

## Symposium

# A Genomic Survey of Alcohol Dependence and Related Phenotypes: Results from the Collaborative Study on the Genetics of Alcoholism (COGA)

Chaired by Theodore Reich, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA

### Introduction to the COGA Study

Henri Begleiter, Department of Psychiatry, SUNY, Downstate Medical Center, Brooklyn, NY, USA

### A Genomic Survey of Alcohol Dependence - Part I

Theodore Reich, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA

### A Genomic Survey of Alcohol Dependence - Part II

Howard J. Edenberg, Department of Biochemistry and Medical Biology, Indiana University School of Medicine, Indianapolis, IN, USA

### Phenotypic Features Derived from Neurophysiological Measures and Their Genotypes

Henri Begleiter, Department of Psychiatry, SUNY, Downstate Medical Center, Brooklyn, NY, USA

### Mapping Genes for Quantitative Personality Traits Related to Alcoholism

C. Robert Cloninger, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA

## INTRODUCTION

The Collaborative Study on the Genetics of Alcoholism (COGA) is a six center program (Cooperative Agreement of the NIAAA) to detect and map susceptibility genes for Alcohol Dependence and related Phenotypes. Participating centers are located at State University of New York Health Science Center (Principal Investigator-Bernice Porjesz, Ph.D.), University of Connecticut (Principal Investigator-Victor Hesselbrock, Ph.D.), Indiana University (Principal Investigator-John Nurnberger, Jr., M.D., Ph.D.), University of Iowa (Principal Investigator-Raymond Crowe, M.D.), University of California/San Diego (Principal Investigator-Marc Schuckit, M.D.), The Scripps Research Institute, CA (Principal Investigator-Floyd Bloom, M.D.), and Washington University, St. Louis (Principal Investigator-Theodore Reich, M.D.). The National Principal Investigator is Henri Begleiter, M.D., Ph.D. at the State University of New York Health Science Center and the NIAAA Collaborator is Robert Karp, Ph.D.

**Design of the Study (H. Begleiter).** Proband for the study are systematically ascertained from inpatient and outpatient treatment units and are invited to participate if their families are sufficiently large, cooperative and have two or more members in one of the COGA catchment areas. Caseness is defined as alcoholic at the definite level using Feighner Criteria and Alcohol Dependent by DSM IIIR criteria. (This

combination is referred to as COGA criteria throughout the study.) Probands are not accepted if they are intravenous drug users, have AIDS, are unable to participate in the protocol or have a terminal illness. Randomly ascertained control families including two parents and three or more offspring over the age of fourteen are also included in the study. A total of 1440 families have participated in COGA including 233 controls. Adults as well as adolescents and children (to age 7) are interviewed and 9475 interviews have been given, including 1081 juveniles.

All subjects receive a personal interview and a battery of Personality Tests, including the Tri-dimensional Personality Questionnaire. Families with two affected first degree relatives, in addition to the proband, are accepted for more detailed study. In these families, adult subjects have blood drawn for cryopreservation and the preparation of DNA. Blood for DNA extraction is drawn from juveniles. Biochemical markers are also studied on a platelet preparation and EEG (electroencephalogram) measurements are recorded along with an extensive battery of auditory and visual evoked potentials.

Pedigrees of densely affected nuclear families are extended to second and third degree relatives when the extensions are informative for genetic linkage. This is accomplished by ascertaining relatives of affected first degree relatives (of the proband). Additionally pedigrees are extended if they include affected second or third degree members (Affectational status

determined initially by the family history method). Members of extended pedigrees are personally interviewed and receive the same detailed protocol used with densely affected nuclear families.

Lifetime assessment of Psychiatric Diagnosis in adults is being made using the "Semi-Structured Assessment for the Genetics of Alcoholism" (SSAGA) interview schedule. This is a polydiagnostic instrument, created for this study, that enables diagnoses of Alcoholism or Alcohol Dependence to be made using Feighner; DSM III-R, DSM IV and ICD 10 criteria. Medical records and family history information are also available to assist in diagnosis. The within and between center test-retest reliability of the SSAGA is high, assuring that diagnoses of Alcohol Dependence from different centers can be safely combined.<sup>1</sup>

The lifetime prevalence of Alcohol Dependence using the COGA criteria among the male first degree relatives (father, brother, son) of probands is approximately 38-50% (N= 1733) and among female first degree relatives (mother, sister, daughter) is approximately 10-24% (N= 2322). This contrasts with rates in control male and female subjects of 18% (N= 485) and 4% (N= 495). For ICD 10 the rate of Alcohol Dependence in male and female first degree relatives is approximately 19-32% and 7-14% respectively. The comparable rates in male and female controls is 7% and 1.5%. The number of ICD 10 controls is small so the population rates have large standard errors.

One-hundred and five densely affected families were chosen for linkage analysis as they were informative using alcohol dependence, personality and neurophysiological measurements as phenotypes. Uninformative parts of the pedigrees were trimmed resulting in a data set for genotyping and analysis that includes 987 subjects. The proportion of these subjects with a life time diagnosis of alcohol dependence is 28% using ICD 10 and 45.7% using COGA criteria. 32.8% are unaffected and do not have a diagnosis of alcohol dependence, abuse or harmful use although they may have one or more symptoms of these disorders. Some level of diagnosis of dependence or abuse occurs in 17%.

The non-parametric methods of analysis that were chosen to analyze these data primarily use affected sibling pairs. Accordingly, the 105 pedigrees were subdivided into all possible subsets of non-redundant nuclear families. In this newly configured sample there are 177 nuclear families. In order to provide a contrast with the category of "affected", a group of subjects who were not abstinent and who had NO symptoms of dependence or abuse by any diagnostic category were designated "unaffected". Individuals who had one or more symptoms but did not have a sufficient number for a criteria based diagnosis were designated "unknown." Similarly, individuals who denied ever being exposed to alcohol were designated "unknown." Using these definitions there are 382 doubly affected sib pairs with the COGA diagnosis and 145 with ICD 10. There are 182 discordant sibpairs where one is affected with the COGA diagnosis and 100 discordant pairs with ICD 10. There are 47 concordant unaffected pairs.

## DISCUSSION

**Methods of Linkage Analysis (T. Reich).** Since the genetic factors contributing to alcohol dependence are probably the interaction of multiple genes of small to moderate effect, linkage analysis preferentially uses "model free" nonparametric methods. These methods do not require an estimate of the parameters of the underlying genetic model, which is likely to be complex and not well understood. The model free linkage approach has been successful in studies of diabetes and multiple sclerosis and is ideally suited to psychiatric phenotypes. By contrast parametric approaches, are less robust and more powerful. Much research is being conducted to increase analytic power and several new methods have recently been introduced.

The SIBPAL module of the SAGE program<sup>2</sup> was used to estimate the probability of identity-by-descent (IBD) for affected, unaffected and discordant sib pairs. If the marker and disease loci are close to each other (linked), doubly affected and unaffected sibpairs are expected to be more alike and discordant pairs less alike than if the loci are unlinked. This was tested by regression. Additionally, the IBD scores of affected pairs were also tested for excess sharing.

Two other methods of analysis were also used. The "Modscore" method is a non-parametric modification of a parametric approach which is both robust and powerful.<sup>3</sup> In this application only affected subjects were included. This method proceeds one marker at a time. It differs from other affected sibpair methods in that sibships are not subdivided into pairs but studied jointly even if the sibship size is large. Finally, several multipoint affected sibpair methods were used. The Aspex program;<sup>4</sup> uses the entire pedigree, including unaffected and unknown individuals to determine the genotype of missing or uninformative parents in order to maximize power. Rather than studying individual marker loci, parental and offspring multipoint haplotypes are considered so the likelihood of linkage is tested along the entire chromosome using information simultaneously from all loci. The affectational status of unaffected or "unknown" individuals is not considered in this method. The program MAPMAKER/Sibs<sup>5</sup> was also used in some of the multipoint affected sibpair analyses.

**Laboratory Methods (H. Edenberg).** Genotyping has been carried out at Indiana (H. Edenberg) and Washington (A. Goate) Universities. The entire genome has been scanned at an approximately 20 centimorgan (cM) spacing. COGA has pursued a survey of the genome rather than a candidate gene approach to explore the possibility that novel unexpected regions of the genome are significant. This does not require guesses about the importance of various candidate genes. Candidate genes have been tested as parts of regions where highly polymorphic markers are studied.

Microsatellite repeat markers, detected by PCR have been used. We have favored tetra- and tri-nucleotide repeats because of their greater stability and ease of reading when compared with di-nucleotides. Alleles are determined by PCR amplification of the DNA containing a repeat segment and

measurement of the size of the alleles by gel electrophoresis. The Indiana University laboratory has used manual genotyping methods and the Washington University laboratory has used automated approaches. In both cases there is an emphasis on accuracy of measurement and laboratories are blind to phenotypic status. Error checking is accomplished by double reading of autoradiograms and multiple rechecking of discrepancies. Checks for Mendelian inheritance are then carried out in the genotyping database using the Genemaster suite of programs.<sup>6</sup> The programs CRIMAP<sup>7</sup> and User M13<sup>8</sup> used to estimate the interval between markers and marker allele frequencies. These programs also conduct further checks for Mendelian inheritance. Discrepant data are deleted after databasing and the data is formatted for genetic analysis. Data for this symposium includes more than 277,000 genotypes and 275 markers.

**Linkage Results from the Washington University Laboratory (T. Reich).** Nine autosomes and the X chromosome were analyzed in the Washington University COGA laboratory. One hundred thirty-nine marker loci were studied for presentation at the RSA/ISBRA meeting. The average number of alleles per marker per chromosome varied between 8.5 to 10.4. The average heterozygosity of the markers per chromosome varied between 0.68 to 0.78. The distance between markers averaged approximately 16 cM overall and between 10 to 21 cM. by chromosome.

The analysis of affected sib pairs revealed significant deviations in the probability of the IBD score on chromosome 1 using the COGA diagnosis as the affected phenotype and on two regions on chromosome 8 using the ICD 10 diagnosis as the affected phenotype. The chromosome 1 finding was less striking when the diagnosis was changed to ICD 10 but was still significant. By contrast, when the diagnosis was changed to the COGA phenotype the findings on chromosome 8 were no longer significant.

The linear regression approach (SIBPAL/SAGE) which includes unaffected as well as discordant pairs in the analysis confirmed the findings reported above. Similarly, the MODSCORE method found significant evidence for linkage at the same locations. Although other locations also had interesting findings these initial observations seemed the most promising.

Using the program ASPEX, the proportion of alleles IBD at the marker loci was computed and reviewed. A version of this program was used that calculated the IBD score of a subset of affected sibpairs unambiguously using marker information from the parents. This eliminated possible sources of bias due to families of varied ethnic background. Thus the findings reported above are not the consequence of misspecification of marker frequencies nor the result of population stratification using these criteria.

Multipoint analyses were conducted using the ASPEX Program. These analyses use marker information from all members of the family and from the entire chromosome to measure IBD continuously along the chromosome. The chromosome 1 and one of the chromosome 8 findings were

supported by the multipoint analysis. One finding on chromosome 8 was not supported.

The next steps in the search for the susceptibility genes for alcohol dependence includes fine mapping of the candidate areas that have been described in this presentation. We are also continuing to map less densely genotyped regions and the number of candidate regions may thus increase. We are also planning to genotype candidate regions suggested by animal and biological marker studies.

**Linkage Results from the Indiana University Laboratory (H. Edenberg).** Thirteen autosomes were genotyped at the Indiana University Laboratory and analyzed by Dr. T. Forou. There were 136 markers with an average of 10.5 alleles each and an average P.I.C. (polymorphism information content) greater than 0.71. The map generated from these data had an average spacing of 14 cM. The primary method of analysis of these data for both excess allele sharing and for regression analysis in sibling pairs used the SIBPAL program of the SAGE package. Three definitions of alcohol dependence were studied including the COGA, DSM IV and ICD 10 criteria. Two definitions of "unaffected" were used. A broad definition which was simply "not affected by a specific criterion" and a narrow definition which only included individuals who had no symptoms of abuse or dependence by any criterion.

Two regions of interest were reported in this symposium. The first is a region on 16 p where two adjacent markers gave evidence of increased allele sharing. One of these was also significant using regression analysis. This region also gave a strong signal using the MODSCORE approach and the multi point (MAPMAKER/SIBS) approach as well. Thus, all of the analytic methods supported the suggestion and make it a prime candidate for further research.

A different type of interesting result was found on chromosome 4. In this case evidence for linkage gained most of its statistical power from an excess of allele sharing among unaffected sibpairs and a decrease in allele sharing among discordant (affected/unaffected) sibpairs. This may reflect the influence of a protective gene in these families with a high density of alcohol dependent members. It is interesting that this region of chromosome 4 includes the alcohol dehydrogenase loci for which there is strong evidence of a protective effect of certain alleles in Oriental populations.

The next step in these analyses is extension and confirmation of these results using flanking markers and additional family data. These studies will provide increased evidence in regions where there is a gene affecting risk and allow the region to be narrowed somewhat. The genetic complexities of a disease such as alcoholism, however, limit how much the region can be narrowed. To put this into perspective, a 10 cM region covers 10,000,000 base pairs of DNA and includes approximately 300 genes. We are fortunate that the Human Genome project will provide tools for the more rapid localization and characterization of susceptibility genes.

**Linkage Analysis of ERP Data (H. Begleiter).** It has been amply demonstrated that the P3 component of the

event related potential (ERP) is significantly reduced in abstinent alcoholics compared to matched controls. This component is the third positive component of the ERP which is typically elicited by infrequent relevant stimuli. This deficit is striking in abstinent alcoholics and does not change with abstinence.<sup>9</sup> The protracted nature of this deficit was hypothesized to be due to the neurotoxic effects of alcohol. However, we first hypothesized that this deficit interceded the development of alcoholism, and should be present in young subjects at high risk for the disorder. We tested the hypothesis in the sons of alcoholic fathers who had no experience with alcohol and were matched on several demographic variables with controls. Our study demonstrated in compelling fashion that young sons of alcoholic fathers demonstrated deficits identical to those observed in their fathers.<sup>10</sup> These findings have been widely replicated.<sup>11-15</sup>

In the COGA project we have collected EEG/ERP data from alcoholics and their families as well as control families. In keeping with prior observations our data confirm that probands, males and females manifest significantly reduced P3 amplitude. This is also the case for unaffected relatives in the families of alcoholic probands. Similarly these results obtain when the children of alcoholics are compared to controls. Since the ERP P3 amplitude is heritable,<sup>16</sup> we began linkage analysis using DNA collected from 987 individuals in 105 COGA families. Using the ERP phenotype this included approximately 337 sibpairs.

Two methods of linkage analysis are reported in this presentation. The first method used a two point regression approach implemented the SIBPAL program from the SAGE package.. The second used the program SOLAR.<sup>17</sup> This is a multipoint method which estimates the genetic variance attributable to a genomic region using information from all individuals in a nuclear family and all loci on the chromosome. Thus quantitative loci of moderate effect can be detected.

The preliminary and partial results of these analyses indicate that there are several genetic loci that yield significant findings. In particular we noted that chromosomes 2 and 6 yielded lod scores of 3.28 and 3.41 respectively. Other chromosomal regions appear to be quite promising and are still being analyzed. We hope that these preliminary genetic findings will provide us with some new insights into the genetics of alcoholism.

**Linkage Analysis of Personality Traits (R. Cloninger).** As part of the Consortium on Genetics of Alcoholism (COGA), a 20 centimorgan screen of personality traits was carried out in multiplex families of alcoholics. The Tridimensional Personality Questionnaire was used to measure temperament traits of harm avoidance, novelty seeking, reward dependence, and persistence. Factor analysis in the full COGA sample of 6162 individuals confirmed that individual differences in each of these four dimensions were substantial, independent of one another, and strongly predictive of susceptibility to alcoholism. Elston's Sibpal program 2.7.2 was used to screen for possible linkage between these four traits and individual markers on each of the autosomes in over 400 sib pairs from

175 sibships. For each of the four temperament factors tested, six to twelve chromosomes were identified as having one or more marker loci with possible linkage ( $p < .05$ ). These regions of possible linkage were further evaluated for linkage in multipoint analyses using a program developed by Blangero.<sup>17</sup> The strongest evidence of linkage was for harm avoidance on chromosome 8p (LOD 3.5). Further analysis is underway with additional flanking markers.

## SIGNIFICANCE

This has been a presentation of the first genomic survey for alcohol dependence, the P3 component of the ERP and the personality traits measured by the TPQ. In each case there have been several regions of interest that warrant follow up by flanking markers which saturate the region. The multipoint strategy will then be used to eliminate or reduce false positive results. In addition, a replication sample of approximately the same size and composition as the current sample will shortly be available. This will provide a further stringent test of the linkage hypotheses generated in the initial survey and follow-up.

In addition to linkage studies we are also exploring the possibility of association or "disequilibrium" between the linked markers and the alcohol dependence phenotype. This mapping strategy compliments the linkage strategy and is useful if the marker and disease loci are in close proximity. Thus, this technique will gain greater importance as we saturate regions where linkage may be present. In addition although we have not set out to study candidate loci suggested by the literature, many have been tested since they are located near one of our highly polymorphic markers. Examples include the structural loci for DrD2, MAOB and ADH.

There are many exciting findings in this study and we are looking forward to continued significant progress in finding genes that affect the risk for alcohol dependence and modify correlated phenotypes. The rapid progress of the Human Genome Project is providing tools that will make these findings much more likely. It must be remembered however that genetics does not provide the whole answer to the cause of alcohol dependence. Although genes may increase or decrease the risk they do not determine ones fate. The presence of "genetic factors" in no way invalidates treatment or prevention. Indeed, characterizing the genetic factors will probably lead to a much clearer understanding of psychological and social factors as well.

## SELECTED REFERENCES

1. Bucholz KK, Hesselbrock VM, Shayka JJ, Numberger JJ, Schuckit MA, Reich T: Reliability of individual diagnostic criterion items for psychoactive substance dependence and the impact on diagnosis. *J Stud Alcohol* 56:500-504, 1995
2. S.A.G.E.: Statistical Analysis for Genetic Epidemiology, Release 2.2. Computer program package available from the Department of Biometry and

Genetics, LSU Medical Center, New Orleans, 1994

3. Rice JP, Neuman RF, Hoshaw SL, Daw ED, Gu C: TDT with covariates and genomic screens and mod scores: Their behavior on simulated data. *Genet Epidemiol* 12: 659-664, 1995

4. Risch N: Exclusion mapping for complex disease. *Am J Hum Genet* 53: A185, 1993 (abstr)

5. Kruglyak L, Lander ES: High-resolution genetic mapping of complex traits. *Am J Hum Genet* 56:1212-1223, 1995

6. Rice JPR, Goate A, Bierut L, et al: Initial genome scan of the NIMH genetics initiative bipolar pedigrees: Chromosomes 1,6,8,10 and 12. 1996 (submitted)

7. Lander ES, Green P: Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci USA* 84:2363-2367, 1987

8. Boehnke M: Allele frequency estimation from pedigree data. *Am J Hum Genet* 48:22-25, 1991

9. Porjesz B, Begleiter H: Human brain electrophysiology and alcoholism, in: Tarter DV, Thiel DV (eds.): *Alcohol and the Brain*. New York, Plenum, 1985, pp193-182

10. Begleiter H, Porjesz B, Bihari B, Kissin B: Event-related potentials in

boys at risk for alcoholism. *Science* 225:1493-1496, 1984

11. O'Connor SJ, Hesselbrock V, Tashman A, DePalma N: P3 amplitudes in two distinct tasks are decreased in young men with a history of paternal alcoholism. *Alcohol* 4:323-330, 1987

12. Whipple SC, Berman MS, Noble EP: Event-related potentials in alcoholic fathers and their sons. *Alcohol* 8:321-327, 1991

13. Berman MS, Whipple SC, Fitch RJ, Noble EP: P3 in young boys as a predictor of adolescent substance abuse. *Alcohol* 10:69-76, 1993

14. Hill SY, Steinhauer SR, Zubin J, Baughan T: Event-related potentials as markers for alcoholism risk in high density families. *Alcohol Clin Exp Res* 12: 545-555, 1988

15. Benegal V, Jain S, Subbukrishna DK, Channabasavanna SM: P300 amplitudes vary inversely with continuum of risk in first degree male relatives of alcoholics. *Psychiatr Genet* 5:1-8, 1995

16. O'Connor SJ, Morzorati S, Christian JC, Li TK: Heritable features of the auditory oddball event-related potential: Peaks, latencies, morphology and topography. *Electroencephalogr Clin Neurophysiol* 92:115-125, 1994

17. Blangero J: Multivariate oligogenic linkage analysis of quantitative traits in general pedigrees. *Am J Hum Genet* 57:11, 1995 (abstr)