A Genome-Wide Linkage Analysis for the Personality Trait Neuroticism in the Irish Affected Sib-Pair Study of Alcohol Dependence

Po-Hsiu Kuo,1* Michael C. Neale,1 Brien P. Riley,1 Diana G. Patterson,2 Dermot Walsh,3 Carol A. Prescott,4 and Kenneth S. Kendler1
1Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia
2Shaftsbury Square Hospital, Belfast, Northern Ireland
3Health Research Board, Dublin, Ireland
4Department of Psychology, University of Southern California, Los Angeles, California

Neuroticism is a personality trait which reflects individual differences in emotional stability and vulnerability to stress and anxiety. Consistent evidence shows substantial genetic influences on variation in this trait. The present study seeks to identify regions containing susceptibility loci for neuroticism using a selected sib-pair sample from Ireland. Using Merlin regress, we conducted a 4 cM whole-genome linkage analysis on 714 sibpairs. Evidence for linkage to neuroticism was found on chromosomes 11p, 12q, and 15q. The highest linkage peak was on 12q at marker D12S1638 with a Lod score of 2.13 \( (\log p = 2.76, \text{ empirical } P\text{-value } < 0.001) \). Our data also support gender specific loci for neuroticism, with male specific linkage regions on chromosomes 1, 4, 11, 12, 15, 16, and 22, and female specific linkage regions on chromosomes 2, 4, 9, 12, 13, and 18. Some genome regions reported in the present study replicate findings from previous linkage studies of neuroticism. These results, together with prior studies, indicate several potential regions for quantitative trait loci for neuroticism that warrant further study. © 2007 Wiley-Liss, Inc.

KEY WORDS: negative affect; emotional stability; internalizing; quantitative genetics; susceptibility loci


INTRODUCTION

Neuroticism (N), a personality trait which measures individual differences in emotional stability, has long been linked to risk for anxiety, depression, and other forms of psychopathology. Many studies suggest that N predicts the development of a range of internalizing disorders [Martin et al., 1988; Kendler et al., 1993; Clark et al., 1994; Muris and Ollendick, 2005]. Furthermore, the substantial comorbidity seen between internalizing disorders, especially between anxiety and depressive disorders [Middeldorp et al., 2005], may be largely explained by N. Using structural equation modeling, Khan et al. [2005] determined that N alone could account for 20–45% of the covariance between pairs of internalizing disorders and 19–88% of the comorbidity between pairs of internalizing and externalizing disorders.

The association between N and risk for psychopathology appears to be in part mediated by shared genetic factors. Two studies have consistently documented heritability estimates for N in the range of 30–50% [Bouchard and Loehlin, 2001], similar in magnitude to that seen for many common complex disorders. In addition, twin studies using general population samples suggest that there are common sets of genes underlying N and both major depression and anxiety disorders [Jardine and Martin, 1984; Jardine et al., 1984; Kendler et al., 1993, 2006; Fanous et al., 2002]. For instance, there is a positive genetic correlation between N and major depression which has been estimated to be 0.41–0.68 [Fanous et al., 2002; Kendler et al., 2006]. Recent evidence has documented substantial sharing of genetic risk factors for N and internalizing disorders [Middeldorp et al., 2005; Hettema et al., 2006]. Because N is a quantitative trait and easily assessed, it may be a useful phenotype for genetic analyses and a potentially important endophenotype for mood and anxiety disorders. Thus elucidating the genetic substrate of N could not only clarify the genetic basis of personality but may also shed light on the genetic nature of complex psychiatric disorders such as major depression and generalized anxiety disorder.

Women have been consistently shown to have higher mean levels of N than men [Jardine et al., 1984; Jorm, 1987]. Sex-specific genetic factors for N have been reported in some studies [Eaves et al., 1998; Martin et al., 2000; Fanous et al., 2002], but not in others [Lake et al., 2000]. In addition, N declines modestly with age [McCrae et al., 1996], especially for women [Viken et al., 1994]. One study reported genetic effects become more pronounced with age in females [Jardine et al., 1984], while another found that genetic influences on N were relatively stable after age 30 [Viken et al., 1994].

Although substantial evidence has accumulated that genetic factors are etiologically important in N, we are aware of only three previous studies reporting genome-wide linkage results. Two studies used an extreme selected sib-pair design to increase power for linkage analysis in general population samples. One of them found significant linkage to regions on chromosomes 1p, 4q, 7p, 12q, and 13q with suggestive linkage.

Grant sponsor: NIH; Grant number: MH 65322.
*Correspondence to: Po-Hsiu Kuo, Ph.D., Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, P.O. Box 980126, Richmond, VA 23298-0126.
E-mail: pkuo@vcu.edu
Received 31 July 2006; Accepted 30 October 2006
DOI 10.1002/ajmg.b.30478

© 2007 Wiley-Liss, Inc.
to 11q [Fullerton et al., 2003]. In the same study, they also reported sex-specific linkage regions for N. The second study reported suggestive evidence for linkage on N on chromosomes 1p and 1q [Nash et al., 2004]. The third study, based on a sample selected for nicotine dependence showed suggestive evidence for linkage on 1p and 11q, replicating previous findings, and reported a new region on 12p [Neale et al., 2005].

In this article, we aim to replicate previous linkage findings and further localize susceptibility loci for N. We report results of a genome-wide linkage study for N based on a large sample of sibling pairs with alcohol dependence collected in Ireland. Given prior evidence of potential differences in genetic risk factors for N in men and women, we also attempt to determine whether there are sex specific loci for N in our sample.

MATERIALS AND METHODS

Subjects

Participants in this study were recruited in Ireland and Northern Ireland between 1998 and 2002. More details of the study design, sample ascertainment, and clinical characteristics are described elsewhere [Prescott et al., 2005]. In brief, ascertainment of probands was mainly conducted in community alcoholism treatment facilities and public and private hospitals. Probands were eligible for study inclusion if they met the DSM-IV (1994, APA) criteria for alcohol dependence and if all four grandparents were born in Ireland, Northern Ireland, Scotland, Wales or England. After a prospective family was identified through probands, parents and potentially affected siblings whom the probands provided permission to contact were recruited. We attempted to enroll all living, biological parents. Probands, siblings and parents were all interviewed by clinically trained research interviewers.

Phenotype Assessment

While our evaluations focused on alcohol-related phenotypes, we also assessed lifetime history of other comorbid conditions, alcohol-related traits, and self-reported personality features, including N, extraversion, and novelty seeking. N was measured by the short-form Eysenck Personality Questionnaire, consisted of 12 yes–no items [Eysenck et al., 1985]. We summed these items to obtain a sum score of N for each individual.

There were 1248 individuals from 591 families who completed the interview. Age in this sample ranged from 17 to 71 years with mean (SD) of 41.9 (9.7). Of these, 743 men (59.5%) and 420 women (33.7%) responded to the N items in the present study. Among these individuals, two did not answer more than half of the scale items and were deleted from analysis; 23 people had one to four missing items, for them we used their mean score from the items answered and rescaled to a sum score based on 12 items. Due to the selection scheme for alcohol dependence in our sample, we noted that the distribution of N was negatively skewed, and the mean (7.9) was higher than that reported for general population samples (mean = 5.0 in men and 5.9 in women) [Eysenck and Eysenck, 1991]. In addition, there was no significant gender difference for N sum score; the mean (SD) for men was 7.8 (3.7) and for women was 8.2 (3.6). Because age was significantly and negatively correlated with N (r = –0.16, P < 0.0001), we regressed the N sum score on age and age-squared. The distribution of the studentized residual was still negatively skewed (skewness = –0.6, kurtosis = –0.74) so we normalized it using the RANK procedure in SAS [SAS, 2002], which produced a distribution with skewness = 0 and kurtosis = –0.05.

Laboratory Methods

Details of the laboratory methods and error-checking are provided elsewhere [Prescott et al., 2006]. DNA was extracted using standard techniques. A 1020-marker autosomal genomewide screen was conducted by deCODE Genetics (http://www.decode.com/) using their standard panel of polymorphic microsatellite markers. The average marker spacing was 4 cm, and the average heterozygosity in our sample was 72.5% (min 6.3%, max 91.8). Genotyping reliability was assessed in 17 individuals for whom we submitted duplicate blinded samples for genotyping. The between-sample agreement ranged from 98.44 to 100%, with an average of 99.68%.

We used GRR [Abecasis et al., 2001] to check for familial relationships/non-paternity and marker-error. There were 39 families whose samples were excluded from analyses because one or more individuals had genotypes inconsistent with the reported family structure. The genome scan was conducted with samples from 1,423 individuals, including 1,357 family members and 66 control subjects. Blood samples from 66 volunteer control subjects recruited in Ireland were used to obtain allele frequency estimates. Controls were screened and excluded if they reported a history of heavy drinking or problem alcohol use. After data cleaning, 1,289 samples were included in the genotyping. The genotyping success rate among these individuals was 90.2%, reflecting 1,185,544 of a possible 1,314,780 genotypes (1,289 subjects x 1,020 markers).

Statistical Methods

The final sample used for linkage analysis for N included 1,022 individuals from 486 informative families, ranging from 2 to 12 members with the majority of family size of 2 (81.1% affected sib-pair), 3 (14.2%), or 4 (1.9%) persons. In total this sample formed 714 all possible sib-pairs. To explore sex-specific genetic effects on N, we used only the same sex sib-pairs. There were 361 female siblings provided phenotype data and formed 139 possible female–female pairs, while 665 male siblings provided phenotype data and formed 413 possible male–male pairs.

We performed genome-wide multipoint linkage analyses using a regression method [Sham et al., 2002] implemented in the Merlin program [Abecasis et al., 2002] to analyze our quantitative traits. Merlin-Regress is robust for a selected sample in which a population distribution for the phenotype was provided. In this case, after adjustment for age variables, we specified the model with the expected mean for studentized residual of N as 0 and variance as 1, and an estimated genetic variance of 40%. Both transformed and non-transformed but age-adjusted studentized residuals of N were adapted in linkage analysis. Results were very similar in terms of significance levels and chromosome positions. We report linkage results based on the transformed measure. The statistic shown in the figures is the negative logarithm (base 10) of the \( P \)-value \((−\log P)\) corresponding to the Lod scores obtained from the regression analysis.

To evaluate the possible bias associated with selection on alcohol dependence and trait distribution, we used the simulate option in Merlin to run 2,000 null genome scan in the total sample. Data were simulated based on our original family structure, marker informativeness, spacing, and missing data. In addition, phenotypic measurements and affection status were preserved. The empirical \( P \)-values were obtained by calculating the probability of getting the same or a more extreme Lod score under the null hypothesis than our observed value.
RESULTS

The results of the whole-genome linkage scan for N are shown in Figure 1. According to the results of our 2,000 simulations, a Lod score of 1.29 corresponds to a genome-wide significance level of 0.05. There was only one linkage peak exceeding the genome-wide significance level, which was on 12q close to the telomere at approximately 175 cM with a Lod score of 2.13 (log p = 2.76, empirical P-value = 0.0065). Two other suggestive linkage peaks were on 11p and 15q, produced Lod scores above 1. Details are shown in Table I. In general, the empirical results from simulations agreed with the initial findings, with a tendency to higher significance levels than the P-values obtained from initial linkage analysis. The empirical P-values were only reported for linkage peaks in Table I.

Results for sex-specific genetic effects on N are shown in Table II. Because our same-sex analyses had reduced power to detect significant effects due to a smaller sample size, we used the criterion of log p greater than 1.3 for suggestive linkage signals. In these sex-limited analyses, several regions showed potential sex-specific signals for N. The strongest linkage signal for males was on 11p (log p = 2.68, with a peak Lod score of 2.06, extending from 38 to 60 cM in a one Lod score interval). Two close regions on 1p exhibited male-specific signals, one was at approximately 49 cM (log p = 2.36, with a Lod score of 1.77), and the other was at 83 cM (log p = 1.93, with a Lod score of 1.38). In addition, there were also a few male-specific peaks for which the log p ranged from 1.3 to 1.5 on chromosomes 4, 12, 15, 16, and 22 (see Fig. 2). For females, there were four chromosome regions with Lod scores in excess of 1: 2q, 9q, 12q, and 18q. Among them, the more pronounced signals were on 18q (log p = 1.94, with a Lod score of 1.4) and 12q (log p = 1.61, with a Lod score of 1.3). Another two regions exhibited weaker female-specific linkage signals on chromosomes 4 and 13, with log p ranging from 1.3 to 1.4.

DISCUSSION

In our analyses, we found, in the overall sample, a significant evidence for linkage for N on 12q and other positive evidence on chromosome regions 11p and 15q. We also found several regions that may contain susceptibility genes for N only in men (e.g., regions 1p and 11p) or women (e.g., regions 12q and 18q). Our findings replicate some results from previous linkage studies of N [Fullerton et al., 2003; Nash et al., 2004; Neale et al., 2005]. The most pronounced finding was on chromosome 12, the region with the best finding in both ours and the Fullerton and colleagues study. Although linkage signals in the two studies were about 15–30 cM away from each other, these findings may actually be replications of one another given the imprecision of linkage peak position [Roberts et al., 1999]. More interestingly, our sex-limited analysis showed that the chromosome 12 signal mainly comes from female subjects and a similar finding also emerged in this region from the Fullerton and colleagues study. These results suggest that this region might contain genes which impact on the liability to N specifically in females.

For chromosome 11, our linkage signals were mainly driven by males (Fig. 2). Two previous studies reported linkage signals on chromosome 11, though in different regions (one at 99 cM, log p = 3.7; the other at 132 cM, log p = 1.97). In the total sample, we found one linkage signal on this chromosome at 55 cM and a second at 105 cM, the latter of which is about 5 cM away from a suggestive peak in the Fullerton and colleagues study. The third noteworthy region in our analyses, 15q, also appeared to derive from largely the male siblings in our sex-limited analysis, and has not been reported in any of previous three linkage studies. However, 15q has been implicated in linkage studies of major depression. Holmans et al. [2004] reported strong linkage for recurrent early onset major depression in a region less than 10 cM apart from our peak on 15q (Lod score = 3.73). Camp et al. [2005] also found linkage in this region for major depression in men.

There were two regions linked to N in prior studies on chromosome 1, one at 43–70 cM with a Lod score of 1.6 [Nash et al., 2004] and the other at 126–137 cM (log p ranged from 2.52 to 3.95) [Fullerton et al., 2003; Neale et al., 2005]. In our overall sample, no substantial signal was seen in the former region. However, this region contained a Lod score of 1.77 (log p = 2.36) in our male-specific linkage analysis and a female driven signal in the Fullerton and colleagues study (log p = 3.31). We did not find strong evidence implicating the region on chromosome 1 at 126–137 cM.

Our results showed two noteworthy female driven linkage peaks on chromosomes 13 and 18. Interestingly, the significant finding on 13q in the Fullerton and colleagues study was
observed mainly in female pairs, and their linkage peak was approximately 20 cM away from ours. The region on 18q contained the strongest female specific linkage signal seen in our study. The same region was reported in a linkage study of recurrent early onset major depression/anxiety disorder [Camp et al., 2005]. Because N is an important predictive factor for major depression [Clark et al., 1994] and twin studies suggest that they share genetic liability [Fanous et al., 2002; Kendler et al., 2006], the similarity of linkage findings across these phenotypes suggests this region contains susceptibility loci to internalizing traits.

In the past, genetic studies to locate susceptibility genes for neuroticism or emotionality in animal studies have yielded a range of potential findings. Several linkage regions have been identified for a variety of emotionality tests on mouse chromosomes 1, 3, 5, 6, 10, 15, 19, and X [Flint, 2004]. These consistent findings from rodent studies provide important potential information to assist in locating comparable genes for emotion-related traits in humans, as genes for neuroticism/emotionality may be conserved across species. Despite the more complex genetic nature of human behavior or traits compared to those found in the rodent, there have been some clues from recent studies showing possible replicated linkage regions for emotionality related traits across species [Flint, 2004]. Likewise, our linkage findings on chromosome 1p at 44–57 cM might be syntenic with the locus reported in the middle of rat chromosome 5 for anxiety trait in laboratory rat [Fernandez-Teruel et al., 2002]. However, due to imprecision of both human and rodent gene mapping studies, it is still impossible to say whether our linkage signals arise from the same genetic effects.

The sample in this study consisted of sibling pairs concordant for alcohol dependence. Previous studies suggested that high N predicts the risk of alcohol dependence [Prescott et al., 1997] and relapse from treatment for drinking problems [Fisher et al., 1998]. In addition, consistent with the high levels of comorbidity observed between alcoholism and major depression [Grant and Harford, 1995; Lynskey, 1998], we found an elevated lifetime prevalence of major depression in our sample (~70%), which also links with high N. Thus, in our selected sample, we expected to see a distribution of N considerably skewed to the right from that observed in the general population. Indeed, that is what we obtained, having found a mean score of N of 7.9 in our sample, compared to a general population sample mean of 5.0 in men and 5.9 in women. Nevertheless, the unusual distribution of N in our sample should not substantially attenuate our ability to locate susceptibility genes for N for at least two reasons. First, for the present study, we regressed out the age effect and normalized the residual scores before running whole-genome linkage analysis. Thus, unlike the original phenotypic distribution, the residual distribution is no longer skewed. Second, the statistical package we applied, Merlin-Regress, is robust when the trait distribution is non-normal [Sham et al., 2002]. This method is also robust for a selected sample if the expected trait distribution in the general population is provided. Using the studentized residual, the nature of the mean and variance are well understood and correctly specified in our linkage model. Our quantitative linkage analysis results for N should not be substantially affected by our ascertainment methods. Nevertheless we draw our conclusions cautiously.

### Table I. Markers and Positions of Multipoint Linkage Analysis for Suggestive Regions in the Total Sample

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Marker</th>
<th>Position (cM)</th>
<th>Lod score</th>
<th>(-\log p)</th>
<th>Empirical P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>D2S126</td>
<td>228</td>
<td>0.90</td>
<td>1.37</td>
<td>0.022</td>
</tr>
<tr>
<td>9</td>
<td>D9S1690</td>
<td>107</td>
<td>0.91</td>
<td>1.39</td>
<td>0.026</td>
</tr>
<tr>
<td>11</td>
<td>D11S4102</td>
<td>55</td>
<td>1.02</td>
<td>1.52</td>
<td>0.017</td>
</tr>
<tr>
<td>12</td>
<td>D12S1638</td>
<td>175</td>
<td>2.13*</td>
<td>2.76</td>
<td>0.00065</td>
</tr>
<tr>
<td>15</td>
<td>D15S1014</td>
<td>124</td>
<td>1.03</td>
<td>1.53</td>
<td>0.019</td>
</tr>
</tbody>
</table>

*The Lod score at this linkage region exceeds a genome-wide significance level of 0.05 in the present study, extending from 152 to 175 cM in a one Lod score interval.

### Table II. Markers and Positions of Multipoint Linkage Analysis for Suggestive Sex-Specific Regions

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Marker</th>
<th>Position (cM)</th>
<th>Lod score</th>
<th>(-\log p)</th>
<th>Gender specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D1S470</td>
<td>49</td>
<td>1.77</td>
<td>2.36</td>
<td>Male</td>
</tr>
<tr>
<td>2</td>
<td>D1S476</td>
<td>83</td>
<td>1.38</td>
<td>1.93</td>
<td>Male</td>
</tr>
<tr>
<td>4</td>
<td>D4S1572</td>
<td>110</td>
<td>0.88</td>
<td>1.36</td>
<td>Female</td>
</tr>
<tr>
<td>9</td>
<td>D9S1826</td>
<td>162</td>
<td>1.02</td>
<td>1.32</td>
<td>Female</td>
</tr>
<tr>
<td>11</td>
<td>D11S4080</td>
<td>43</td>
<td>2.06</td>
<td>2.68</td>
<td>Male</td>
</tr>
<tr>
<td>12</td>
<td>D12S1616</td>
<td>130</td>
<td>1.30</td>
<td>1.84</td>
<td>Female</td>
</tr>
<tr>
<td>13</td>
<td>D13S279</td>
<td>66</td>
<td>0.86</td>
<td>1.32</td>
<td>Female</td>
</tr>
<tr>
<td>15</td>
<td>D15S1005</td>
<td>86</td>
<td>0.95</td>
<td>1.44</td>
<td>Male</td>
</tr>
<tr>
<td>16</td>
<td>D16S3066</td>
<td>91</td>
<td>0.98</td>
<td>1.47</td>
<td>Male</td>
</tr>
<tr>
<td>18</td>
<td>D18S68</td>
<td>91</td>
<td>1.39</td>
<td>1.94</td>
<td>Female</td>
</tr>
<tr>
<td>22</td>
<td>D22S1174</td>
<td>20</td>
<td>0.85</td>
<td>1.32</td>
<td>Male</td>
</tr>
</tbody>
</table>

*This table lists markers with \(-\log p\) greater than 1.3.
The results of this study should be interpreted in the context of at least three methodological limitations. First, this linkage study is based on a sample selected for alcohol dependence with higher levels of N than are seen in general population samples. Although we incorporated methods to reduce this potential bias, it remains possible that some residual ascertainment bias influenced the results of our linkage analysis. Although we cannot directly address issues for false negatives, on the other hand, it is noteworthy that the previous linkage evidence for alcohol dependence in this sample was mainly on 4q [Prescott et al., 2006], and we found no evidence for linkage for N to that region. Second, we cannot rule out the possibility that the genetic factors that impact on N in individuals with alcohol dependence differ from those factors that influence N in the general population. Third, our statistical power was probably too small to detect modest sized susceptibility genes that affect N. Our power was even more limited in our analyses of sex-specific linkages, especially for female-specific loci given the limited number of female–female sib-pairs in this sample. Thus, we may have missed some regions that have susceptibility genes for N.

In summary, although we did not replicate all previous linkage findings for N, we provide supportive evidence for some genome linkages from both our overall sample and our sex-limited analysis, especially on 1p, 2q, 11q, 12q, and 13q. Furthermore, we found evidence for a potential male-specific locus for N on 15q and a female-specific locus on 18q, regions reported as positive in prior linkage studies of major depression. Whether there are pleiotropic gene effects in those regions for emotionality warrant follow-up studies.

ACKNOWLEDGMENTS

Data collection was funded by U.S. National Institutes of Health grant R01-AA-11408 with support from the Irish Health Research Board. Ruth Barrington, Ros Moran and Carol Cronin of the Irish Health Research Board provided administrative support. F.A. O’Neill, the Northern Ireland Blood Transfusion Service of the British National Health Service and the Irish Gardaí provided assistance in obtaining control blood samples. Bradley Todd Webb provided helpful discussion. John Myers, Indrani Ray, and Cheryl Smith assisted with data management, Jen Vittum, Brandon Wormley, and Dawn Thielson assisted with genotyping and Stacey Garnett, Jill Opalesky, and Rebecca Ortiz provided administrative support at VCU. Preparation of this manuscript was supported by a Young Investigator award from the National Alliance for Research on Schizophrenia and Depression to Po-Hsiu Kuo. Michael Neale was supported by NIH grant MH 65322.

REFERENCES


