Linkage analysis of anorexia nervosa incorporating behavioral covariates

Bernie Devlin1,*, Silviu-Alin Bacanu1, Kelly L. Klump2, Cynthia M. Bulik3, Manfred M. Fichter4, Katherine A. Halmi5, Allan S. Kaplan6,7, Michael Strober8, Janet Treasure9, D. Blake Woodside7, Wade H. Berrettini10 and Walter H. Kaye1

1Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA 15213-2593, USA, 2Department of Psychology, Michigan State University, East Lansing, MI, 48824, USA, 3Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, VA 23298-0126, USA, 4Klinik Rosenbeck, Hospital for Behavioral Medicine, affiliated with the University of Munich, Prien, Germany, 5New York Presbyterian Hospital, Weill Medical College of Cornell University, White Plains, NY 10605, USA, 6Program for Eating Disorders and 7Department of Psychiatry, Toronto General Hospital, Toronto, Ontario, M5G 2C4, Canada, 8Department of Psychiatry and Behavioral Science, University of California at Los Angeles, Los Angeles, CA 90024-1759, USA, 9Eating Disorders Unit, Institute of Psychiatry and South London and Maudsley National Health Service Trust, UK and 10Center of Neurobiology and Behavior, University of Pennsylvania, Philadelphia, PA 19104, USA

Received November 26, 2001; Revised and Accepted January 18, 2002

Eating disorders, such as anorexia nervosa (AN) and bulimia nervosa (BN), have genetic and environmental underpinnings. To explore genetic contributions to AN, we measured psychiatric, personality and temperament phenotypes of individuals diagnosed with eating disorders from 196 multiplex families, all accessed through an AN proband, as well as genotyping a battery of 387 short tandem repeat (STR) markers distributed across the genome. On these data we performed a multipoint affected sibling pair (ASP) linkage analysis using a novel method that incorporates covariates. By exploring seven attributes thought to typify individuals with eating disorders, we identified two variables, drive-for-thinness and obsessionality, which delimit populations among the ASPs. For both of these traits, or covariates, there were a cluster of ASPs who have high and concordant values for these traits, in keeping with our expectations for individuals with AN, and other clusters of ASPs who did not meet those expectations. When we incorporated these covariates into the ASP linkage analysis, both jointly and separately, we found several regions of suggestive linkage: one close to genome-wide significance on chromosome 1 (at 210 cM, D1S1660; LOD = 3.46, P = 0.00003), another on chromosome 2 (at 114 cM, D2S1790; LOD = 2.22, P = 0.00070) and a third region on chromosome 13 (at 26 cM, D13S894; LOD = 2.50, P = 0.00035). By comparing our results to those implemented using more standard linkage methods, we find the covariates convey substantial information for the linkage analysis.

INTRODUCTION

Eating disorders are classic examples of complex psychiatric phenotypes having both genetic and environmental determinants. Some of the environmental underpinnings are well understood (1,2), being rooted in the reigning western view of the perfect body; exceptionally thin and fit. Apparently individuals with anorexia nervosa (AN) are quite sensitive to these cultural messages and exhibit a dread of body fat that drives, then sustains an unrelenting avoidance of normal body weight (3). Still, despite widespread dieting in the western world, only a small fraction of women (0.1–0.7%) (4–6) and an even smaller fraction of men ever develop eating disorders.

In addition to environment, liability to eating disorders has a familial (7–10) and genetic basis. Estimated heritabilities of liability to eating disorders, based on twin studies, range from 54 to 80% (11–16). Moreover, studies of clinically unaffected twins find that 40–60% of the variance in liability to abnormal eating attitudes is attributable to additive genetic effects (17–19). First-degree relatives of eating disorder probands have an ~3% lifetime risk of AN, whereas none of the approximately 1000 first-degree relatives of control probands received the diagnosis (7,9). Based on these rates, the recurrence risk for first-degree relatives of an AN proband is likely to be at least 10-fold higher than the population prevalence.

Psychometric studies (20–32) have consistently linked AN to a cluster of moderately heritable (33) personality and temperamental traits, specifically obsessionality, perfectionism, neophobia and harm avoidance. These traits persist after the long-term normalization of body weight and menses,

*To whom correspondence should be addressed. Tel: +1 412 624 1432; Fax: +1 412 624 8181; Email: devlinb@msx.upmc.edu
supporting the possibility that they influence susceptibility. Pharmacologic and physiologic studies in people with AN suggest that these symptoms, as well as pathologic feeding behavior, are associated with a dysregulation of central nervous system serotonin pathways (24).

In our ongoing multinational study to determine the genetic underpinnings of eating disorders (34), we have measured a wide variety of psychiatric, personality and temperamental phenotypes on more than 200 affected relative pairs (ARPs), mostly affected sibling pairs (ASPs). Families in this first cohort have been recruited through a proband diagnosed with AN. Affected relatives of the proband have a wider range of eating disorder diagnoses [i.e. AN, bulimia nervosa (BN) or eating disorder not otherwise specified (NOS)], although most individuals have AN. By using a combined approach of genome-wide linkage analysis, teamed with the analysis of phenotypes related to eating disorders, we hope to identify some of the polymorphisms that contribute to eating disorder liability.

In this paper, we analyze the ASP data using a new method of linkage analysis that incorporates covariates (35). Whereas the larger study included numerous potential covariates, including personality, behavioral and psychiatric traits, we limit our analysis a priori to attributes thought to typify individuals with AN. We use several criteria to select these potential covariates: they must be consistently related to eating pathology, they must be heritable and they must be indicators of severity of some aspect of the disorder, or enduring traits rather than states resulting from the illness itself. Thus, these traits are unlikely to result from nutritional deficiencies and, in fact, could underlie the etiology of eating pathology.

Based on these criteria, seven traits have been initially selected for analysis: harm avoidance and novelty seeking from the Temperament and Character Inventory (TCI) (36), trait anxiety from the State Trait Anxiety Inventory (37), overall perfectionism from the Multidimensional Perfectionism Scale (MPS) (38), overall obsessiveness/compulsivity from the Yale–Brown Obsessive Compulsive Scale (Y-BOCS) (39), overall eating disorder-related obsessiveness/compulsivity from the Yale–Brown–Cornell Eating Disorder Scale (YBCEDS) (40) and drive-for-thinness from the Eating Disorders Inventory-2 (EDI-2) (41). All of these traits show phenotypic relationships with eating disorders. Two of these traits are core measures of eating disorder pathology (i.e. YBCEDS total score and EDI drive-for-thinness), whereas the others capture features commonly associated with AN (25,27,30,42–44).

RESULTS

To estimate marker allele frequency differences among the families from the seven different recruitment sites, we fit a Bayesian hierarchical model (45) to the site-specific data. The estimated degree of differentiation, expressed as a standardized measure like Wright’s (46) Fst, is 0.0032 for 316 autosomal loci showing any differentiation. Another 50 autosomal loci showed no detectable differentiation (i.e. estimated differentiation is zero). Limited differentiation is to be expected (47) because almost all participants identified their ancestry as European. Nonetheless, for our linkage analyses, we use the site-specific estimates of allele frequencies for each marker, which arise naturally from the hierarchical model (45) (see http://wpicr.wpic.pitt.edu/WPICCompGen for complete, population-specific allele frequencies). Because we use a likelihood ratio statistic for linkage, log-likelihoods can be summed over sites to produce a summary statistic.

To implement the ASP linkage analysis with covariates using the pre-clustering mixture model (35), we sought one or more covariates that separate ASPs into subpopulations, which plausibly have different origins for eating disorder liability. To do so, we first normalize the seven chosen covariates by their respective normative distributions, derived from convenience samples from the population (Materials and Methods). For each trait, we plot the transformed sibling-pair values for each trait against each other as a diagnostic to identify clusters. Five of the traits appear to display little discriminatory power; as an example, we display the values for perfectionism in Figure 1. While it is clear that individuals with eating disorders generally are perfectionists, there are no clear-cut populations among ASPs. Two traits tend to cluster ASPs, however; namely EDI-2 drive-for-thinness (EDI-DT) and Y-BOCS obsessiosity (Fig. 1).

For both traits, there is a group of ASPs who are extreme, relative to most individuals in the population, and highly concordant. This group fits our expectations for individuals with AN. In contrast a substantial number of ASPs either have similar but not extreme values for these traits, or the ASPs show little similarity for these traits. We suspect these ASPs have different origins for their vulnerability to eating disorders, compared to those ASPs who are both extreme and concordant. For our linkage analyses, then, we assume the extreme concordant ASPs for either trait are typical for the linked group.

Figure 1. Trait values for ASPs for perfectionism, drive-for-thinness (EDI-DT) and obsessiosity. Trait values are transformed into standard normal deviates, so the scales are in SD from the population mean.
For the pre-clustering model, the probability of membership in the linked group must be determined. To form two clusters, we use the mclust function in Splus (48) and both covariates jointly, as described elsewhere (35); then, to estimate probability of group membership, we use the Splus quadratic discriminant analysis (QDA) function. (QDA discriminates between two or more known groups by assuming that their covariate values follow normal distributions that differ in their means and variances.) We also compute these probabilities using one covariate at a time. If either individual of the ASP had a missing value for a trait, the ASP is excluded from the linkage analysis. Missing values are uncommon for obsessionality; after dropping ASPs due to missing values, the total number of ASPs we analyzed was 180. Unfortunately, EDI-DT was not measured in the same way in Munich, Germany, as it was in the other sites, and therefore we omit all ASPs from Munich for the obsessionality-based linkage analysis; for obsessionality we analyzed 143 ASPs. For the analysis incorporating both EDI-DT and obsessionality, 116 ASPs are available for analysis.

The familial intraclass correlations for EDI-DT and obsessionality are not large; \( r = 0.17 \). These correlations, however, result from a specially selected population, specifically eating disorder ASPs, and are therefore likely to be different from those obtained from a random sample from the population. EDI-DT and obsessionality are weakly correlated \( (r = 0.21) \), but the weights of membership in the linked group are less so \( (r = 0.07) \). Thus we expect the covariates to contribute largely independent information to the linkage analysis.

Using standard criteria for declared significance of genome-wide linkage analysis (49) (i.e. LOD > 3.6 for significant linkage, LOD > 2.2 for suggestive linkage), linkage analysis with covariates finds three suggestive linkages (Fig. 2): one on chromosome 1 that approaches genome-wide significance (at 210 cM, D1S1660; LOD = 3.46, \( P = 0.00003 \)); another on chromosome 2 (at 114 cM, D2S1790; LOD = 2.22; \( P = 0.00070 \)); and a third region on chromosome 13 (at 26 cM, D13S894; LOD = 2.50; \( P = 0.00035 \)). These results were obtained for different combinations of covariates: for chromosome 1, both EDI-DT and obsessionality, for chromosome 2, only obsessionality, and for chromosome 13, only EDI-DT.

Weighting of the ASPs could not be related to any obvious diagnostic or familial characteristics, regardless of whether the weightings were derived from both covariates jointly or from either covariate alone. When we contrast ASPs receiving weights greater than 0.5 with those receiving weights less than 0.5, there are no significant differences for gender combination of the ASPs (female–female versus male–female), diagnostic category (AN–AN, AN–BN or AN–NOS) or number of affected individuals per family (data not shown). Likewise, selection of ASPs with complete covariate data cannot account for the results. Linkage analysis for the sample of 116 ASPs who have complete data for EDI-DT and obsessionality yields a maximum LOD = 1.52 in the same q1 region as before. For the same sample and weights derived from values of EDI-DT or obsessionality alone, the LOD scores are 1.29 and 1.31 respectively. Thus, insofar as we can discern, the combination of covariates selects for a critical and presumably homogeneous subset of ASPs from the larger sample.

The results in Figure 2 are derived by transforming the \( P \)-values from the original likelihood ratio statistic into LOD scores. The original statistic has a one-sided \( \chi^2 \) distribution, asymptotically, under the null hypothesis of no linkage (35). Thus, under the null hypothesis, 50% of the \( P \)-values equal 0.5, and the remaining \( P \)-values are distributed uniformly between 0 and 0.5. Our results (Fig. 2) roughly meet this expectation, except for a slight excess of small \( P \)-values, which is the desired result. On the other hand, while analysis using either covariate yielded ~52% \( P \)-values at 0.5, only 42% of the linkage analysis using both covariates did likewise. This result is not likely to be significant, and in fact we find an even bigger deviation—in the opposite direction—for standard linkage analyses on the same sample using GeneHunterPlus (50). Nonetheless, the tendency for positive results for two covariates, versus one, could be partially due to the fact that the likelihood ratio statistic for the pre-clustering model is more variable for small samples (35), and the use of additional covariates reduces the effective sample size.

Because chromosome X requires special treatment for linkage analysis compared to the autosomes, and this treatment is not yet coded into a computer program for the pre-clustering model, we took a slightly simpler approach to linkage analysis for this chromosome. We clustered as described previously;
however, the probabilities of group membership were used as weights for whole families, whose data were analyzed by GeneHunterPlus (50). Because GeneHunterPlus’ NPL scores are standard normal deviates, it is straightforward to derive the variance of the sum of weighted deviates.) For both covariates taken together, the best linkage result occurred at GATA172D05 (LOD = 0.93; \( P = 0.039 \)). For either covariate alone, the best linkage results occurred at the same marker, GATA10C11, which is \( \approx 3 \) cM from GATA172D05 on Xq. The LOD score for obsessiosity (LOD = 0.91; \( P = 0.040 \)) was slightly larger than EDI-DT (LOD = 0.59; \( P = 0.099 \)).

In summary, the best linkage signals occurred on the autosomes, and varied by the covariate(s) used in the linkage analysis (Table 1). Nonetheless, linkage signals for different covariates overlap in some regions, such as chromosome 1q31 and chromosome 13q11–12 (Table 1 and Fig. 2). In the 1q region, information on drive-for-thinness and obsessiosity combine to produce an even more compelling linkage signal. Disappointingly, the converse is true for the region of overlap on 13q (Fig. 2).

**DISCUSSION**

We present the results of a genome-wide scan for loci affecting liability to eating disorders. While we use a standard sampling scheme, ASPs and a standard set of genetic markers, our approach deviates from the typical genome scan in several ways. First, because our study is a multinational collaboration and ASPs are drawn from disparate locations, we estimate marker allele differentiation among these populations and account for the differentiation in the linkage analysis by using population-specific allele frequency estimates. To do so, we implement a Bayesian hierarchical model for allele frequencies (45). As might be expected based on the fact that the majority of families are of European ancestry (47), the average differentiation among the populations is not large. Thus, the adjustment for population heterogeneity did not have a major impact on the linkage analysis (unpublished data).

Secondly, we explore various attributes thought to typify individuals with eating disorders to determine if any of them might contribute to the linkage analysis. We identify two variables, EDI-DT and Y-BOCS total score (measuring obsessiosity), which delimit populations among the ASPs in this study, whereas five others show little discriminatory power (Fig. 1). The cluster of ASPs who are extreme and concordant with respect to EDI-DT and obsessiosity have prototypical features for individuals diagnosed with AN.

Thirdly, we combine the information on EDI-DT and obsessiosity with linkage analysis using a new method of multipoint ASP linkage analysis with covariates (35), specifically the pre-clustering model. Based on standard criteria for significant (LOD > 3.6) and suggestive (LOD > 2.2) linkage (49), we find three regions showing suggestive linkage, with the results from one region close to genome-wide significance (LOD = 3.46). It is worth noting that these criteria are based on complete, pointwise IBD information; for partial information on identical by descent (IBD), such as that obtained by this genome scan, the criterion for significant linkage is lower (51,52).

Our results are substantially better than the results using more standard methods of analysis. For example, when GeneHunter (53) was used to analyze the entire data set reported herein, additional families with ARPs and some larger, multiplex families, the three best \( P \)-values were 0.036, 0.078 and 0.109 (54), which pale in comparison to the results in Figure 2 for any combination of covariates. Likewise, for linkage analysis using all ASPs and our model, setting weights for all families equal 0.99; the best \( P \)-values are 0.0010, 0.0076 and 0.0222. Analysis of 38 ASPs who are concordant for AN, restrictor subtype, produces a finding of suggestive linkage on chromosome 1, but quite far from our best signal (54). This suggestive linkage signal \( (P = 0.00028) \) at \( \approx 65 \) cM, obtained for a dense grid of short tandem repeat (STR) markers across the region, corresponds to a weaker signal on chromosome 1 for the covariate analyses (Fig. 2). Individuals with AN, restricting subtype, typically have high EDI-DT and obsessiosity. As a group, they are well known for their highly homogeneous presentation of both illness and behavioral characteristics, something we have also observed in our data (55).

Why do the covariates convey information regarding linkage? At this time we can only speculate about ultimate causes, but we assume that drive-for-thinness and obsessiosity differentiate both eating disorder patients themselves and the genetic loci affecting eating disorder liability. It has long been observed that certain features typify eating disorder patients, such as obsession with order, symmetry and exactness, rigidity, perfectionism, anxiety, and emotional and behavioral overcontrol. Yet not all individuals display these features, and diagnostic categories like AN and BN cannot explain the wide variety of behavioral phenotypes observed in eating disorders.
or vice versa (55). Thus, as befits a general model for a complex disorder, it would not be surprising to find that various loci contribute in different ways to eating disorder liability, and that those loci also affect other traits, such as compulsive drive-for-thinness and obsessionality. In this regard, the use of covariates to amplify linkage signals is becoming increasingly commonplace (56–59).

Given the suggestive linkage findings, it is reasonable to ask if any of these regions harbor excellent candidate genes for liability to eating disorders. Like many psychiatric disorders, the origins of eating disorders are unclear. Nonetheless, ample research on humans and model organisms (60–71) points to serotonergic dysregulation as at least one important cause. In light of our results, recent brain imaging studies are especially intriguing because they suggest that eating disorders and obsessive-compulsive disorder (OCD) share frontal-cingulate-temporal-subcortical pathways served, in part, by serotonin neuronal systems (69,70). The common phenotypic characteristics shared by many AN and OCD patients suggest that these disorders may share common brain behavioral pathways; however, the lack of complete overlap suggests that they have different foci of pathology within those pathways. Thus, we believe that genes involved in serotonergic regulation are excellent candidates for liability to eating disorders, and are plausible targets for follow-up analyses.

We plan to continue our research to identify genes affecting eating disorder liability in two directions. We will evaluate candidate genes falling in the regions revealed by our linkage analyses (Fig. 2), with emphasis on genes having a direct impact on serotonin regulation. At the same time, we will begin analyzing a new cohort of almost 400 eating disorder families. While the new cohort will have some different properties to the first cohort, largely because families were accessed through a bulimic rather than anorexic proband, we expect that covariates common to both cohorts can be used to extract information about etiology in much the same way as the analyses reported here. Through common covariates, these two samples should be a powerful means to refine our search for polymorphisms conferring risk for eating disorders.

MATERIALS AND METHODS

Participants and family structure

In a previous report (34), we gave a detailed description of the sample to be used in our analyses. In brief, 196 probands meeting modified diagnostic and statistical manual of mental disorders-fourth edition (DSM-IV) criteria (amenorrhea not required) for AN, and 237 affected relatives with modified DSM-IV AN, DSM-IV BN or DSM-IV NOS, were recruited from seven sites across North America and Europe: Pittsburgh, New York, Los Angeles, Toronto, London, Munich and Philadelphia. Because we restricted our analyses to ASPs, our initial sample included 176 families, 61% of which included sibling pairs who had AN. The remaining 39% included AN–BN (19%) and AN–NOS (20%) pairs. Of the 176 families with ASPs, 14 had more than two affected siblings. Of these, nine had three siblings diagnosed with eating disorders and five had four siblings diagnosed with eating disorders. Most families consisted of female ASPs; only four ASPs were of opposite gender.

Behavioral and personality measures

Below we give a brief description of the measures. Please see our previous report (34) for more details. All probands and affected relatives were assessed for these measures.

Temperament and Character Inventory. The personality constructs harm avoidance (i.e. the tendency to inhibit behavior to avoid punishment or negative consequences) and novelty seeking (i.e. the tendency toward behavioral activation to pursue rewards) were assessed by the 240-item TCI Version 9 (36).

State Trait Anxiety Inventory. This was used to evaluate trait anxiety (37).

Yale–Brown Obsessive Compulsive Scale. We used the Y-BOCS (39) to assess participants’ ‘worst lifetime’ obsessions and compulsions in order to capture lifetime symptoms of greatest severity.

Yale–Brown–Cornell Eating Disorder Scale. Through the YBCEDS (40), participants were asked to report their ‘worst lifetime’ obsessions and compulsions related to eating disorders.

Multidimensional Perfectionism Scale. The MPS (38) measured overall levels of perfectionism; ‘worst lifetime’ symptom expression was used.

Eating Disorder Inventory-2. The EDI-2 (41) measured the drive-for-thinness (i.e. preoccupation with dieting and the pursuit of thinness), using ‘worst lifetime’ expression.

We transformed these trait values to standard normal deviates using their corresponding distributions from normative populations, subtracting the observed means and dividing by the observed SD from the normative population. Normative values for harm avoidance and novelty seeking (72), perfectionism (43), trait anxiety (37), drive-for-thinness (41), total Y-BOCS score (73) and total YBC-EDS score (40) were reported previously.

Genotyping methods

Genomic DNA was extracted from leucocytes using standard salting procedures (74). DNAs were genotyped for the Weber/CHLC Screening Set 9 (http://research.marshfieldclinic.org/genetics/), which consists primarily of tri- and tetranucleotide repeat markers separated by, on average, ~10 cM (75). All 387 markers were fluorescently tagged for use with the ABI sequencer.

Statistical analyses

Error checking. We evaluated markers and pedigrees for Mendelian errors using the PedCheck program (76). Genotyping errors were set to missing. Nominal and imputed genetic relationships among individuals from the same family were contrasted using the Relpair program (77). We excluded from further analysis three families who appeared to be composed of genetic half-siblings and three families with ambiguous relationships among parents and progeny.

Allele frequency estimation. To estimate and account for heterogeneity of marker allele frequencies across the recruitment sites, we used a Bayesian hierarchical model (45) and implemented in the program AllDist (http://wpicr.wpic.pitt.edu/WPICCompGen). To estimate population-specific allele frequency distributions, the observed allele frequencies, computed by counting genotypes over all individuals from
each site, were ‘shrunk’ toward the distribution obtained by pooling subpopulations and toward a prior distribution. For a prior distribution, we used the CEPH allele distributions available from the Mammalian Genotyping Service (http://research.marshfieldclinic.org/genetics/). Using this model, the level of shrinkage was determined by the data for all STR loci and prior information on the amount of subpopulation divergence. Because most participants identified themselves as being of European ancestry, and a few of mixed European and Asian or Amerindian ancestry, we used 0.003 as an estimate of population divergence (78). See http://wpicr.wpic.pitt.edu/ WPICCompGen for population-specific allele frequency estimates.

**Linkage analysis.** To incorporate covariate information into the linkage analysis, we implemented a new method of analysis (35). In brief, ASP data were modeled as a mixture distribution, which assumes a disease mutation is segregating in only a fraction $\alpha$ of the sibships, with $1 - \alpha$ sibships being unlinked. Covariate information is used to predict membership within groups. For an ASP with covariate(s) $Z = z$ and multilocus genotype $X = x$, the mixture model is $g(zg(x;\lambda) + [1 - g(z)]g_0(x)$, in which $g_0(x)$ follows the distribution of genotypes under the null IBD distribution and $g(x;\alpha)$ allows for increased IBD sharing. We used the pre-clustering version of this model (35), which uses only covariate information to assign weights of membership in the linked and unlinked groups for each ASP prior to the linkage analysis. Once these weights are determined, they are combined with multipoint IBD estimates, such as those from standard linkage packages [e.g. GENEHUNTER (53), Aspex (79)], to test for linkage. As the test statistic for linkage, we used a likelihood ratio, which has an asymptotic, one-sided $\chi^2$ limiting distribution for the pre-clustering method under the null hypothesis of no linkage (35).

Fourteen families had more than two affected siblings. For these families, we formed all possible ASPs for our analyses after determining the joint IBD status using GENEHUNTER (53) and after determining the probability of membership in the linked group. The latter was calculated as described elsewhere (35). We did not correct for these dependent observations because they have little impact on the distribution of the test statistic under the null hypothesis (80).

**ACKNOWLEDGEMENTS**

The authors are indebted to the participating families for their contribution of time and effort in support of this study. The authors wish to thank the Price Foundation for the support of the clinical collection of subjects and genotyping, and contribution to the support of data analysis. Data analysis was also supported by a grant from the National Institute of Health (MH57881; B.D.).

**REFERENCES**


