the 26 BPI individuals studied; (ii) BPI is associated with particular marker alleles at several loci in this region, as shown by analysis of marker allele frequencies in affected individuals and in the general Costa Rican population; and (iii) analyses jointly considering evidence for linkage and association provide further corroboration.

As with any putative gene localization, ours requires confirmation in an independent sample. Our haplotype and association findings suggest that LD, within 18q23, should be observed among unrelated BPI patients from the same population. We have nearly completed collection of a sample of such patients, which we will genotype using 18q23 markers to attempt to confirm our findings. Additionally, as has been shown for other complex traits, the area of maximum LD observed in this sample should contain the putative disease gene[30, 31]; pinpointing its location may enable us to clone it[32, 33]. Our genome screen and other mapping studies suggest that several different genes may be involved in susceptibility to BP. A multistep approach such as we have initiated in Costa Rica (defining a narrow phenotype, screening the genome and evaluating LD in highlighted regions) could be used in similarly isolated populations to identify additional BP genes as well as genes associated with other psychiatric disorders.

Methods

Pedigrees. As described elsewhere[2,3], we studied two Costa Rican pedigrees heavily loaded for BPI. Extensive church and civic records permitted us to trace genealogies in the study pedigrees, to the 16th–18th centuries[2, 3]. The population of the Costa Rican Central Valley (from which these families derive) is mainly descended from a small number of founders in the 16th and 17th centuries. It grew to its current size (> 2,000,000) in the absence of substantial additional immigration[3], indicating that present-day BP affected individuals are likely to share disease susceptibility mutations identical by descent (IBD) from one or more common ancestors. Additionally, extensive genealogical records dating to the founding of Costa Rica facilitated identification of such common ancestors of BPI affected individuals (as described below). These genealogical studies indicated that the members of pedigrees CR001 and CR004 are descended from the ancestral Central Valley population but have not been genetically isolated within this population.

Genotyping. We previously described[2] the basis for choosing for genotyping studies a set of individuals from pedigrees CR001 and CR004. The Costa Rican population (control) sample used for the association analyses consisted of unrelated individuals descended from the ancestral Costa Rican Central Valley population who had previously been collected for a population genetic study. In that study these individuals had been genotyped using a set of microsatellites from several random genome regions[12, 13]. No significant deviations from Hardy-Weinberg equilibrium or significant differences in allele frequencies in comparison with other human populations were noted (E.R., P.L., N.B.F., unpublished observations). As only small amounts of DNA had been obtained from these individuals, we were unable to genotype them using 18q23 markers that we recently obtained (*D18S1121* and *D18S1161*). The markers used for genotyping were from the most recent published maps of Genethon[21], the Cooperative Human Linkage Centers (CHLC)[34] and the public databases of Genethon and the University of Utah Genome Center. The genotyping procedures used for all experiments were as previously described[12].

Haplotype construction. Haplotypes were constructed by hand using a minimum-recombination strategy. When parents were not available for genotyping, haplotypes were inferred from children, if possible. Often it was not known which of the four inferred haplotypes belonged to which parent; in these instances the two sets of inferred haplotypes were offset from the parental symbols in Fig. 2.

Statistical analyses. For all analyses, only individuals with diagnoses of BPI were considered affected, as well as one individual with a diagnosis of schizoaffective disorder, manic type (SAD-M). The rationale for designating individuals with this diagnosis as 'affected' has been previously discussed[2]. All other individuals were designated as having an unknown phenotype. Marker allele frequencies were estimated on a combined data set which included both the population sample and the BPI affected individuals from the pedigrees (with correction for dependency due to family relationships[10]). Rare alleles were collapsed until no alleles remained with a frequency in the combined data set of less than 4%. The frequency of the remaining alleles was also estimated separately for the population sample and for the patients from the families. The likelihood obtained on the combined data set was compared with the product of the likelihoods obtained in the separate analyses, to yield a likelihood ratio statistic approximately following a chi-square distribution with n-1 degrees of freedom, where n is the number of alleles for a given marker. A similar procedure was used by Schellenberg *et al.*[35].

From the results of these association analyses, the frequency of the BPI gene in the family sample was estimated to be 0.115, by averaging the excess frequency of all alleles of 18q23 markers that showed a significantly (P<0.01) increased frequency among the patients from the pedigrees. This BPI gene frequency was used in the subsequent combined analysis of linkage and association. For this analysis, we utilized a recently published likelihood ratio test for association[36], which includes only a single estimated parameter, λ. This test assumes that some marker allele will be over-represented on chromosomes that carry the disease mutation, when many of those chromosomes descend from a single ancestor. The proportion of disease chromosomes with this ancestral allele is represented by the parameter λ. It is not a-priori known which marker allele will be the over-represented allele, and therefore the procedure considers each of the marker alleles separately as potential founder alleles. Consequently, a total likelihood is obtained for a given value of λ by computing the likelihood on the data for each potential founder allele, and summing those likelihoods, weighted for the population frequency of the respective founder allele (see equations 1 and 2

© 1996 Nature Publishing Group http://www.nature.com/naturegenetics

in ref. 36). While the procedure was originally applied to genotype data in samples of affected and unaffected individuals[36], it can also be used to calculate likelihoods for pedigree data, where the population frequencies are required to compute genotype probabilities of founder individuals. We modified the ILINK option of the LINKAGE package, version 5.04 (ref. 37), to estimate simultaneously for each marker, the parameter λ and the maximum likelihood of the recombination frequency (θ) within the families. This likelihood ratio test has two degrees of freedom. In this estimation procedure, the disease gene frequency was kept constant at 11.5%, as in our earlier analyses, and the marker allele frequencies obtained in the Costa Rican reference sample were included as population frequencies. By comparing the likelihood obtained for the best fitting values for λ and θ with the likelihood obtained for a value of 0.0 for λ (no over-represented marker allele) and 0.5 for θ (no linkage) the joint evidence for linkage and association was evaluated. This method can be considered a combination of an admixture and a linkage test (with a certain proportion of disease chromosomes representing descendants of a single common founder). It is conceptually similar to the combined linkage and admixture test used in the analysis of locus heterogeneity[15, 38]. For that test it has been proposed that a lod score of 3.28 corresponds (in terms of statistical significance) to a lod score of 3.0 in the regular test for linkage. We suggest that the same threshold be adhered to in the joint association and linkage analysis used here.

Acknowledgements

We thank S. Blower, L. Bull, J. Gitschier and I. Herskowitz for helpful comments, C. Araya for administrative efforts and H. Consengco for technical assistance. This work was supported by the National Institutes of Health (MH00916, MH49499, MH48695, MH47563), a Veterans Administration Research Psychiatrist Award (to LAM), a Young Investigator Award (to NBF) from the National Alliance for Research on Schizophrenia and Depression, and a grant from the American Psychiatric Association's Program for Minority Research Training in Psychiatry (to MAE). Above all we would like to thank the members of Families CR001 and CR004, and the Costa Rican institutions that made this work possible: The Hospital Nacional Psiquiatrico, The Hospital Calderon Guardia, The Ministry of Science, and The Caja Costarricense de Seguro Social.

Received 13 February; accepted 8 January 1996.

1. Goodwin, F.K. & Jamison, K.R. *Manic Depressive Illness*. (New York: Oxford University Press, 1990).
2. Freimer, N.B. *et al*. An approach to investigating linkage for bipolar disorder using large Costa Rican pedigrees. *Neuropsychiat. Genet.* (in the press).
3. Escamilla, M.A. *et al*. The use of linkage disequilibrium approaches to map genes for bipolar disorder in the Costa Rican Population. *Neuropsychiat. Genet.* (in the press).
4. Houwen, R.H.J. *et al*. Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nature Genet.* **8**, 380-388 (1994).
5. Puffenburger E.G., Lynn, A., Kunst, C., Warren, S. & Chakravarti, A. Genetic variation in the old order Mennonites of Lancaster County, PA. *Am J. Hum. Genet.* **57**, A42 (1995).
6. Ozelius, L.J. *et al*. Strong allelic association between the torsion dystonia gene (DYT1) and loci on chromosome 9q34 in Ashkenazi Jews. *Am. J. Hum. Genet.* **50**, 619-628 (1992).
7. Puffenberger, E.G. *et al*. Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. *Hum. Mol. Genet.* **3**, 1217-1225 (1994).
8. Copeman, J.B. *et al*. Linkage disequilibrium mapping of a type 1 diabetes susceptibility gene (*IDDM7*) to chromosome 2q31-q33. *Nature Genet.* **9**, 80-85 (1995).
9. Friedman, T. B. *et al*. A gene for congenital recessive deafness DFNB3 maps to the pericentromeric region of chromosome 17. *Nature Genet.* **9**, 86-91(1995).
10. Boehnke, M. Allele frequency estimation from data on relatives. *Am. J. Hum. Genet.* **48**, 22-25 (1991).
11. Dausset, J. *et al*. Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* **6**, 575-577 (1990).
12. Di Rienzo, A. *et al*. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* **91**, 3166-3170 (1994).
13. Garza, J.C., Slatkin, M., & Freimer, N.B. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.* **12**, 594-603 (1995).
14. Gruis, N.A. *et al*. Homozygotes for the CDKN2 (p16) germline mutation in Dutch familial melanoma kindreds. *Nature Genet.* **10**, 351-353 (1995).
15. Ott, J. *Analysis of Human Genetic Linkage*. 2nd edn. (John Hopkins University Press, Baltimore, 1991).
16. O'Connell, J. R. & Weeks, D. E. The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nature Genet.* **11**, 402-408 (1995).
17. Berrettini, W. H. *et al*. Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. *Proc. Natl. Acad. Sci. USA* **91**, 5918-5921 (1994).
18. Stine, O.C. *et al*. Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. *Am. J. Hum. Genet.* **56**, 1384-1394 (1995).
19. Weeks, D. E. & Lange, K. The affected-pedigree-member method of linkage analysis. *Am. J. Hum. Genet.* **42**, 315-326 (1988).
20. Hudson, T.J. *et al*. An STS-based map of the human genome. *Science* **270**, 1945-1954 (1995).
21. Gyapay, G., *et al*. The 1993-94 Genethon human genetic linkage map. *Nature Genet.* **7**, 246-339 (1994).
22. Pauls, D.L., Bailey, J.N., Carter, A.S., Allen, C.R., & Egeland, J.A. Complex segregation analyses of old order Amish families ascertained through bipolar I individuals. *Am. J. Med. Genet.* **60**, 290-297 (1995).
23. Spence, M.A. *et al*. Bipolar disorder: evidence for a major locus. *Am. J. Med. Genet.* **60**, 370-376 (1995).
24. Puffenberger, E.G. *et al*. A missense mutation of the endothelin-B receptor in multigenic Hirschsprung's disease. *Cell* **79**, 1257-1266 (1994).
25. Baron, M. *et al*. Genetic linkage between X-chromosome markers and bipolar affective illness. *Nature* **326**, 289-292 (1987).
26. Egeland, J.A. *et. al*. Bipolar affective disorders linked to DNA markers on chromosome 11. *Nature* **325**, 783-755 (1987).
27. Straub, R.E. *et al*. A possible vulnerability locus for bipolar affective disorder on chromosome 21q22.3. *Nature Genet.* **8**, 291-296 (1994).
28. Kelsoe J. R. *et al*. Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. *Nature* **342**, 238-243 (1989).
29. Baron, M. *et al*. Diminished support for linkage between manic depressive illness and X-chromosome markers in three Israeli pedigrees. *Nature Genet.* **3**, 49-55 (1993).
30. Nystrom-Lahti, M. *et al*. Close linkage to chromosome 3p and conservation of ancestral founding haplotype in hereditary nonpolyposis colorectal cancer families. *Proc. Natl. Acad. Sci. USA* **91**, 6034-6058 (1994).
31. Sherrington, R. *et al*. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature* **375**, 754-760 (1995).
32. Hastbacka, J. *et al*. The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium. *Cell* **78**, 1073-1087 (1994).
33. Uhrhammer, N. *et al*. Sublocalization of an ataxia-telangiectasia gene distal to D11S384 by ancestral haplotyping in Costa Rican families. *Am. J. Hum. Genet.* **57**, 103-111 (1995).
34. Murray, J. C. *et al*. (CHLC). A comprehensive human linkage map with centimorgan density. *Science* **265**, 2049 (1994).
35. Schellenberg, G.D. *et al*. Association of an apolipoprotein CII allele with familial dementia of the Alzheimer type. *J. Neurogenet.* **4**, 97-108 (1987).
36. Terwilliger, J.D. 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 (1995).
37. Lathrop, G.M., Lalouel, J.M., & Julier, C. Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* **81**, 3443-3446 (1984).
38. Faraway, J.J. Distribution of the admixture test for the detection of linkage under heterogeneity. *Genet. Epidemiol.* **10**, 75-83 (1993).