Am. J. Hum. Genet. 74:886–897, 2004

Genomewide Linkage Analysis of Bipolar Disorder by Use of a High-Density Single-Nucleotide–Polymorphism (SNP) Genotyping Assay: A Comparison with Microsatellite Marker Assays and Finding of Significant Linkage to Chromosome 6q22

F. A. Middleton,[1,2,3] M. T. Pato,[2,3,4] K. L. Gentile,[1,2] C. P. Morley,[2] X. Zhao,[1,2] A. F. Eisener,[2] A. Brown,[1,2] T. L. Petryshen,[6] A. N. Kirby,[5,6] H. Medeiros,[2,4] C. Carvalho,[2] A. Macedo,[8] A. Dourado,[8] I. Coelho,[8] J. Valente,[8] M. J. Soares,[8] C. P. Ferreira,[9] M. Lei,[9] M. H. Azevedo,[4] J. L. Kennedy,[10] M. J. Daly,[5] P. Sklar,[6,7] and C. N. Pato[2,3,4,9]

[1]Department of Neuroscience and Physiology, [2]Center for Psychiatric and Molecular Genetics, and [3]Department of Psychiatry, State University of New York (SUNY), Syracuse, NY; [4]Veterans Administration Medical Center, Washington, DC; [5]Whitehead Institute and [6]Broad Institute, Cambridge, MA; [7]Psychiatric and Neurodevelopmental Genetics Unit, Harvard Medical School, Charlestown, MA; [8]Psicologia Medica, Universidade de Coimbra, Coimbra, Portugal; [9]Psychiatry Service, Ponta Delgada, Azores, Portugal; and [10]Clarke Division, Centre for Addiction and Mental Health, Toronto

We performed a linkage analysis on 25 extended multiplex Portuguese families segregating for bipolar disorder, by use of a high-density single-nucleotide–polymorphism (SNP) genotyping assay, the GeneChip Human Mapping 10K Array (HMA10K). Of these families, 12 were used for a direct comparison of the HMA10K with the traditional 10-cM microsatellite marker set and the more dense 4-cM marker set. This comparative analysis indicated the presence of significant linkage peaks in the SNP assay in chromosomal regions characterized by poor coverage and low information content on the microsatellite assays. The HMA10K provided consistently high information and enhanced coverage throughout these regions. Across the entire genome, the HMA10K had an average information content of 0.842 with 0.21-Mb intermarker spacing. In the 12-family set, the HMA10K-based analysis detected two chromosomal regions with genomewide significant linkage on chromosomes 6q22 and 11p11; both regions had failed to meet this strict threshold with the microsatellite assays. The full 25-family collection further strengthened the findings on chromosome 6q22, achieving genomewide significance with a maximum nonparametric linkage (NPL) score of 4.20 and a maximum LOD score of 3.56 at position 125.8 Mb. In addition to this highly significant finding, several other regions of suggestive linkage have also been identified in the 25-family data set, including two regions on chromosome 2 (57 Mb, NPL = 2.98; 145 Mb, NPL = 3.09), as well as regions on chromosomes 4 (91 Mb, NPL = 2.97), 16 (20 Mb, NPL = 2.89), and 20 (60 Mb, NPL = 2.99). We conclude that at least some of the linkage peaks we have identified may have been largely undetected in previous whole-genome scans for bipolar disorder because of insufficient coverage or information content, particularly on chromosomes 6q22 and 11p11.

Introduction

We recently reported the identification of several chromosomal regions suggestive of linkage, found in an initial genome scan of 16 multiplex Portuguese families who segregated for bipolar disorder (Pato et al., in press). As a follow-up study, we examined 25 extended families from the same population (including 12 of the original 16 families with sufficient DNA available) by use of a new high-density genotyping assay, the GeneChip Human Mapping 10K Array (HMA10K) (Affymetrix). This assay permits the simultaneous genotyping of 11,560 SNPs, spaced throughout the human genome at a median intermarker distance of 210 kb (Kennedy et al. 2003). Each SNP in this assay is interrogated by a distinct set of multiple overlapping probes for each allele and each DNA-strand orientation. SNPs represent variations in DNA sequence that can be mapped to reveal patterns of genetic transmission within pedigrees. Such inheritance patterns can be instrumental in linkage or linkage-disequilibrium analyses of simple or complex phenotypes.

Bipolar disorder affects ~1% of the worldwide population and shows strong evidence of heritability (Tsuang

Received January 20, 2004; accepted for publication February 20, 2004; electronically published April 1, 2004.

Address for correspondence and reprints: Dr. Carlos Pato, Veterans Administration Medical Center, 50 Irving Street NW, Washington, DC 20422. E-mail: cmpato@aol.com

0002-9297/2004/7405-0011$15.00

and Farone 1990). The search for candidate genes in bipolar disorder has traditionally been focused on studies of large extended pedigrees, extensive sib-pair collections, and a variety of family-based and population-based association studies (Ewald et al. 2002; Sklar 2002; Dick et al. 2003; Segurado et al. 2003). Over the past 10 years, we have developed the Portuguese Island Collection to study the genetics of bipolar disorder and schizophrenia in a relatively homogeneous population (Pato et al. 1997). The strategy of using a population isolate and nearly complete ascertainment of multiplex families provides a uniquely powerful and informative study population for complex genetic disorders (Peltonen 2000). Focused on the Azorean and Madeiran islands, this study benefits from the unique parallel history of these two archipelagoes (Pato et al. 2000; Lewis et al. 2003; reviewed by Sklar et al. 2004). For example, we recently reported linkages of schizophrenia to chromosome 5q (Sklar et al. 2004), a region strongly implicated from a meta-analysis of 20 genomewide scans (Lewis et al. 2003).

The linkage results that were reported from the whole-genome scan for bipolar disorder in 16 families from the Portuguese population revealed three regions on chromosomes 2, 11, and 19 with genomewide suggestive linkage and several other regions, including chromosome 6q (marker D6S1021), that also approached suggestive significance (Pato et al., in press). Higher-density mapping with microsatellite markers at 4.2-cM spacing showed a nonparametric linkage (NPL) score of 2.59 at this same marker and an additional peak at D6S1639 (125 Mb), with an NPL score of 3.06 (Pato et al., in press). The finding on 6q was noted in that study since it replicated the findings of the National Institute of Mental Health (NIMH) Genetics Initiative (see Dick et al. 2003). Those authors reported in a study of 250 families with bipolar disorder a maximum LOD score of 2.2 at marker D6S1021 (104.7Mb) on chromosome 6q; this LOD score was subsequently raised to a genomewide significance level of 3.8 in a combined analysis of all 399 pedigrees with bipolar disorder from the authors' collection. Here we report the results of an HMA10K assay in 25 Portuguese families with bipolar disorder, and we present a direct comparison of low- and intermediate-density microsatellite scans with the HMA10K in 12 of these families for the two chromosomes (6 and 11) with the most significant results. Our data indicate that the significant linkage peaks we have identified with the HMA10K assay may have been largely undetected or underestimated in terms of their significance because of insufficient coverage or information content of the microsatellite assays, and we argue strongly in favor of future whole-genome scans with the HMA10K assay.

Material and Methods

Subjects

Families were ascertained from the systematic screening of clinicians, treatment facilities, social services, and extensive family interviews. In the Azores, all four psychiatric hospitals and the two general hospitals participated in the study. Similarly, in Madeira, both psychiatric hospitals and the general hospital participated. On the mainland, families were identified by collaborators at the University of Coimbra. Informed consent was obtained from participants in the genetic and family studies. Collection of blood and family history information was approved by all of the appropriate institutional review boards. On the islands, families were ascertained if both of the proband's parents and all four of the grandparents were native to the islands. Diagnoses were made from data obtained with the Diagnostic Interview for Genetic Studies (DIGS) (Nurnberger et al. 1994). Interviews were performed at the site of the subject's clinical care, in the subject's home, or at an alternate site of the subject's choosing. Interviewers were highly trained psychiatrists, psychologists, or social workers. Interrater reliability was assured by careful training and evaluation of each interviewer prior to the onset of the study and annually thereafter. Reliability was assessed periodically through the independent assessment of 12 subjects by interviewers from different study sites and by a senior supervisor. The chance-corrected interrater agreement, k, was high for the diagnoses of schizophrenia, bipolar disorder, and schizoaffective disorder ($k = 0.9$–0.94). Items related to the diagnosis of schizotypy showed slightly less agreement ($k =\sim 0.75$). Thorough clinical narratives were completed for all subjects. Best-estimate diagnoses according to the *Diagnos tic and Statistical Manual,* fourth edition (DSM-IV), were made by two independent blinded researchers after review of clinical information, the DIGS, the Operational Criteria Checklist for Psychotic Illness (McGuffin et al. 1991), and written narratives. Cases with disparate diagnoses were reviewed by a third senior psychiatrist blind to the case status (M.T.P.).

We used a total of 25 families for linkage analysis in the present study. These families were represented by pedigrees that included 233 individuals (83 founders, 150 nonfounders). The mean family size was 9.32 (range 4–17; mode 5), spanning a mean of 2.88 generations (range 2–4; mode 3). In these families, we ran a total of 148 HMA10K assays, with samples from 88 female and 60 male subjects. The mean number of assays per family was 5.9 (range 3–11; mode 8). Full definitive diagnoses were available for 140 of the 148 subjects, with the remainder of subjects analyzed using the "unknown" phenotype. Of the affected subjects, 73 were diagnosed

with bipolar disorder and 2 were diagnosed with schizoaffective-bipolar disorder, representing a mean of 3.0 affected subjects per family. Of the 25 families, 15 were from the Azores, 5 from Madeira, and 5 from the city of Coimbra on the Portuguese mainland. To compare the HMA10K assay with microsatellite assays, we performed linkage analysis using the *same individuals* in 12 of the 25 families. These 12 families had been used in a previous microsatellite study of bipolar disorder involving 16 families (Pato et al., in press) and represented all families with sufficient DNA available from that study. These 12 families included 107 subjects (36 founders, 71 nonfounders) with a mean family size of 8.92 (range 5–17; mode 7), spanning 2.75 generations (range 2–4; mode 3). A total of 68 subjects were used for the comparisons (29 males, 39 females). Definitive diagnoses were available for 66 of the subjects. A total of 35 subjects were diagnosed with bipolar or schizoaffective-bipolar disorder (mean 2.91 affected individuals per family). Of the 12 families, 8 were from the Azores, 1 was from Madeira, and 3 were from Coimbra.

DNA Isolation

DNA was isolated from 10 ml of venous blood collected from subjects in the clinic or in their homes. Next, 1 vol of blood and 3 vol of ice-cold 1 × Miller's RBCL (155 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM Na_2 EDTA) were added to a 50-ml conical-bottom polypropylene tube and inverted several times. Tubes were placed on ice for 30 min, inverted several times throughout the incubation, and centrifuged (2,500 rpm, 15 min, 40°C). The supernatant was decanted, the sample pellet was resuspended in 30 ml of 1 × Miller's RBCL, and it was then centrifuged again (2,500 rpm, 15 min, 40°C). The supernatant was again decanted, the pellet resuspended, and the sample digested with 30 ml of proteinase K (20 mg/ml) overnight at 42°C, with constant shaking (100 rpm). After digestion, 3 ml of 8M ammonium acetate were added to each sample to precipitate proteins, followed by vortexing for 30 s. Each sample was then centrifuged in a swinging bucket rotor (3,500 rpm, 10 min, 25°C), and the cleared supernatant was transferred to a new 50-ml tube. The purified DNA was precipitated with 2 vol of ice-cold 100% ethanol, centrifuged (2,500 rpm, 10 min, 40°C), and washed with 1 vol of 70% ethanol, followed by centrifugation (2,500 rpm, 5 min, 40°C). After the final wash was decanted, pellets were air dried (30 min, 25°C) and resuspended in 350 μl of Tris-EDTA containing 1 μl RNase A (Sigma). The yield and purity of each sample was assessed using UV spectrophotometry.

DNA Preparation for Human Mapping Assay

Samples were processed according to the GeneChip Mapping Assay Manual (Affymetrix). First, 250 ng of genomic DNA in reduced EDTA TE buffer (10 mM Tris-HCl, pH 8.0; 0.1 mM EDTA, pH 8.0) was digested with *Xba*I (New England Biolabs) for 2 h at 37°C. The enzyme was inactivated by incubation at 70°C for 20 min. The Adaptor *Xba* (Affymetrix) was then ligated to the digested DNA with T4 DNA ligase at 16°C for 2 h and inactivated at 70°C for 20 min. Each ligation reaction was diluted with 75 μl water. For each sample, 10 μl of the diluted ligation reaction was added to four separate 100-μl PCR reactions (1 × PCR buffer II [Applied Biosystems]; 250 μM dNTP [Panvera Takara]; 2.5 mM $MgCl_2$ [Applied Biosystems]; 0.75 μM PCR primer *Xba* [Affymetrix]; 0.1 U/μl Ampli*Taq* Gold [Applied Biosystems]). The PCR was run on an MJ DNA Engine Tetrad thermal cycler (initial denaturation at 95°C for 3 min; 35 cycles of 95°C for 20 s, 59°C for 15 s, 72°C for 15 s; final extension at 72°C for 7 min; hold 4°C). PCR products were run on a 2% Tris-Acetate-EDTA (TAE) gel and then purified with QIAquick spin columns (Qiagen). Each PCR reaction was purified in a separate column, but the same buffer eluate was used for each column to concentrate the samples. Of this purified PCR product, a 20-μg portion was fragmented (1 × fragmentation buffer, 0.24 U fragmentation reagent [Affymetrix]) by incubating at 37°C for 30 min, and then at 95°C for 15 min. The fragmented PCR products were run on a 4% TAE gel to confirm optimal fragmentation (products appear as smears <150 bp in length). The fragmented product was then end-labeled (1 × terminal deoxynucleotidyl transferase [TdT] buffer, 0.143 mM GeneChip DNA labeling reagent, 1.5 U/μl TdT [Affymetrix]) at 37°C for 2 h and inactivated at 95°C for 15 min. A hybridization cocktail containing 0.056 M MES stock (1.22 M MES, 0.89 M [Na^+]), 5% DMSO, 2.5 × Denhardt's, 5.77 mM EDTA, 0.115 mg/ml herring sperm DNA, 11.5 mg/ml human *Cot*I, 0.0115% Tween-20, 2.69 M tetramethyl ammonium chloride, and 2μl oligonucleotide control reagent (Affymetrix) was added to the labeled DNA and then heated at 95°C for 10 min to denature. The reactions were cooled on ice for 10 s and then incubated at 48°C for 2 min. Each sample was then added to a HMA10K and hybridized at 48°C for 16 h in an Affymetrix GeneChip Hybridization Oven at 60 rpm. After 16 h, the probe arrays were washed and stained according to the GeneChip Mapping Assay Manual by use of the DNAARRAY_WS2 protocol on the Affymetrix Fluidics Station 400. Arrays were scanned once with the Agilent GeneArray 2500 scanner and analyzed with Affymetrix GeneChip DNA Analysis Software (GDAS) to generate genotype calls for each of the SNP probes on the array.

Linkage Analysis of HMA10K and Microsatellite Genotype Data

We used MERLIN (Multipoint Engine for Rapid Likelihood Inference) software (Abecasis et al. 2002) to analyze the compiled pedigree structures of the 25-family and 12-family data sets, to examine the information content of the HMA10K and microsatellite assays in these sets, and to perform multipoint NPL analysis. For the SNP assays, output files from GDAS were converted for use in MERLIN using a custom-written program (GenPed, available from F.A.M.). MERLIN implements the Whittemore and Halpern (1994) algorithm to test for allele sharing among affected individuals and also calculates a nonparametric LOD score by use of the Kong and Cox (1997) linear model. Both the NPL *Z* scores and LOD scores were extracted from the MERLIN runs and used to generate graphical plots of the whole-genome scan results. All data presented are plotted using physical positions specified in the July 2003 freeze of the University of California–Santa Cruz (UCSC) human genome assembly (UCSC Genome Bioinformatics). To generate empirical *P* values for the present analysis, we used MERLIN to simulate genotype data for the HMA10K, assay and the 10-cM and 4-cM microsatellite assays, with the same structures as the 25- and 12-family data sets. With each of the four data sets, 100 genomewide simulations were performed, and the resulting simulated data was analyzed with MERLIN; the highest NPL *Z* and LOD scores were recorded for each chromosome for each simulation. For the 25-family HMA10K data set, empirical limits for genomewide significance were established at 3.85 and 3.0 for NPL *Z* and LOD values, respectively, with suggestive levels of 2.89 and 1.92. For the 12-family HMA10K data set, empirical limits for genomewide significance were an NPL score of 4.05 and a LOD score of 2.6, with suggestive level thresholds of 3.05 and 1.84, respectively. For the 10-cM microsatellite data set, empirical limits for genomewide significance were an NPL score of 3.17 and a LOD score of 2.12, with suggestive level thresholds of 2.40 and 1.52, respectively. For the 4-cM microsatellite data set, empirical limits for genomewide significance were an NPL score of 3.40 and a LOD score of 2.11, with suggestive level thresholds of 2.48 and 1.53, respectively. These genomewide significance thresholds represent the NPL and LOD values that could be achieved by chance in our data set at a frequency of 0.05 (1 in 20 simulations), whereas the suggestive thresholds represent the values that could be obtained by chance at a frequency of once per genome scan.

Table 1

Assay Performance Summary across 25 Families (148 Samples)

	Call Rate (%)				
		For Linkage	By Genotype		
	Total	Signal	AA	AB	BB
Mean (%)	96.92	99.06	33.06	34.25	32.69
SD (%)	3.46	1.73	.6	1.1	.63

Results

Assay Performance

In the 25-family data set, the HMA10K assay was performed a total of 148 times on samples from 60 male and

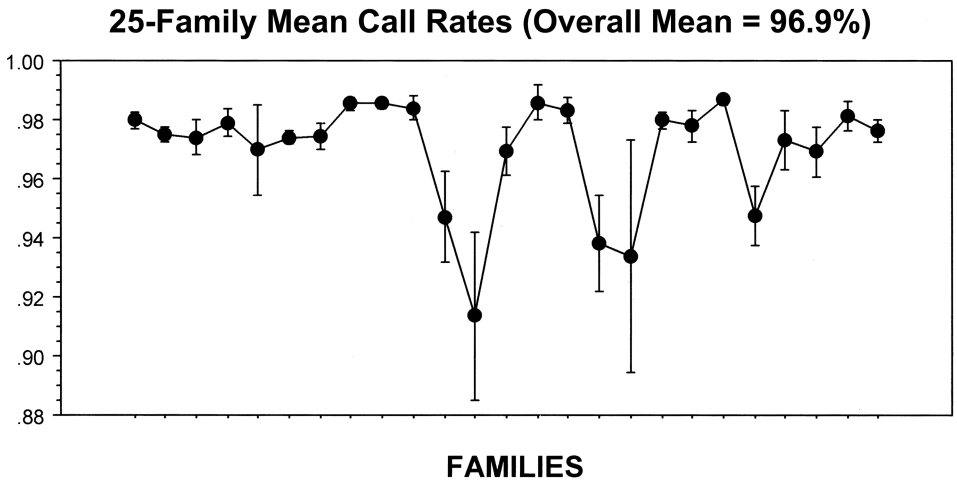

Figure 1 Comparison of SNP call rate across 25 families used in the biopolar study. The mean ± SE of the SNP call rate for each of the families included in our study is shown. The order of the *X*-axis is arbitrary and reflects undisplayed family ID numbers. We note that 20 of the 25 families had a mean call rate of ⩾97%. Analysis of the families with the lowest call rates revealed a potential interaction of the sample age and/or handling with optimal assay performance.

Table 2

Regions of Genomewide Suggestive and Significant NPL Scores in the 12- and 25-Family Data Sets

Chromosome and Position (Mb)	NPL Score in Same 12 Families for			NPL Score in 25 Families for
	10-cM Microsatellite Marker Assay	4-cM Microsatellite Marker Assay	HMA10K	HMA10K
2:				
38:	2.18	ND	**3.15**	1.83
57:	−.44	ND	1.99	**2.98**
145:	−2.19	ND	2.73	**3.09**
4:				
91:	.43	ND	1.33	**2.97**
6:				
104:	1.96	1.88	1.92	1.86
127:	.79	1.16	**3.87**	***4.20***
11:				
45–68	**2.62**	**3.31**	***4.44***	2.51
16:				
20	2.27	ND	1.47	**2.89**
20:				
60:	ND	ND	2.36	**2.99**
Empirical significant threshold	3.17	3.40	4.05	3.85
Empirical suggestive threshold	2.40	2.48	3.05	2.89

Note.—Suggestive NPL scores are in boldface; significant NPL scores are in boldface italics. ND = not determined.

88 female subjects. The "call rate" in the HMA10K assay is the percentage of SNPs for which a definitive genotype call was obtained in any given sample. If the assay is performing optimally, call rates will exceed 95% (Kennedy et al. 2003). Across the 148 samples in our study, the mean call rate was 96.92% (table 1). Of the 25 families we analyzed, 20 had mean call rates of ⩾97%, whereas 5 families had mean call rates ranging from 91%–95%. The distribution of genotype calls across the 25-family data set was approximately equal among AA, AB, and BB genotypes, with slightly more heterozygous calls observed (table 1). The greatest sources of variability in the call rate appeared to be the family origin of the sample (fig. 1). Figure 1 indicates that the lowest call rates we obtained were from the families that originated in the city of Coimbra, where the DNA was extracted before shipment to the United States. These families were some of the first ascertained for bipolar disorder and are thus among the oldest samples in the Portuguese Island Collection. Samples from the same family were generally collected and batch processed. Thus, we interpret these observations to indicate sample age and handling (repeated freeze thawing or variation in experimenter technique) as potential impediments to optimal assay performance. In fact, we highly recommend that researchers employ fresh DNA samples from sources with high quality-control measures for best assay performance.

Error Rates

To estimate the error rates of the assay, we performed two types of analyses. First, in the 60 male subjects used for this analysis, we examined the genotype calls of the X chromosome, which was interrogated by a total of 301 SNP probe sets. Since males inherit only a single X chromosome, they must be completely homozygous for each SNP. Of the total of 18,060 calls, we observed only a single heterozygous call in these 60 males, yielding an error rate of 1 in 18,060, or 0.0055%, for the X chromosome.

To further probe this issue, we also looked for genotyping errors of transmission in the X chromosome, by examining 18 mother-son pairings in the 25 families for inaccurate genotypes (sons not matching mothers at maternal homozygous alleles). On average, <200 alleles were found with homozygous maternal genotype calls among these 18 pairings. Only 2 of >3,500 individual SNPs analyzed in this manner were determined to have incorrect calls. One of these errors was in the maternal genotype (determined because all three male offspring shared the same genotype in this family), whereas the other SNP genotyping error could have been in either the maternal or the offspring genotype. Thus, a reasonable approximation of transmission errors in our data set is 2 in 3,500, or 0.057% per generation. It is important to note that, while performing linkage analysis, MERLIN automatically discards SNPs with these types of errors. Moreover, the data presented in the present study were analyzed after performing the "Pedwipe" command in MERLIN to permanently remove such errors. In our assays, the total number of SNPs included after mapping and removing errors was 11,190.

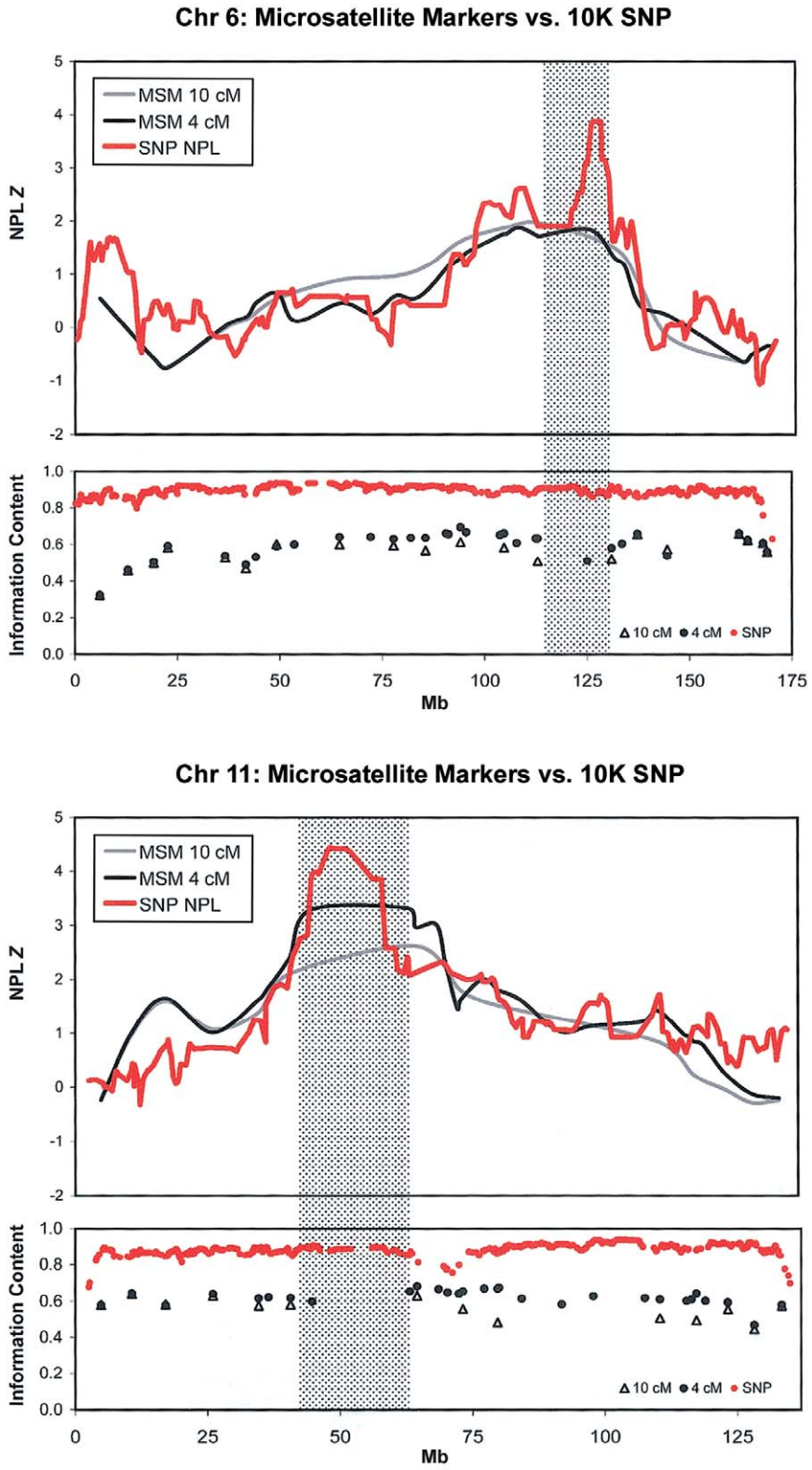

Figure 2 Linkage signals obtained with 10-cM–spaced and 4-cM–spaced microsatellite assays, as well as the HMA10K SNP genotyping assay. These assays were performed on the same individuals from each of the same 12 families. Note the high correlation of the different assays in general, and that for both chromosomes 6 and 11, the SNP assay detected major linkage peaks at locations where the information content and coverage of the microsatellite panels were relatively low. Mb, megabase position; MSM, microsatellite markers.

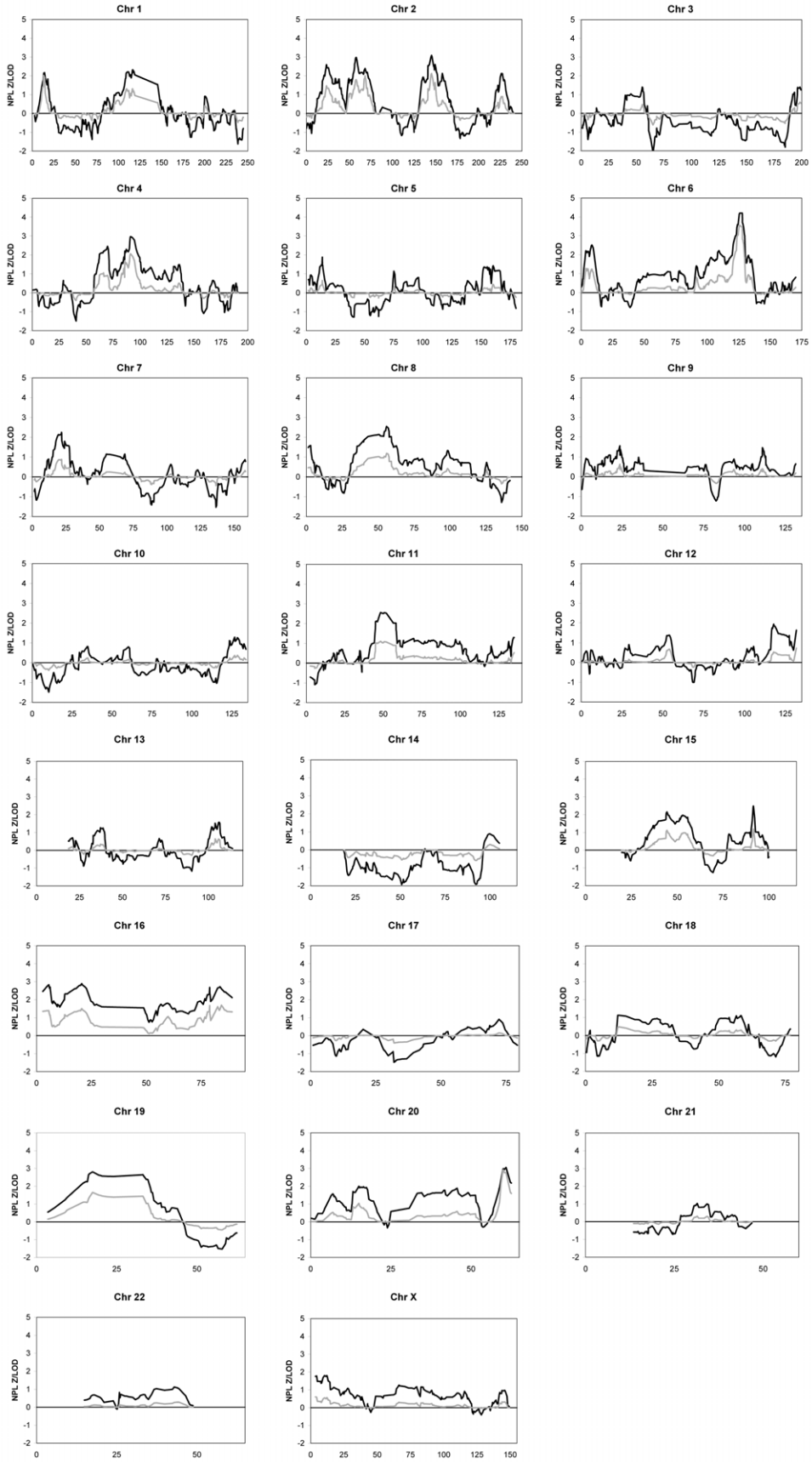
Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
Chr 6
Chr 7
Chr 8
Chr 9
Chr 10
Chr 11
Chr 12
Chr 13
Chr 14
Chr 15
Chr 16
Chr 17
Chr 18
Chr 19
Chr 20
Chr 21
Chr 22
Chr X
NPL Z/LOD

Table 3

Chromosome 6 Linkage Peak Candidate Genes

Position of SNP (Mb)	Gene(s)	Position of Gene (Mb)
125.438	*Tumor protein D52-like 1*	125.505–125.521
125.769	*CGI 130 protein*	126.532–125.538
125.798	*CGI 130 protein*	**126.532–125.538**
	HEY2 hairy/enhancer-of-split related 2	126.006–126.018
	ERAP140/NCOA7	126.142–126.187
126.371	*Hypothetical protein AK094706*	126.176–126.187
	Hypothetical protein AL834442	126.111–126.188

NOTE.—Maximum NPL position appears in boldface.

Linkage Analysis

We present a comparison of the low- and intermediate-density microsatellite scans and the higher-density SNP assay by use of the 12-family data set, and we report the results of the high-density SNP assay in the larger 25-family data set.

Twelve-Family Genomewide Scan

Our results with the HMA10K assay confirm the initial finding of linkages to chromosomes 2, 6, and 11 in 12 of the same families we used for either the 10-cM or the 4-cM microsatellite screening (table 2) (see Pato et al., in press). Moreover, throughout much of chromosomes 6 and 11, the linkage results obtained with the different assays showed a remarkable degree of similarity (fig. 2). However, the increased coverage and information content of the HMA10K assay in these families has now raised the maximum NPL *Z* and/or LOD scores on chromosomes 6 and 11 beyond the level of genomewide significance. These thresholds were empirically determined to be an NPL score of 4.05 and a nonparametric LOD score of 2.60 (see the "Material and Methods" section) (table 2). Specifically, the 12-family data set achieved an NPL score of 3.87 and LOD score of 2.79 at position 125.8 Mb on chromosome 6q22 and an NPL score of 4.44 and LOD scores of 2.62 at position 48.5 Mb on chromosome 11p11. Inspection of the positions of these significant peaks on chromosomes 6 and 11 in the HMA10K assay indicates that they are in regions where the information content and coverage of the microsatellite marker assays were relatively low and potentially insufficient for peak detection. In contrast, the mean information content of the SNP assay was 0.842 (range: 0.739–0.874).

Twenty-five–Family Genomewide Scan

Empirical thresholds for genomewide significance in the 25 families were established at an NPL score of 3.85 and a nonparametric LOD score of 3.0 (see the "Material and Methods" section). The results of the linkage analysis for this data set (table 2; fig. 3) provided further validation of the finding on chromosome 6q22, achieving a maximum NPL score of 4.20 and a maximum LOD score of 3.56 at position 125.8 Mb. This peak and the surrounding region of chromosome 6 that exceeded the threshold for suggestive genomewide significance (NPL > 2.89) were interrogated by a total of 38 SNPs, with an average inter-SNP distance of 0.20 Mb. At the precise location of maximal peak height, there were a total of five individual SNPs interspaced with six genes (table 3). Haplotype-based analysis of 55 SNPs spanning this peak for the 25-family data set with MERLIN failed to indicate any evidence of linkage disequilibrium in this population (not shown). In addition to these findings on chromosomes 6 and 11 in the 12- and 25-family data sets, several other regions of suggestive linkage on chromosomes 2, 4, 16, and 20 have also been identified in these families using the HMA10K assay that were largely undetected with the microsatellite scans of the 12-family set (table 2). The HMA10K results in the two family sets are shown in figure 4 for chromosomes 2, 4, 6, 11, 16, and 20. We note that, in general, the linkage results obtained with the 25 families were very similar to those obtained using only 12 families. Moreover, the linkage signal on chromosome 6q22 at position 125.8 Mb is the only region that achieves genomewide significance for NPL or LOD scores in both scans. The linkage signal on chromosome 11p11 that attained genomewide significance in the 12 families clearly dropped below that level for the full 25-family data set, as did peaks on

Figure 3 NPL analysis of 25 families with bipolar disorder from the Portuguese Island Collection. The number of each chromosome is shown at the top of each plot. The *X*-axis indicates the physical position (Mb) of the SNP marker. The *Y*-axis indicates the NPL *Z* score (*black*) or Kong and Cox LOD score (*gray*). For this scan, the empirical limit for genomewide significance was an NPL score of 3.85 and a LOD score of 3.15. Note that only the peak on chromosome 6 at 125.8 Mb was significant when both NPL *Z* and LOD thresholds were used.

several other chromosomes, with notable exceptions of peaks on chromosomes 4 and 20.

Discussion

A direct comparison of the linkage peaks obtained through the use of microsatellite markers and the HMA10K assay has indicated a high degree of correspondence between the two approaches in general (fig. 2). However, there are clear instances where the sampling provided by traditional microsatellite approaches is largely insufficient to detect highly significant linkage peaks. The peaks on chromosomes 6 and 11 that reached genomewide significant levels in the SNP assay are two examples of this fact. On the basis of the greater information content and coverage of the HMA10K assay, its high level of accuracy, and the ability to rapidly perform and analyze the data obtained from genomewide scans with MERLIN and other software, we believe further use of the assay is warranted.

Major Findings

In a previous study (Pato et al., in press), we reported suggestive genomewide linkage signals for chromosomes 2, 6, 11, and 19, on the basis of the results obtained using a 10-cM–spaced microsatellite scan. Mapping chromosome 6 at 4.2-cM spacing with microsatellite markers in those 16 Portuguese families raised the NPL score to 2.59 at the same marker, D6S1021 (104.7 Mb), on chromosome 6 and revealed an additional peak at marker D6S1639 (125 Mb), with an NPL score of 3.06. The finding on 6q was highlighted in that report because of its relevance to the results of Dick et al. (2003), who reported in a study of 250 families with bipolar disorder a maximum LOD score of 2.2 at marker D6S1021 (104.7 Mb) on chromosome 6q, which was subsequently

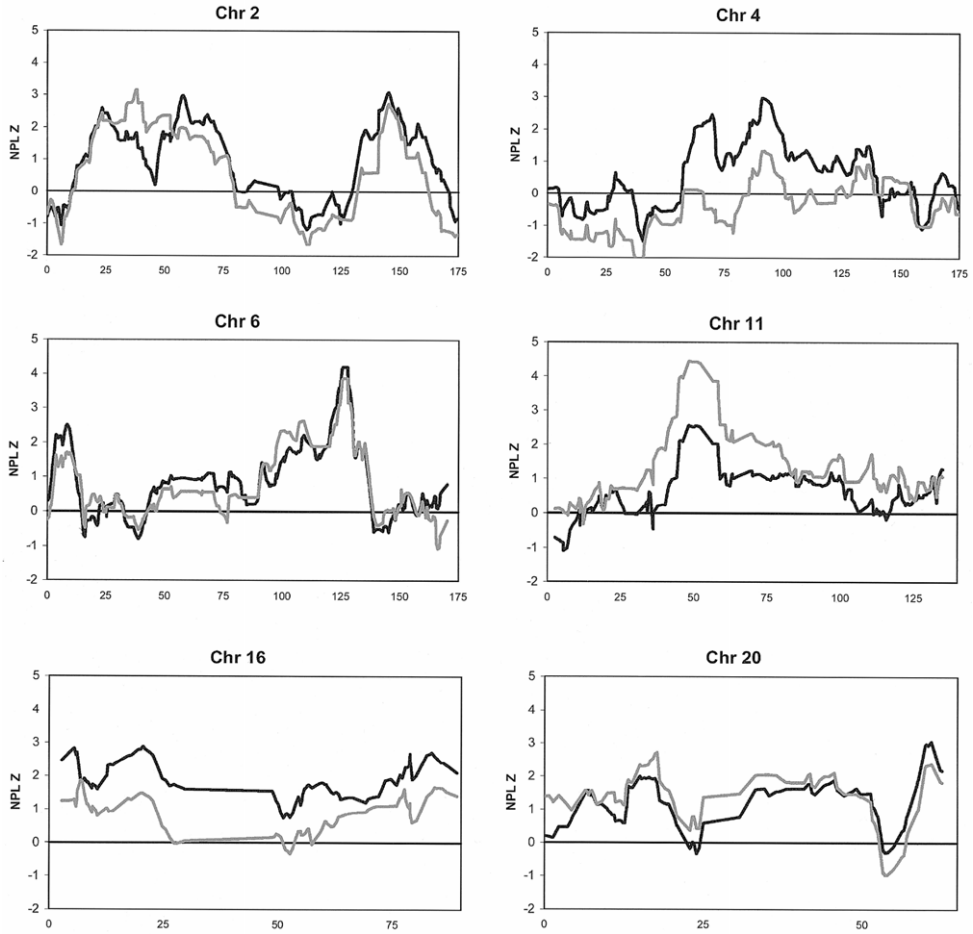

Figure 4 Comparison of the 12-family (*gray*) and 25-family (*black*) genomewide linkage scans for selected chromosomes showing suggestive or significant linkage (see table 1). The *X*-axis indicates physical position (Mb). Note that for both scans, the signal on chromosome 6 at position 125.8 Mb is the only genomic region that achieves genomewide significance (of NPL score and/or LOD score).

raised to a genomewide significance level of 3.8 in a combined analysis of all 399 pedigrees with bipolar disorder from their collection.

In the present report, using a new high-density SNP genotyping array, we have confirmed the initial finding of suggestive linkage to chromosomes 2, 6, and 11 in a subset of the same families used for the microsatellite screening. Moreover, as a result of the greater information content and coverage of the HMA10K assay, we have achieved a significant level of linkage for the peaks on chromosomes 6 and 11, with maximum NPL and LOD scores of 4.44 and 2.62, respectively, on chromosome 11 at position 48.5 Mb, and maximum NPL and LOD scores of 3.87 and 2.79, respectively, on chromosome 6 at position 125.8 Mb. Moreover, with the complete set of 25 families, we can now clearly establish a significant genomewide finding with a maximum NPL score of 4.20 and a maximum LOD score of 3.56 for the peak on chromosome 6 at position 125.8 Mb.

Relationship to Other Studies

In an extensive nonparametric genomewide linkage analysis of 153 pedigrees studied through the NIMH Collaborative Genetics Initiative, McInnis and colleagues (2003) reported a single chromosomal region on 16p13 with genomewide significance, as well as nominally significant linkage at four additional chromosomal regions (20p12, 11p15, 6q24, and 10p12). The present study has reported significant or suggestive linkage on four of these same chromosomes (6, 11, 16, and 20). The location of our significant or suggestive peaks on these four chromosomes is at least 10 Mb distant (range 11–40 Mb) from each of the peaks reported in that study, although, in some cases, we have obtained nominally significant peaks that are closer in proximity. Similarly, in another study of bipolar disorder in families of Danish origin, Ewald and colleagues (2002) also reported significant or suggestive linkage of bipolar disorder with regions of chromosomes 1p14-21, 4p16, 6q14-22, 10q26, 12q24.3,and 16p13.3. However, the locations of their linkage peaks are also somewhat distant from the ones we obtained in the present study. Thus, we believe the present results represent novel localizations of linkage for bipolar disorder. On the other hand, it is possible that mapping of the families in those studies with a more comprehensive assay, such as the HMA10K, may reveal refinement in their localization that results in greater consistency between our results.

The level of significance achieved in this study for the linkage peak on chromosome 6 rivals those obtained in any previously reported linkage study of bipolar disorder (reviewed in Sklar 2002). The fact that we observed this peak to be significant and highly localized in both the 12- and 25-family data sets and that the peak was significant using both NPL and Kong and Cox LOD scores argues in favor of additional research in this genomic region. We briefly discuss below some of the potential candidate genes that might contribute to the bipolar phenotype in our sample population.

Chromosome 6 Candidate Genes

In the maximal NPL and maximal LOD region identified on chromosome 6, position 125.8 Mb, there are only four known genes and two hypothetical genes (table 3). One of these genes (*CGI-130*) actually spans the linkage peak at position 125.8 Mb. The next-closest genes are located ~200 kb upstream (*tumor protein D52-like 1*) or ⩽350 kb downstream of the linkage signal peak (*HEY2* and *ERAP140*). It is interesting that, although the biological role of *CGI-130* has never been studied, both *HEY2* and *ERAP140* appear to have a major role in brain function. Therefore, we briefly discuss some of the potential import of these genes in our findings.

HEY2

HEY2 is the mammalian homologue of the *Drosophila melanogaster hairy/enhancer of split 2* gene. The action of *HEY2* appears to largely revolve around inhibition of the transcriptional activities of the neuronal basic helix-loop-helix (bHLH) factors, particularly Mash1 and Math3 (Sakamoto et al. 2003). Neuronal bHLH genes both promote neuronogenesis and inhibit gliogenesis. In normal mouse brain development, *HEY2* is expressed in the cortical matrix, as well as the retina and cochlear ganglion (Leimeister et al. 1999). Misexpression of the *HEY2* transcript by in utero electroporation in mice at 13.5 d gestation (the time when cortical neurons are being produced) results in a temporary stabilization of the neuronal precursor population that subsequently causes more late-born neurons to arise in the superficial layers of cortex. On the other hand, when *HEY2* is overexpressed at later stages of mouse brain development, it actually inhibits neuronogenesis in favor of gliogenesis (Sakamoto et al. 2003). If there were an alteration in the *HEY2* gene in bipolar disorder, it could have an influence on neural cell fate and circuit formation or the regulation of neurogenesis that occurs in such adult human brain regions as the dentate gyrus (Eriksson et al. 1998). It is interesting to note that high levels of stress have also been shown to reduce the amount of neurogenesis in the adult dentate gyrus (reviewed in Gould et al. 2000), and stress itself is one of the best predictors of relapse in bipolar disorder (Ambelas and George 1986). This raises the possibility that modulation of *HEY2* expression may be one of the mechanisms of regulating neurogenesis and long-term stress adaptation in adult humans.

ERAP140

Estrogen-receptor associated protein 140 (*ERAP140*) is a recently described nuclear receptor coactivator that is highly abundant in the brain, where its expression is exclusive to neurons (Shao et al. 2002). The actions of *ERAP140* are believed to be similar to those of other nuclear receptor coactivators, in that these coactivators directly regulate the effects of steroid hormones on neurons. These effects are known to influence transcription of genes involved in neurodevelopment and neuroprotection. Although purely speculative, one possibility is that alteration of the *ERAP140* gene in bipolar disorder could produce significant differences in the responsivity of neurons to estrogens and steroid hormones. Prediction of the consequences of such an alteration would depend on knowing which regions of the human brain would be most affected by this change in responsivity, information that is lacking at present. The profound influence of stress on circulating steroid levels in the brain, however, is well established (see Gould et al. 2000). Thus, two of the most attractive candidate genes identified through our linkage analysis of 25 bipolar families are known to influence brain development and may play a role in the brain's adaptation to environmental stress.

Conclusions

The HMA10K genotyping assay provides increased power for detection of significant linkage signals, compared with the power of traditional microsatellite marker panels. The ease, speed, and accuracy with which these assays can be performed and analyzed on high-quality DNA samples should serve as an impetus to other researchers searching for clues regarding the pathogenesis of human diseases. In our study, we have now obtained strong indications that linkage to a region of chromosome 6q22 is a fairly common feature in Portuguese families with bipolar disorder. This information will be of great benefit to us in our efforts to identify candidate genes for this debilitating psychiatric illness. We tentatively suggest follow-up studies on three promising candidate genes, *CGI-130, HEY2,* and *ERAP140*.

Acknowledgments

We thank the families for their participation and Dr. Eric Lander for his support of the project. Financial support for this study was provided by awards from the National Alliance for Research on Schizophrenia and Depression (to C.N.P., M.T.P., and J.L.K.), as well as the NIMH (MH52618 and MH058693 [to C.N.P. and M.T.P.]) and the Canadian Institutes for Health Research (to J.L.K.).

References

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin: rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101

Ambelas A, George M (1986) Predictability of course of illness in manic patients positive for life events. J Nerv Ment Dis 174:693–695

Dick DM, Foroud T, Flury L, Bowman ES, Miller MJ, Rau NL, Moe PR, et al (2003) Genomewide linkage analysis of bipolar disorder: a new sample of 250 pedigrees from the National Institute of Mental Health genetics initiative. Am J Hum Genet 73:107–114 (erratum 73:979)

Eriksson PS, Perfilieva E, Bjork-Eriksson T, Alborn AM, Nordborg C, Peterson DA, Gage FH (1998) Neurogenesis in the adult human hippocampus. Nat Med 4:1313–1317

Ewald H, Flint T, Kruse TA, Mors O (2002) A genome-wide scan shows significant linkage between bipolar disorder and chromosome 12q24.3 and suggestive linkage to chromosomes 1p14–21, 4p16, 6q14–22, 10q26 and 16p13.3. Mol Psychiatry 7:734–744

Gould E, Tanapat P, Rydel T, Hastings N (2000) Regulation of hippocampal neurogenesis in adulthood. Biol Psychiatry 48:715–720

Kennedy GC, Matsuzaki H, Dong S, Liu WM, Huang J, Liu G, Su X, Cao M, Chen W, Zhang J, Liu W, Yang G, Di X, Ryder T, He Z, Surti U, Phillips MS, Boyce-Jacino MT, Fodor SP, Jones KW (2003) Large-scale genotyping of complex DNA. Nat Biotechnol 21:1233–1237

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Leimeister C, Externbrink A, Klamt B, Gessler M (1999) Hey genes: a novel subfamily of hairy- and Enhancer of split related genes specifically expressed during mouse embryogenesis. Mech Dev 85:173–177

Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, et al (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: schizophrenia. Am J Hum Genet 73:34–48

McInnis MG, Dick DM, Willour VL, Avramopoulos D, MacKinnon DF, Simpson SG, Potash JB, et al (2003) Genome-wide scan and conditional analysis in bipolar disorder: evidence for genomic interaction in the National Institute of Mental Health genetics initiative bipolar pedigrees. Biol Psychiatry 54:1265–1273

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

Nurnberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T (1994) Diagnostic interview for genetic studies: rationale, unique features, and training. NIMH Genetics Initiative. Arch Gen Psychiatry 51:849–859 (discussion 51: 863–864)

Pato CN, Azevedo MH, Pato MT, Kennedy JL, Coelho I, Dourado A, Macedo A, Valente J, Ferreira CP, Madeira J, Gago da Camara J, Moniz M, Correia C (1997) Selection of homogeneous populations for genetic study: the Portugal genetics of psychosis project. Am J Med Gen 74:286–288

Pato CN, Macedo A, Ambrosio A, idsVincent JB, Bauer A, Schindler K, Xu J, Coelho I, Dourado A, Valente J, Azevedo MH, Kennedy JL, Pato MT (2000) Detection of expansion regions in Portuguese bipolar families. Am J Med Gen 96: 854–857

Pato CN, Pato MT, Kirby A, Petryshen TL, Medeiros H, Carvalho C, Macedo A, Dourado A, Coelho I, Valente J, Soares MJ, Ferreira CP, Lei M, Verner A, Hudson TJ, Morley CP, Kennedy JL, Azevedo MH, Daly MJ, Sklar P. Genome-wide scan in Portuguese island families implicates multiple loci in bipolar disorder. Am J Med Gen (in press)

Peltonen L (2000) Positional cloning of disease genes: advantages of genetic isolates. Hum Hered 50:66–75

Sakamoto M, Hirata H, Ohtsuka T, Bessho Y, Kageyama R (2003) The basic helix-loop-helix genes Hesr1/Hey1 and Hesr2/Hey2 regulate maintenance of neural precursor cells in the brain. J Biol Chem 278:44808–44815

Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger JI, Craddock N, et al (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: bipolar disorder. Am J Hum Genet 73:49–62

Shao W, Halachmi S, Brown M (2002) ERAP140, a conserved tissue-specific nuclear receptor coactivator. Mol Cell Biol 22: 3358–3372

Sklar P (2002) Linkage analysis in psychiatric disorders: the emerging picture. Annu Rev Genomics Hum Genet 3:371–413

Sklar P, Pato MT, Kirby A, Petryshen TL, Medeiros H, Carvalho C, Macedo A, Dourado A, Coelho I, Valente J, Soares MJ, Ferreira CP, Lei M, Verner A, Hudson TJ, Morley CP, Kennedy JL, Azevedo MH, Lander E, Daly MJ, Pato CN (2004) Genome-wide scan in Portuguese island families identifies 5q31–5q35 as a susceptibilty locus for schizophrenia and psychosis. Mol Psychiatry 9:213–218

Tsuang MT, Farone SV (1990) Genetics of mood disorders. Johns Hopkins University Press, Baltimore

Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. Biometrics 50:118–127