A Systematic Single Nucleotide Polymorphism Screen to Fine-Map Alcohol Dependence Genes on Chromosome 7 Identifies Association With a Novel Susceptibility Gene ACN9


Background: Chromosome 7 has shown consistent evidence of linkage with a variety of phenotypes related to alcohol dependence in the Collaborative Study on the Genetics of Alcoholism (COGA) project. With a sample of 262 densely affected families, a peak logarithm of odds (LOD) score for alcohol dependence of 2.9 was observed at D7S1799. The LOD score in the region increased to 4.1 when a subset of the sample was genotyped with the Illumina Linkage III panel for the Genetic Analysis Workshop 14 (GAW14). To follow up on this linkage region, we systematically screened single nucleotide polymorphisms (SNPs) across a 2 LOD support interval surrounding the alcohol dependence peak.

Methods: The SNPs were selected from the HapMap Phase I CEPH data to tag linkage disequilibrium bins across the region. Across the 18-Mb region, genotyped by the Center for Inherited Disease Research (CIDR), 1340 SNPs were analyzed. Family-based association analyses were performed on a sample of 1172 individuals from 217 Caucasian families.

Results: Eight SNPs showed association with alcohol dependence at p < .01. Four of the eight most significant SNPs were located in or very near the ACN9 gene. We conducted additional genotyping across ACN9 and identified multiple variants with significant evidence of association with alcohol dependence.

Conclusions: These analyses suggest that ACN9 is involved in the predisposition to alcohol dependence. Data from yeast suggest that ACN9 is involved in gluconeogenesis and the assimilation of ethanol or acetate into carbohydrate.

Key Words: ACN9, alcohol dependence, association, genetics, linkage disequilibrium

Alcohol dependence is a common complex disorder that affects millions of people worldwide and causes considerable burden in terms of personal, interpersonal, and societal costs (1). Results from the National Comorbidity Study indicate that over 14% of adults in the United States have a lifetime history of alcohol dependence, making it one of the most prevalent adult psychiatric disorders (2). Family, twin, and adoption studies have convincingly demonstrated that genes play an important role in the development of alcohol dependence, with heritability estimates in the range of 50%–60% for both men and women (3,4). Efforts are now underway to identify specific genes involved in the development of the disorder.

The Collaborative Study on the Genetics of Alcoholism (COGA) is a multi-site collaboration aimed at identifying genes contributing to alcohol dependence. COGA ascertained families densely affected with alcohol dependence from treatment centers at multiple sites across the United States. Initially, an approximately 10-cM genome-wide microsatellite survey was conducted, and linkage analyses were performed to detect chromosomal regions likely to harbor genes contributing to a variety of phenotypes related to alcohol dependence (5–7). In those regions with evidence of linkage, more extensive genotyping was performed, and association analyses were employed to identify specific genes involved in the predisposition to alcohol dependence and related phenotypes. COGA has also made use of electrophysiological endophenotypes (8–10), as a complement to clinical diagnoses in genetic analyses.

One region that has consistently emerged with significant evidence of linkage in the COGA project is chromosome 7q. In the initial COGA sample of 105 pedigrees, chromosome 7 provided the strongest evidence of linkage to alcohol dependence (11). With the alcohol dependence criterion of meeting DSM-III-R alcohol dependence and Feighner definite alcoholism, the maximum multipoint LOD score on chromosome 7 was 3.49 near the marker D7S1793. An independent sample consisting of an additional 157 extended families also showed modest, consistent evidence of linkage to chromosome 7, with a LOD score of 1.3 (7). Additional microsatellite markers were genotyped on chromosome 7, and linkage analyses employing the full sample
of 262 extended pedigrees yielded a peak LOD score of 2.9 at D7S1799 (12). Further genotyping was conducted as part of the Genetic Analysis Workshop 14 (GAW14) on a densely affected subset of the sample (n = 143 pedigrees) with both the Affymetrix 10K Mapping SNPs and the Illumina Linkage Panel III (13). The LOD score at the peak increased to 4.1 with a reduced set of single-nucleotide polymorphisms (SNPs) not in linkage disequilibrium (LD) with adjacent markers (14). In addition to the linkage observed in the COGA sample, linkage has been observed to this region of chromosome 7q in an Australian sample with P3 amplitude (15), a phenotype thought to index genetic vulnerability to alcohol dependence (9). In addition, there has been a recent report of linkage to this region with alcohol consumption phenotypes in the Nicotine Addiction Genetics project (16). Modest evidence of linkage to this region was also reported for an alcoholism phenotype with age and gender as covariates in an independent sample of multiplex families ascertained at Pittsburgh (17). Here, we report results from a systematic screen of SNPs across the chromosome 7 alcohol dependence linkage peak in the COGA sample in an effort to identify the gene(s) contributing to the observed linkage peak.

Methods and Materials

Sample

COGA is a multisite project, in which families were collected by six centers across the United States: Indiana University, State University of New York Health Science Center, University of Connecticut, University of Iowa, University of California/San Diego, and Washington University at St. Louis. Probands identified through inpatient or outpatient alcohol treatment programs by each of these six sites were invited to participate if they had a sufficiently large family (usually sibships > 3 with parents available) with two or more members in a COGA catchment area (6). Multiplex alcoholic families that had at least two biological first-degree relatives affected with alcohol dependence in addition to the proband were invited to participate in the more intensive stage of the study, which included obtaining blood for genetic analyses. Second- and third-degree relatives in the families were assessed when they were considered to be informative for the genetic linkage studies. The institutional review boards of all participating centers approved the study. Additional details about the study have been published previously (5,6).

We analyzed a set of 217 Caucasian families here because: 1) the marker selection strategy (detailed in the following text) was based on patterns of LD among Caucasians, and allele frequencies often differ between races; and 2) the maximal LOD score with alcohol dependence in the region (4.1) was based on a sample of Caucasian families. These 217 families contained a total of 1172 individuals with genotype and phenotype data: 554 women and 618 men. There were 855 affected individuals in the sample (288 women, 567 men). The mean age of affected women was 34.5 years (SD 9.7), unaffected women: 50.4 years (SD 9.7), unaffected men: 50.5 years (SD 9.7), and 509 avuncular pairs. All individuals in the genetic analysis sample were interviewed as adults (≥ 18 years of age) with the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), a highly reliable, psychiatric interview (18,19). The definition of alcohol dependence used in association analyses required individuals to meet criteria for DSM-III-R alcohol dependence and Feighner definite alcoholism (20). This was the phenotype that yielded the linkage peak from which the region was delineated for the SNP screen (14). Seventy-three percent of the sample used in genetic analyses was affected with alcohol dependence. There are also high rates of comorbid psychiatric disorders in the sample: 37% meet criteria for dependence on an illicit drug, 41% report a major depressive episode, 19% meet criteria for childhood conduct disorder, and 14% have a diagnosis of adult antisocial personality disorder. Additional information about comorbidity in the COGA sample has previously been reported elsewhere (21,22). In the sample analyzed here, the highest level of educational attainment was < a high school degree in 23%, a high school education in 31%, some college in 30%, a college degree in 11%, and a postgraduate degree in 5%. At the time of interview, 35% of the sample was unemployed, and 65% was currently employed. The modal current household gross income was $20,000–$29,000/year.

SNP Selection Procedure

The SNPs were selected to cover a 2-LOD support interval on either side of the linkage signal at 7q22, on the basis of the peak observed in the GAW sample (14); that peak was narrower than the peak in the original, approximately 10-cM microsatellite linkage screen (12). The region was bounded by rs194506 (89.487 Mb) and rs441534 (107.423 Mb) (dbSNP 124/NCBI Human Build 35.1), and covered approximately 18 Mb (Figure 1). HapMap Phase I CEPH data (build 16c.1, June 2005) was used to select SNPs; only common SNPs (minor allele frequency ≥ 10%) were considered. A total of 4067 SNPs meeting this criteria were identified across the region. The SNPs were grouped into LD bins (23), on the basis of the “greedy” algorithm (24,25). With this method, each bin had at least one SNP that satisfied r² ≥ 0.8 with all other SNPs in the bin. Tag SNPs were selected for each bin in an 8:1 ratio (e.g., bins with 1–8 SNPs get 1 tag, 9–16 get 2). The SNPs with the highest r² with other SNPs in the bin were chosen to be tag SNPs. The tagSNP selection method allowed for a reduction in SNPs of 61%, yielding 1581 SNPs. An additional 55 nonsynonymous polymorphic HapMap SNPs were added to the set of SNPs (23 of 55 had minor allele frequency < 10%). Selected SNPs were scored by the Center for Inherited Disease Research (CIDR) for expected performance on the Illumina platform. Failed tag SNPs were replaced by the next best tag. If there were no passing tags, then the entire LD bin was selected for genotyping. The final list of 1536 SNPs covered 221 genes. It consisted of 883 intra-genic SNPs (654 in introns, 68 in exons, 128 in untranslated regions, and 33 SNPs within 2 kb of the first 5' promoter and 500 bp of the 3' end of the largest known transcript [known as "locus" SNPs in dbSNP]) and 653 inter-genic SNPs. Note that although dbSNP build 124/NCBI Human Build 35.1 was used for SNP selection, information about the location of SNPs presented in the article tables is based on the updated dbSNP 126/NCBI 36.1 data.

Genotyping and Analysis

Genotyping was conducted by CIDR with the Illumina technology on a BeadLab station with GoldenGate chemistry. One thousand five hundred thirty-six SNPs were attempted across the region, with 3,539,740 genotypes released for 1436 SNPs. CIDR
cited the following reasons for dropping loci: poorly defined clusters; excessive replicate and/or Mendelian errors; more than 50% missing data; or all samples genotyping as heterozygous. An additional 96 SNPs were flagged with atypical clustering, and these SNPs were also omitted from analyses. Accordingly, 1340 SNPs passed all quality control checks and were used in analyses. The missing data rate was .056%. The program Pedcheck (26) was used by CIDR to check for Mendelian inconsistencies; the Mendelian consistency rate was 99.91%. Further checks were performed by COGA collaborators with the Prest program (27), and a small number of questionable relationships were removed, such that there were no Mendelian inconsistencies remaining in the data.

Findings for the most promising gene were followed up by conducting additional genotyping to thoroughly evaluate the evidence for association. Publicly available databases, dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and HapMap (http://www.hapmap.org), were used to identify SNPs within and flanking the gene. The SNPs with an $r^2 < .9$ in HapMap with any of the associated CIDR SNPs were selected for genotyping. An additional 16 SNPs within and flanking ACN9 were genotyped with a modified single nucleotide extension reaction, with allele detection by mass spectroscopy (Sequenom MassArray system; Sequenom, San Diego, California). All genotypic data were checked for Mendelian inheritance of marker alleles with the USERM13 (28) option of the MENDEL linkage computer programs. Trio data from Caucasian individuals genotyped in the COGA dataset were entered into the program Haplovie (29) to examine the LD structure of the genotyped SNPs. Figure 2 shows the LD structure across the region.

Hardy-Weinberg equilibrium was assessed for all SNPs. Among 601 unrelated individuals used in checks of the 1340 CIDR SNPs, 68 SNPs significantly deviated from Hardy-Weinberg at $p < .05$, and 12 SNPs were significant at $p < .01$. These numbers are very close to that expected by chance (67 and 13, respectively, on the basis of 1340 SNPs). None of the SNPs most significantly associated with alcohol dependence ($p < .01$) had significant deviations from Hardy-Weinberg equilibrium. Those SNPs with potential Hardy-Weinberg problems are indicated in Table 1, so that this information can be considered in interpretation of results. None of the additional SNPs genotyped by COGA showed significant deviation from Hardy-Weinberg.

Multiplex families of alcoholics were used in tests of association between each of the SNPs and each of the phenotypes studied, with the Pedigree Disequilibrium Test (PDT) (30). The PDT uses all available trios in a family (two parents plus child genotyped) as well as discordant siblings. The PDT-ave statistic was computed, which averages the association statistic over all families (30). No corrections were made for multiple testing in the initial analyses of the 1340 SNP screen set, because these analyses were considered the first stage to consider genes for further follow-up. In the second stage of analysis, SNPs were tested across the ACN9 gene. A Nyholt correction was applied to the data to determine significance after taking into account multiple testing in the presence of correlated SNPs (31). This method uses information on the pairwise LD between the genotyped SNPs to compute the number of “effectively independent” SNPs. With the updated method of Li and Ji (32), the effective number of SNPs ($M_{\text{eff}}$) on the basis of the 23 SNPs genotyped in this study was 10. With $M_{\text{eff}} = 10$, the Bonferroni corrected significance threshold required across the gene is $p = .005$.

**Results**

Table 1 shows the SNPs from the CIDR screen yielding evidence of association at $p < .050$, with $p$ values < .01 differentiated with a
horizontal rule on the table. The SNPs are listed in order of ascending $p$ values, with the most significant SNPs at the top of the table. (The results for all 1340 genotyped SNPs are available in Supplement 1.) Of the eight SNPs listed in Table 1 with $p < .01$, four were located in or very near the gene ACN9: rs10499934, rs7794886, rs12056091, and rs1917939 (only rs7794886 and rs12056091 are listed as in ACN9 in Table 1, because these are the notations as listed in NCBI). Although rs7794886 and rs12056091 are in the same bin and display high LD (HapMap $r^2 = .93$), the other two SNPs are in separate LD bins and provide independent evidence for association. Accordingly, additional genotyping was undertaken to more thoroughly investigate this gene. The results from association analyses of all SNPs genotyped across ACN9, including those genotyped by CIDR, are shown in Table 2. Twelve of the 23 genotyped SNPs in ACN9 were significant at $p < .05$. Eight SNPs surpassed the Bonferroni corrected significance level suggested by the Nyholt correction of $p < .005$. The top two most significant SNPs, rs10246622 ($p = .000008$) and rs13475 ($p = .00014$), showed substantial LD ($r^2 = .88$), making haplotype analyses uninformative, because only the 1 1 and 2 2 haplotypes were observed with considerable frequency. Figure 1 shows the location of the ACN9 gene with reference to the linkage peak observed in the GAW data from which the region for the SNP screen was selected. ACN9 is located centromeric of the linkage peak, with a LOD score of 3.08. The peak LOD score was 4.08 in the region; thus, ACN9 is located at a 1-LOD distance from the peak.

**Discussion**

In this study, we undertook a systematic screen of SNPs covering a 2-LOD support interval around a linkage peak previously reported for alcohol dependence (12,14). This screen led to the identification of a novel gene that seems to affect susceptibility to alcohol dependence, ACN9. ACN9 was originally identified in a collection of respiratory-competent yeast mutants that were unable to use acetate as a carbon source (33). ACN9 seems to be involved in gluconeogenesis and is required for the assimilation of ethanol or acetate into carbohydrate (34). Little is known about the human homolog of the gene. Accordingly, this gene likely would not have been prioritized for investigation, were it not for the results from the systematic screen of the linked region.

The most significant SNP in the CIDR screen, rs2157745, is located in a region with no genes listed in NCBI. It is approximately 20 kb past the 3' end of SAMD9, making it unlikely that it would be involved in the function or expression of SAMD9.
There is a large gene upstream (CDK6, a cyclin-dependent kinase), whose promoter might be in this region. One of the other top eight SNPs is located in SAMD9. This gene is believed to be involved in the regulation of extraosseous calcification and has been associated with familial tumoral calcinosis (FTC), a rare autosomal recessive disorder characterized by the progressive deposition of calcified masses in cutaneous and subcutaneous tissues, resulting in ulcerative lesions and severe skin and bone infections (35). There is little evidence for expression of SAMD9 in the brain (35); thus, it was not considered a high priority for further follow-up for potential association with alcohol dependence.

We have previously reported association between alcohol dependence and related phenotypes and several genes located elsewhere on chromosome 7, including a muscarinic cholinergic receptor gene, CHRM2 (12), and two taste receptor genes, TAS2R16 (36) and TAS2R38 (37). Those genes were selected on the basis of proximity to a more distal linkage peak with an electrophysiological endophenotype (38), and none are located within the 2-LOD interval of the linkage peak for alcohol dependence studied here.

Finally, we note that at the time when the SNP selection took place for this study, only the Phase I build 16c.1 data were available. Table 1.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Allele</th>
<th>HMMAF</th>
<th>p</th>
<th>Bin ID (size)</th>
<th>Position</th>
<th>Gene Location</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2157745</td>
<td>[A/C]</td>
<td>.108</td>
<td>.0010</td>
<td>92,543,595</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>rs10499934</td>
<td>[A/G]</td>
<td>.267</td>
<td>.0016</td>
<td>96,575,512</td>
<td>[ACN9]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>rs7794886</td>
<td>[T/C]</td>
<td>.442</td>
<td>.0033</td>
<td>76 (2)</td>
<td>ACN9</td>
<td>INTRON</td>
<td>—</td>
</tr>
<tr>
<td>rs12056091</td>
<td>[T/C]</td>
<td>.442</td>
<td>.0033</td>
<td>76 (2)</td>
<td>ACN9</td>
<td>INTRON</td>
<td>—</td>
</tr>
<tr>
<td>rs1917939</td>
<td>[T/C]</td>
<td>.319</td>
<td>.0035</td>
<td>96,554,800</td>
<td>[ACN9]</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>rs10953089</td>
<td>[T/C]</td>
<td>.158</td>
<td>.0050</td>
<td>92,566,363</td>
<td>SAMD9</td>
<td>LOCUS</td>
<td>—</td>
</tr>
<tr>
<td>rs13228694</td>
<td>[T/C]</td>
<td>.142</td>
<td>.0050</td>
<td>99,778,243</td>
<td>PLIR8</td>
<td>INTRON</td>
<td>—</td>
</tr>
<tr>
<td>rs7778571</td>
<td>[A/G]</td>
<td>.233</td>
<td>.0071</td>
<td>99,502,753</td>
<td>ZNF3</td>
<td>INTRON</td>
<td>—</td>
</tr>
</tbody>
</table>

Position, gene location, and function as listed in NCBI. Single nucleotide polymorphisms (SNPs) without a Bin ID listed were the only SNP genotyped in that bin. Bin size refers to the number of SNPs in the bin. HMMAF, HapMap minor allele frequency; CIDR, Center for Inherited Disease Research.

*These SNPs are not annotated as ACN9 in NCBI but are very near the gene.

SNPs showing significant deviations from Hardy-Weinberg.
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 (S. Kuperman, R. Crowe); SUNY HSCB (B. Porjesz, H. Begleiter); Washington University in St. Louis (L. Bierut, A. Goate, J. Rice); University of California at San Diego (M. Schuckit); Howard University (R. Taylor); Rutgers University (J. Tischfield); and the Southwest Foundation (L. Almasy). Zhaoxia Ren serves as the NIAAA Staff Collaborator. This national collaborative study is supported by National Institutes of Health (NIH) Grant U10AA08401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). In memory of Henr Begleiter, Ph.D., and Theodore Reich, M.D., Principal and Co-Principal Investigators of COGA, we acknowledge their immeasurable and fundamental scientific contributions to COGA and the field. The author SS was supported by American Cancer Society grant IRG-58-010-50; NLS was supported by NIDA grant K01DA015129. Genotyping services were provided by CIDER. CIDER is fully funded through a federal contract from the NIH to The Johns Hopkins University, Contract Number N01-HG-65403. All authors were provided with the Biological Psychiatry guidelines from the web for Disclosure of Biomedical Financial Interests and Potential Conflicts of Interest. Dr. Raymond Crowe reported that he has consulted with a law firm that is defending the Pfizer company in lawsuits against its product Zoloft. All other authors reported no interests to disclose related to this report.

Supplementary material cited in this article is available online.


