Brief Research Communication

A Novel Non-Parametric Regression Reveals Linkage on Chromosome 4 for the Number of Externalizing Symptoms in Sib-Pairs


1Human Genetics Unit, Indian Statistical Institute, Kolkata, India
2Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri
3Department of Psychiatry, SUNY HSC at Brooklyn, Brooklyn, New York
4Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana
5Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana
6Department of Psychiatry, University of Connecticut School of Medicine, Farmington, Connecticut
7Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana
8Department of Psychiatry, University of Iowa, Iowa City, Iowa
9Department of Biostatistics, Washington University School of Medicine, St. Louis, Missouri

In this report, we present results of a genome-wide linkage scan using as a phenotype the number of externalizing symptoms associated with alcohol use disorders. Subjects were collected by the Collaborative Study on the Genetics of Alcoholism project from families in which at least three first degree relatives were affected by alcohol dependence. We use a novel non-parametric regression method based on kernel smoothing for our analysis. We report a statistically significant linkage close to the ADH gene cluster on Chromosome 4. We also obtain evidence for epistatic interaction between a region on Chromosome 1 and one on Chromosome 15. Although alcoholism as a covariate does not have any effect on the linkage scan, it has an effect on the epistatic interaction. © 2008 Wiley-Liss, Inc.

KEY WORDS: non-parametric regression; linkage; externalizing symptoms


Alcohol dependence is a complex disorder that is influenced by genetic and environmental factors. A well-known link exists between externalizing behavior disorders such as conduct disorder and alcohol use disorders. For example, adults with alcohol dependence report an increased rate of externalizing behaviors [Cloninger et al., 1981; Hesselbrock et al., 1988; Babor et al., 1982]. In addition, children and adolescents who manifested conduct disorder are at heightened risk for the subsequent development of alcohol problems [Robins, 1966; Myers et al., 1993]. Twin studies have demonstrated that shared genetic factors are largely responsible for much of the co-occurrence childhood conduct disorders and adult alcohol dependence [Kendler et al., 1997; Slutske et al., 1998]. This relationship between externalizing symptoms and alcoholism is also seen in the Collaborative Study on the Genetics of Alcoholism (COGA) dataset. A higher prevalence of alcohol dependence and more severe alcohol problems are correlated with the most disruptive behavior [Bucholz et al., 2000].

The Collaborative Study on the Genetics of Alcoholism is a multicenter research program established to detect and map susceptibility genes for alcohol dependence and related phenotypes. Studies have revealed that many regions of the genome exhibit significant linkage with alcohol dependence [Reich et al., 1998; Foroud et al., 2000] as well as alcohol-related phenotypes such as a severity phenotype [Foroud et al., 1998], maximum number of drinks in a 24 hr period [Reich et al., 2000], a combined phenotype of alcohol dependence or depression [Nurnberger et al., 2001] and habitual smoking in families of alcoholics [Bierut et al., 2004]. In addition, electro-physiological phenotypes related to alcoholism like event-related potentials (ERP) [Begleiter et al., 1998; Ainsley et al., 2001; Porjesz et al., 2002b], electroencephalogram (EEG) waves [Porjesz et al., 2002a; Ghosh et al., 2003], event-related oscillations (ERO) [Jones et al., 2004] and a bivariate phenotype combining alcohol dependence and ERP [Williams et al., 1999] have also shown evidence of linkage on multiple chromosomes.

The COGA project systematically ascertainment through probands from alcoholism treatment facilities [Begleiter et al., 1998; Reich, 1998]. COGA sites providing data for this study include SUNY Downstate Medical Center, New York; University of Connecticut Health Science Center; Indiana University School of Medicine, University of Iowa School of Medicine, University of California School of Medicine, San Diego and Washington University School of Medicine, St. Louis. The data collection procedures were standardized at all sites [Begleiter et al., 1998]. The extensive SSAGA interview instrument was developed to obtain detailed information on symptoms related

Grant sponsor: National Institute on Alcohol Abuse and Alcoholism (NIAAA); Grant number: U10AA08403; Grant sponsor: Fogarty International Center; Grant number: R01-TW-6004.

Correspondence to: Dr. Saurabh Ghosh, Human Genetics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700 108, India. E-mail: saurabh@isical.ac.in

Received 11 August 2006; Accepted 22 January 2008
DOI 10.1002/ajmg.b.30783
Published online 2 May 2008 in Wiley InterScience (www.interscience.wiley.com)

© 2008 Wiley-Liss, Inc.
to alcohol use, abuse and dependence, other substance abuse and dependence, and different psychiatric conditions [Bucholz et al., 1994, 1995; Hasselbrock et al., 1999]. The genetic dataset comprises 262 extended families in which at least three members met the COGA (DSM-III-R and Feighner [Feighner et al., 1972]) criterion of alcohol dependence. Data on externalizing symptoms were available for 1,553 subjects ranging from 7 to 70 years of age.

Genotyping was performed at laboratories in Indiana University and in Washington University [Foroud et al., 2000]. For this study, a map comprising 405 markers, mostly microsatellites, was used. The average intermarker distance and the average heterozygosity of the markers were 10.9 cM and 0.74, respectively, although the marker coverage in some regions was denser than 5 cM. Most markers were tri- and tetranucleotide repeat polymorphisms. In order to be consistent and be able to compare with previous linkage studies performed in the COGA project, the program CRI-MAP [Lander and Green, 1987] was used to calculate marker order and distances. The maximum likelihood estimates of the allele frequencies were obtained using USERM/M [Boehnke, 1991].

Data on three covariates were used in the analyses: age, sex and diagnosis of alcoholism by the COGA criteria noted above. Our analysis was based on 171 independent sib-pairs belonging to distinct families and having at least one parent genotyped [i.e., sibs with both parents missing were excluded from the analysis for obtaining more precise estimates of identity-by-descent (i.b.d.) scores]. To ensure independence of observations, only one sib-pair was chosen from each family (the sib-pair with the maximum number of marker genotypes available for both sibs). The sample comprises 49 male–male pairs, 39 male–female pairs and 33 female–female pairs. Most of the families (148, 86.5%) were Caucasian while 15 families (9%) were African-American.

A set of 24 externalizing symptoms (see Appendix A) related to anti-social behavioral traits was chosen. The number of these symptoms endorsed by each individual was summed to obtain a quantitative scale ranging from 0 to 24. In this study, we present results on a genome scan of the number of externalizing symptoms. Since there are gender and age differences with externalizing behaviors, we regress out the effects of sex and age from the phenotypic values using least squares linear regression. We analyze the data both with and without regressing out the effect of alcoholism.

The probability distribution of the number of externalizing symptoms is very skewed (see Fig. 1) and hence deviant from normal. A possible better model fit is a truncated Poisson distribution. Thus, using likelihood based linkage methods which assume normality of traits may lead to misleading inferences. The classical Haseman-Elston regression [1972] and its extensions do not assume any specific probability distribution for the trait but assume a linear relationship between a function of the sib-pair trait values and the identity-by-descent scores at a marker locus. Since this relationship deteriorates with increase in dominance at the trait locus, a more prudent strategy is to estimate empirically the nature of the relationship between the regression variables [Ghosh and Majumder, 2000]. We use a modified version of the non-parametric regression proposed by Ghosh et al. [2003]. The major advantage of the non-parametric regression is that it does not involve any likelihood based framework which requires explicit specification of the probability distribution of the quantitative trait values, and hence is robust to violations in distributional assumptions like normality. The details of the non-parametric regression are provided in Appendix B.

We performed the proposed non-parametric regression at 1 cM intervals throughout the genome. The only statistically significant (nominal P-value < 0.0001, a conservative correction equivalent to 100 independent tests would give genome-wide significance < 0.01) peak (when alcoholism is not regressed out) is located at 120 cM on Chromosome 4 (see Fig. 2), near a marker in the ADH1C gene (124.9 cM). When the effect of alcoholism is regressed out as a covariate, we find that the linkage finding is hardly altered. The position of the peak is almost identical (within 3 cM) to that obtained when the effect of alcoholism is not regressed out. Moreover, the P-value corresponding to the peak is only changed marginally.

There has been some evidence in previous COGA studies that a region (near 100 cM) on Chromosome 1 and a region on Chromosome 4 (around 57 cM) interact epistatically with a region (near 50 cM) on Chromosome 15 [Ghosh et al., 2003] for different alcohol-related phenotypes. Although the only statistically significant linkage was on Chromosome 4, recent theoretical and simulation studies [Culverhouse et al., 2002] have shown that epistatic effects may be present even in the absence of main effects. Hence we investigated the presence of epistasis between two pairs of regions: (i) 95–105 cM on Chromosome 1 and 45–65 cM on Chromosome 15; and (ii) 52–82 cM on Chromosome 4 and 45–55 cM on Chromosome 15. The details of the procedure for testing the presence of epistatic interactions are provided in Appendix B.

![Fig. 1](image1.png)

Fig. 1. Relative frequencies of the number of externalizing symptoms; N = number of symptoms; f = relative frequency.

![Fig. 2](image2.png)

Fig. 2. Linkage results on Chromosome 4 for the phenotype defined as the number of externalizing symptoms.
Linkage Using Non-Parametric Regression

When alcoholism is not regressed out, we observe that the epistatic interaction: between each of the integer points on Chromosome 1 (95–105 cM) and each of the integer points on Chromosome 18 (45–55 cM) is significant at $P < 0.01$ (adjusted for multiple tests). In all these cases, epistasis (in terms of increase in $\Delta$ for Model II over Model I) explains about 8% of the variance in the squared difference in sub-pair externalizing symptoms values. However, none of these interactions are significant at $P < 0.01$ when the effect of alcoholism is regressed out. There was no evidence of epistatic interaction between the above mentioned regions on Chromosomes 4 and 15 either with or without regressing out the effect of alcoholism. We note here that since our linkage finding is around 120 cM on Chromosome 4, we tested for possible epistatic interactions between that region and the above mentioned regions on Chromosomes 1 and 15. However, none of these were statistically significant.

The linkage findings on Chromosome 4 is in the q42.3 region. This region contains the alcohol dehydrogenase gene cluster (ADHA1, ADHB, ADHIC, ADH4–ADH7). Variants in the class I ADH genes affect enzyme kinetic properties and have been reported to reduce the risk of alcoholism in Chinese and Japanese populations [Edenberg and Boaron, 1997]. This region has produced a linkage finding with alcohol dependence in a Mexican American population (unpublished data). Previous COGA studies have also found evidence of linkage in this region with various alcohol-related phenotypes. Reich et al. [1998] obtained suggestive linkage evidence of a protective factor in the risk of alcoholism among unaffected sibs of alcoholic probands. A phenotype defined as the maximum number of drinks consumed in a 24 hr period [Saconce et al., 2000] which provides a quantitative measure to grade moderate drinkers and is closely related to alcoholism diagnosis also produced a linkage peak in this region. In individuals at risk for alcoholism, the amplitude of the visual P300 component of event-related potentials (ERP) is significantly decreased [Porjesz et al., 1998]. Williams et al. [1999] analyzed a bivariate phenotype combining DSM-IV diagnosis and the amplitude of the P300 component of the Cz event-related potential. Their linkage scan gave a significant finding near the class I alcohol dehydrogenase subunit ADHIC.

Oscillations within the electroencephalogram (EEG) gamma and beta bands have been linked to sensory perception and memory and have been shown to be modified by anesthetic agents. Since these oscillations appear to be trait as opposed to state-related, Ghosh et al. [2003] used Beta 2 EEG waves as an endophenotype for studying the risk of developing alcohol dependence. They obtained a linkage signal in the same region on Chromosome 4.

We have previously conducted a genome-wide linkage scan of conduct disorder diagnoses in the COGA dataset, using standard non-parametric, multipoint methods of linkage analyses (e.g., as implemented in Asperx) for affected sibling pairs [Dick et al., 2004]. The chromosomal regions yielding evidence of linkage in that study (on chromosomes 2 and 19) were not overlapping with the significant linkage peak from this study. We note here that the conduct disorder variables contained in the set of externalizing symptoms are identical to the SSAGA diagnosis protocol. Accordingly, our data illustrate that different findings can emerge when analyzing phenotypes capturing different information, and when utilizing different methods of analysis (that is, methods for analyzing quantitative traits versus binary traits).

The sample comprised Caucasian as well as African-American families. While population stratification cannot result in an inflated rate of false positives in sub-pair linkage analyses, it may lead to false negatives. To investigate this possibility, we also performed our linkage scan on the subset of 148 Caucasian sib-pairs, but found that the linkage findings were identical. Although we used CRI-MAP for the purpose of consistency with previous COGA studies, we note that the Decode map is estimated from a larger number of meioses and hence, likely to be more precise. We found that the COGA map was marginally expanded compared to the Decode map, but the map orders were identical. When we repeated our linkage analyses with the Decode map, the results remained unchanged with the only linkage finding being in the same region on Chromosome 4, albeit with a marginally lower significance level. We also realized that larger sibships provide more information on linkage than independent sub-pairs. However, a direct extension of the proposed non-parametric method to sibships involves modeling of the covariance structure between the squared phenotype difference of different sib-pairs within a sibship, which compromises on the model-free framework of the method. We are currently exploring a non-parametric modification of the linear regression procedure proposed by Ghosh and Reich [2002] based on a so-called "contrast function" of the quantitative trait values of the sibs within each sibship.

As discussed earlier, a possible parametric probability model for the phenotype is Poisson. Hence, it may be more appropriate to perform a Poisson regression instead of a linear regression while eliminating the effects of covariates.

Previous COGA studies have also found evidence of linkage in this region with various alcohol-related phenotypes. Reich et al. [1998] obtained suggestive linkage evidence of a protective factor in the risk of alcoholism among unaffected sibs of alcoholic probands. A phenotype defined as the maximum number of drinks consumed in a 24 hr period [Saconce et al., 2000] which provides a quantitative measure to grade moderate drinkers and is closely related to alcoholism diagnosis also produced a linkage peak in this region. In individuals at risk for alcoholism, the amplitude of the visual P300 component of event-related potentials (ERP) is significantly decreased [Porjesz et al., 1998]. Williams et al. [1999] analyzed a bivariate phenotype combining DSM-IV diagnosis and the amplitude of the P300 component of the Cz event-related potential. Their linkage scan gave a significant finding near the class I alcohol dehydrogenase subunit ADHIC.

Oscillations within the electroencephalogram (EEG) gamma and beta bands have been linked to sensory perception and memory and have been shown to be modified by anesthetic agents. Since these oscillations appear to be trait as opposed to state-related, Ghosh et al. [2003] used Beta 2 EEG waves as an endophenotype for studying the risk of developing alcohol dependence. They obtained a linkage signal in the same region on Chromosome 4.

We have previously conducted a genome-wide linkage scan of conduct disorder diagnoses in the COGA dataset, using standard non-parametric, multipoint methods of linkage analyses (e.g., as implemented in Asperx) for affected sibling pairs [Dick et al., 2004]. The chromosomal regions yielding evidence of linkage in that study (on chromosomes 2 and 19) were not overlapping with the significant linkage peak from this study. We note here that the conduct disorder variables contained in the set of externalizing symptoms are identical to the SSAGA diagnosis protocol. Accordingly, our data illustrate that different findings can emerge when analyzing phenotypes capturing different information, and when utilizing different methods of analysis (that is, methods for analyzing quantitative traits versus binary traits).

The sample comprised Caucasian as well as African-American families. While population stratification cannot result in an inflated rate of false positives in sub-pair linkage analyses, it may lead to false negatives. To investigate this possibility, we also performed our linkage scan on the subset of 148 Caucasian sib-pairs, but found that the linkage findings were identical. Although we used CRI-MAP for the purpose of consistency with previous COGA studies, we note that the Decode map is estimated from a larger number of meioses and hence, likely to be more precise. We found that the COGA map was marginally expanded compared to the Decode map, but the map orders were identical. When we repeated our linkage analyses with the Decode map, the results remained unchanged with the only linkage finding being in the same region on Chromosome 4, albeit with a marginally lower significance level. We also realized that larger sibships provide more information on linkage than independent sib-pairs. However, a direct extension of the proposed non-parametric method to sibships involves modeling of the covariance structure between the squared phenotype difference of different sib-pairs within a sibship, which compromises on the model-free framework of the method. We are currently exploring a non-parametric modification of the linear regression procedure proposed by Ghosh and Reich [2002] based on a so-called "contrast function" of the quantitative trait values of the sibs within each sibship.

As discussed earlier, a possible parametric probability model for the phenotype is Poisson. Hence, it may be more appropriate to perform a Poisson regression instead of a linear regression while eliminating the effects of covariates.
ACKNOWLEDGMENTS

The Collaborative Study on the Genetics of Alcoholism (COGA) Co-Principal Investigators: L. Bierut, H. Edenberg, V. Hesselbrock, B. Porjesz) includes nine different centers where data collection, analysis, and storage take place. The nine sites and Principal Investigators and Co-Investigators are: University of Connecticut (V. Hesselbrock); Indiana University (H. Edenberg, J. Nurnberger Jr., P.M. Conneally, T. Foroud); University of Iowa (R. Crowe, S. Kuperman); SUNY HSCB (B. Porjesz); Washington University in St. Louis (L. Bierut, J. Rice, A. Gottle); University of California at San Diego (M. Schuckit); Howard University (R. Taylor); Rutgers University (J. Tischfield); Southwest Foundation (L. Almasy). Zhaoxun Ren serves as the NIAAA Staff Collaborator. This national collaborative study is supported by the NIH Grant U10AA08403 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA). This work has also been supported by the NIH grant R01-TW-6604 from the Fogarty International Center. The authors are grateful to Professor Partha Majumder for useful statistical inputs and to Joseph Mallaney for management of the data on externalizing symptoms. In memory of Henri Begleiter and Theodore Reich, Principal and Co-Principal Investigators of COGA since its inception and one of the founders of modern psychiatric genetics, we acknowledge his immeasurable and fundamental scientific contributions to COGA and the field.

APPENDIX A

List of the Externalizing Symptoms

1 = 'hit or throw things while drinking'  
2 = 'physical fights while drinking'  
3 = 'play hooky from school (truancy)'  
4 = 'suspended or expelled from school'  
5 = 'run away from home overnight'  
6 = 'tell many lies or used alias'  
7 = 'steal (3+) from family/home'  
8 = 'shoplift (3+) from store/other people'  
9 = 'stolen/forged (3+) since age 15'  
10 = 'vandalism'  
11 = 'start fights'  
12 = 'get into fights (3+) since age 15'  
13 = '4 or more traffic tickets'  
14 = 'non-traffic arrests'  
15 = 'jail for something other than drugs/alcohol'  
16 = 'challenge parents when young'  
17 = 'enjoy outsmarting authority'  
18 = 'breaking/entering into car/house'  
19 = 'fencing/selling/run numbers since age 15'  
20 = 'hit/physic attack/throw things since age 15'  
21 = 'quit, no job, new job lined up (3+) since age 15'  
22 = 'ignored feelings of others since age 15'  
23 = 'frequently lost temper or often felt irritable/anxious/resentful  
24 = 'often blame others for your troubles'  

APPENDIX B

Non-Parametric Regression

The non-parametric regression model of $y_j$ on $n_{ijp}$ at any arbitrary point $p$ on the genome is given by

$$y_j = \Psi(n_{ijp}) + \epsilon_j, j = 1, 2, \ldots, 171$$

where $y_j$ is the squared difference in the number of externalizing symptoms value for the $j$th sib-pair, $n_{ijp}$ is the multipoint estimated i.b.d. score at a point $p$ on the genome for the $j$th sib-pair computed using the linear regression method proposed by Fulker et al. [1996] and $\Psi$ is a real valued function of $n_{ijp}$ and $\epsilon_j$ are random errors.

The functional form of $\Psi$ is estimated using a kernel-smoothing technique [Silverman, 1986]. The kernel function used is

$$k(t) = \frac{1}{2} (1 - t^2)^2 \quad \text{if } |t|<1$$
$$0 \quad \text{otherwise}$$

The predictor of $y_j$ is given by

$$\frac{\sum_i k((n_{ijp} - n_{ijq})/h) y_i}{\sum_i k((n_{ijp} - n_{ijq})/h)}$$

where $h$ is the "optimal" window length in the kernel smoothing procedure.

To assess the significance of our regression, we use a diagnostic measure [Ghosh et al., 2003] as follows: $\Lambda = 1 - \sum y_i - \Psi(n_{ijp})^2/\sum y_i - y_{mean}^2$ where $y_{mean}$ is the mean of the $y$s. One has to use resampling techniques such as bootstrap to obtain empirical thresholds under the null hypothesis of no linkage.

Testing for Epistatic Interactions

We consider two models:

Model I: $y_j = \Psi_1(n_{ijp}) + \Psi_2(n_{ijq}) + \epsilon_j, j = 1, 2, \ldots, 171$

Model II: $y_j = \Psi_1(n_{ijp}) + \Psi_2(n_{ijq}) + \Psi_3(n_{ijp}n_{ijq}) + \epsilon_j, j = 1, 2, \ldots, 171$

where $n_{ijp}$ and $n_{ijq}$ are estimated marker i.b.d. scores at two positions on the genome, $p1$ and $p2$, respectively, and $\Psi_1$, $\Psi_2$, and $\Psi_3$ are real-valued functions, which are estimated sequentially using kernel smoothing [see Ghosh and Majumder [2000] for details] based on the kernel function given in Equation (1). A significant increase in $\Lambda$ for Model II over Model I would indicate the presence of epistatic interaction. Empirical null thresholds are determined using resampling techniques.

REFERENCES


Linkage Using Non-Parametric Regression 1305


