American Journal of Medical Genetics Part B (Neuropsychiatric Genetics) 141B:844–853 (2006)

Pooled Association Genome Scanning for Alcohol Dependence Using 104,268 SNPs: Validation and Use to Identify Alcoholism Vulnerability Loci in Unrelated Individuals From the Collaborative Study on the Genetics of Alcoholism

Catherine Johnson,[1] Tomas Drgon,[1] Qing-Rong Liu,[1] Donna Walther,[1] Howard Edenberg,[2] John Rice,[3] Tatiana Foroud,[4] and George R. Uhl[1]*

[1]*Molecular Neurobiology Branch, NIDA-IRP, NIH, Baltimore, Maryland*
[2]*Department of Biochemistry and Molecular Biology, Indiana University, Indianapolis, Indiana*
[3]*Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri*
[4]*Department of Medicine, Indiana University, Indianapolis, Indiana*

Association genome scanning can identify markers for the allelic variants that contribute to vulnerability to complex disorders, including alcohol dependence. To improve the power and feasibility of this approach, we report validation of "100k" microarray-based allelic frequency assessments in pooled DNA samples. We then use this approach with unrelated alcohol-dependent versus control individuals sampled from pedigrees collected by the Collaborative Study on the Genetics of Alcoholism (COGA). Allele frequency differences between alcohol-dependent and control individuals are assessed in quadruplicate at 104,268 autosomal SNPs in pooled samples. One hundred eighty-eight SNPs provide (1) the largest allele frequency differences between dependent versus control individuals; (2) t values ≥ 3 for these differences; and (3) clustering, so that 51 relatively small chromosomal regions contain at least three SNPs that satisfy criteria 1 and 2 above (Monte Carlo $P = 0.00034$). These positive SNP clusters nominate interesting genes whose products are implicated in cellular signaling, gene regulation, development, "cell adhesion," and Mendelian disorders. The results converge with linkage and association results for alcohol and other addictive phenotypes. The data support polygenic contributions to vulnerability to alcohol dependence. These SNPs provide new tools to aid the understanding, prevention, and treatment of alcohol abuse and dependence.

KEY WORDS: microarray; complex genetics; substance dependence; addiction

Please cite this article as follows: Johnson C, Drgon T, Liu Q-R, Walther D, Edenberg H, Rice J, Foroud T, Uhl GR. 2006. Pooled Association Genome Scanning for Alcohol Dependence Using 104,268 SNPs: Validation and Use to Identify Alcoholism Vulnerability Loci in Unrelated Individuals From the Collaborative Study on the Genetics of Alcoholism. Am J Med Genet Part B 141B:844–853.

INTRODUCTION

Substance abuse vulnerabilities are complex traits with strong genetic influences documented by family and twin studies [Kaprio et al., 1982; Cadoret et al., 1986, 1995; Grove et al., 1990; Goldberg et al., 1993; Gynther et al., 1995; Uhl et al., 1995, 1997; Tsuang et al., 1996; Woodward et al., 1996; Kendler and Prescott, 1998; Merikangas et al., 1998; Tsuang et al., 1998; Kendler et al., 1999; Maes et al., 1999; Uhl, 1999; Karkowski et al., 2000]. Much of the genetic vulnerability to abuse of different legal and illegal addictive substances is shared; many abusers use multiple addictive substances [Kendler and Prescott, 1998; Tsuang et al., 1998, 1999; Kendler et al., 1999; Karkowski et al., 2000]. Identifying the allelic variants that contribute to vulnerability to alcohol dependence and comparing them to the variants that predispose to other addictions can improve understanding of human addictions and assist efforts to match vulnerable individuals with the prevention and treatment strategies most likely to work for them.

Association genome scanning can help determine which chromosomal regions and genes contain allelic variants that predispose to dependence on alcohol and other substances. This approach does not require family member participation, gains power as genomic marker densities increase [Risch and Merikangas, 1996; Cervino and Hill, 2000; Schork et al., 2000; Sham et al., 2000], identifies smaller chromosomal regions than linkage-based approaches, fosters pooling strategies that preserve confidentiality and reduce costs [Barcellos et al., 1997; Hacia et al., 1999; Germer et al., 2000; Uhl et al., 2001] and provides ample genomic controls that can minimize the chances of unintended ethnic mismatches between disease and control samples. We have used these approaches to assess allelic frequencies at 1,494 and then 11,522 SNPs in unrelated control versus polysubstance abusing individuals who report dependence on at least one illegal substance [Uhl et al., 2001; Liu et al., 2005]. SNPs that displayed nominally "reproducibly-positive" allele frequency differences between abuser and controls in both European- and African-American samples [Uhl et al., 2001] cluster closer to each other and to positive

Catherine Johnson, Tomas Drgon, and Qing-Rong Liu contributed equally to this work.

Grant sponsor: Intramural Research Program of the NIH; Grant sponsor: NIDA DHSS; Grant sponsor: NIDA-IRP.

*Correspondence to: George R. Uhl, Molecular Neurobiology, Suite 3510, 333 Cassell Drive, Baltimore, MD 21224.
E-mail: guhl@intra.nida.nih.gov

Received 2 December 2005; Accepted 3 April 2006
DOI 10.1002/ajmg.b.30346

markers from linkage studies of addictions than anticipated by chance [Uhl et al., 2002; Uhl, 2004]. However, this density of SNP markers provides information about possible associations with addiction for only a modest number of the blocks of restricted haplotype diversity found in these subjects' genomes. Association genome scanning has not yet been employed to study alcohol dependence, to our knowledge.

Unrelated individuals sampled from pedigrees collected by the Collaborative Study on the Genetics of Alcoholism provide an interesting sample for this approach for several reasons. Dependence on alcohol and other substances has been carefully characterized in these individuals using validated instruments. Unrelated control individuals free from substance abuse or dependence diagnoses, largely individuals who marry into these pedigrees, are available. Linkage work with these pedigrees has identified a number of interesting loci [Reich et al., 1998; Bierut et al., 2002]. In addition, Genetics Analysis Workshop data provides individual genotypes for over 14,000 SNPs for a subset of COGA individuals (http://www.gaworkshop.org).

"100k" SNP microarrays (Centurion™, Affymetrix) use size-selected PCR products of genomic restriction fragments that have been ligated to universal linker sequences and amplified using single PCR primer pairs. Early access versions of these arrays allow assessment of 104,268 SNPs that can be localized to autosomes and display minor allele frequencies $\geq$2%. These arrays thus allow studies of many more SNP markers for more unrelated individuals than previously available. The data also overlaps with genotypes obtained in some of these same individuals as part of the Genetics Analysis Workshop, providing a rich set of comparisons of individual versus pooled genotypes that help validate use of pooling with these samples.

We thus now report validation and use of pooled association genome scanning using 100k arrays hybridized with size-selected amplicons from end-ligated *Xba* I and *Hind* III genomic DNA restriction fragments of pooled genomic DNAs. DNAs come from unrelated COGA individuals who report (1) dependence on alcohol versus (2) control individuals free from any alcohol dependence, largely those who have married into these pedigrees. We use this approach to generate more than 29 million person/genotype equivalents, determined in quadruplicate. We discuss the convergence that these results provide with association and linkage studies for alcohol and other addictive phenotypes, the genetic architecture for alcohol dependence that the results support, the classes of candidate genes that they nominate for roles in human alcohol dependence and the implications of these findings for pooled association genome scanning approaches to complex genetic disorders.

MATERIALS AND METHODS

Research Volunteers

We searched COGA pedigrees to identify unrelated individuals who displayed phenotypes 2–3 ("pure unaffected" or "unaffected with some symptoms") or 4 ("affected," e.g., alcohol dependent). We identified 120 unrelated alcohol-dependent individuals and 160 unrelated unaffected controls who self-reported European-American ethnicities. Information was available for Genetics Analysis Workshop (GAW) genotypes for 120 of these individuals. DNAs from these 120 individuals were placed into four of the control pools and two of the pools of alcohol-dependent individuals. DNAs from other COGA subjects who were not included in GAW formed eight additional pools.

Genomic DNA

Genomic DNA was prepared from lymphoblastoid cell lines (Corriel Institute), requantitated by spectrophotometry, picogreen and Heochst dye fluorescence and diluted to 10 ng/μl. Validation studies compared: (1a) allelic determinations from individual CEPH DNAs versus (1b) results from pools (n = 2) of the same DNAs; and (2a) allelic determinations from individual COGA DNAs versus (2b) results from pools (n = 20) of the same DNAs. Other validation studies examined pool-to-pool variation and test-retest variation for each pool tested on four different sets of microarrays.

Allelic Frequencies in Polysubstance Abusers and Controls

Allelic frequencies in polysubstance abusers and controls were compared using pools made by carefully combining equal amounts of DNA from 20 individuals of the same phenotypes. We used hybridization probes prepared from genomic DNA as described (Affymetrix Genechip Mapping Assay Manual) with precautions to avoid contamination. Fifty nanograms of pooled genomic DNA was digested by *Xba* I or by *Hind* III, ligated to appropriate adaptors, amplified by PCR using 3 min 95°C hot start, 35 cycles of 20 sec 95°C/15 sec 59°C/15 sec 72°C, and a final 7 min 72°C extension. PCR products were purified (MinElute™ 96 UF kits, Qiagen, Valencia, CA), digested for 30 min with 0.04 U/μl DNase I to produce 30–200 bp fragments, end-labeled using terminal deoxynucleotidyl transferase and biotinylated dideoxynucleotides and hybridized to 100k arrays (Centurion, Affymetrix, Santa Clara, CA) which were stained and washed as described (Affymetrix Genechip Mapping Assay Manual) using immunopure strepavidin (Pierce, Milwaukee, WI), biotinylated antistreptavidin antibody (Vector Labs, Burlingame, CA), and R-phycoerythrin strepavidin (Molecular Probes, Eugene, OR). Arrays were scanned and fluorescence intensities quantitated using an Affymetrix array scanner as described [Uhl et al., 2001].

Chromosomal positions for each SNP were sought using NCBI and NETAFFYX (Affymetrix) data. Allele frequencies for each SNP in each DNA pool were assessed based on hybridization intensity signals from four arrays, allowing assessment of hybridization to the 20 "perfect match" cells on each array that are complementary to the PCR products from alleles "A" and "B" for each diallelic SNP. Each array was analyzed as follows: (1) "Background" values, the average fluorescence intensity from the 5% of cells with the lowest values, were subtracted from the fluorescence intensity of every cell; (2) background-subtracted values were normalized by division by the average value obtained from the 5% of cells with the highest values; (3) normalized hybridization intensities from the 20 array cells that corresponded to the perfect match "A" and "B" cells for each SNP were averaged; (4) "A/B ratios" were determined by dividing average normalized A values by average normalized B values; (5) arctangent transformations were applied to each ratio to aid combination of data from arrays hybridized and scanned on different days; (6) average arctan values from the four replicates of each experiment were determined; (7) mean and standard deviations of average arctan values for each diagnostic group were calculated; (8) SNPs that displayed any of three criteria were eliminated from further analyses: (i) SNPs with minor allele frequencies <0.02, determined using Affymetrix data from analyses of European-American chromosomes; (ii) SNPs on sex chromosomes; and (iii) SNPs whose chromosomal positions could not be adequately determined; (9) for the remaining 104,268 SNPs, mean arctan A/B ratios for abusers were divided by mean arctan A/B ratios for controls for the rest of the SNPs to form abuser/control ratios; (10) A "t" statistic for the differences between abusers and controls was generated using the formula:

$$t = \frac{\overline{X}_{\text{abuser}} - \overline{X}_{\text{control}}}{\sqrt{\frac{(n_{\text{abuser}}-1)\sigma^2_{\text{abuser}} + (n_{\text{control}}-1)\sigma^2_{\text{control}}}{n_{\text{abuser}}+n_{\text{control}}-2}} \times \sqrt{\frac{1}{n_{\text{abuser}}} + \frac{1}{n_{\text{control}}}}}$$

where $\overline{X}_{abuser}$ and $\overline{X}_{control}$ are means of "arctan A/B" values for pools of the same diagnostic group, n_{abuser} and $n_{control}$ are number of pools in corresponding diagnostic group and σ^2 is the variance of the mean of arctan A/B values for pools of the same diagnostic group.

Although there is no universally accepted method for analyzing association genome scanning data, we used a preplanned analysis (for favorable mention of similar approaches see [Bansal, 2001]). We identified SNPs with abuser/control ratios in the top or bottom 2.5% of all abuser versus control comparisons that also displayed t statistics ≥ 3 for the abuser versus control differences. We then sought evidence for clustering of these SNPs by focusing on chromosomal regions in which at least three of these outlier SNPs lay within 1 Mb of each other; we note that this somewhat arbitrary distance may or may not reflect the entire extent of long range linkage disequilibrium, which varies from chromosomal region to chromosomal region. We term these clustered, nominally positive SNPs "clustered positive SNPs", and focus our analyses on regions in which they lie (Table I).

To seek convergence between current and other association data, we compared the locations of the current clustered positive SNPs with SNPs that met criteria for reproducibly positive association in analyses of (1) European-American and African-American NIDA samples [Uhl et al., 2001], Liu et al., 2005; Liu, Uhl et al., in preparation] and (2) Japanese unrelated methamphetamine dependent and control individuals sampled from the Japanese Genetics Initiative on Drug Abuse (JGIDA), [Drgon et al., submitted].

We assessed the statistical power of our analyses. We used (1) the observed control or abuser pool to pool, standard deviations from the current datasets, (2) the mean abuser/control differences for the SNPs that provided the largest abuser/control differences from the current datasets, (3) $a = 0.05$, (4) sample sizes from the current datasets, (5) abuser/control ratios from the current dataset and the program PS v2.1.31 [Dupont and Plummer, 1990].

Observed results were compared to those expected by chance using 100,000 Monte Carlo simulation trials that sampled from a Microsoft SQL server database that contained the results from the current study: 14 pools x 4 arrays/pool x 20 perfect match cells/array/SNP x 104,268 SNPs = 116,780,160 cells (1120 cells/SNP) (see also Uhl et al. [2001]). For each of 100,000 simulation trials, a randomly selected set of SNPs was chosen and the same procedure that had been followed for the actual data was run. The number of trials for which the results from the randomly selected set of SNPs matched or exceeded the results actually observed from the SNPs identified in the current study was tabulated. Empirical *P*-values were calculated by dividing the number of trials for which the observed results were matched or exceeded by the total number of Monte Carlo simulation trials performed. Since this method examines the properties of the SNPs in the current dataset, it should be relatively robust in the face of a number of features that include the uneven distribution of Affymetrix SNP markers across the genome.

To provide insights into some of the genes that we nominate for further study since they might harbor variants that contribute to individual differences in addiction vulnerability, we sought an identifiable candidate gene(s) for each cluster of positive SNPs. We selected candidate gene nominees when multiple clustered-positive SNPs lay (1) within the gene or (2) in 3′ or 5′ flanking sequences that were contained on a block of high-restricted haplotype diversity along with exon sequences from that gene. We defined the blocks of high-restricted haplotype diversity using Haploviewer and data from CEPH individuals. Clusters that did not identify genes that meet these criteria are annotated as "intergene" in Table I. To assess the nominal false discovery rates for these genes, we obtained the joint false discovery rates for the clustered positive SNPs based on their individual q values, derived from the 104,268 t values and QVALUE software [Storey, 2002; Storey and Tibshirani, 2003]. To provide one of several possible controls for the possibility that observed abuser-control differences might reflect occult stratification and correspondingly different allelic frequencies at the SNPs that display these abuser/control differences, we note the SNPs for which European-American versus African-American ethnicity difference scores from MNB/NIDA control individuals [Liu et al., submitted] lie in the outlying 2.5% of all such differences (Table I).

RESULTS

There were 122,828 SNPs assigned to chromosomes 1–22 that were assessed using these arrays. Of these 104,268 SNPs displayed minor allele frequencies of >2% in European-American samples (Netaffyx), could be assigned reasonably accurate chromosomal locations and were thus used for subsequent analyses.

Pooled genotyping using 100k arrays displays features that support the validity of our results. Regression analyses examined the relationships between (a) "observed allele ratios", background-subtracted, normalized, arctangent transformed hybridization intensity ratio values obtained from six pools of COGA DNAs and (b) "expected allele ratios" the fraction of A and B alleles obtained from individual genotypes obtained for 7393 SNPs from these same individuals in work performed for the Genetics Analysis Workshop using Affymetrix "10k" arrays. Pearson correlation coefficients were 0.91–0.92 for each of these pools ($P < 0.001$ for each). There were 31 "quality control" SNPs that were assessed using both *Xba*I and *Hind*III arrays. Correlations between the intensity ratios for these SNPs yielded a Pearson correlation coefficient of 0.94.

Abuser/control hybridization ratios for the 104,268 SNPs examined here fell into nearly Gaussian distributions with mean values close to one (Fig. 1). There was modest variability of these assessments. Mean arctan A/B allele hybridization ratios for all SNPs assessed here ±standard errors of the mean (SEM) for pool-to-pool differences were 0.79 ± 0.028 for abusers and 0.79 ± 0.024 for controls. SEMs for the four replicate arrays that assessed each sample were 0.036 and 0.034 for abusers and controls, respectively.

For analyses, we selected (1) the 5,216 candidate positive markers that represented the 2.5% of SNPs with greatest and the 2.5% of SNPs with the smallest abuser/control ratios and (2) the 1474 SNPs for which abuser/control differences yielded t values ≥ 3. Six hundred sixty-seven SNPs satisfied both of these criteria; we note that these two criteria are neither totally dependent nor totally independent of each other, and we term the SNPs that satisfy both criteria "candidate positive SNPs". Chance findings of 667 SNPs that satisfy both criteria are rare. We performed 100,000 Monte Carlo simulations, each of which a sampled a random set of 5,216 SNPs. None of these simulation trials identified as many as 667 randomly selected SNPs that shared the properties (abuser/control differences and t values for these abuser control differences) found in the true results of these experiments, yielding Monte Carlo $P < 0.00001$.

These candidate positive SNPs clustered together in ways that would also not be expected by chance, although such clustering would be expected if they identified loci that contain allelic variants that distinguished alcohol-dependent subjects from control subjects. Three hundred sixty-two of these 667 candidate positive SNPs lay in 138 clusters in which they were positioned within 1 Mb of at least one other candidate positive SNP ($P = 0.03715$). One hundred eighty-eight of these candidate-positive SNPs lay in 51 clusters in which at least

TABLE I. SNPs That Meet Criteria for "Clustered Positives" Based on Allele Frequency Differences Between Alcohol-Dependent and Control Individuals, the Nominal Significance of These Differences and Their Proximity to Other Positive SNPs

#	SNP ID	Chr	Location	A/C	t value	Gene
1	rs10492925	1	5,105,744	0.84	3.04	Intergene
	rs10492948	1	5,672,766	0.81	4.02	Intergene
	rs10489535	1	6,278,091	1.20	3.10	BACH brain acyl-CoA hydrolase
2	rs7534135	1	33,557,463	1.38	4.32	Intergene
	rs10493053	1	33,557,901	1.33	3.21	Intergene
	rs4393118	1	33,877,235	1.19	4.30	CSMD2 CUB and Sushi multiple domains 2[c]
3	rs127600	1	57,435,105	1.31	3.24	**DAB1 disabled homolog 1**
	rs10493244	1	58,318,487	0.76	4.28	**DAB1 disabled homolog 1**
	rs7367176	1	58,561,256	1.19	3.70	**5′ flank DAB1 disabled homolog 1**
	rs338932	1	58,570,780	1.23	3.56	**5′ flank DAB1 disabled homolog 1**
4	rs1506700	1	84,588,752	1.23	3.53	**DLAD DNase II-like acid DNase**
	rs3121147	1	84,589,004	1.18	5.57	**DLAD DNase II-like acid DNase**
	rs10493747	1	84,615,307	1.24	3.53	**3′ flank DLAD DNase II-like acid DNase**
5	rs1339405	1	211,745,511	0.79	4.14	**KCNK2 potassium channel, subfamily K, member 2**
	rs2211127	1	211,862,409	0.78	4.32	**Intergene**
	rs2225974	1	211,862,601	0.83	3.09	**Intergene**
	rs4129019	1	212,381,255	1.16	3.05	USH2A Usher syndrome 2A
6	rs10495372	1	232,621,144	1.18	4.56	5′ flank TM7SF1transmembrane 7 superfamily member 1
	rs1528776	1	233,478,279	0.80	4.89	LOC440737
	rs1252144	1	233,478,438	0.74	3.52	LOC440737
7	rs10495616	2	13,567,962	0.81	3.21	Intergene
	rs2077411	2	14,316,043	1.17	3.47	Intergene
	rs10495636[a]	2	14,395,237	1.23	3.83	Intergene
	rs10495638	2	14,395,750	0.80	3.70	Intergene
8	rs2176221	2	51,898,640	1.23	3.19	Intergene
	rs2355813	2	52,298,006	1.22	5.72	Intergene
	rs350732	2	52,898,473	1.23	3.19	5′ flank LOC402072
	rs10496026	2	53,654,834	0.85	3.05	3′ flank LOC388949
9	rs2901780	2	76,801,066	1.17	3.27	**LRRTM4 leucine rich repeat transmembrane neuronal 4**[b,c]
	rs985343	2	76,823,437	1.25	3.76	**LRRTM4 leucine rich repeat transmembrane neuronal 4**[b,c]
	rs1921627	2	77,084,348	1.17	3.88	**LRRTM4 leucine rich repeat transmembrane neuronal 4**[b,c]
10	rs2885467	2	83,812,464	1.44	3.61	Intergene
	rs10496300	2	83,841,704	1.17	3.20	Intergene
	rs6729553	2	84,821,493	0.82	3.25	LOC200383
	rs10496316	2	85,659,029	0.84	3.46	5′ flank MAT2A methionine adenosyltransferase II, alpha
11	rs7566329	2	141,387,286	0.81	3.18	**LRP1B low density lipoprotein-related protein**[c]
	rs10496858	2	141,473,377	1.16	3.28	**LRP1B low density lipoprotein-related protein**[c]
	rs10496905	2	142,470,192	0.80	3.34	**LRP1B low density lipoprotein-related protein**[c]
12	rs10497494	2	178,332,390	1.25	3.55	PDE11A
	rs334005	2	178,953,076	1.27	3.38	OSBPL6 oxysterol binding protein-like 6
	rs1961416	2	179,696,678	1.40	3.40	LOC285026
	rs2008989[a]	2	179,796,596	1.28	3.14	**3′ flank SESTD1 SEC14 and spectrin domains 1**
	rs6736098	2	179,861,873	1.26	3.04	**SESTD1 SEC14 and spectrin domains 1**
	rs1158872[a]	2	179,953,209	0.77	4.00	**5′ UTR SESTD1 SEC14 and spectrin domains 1**
	rs932117	2	180,380,554	1.18	3.26	**ZNF533 zinc finger protein 533**
	rs4894122	2	180,469,360	1.24	3.70	**5′ flank ZNF533 zinc finger protein 533**
13	rs1369843	2	200,913,677	1.32	4.40	Intergene
	rs1347553	2	200,962,288	1.17	3.44	5′ flank DNAPTP6 DNA polymerase-transactivated protein 6
	rs719125	2	201,843,366	1.34	3.00	CFLAR CASP8 and FADD-like apoptosis regulator
	rs2349733	2	202,448,571	1.19	3.43	ALS2 amyotrophic lateral sclerosis 2
14	rs956270	3	21,018,625	0.77	4.53	Intergene
	rs2173234	3	21,045,946	1.31	3.07	Intergene
	rs958635	3	21,047,047	1.22	3.02	Intergene
	rs9310635	3	21,056,888	1.21	3.53	Intergene
	rs1370116	3	21,187,137	0.75	3.20	Intergene
15	rs10510945	3	65,792,243	0.76	4.52	**BAIAP1/MAGI1 membrane associated guanylate kinase, WW and PDZ domain containing 1**
	rs10510946	3	65,792,731	0.77	3.51	**BAIAP1/MAGI1 membrane associated guanylate kinase, WW and PDZ domain containing 1**
	rs10510947	3	65,804,384	1.19	3.04	**BAIAP1/MAGI1 membrane associated guanylate kinase, WW and PDZ domain containing 1**
16	rs718424[a]	3	149,863,241	0.73	3.90	**5′ flank AGTR1 angiotensin II receptor, type 1**
	rs275707	3	149,886,571	0.71	3.40	**5′ flank AGTR1 angiotensin II receptor, type 1**
	rs2639375	3	149,887,816	0.67	3.13	**5′ flank AGTR1 angiotensin II receptor, type 1**
	rs10513338[a]	3	149,952,288	0.76	5.03	**3′ flank AGTR1 angiotensin II receptor, type 1**
17	rs4305478	4	166,397,584	1.19	3.42	FLJ38482
	rs1370687	4	166,747,960	1.23	3.08	**CPE carboxypeptidase E**
	rs1837171	4	166,748,122	1.23	3.57	**CPE carboxypeptidase E**

(*Continued*)

TABLE I. (*Continued*)

#	SNP ID	Chr	Location	A/C	t value	Gene
18	rs1643658	5	79,972,097	0.83	3.54	DHFR dihydrofolate reductase
	rs2897262	5	79,998,432	0.79	3.62	**MSH3 mutS homolog 3 (*E. coli*)**
	rs1650663	5	79,998,953	1.17	3.37	**MSH3 mutS homolog 3 (*E. coli*)**
19	rs152608	5	106,799,108	0.83	3.18	**EFNA5 ephrin-A5**
	rs164838	5	107,008,867	0.80	3.42	**EFNA5 ephrin-A5**
	rs770167	5	107,154,817	0.71	3.09	Intergene
	rs770166	5	107,155,126	0.68	3.26	Intergene
	rs10515391	5	108,032,094	1.20	3.35	Intergene
20	rs1318774	5	123,246,619	0.80	3.23	Intergene
	rs2129846	5	123,658,466	0.81	3.99	Intergene
	rs696479	5	123,699,995	1.16	3.36	Intergene
	rs10519804	5	124,560,322	1.16	3.06	Intergene
21	rs4091539	5	128,423,826	0.78	4.83	Intergene
	rs42562	5	128,432,225	0.75	4.65	Intergene
	rs10520031	5	128,529,519	1.26	3.99	Intergene
22	rs1549920	5	151,121,037	0.87	3.33	5′ flank ATOX1 ATX1 antioxidant protein 1 homolog
	rs10515675	5	151,993,982	1.16	3.39	5′ flank NMUR2 neuromedin U receptor 2[c]
	rs4310018	5	152,266,397	1.17	4.08	Intergene
	rs10515678	5	152,303,429	0.81	3.49	Intergene
23	rs261612	5	169,151,419	1.25	3.31	**DOCK2 dedicator of cytokinesis 2**
	rs1477316	5	169,372,306	1.26	3.22	**DOCK2 dedicator of cytokinesis 2**
	rs1477317	5	169,372,320	1.28	3.14	**DOCK2 dedicator of cytokinesis 2**
24	rs10498762	6	45,851,264	1.26	3.43	Intergene
	rs953062	6	46,734,312	0.77	4.06	SLC25A27 solute carrier family 25, member 27
	rs1490296	6	47,228,712	1.27	3.63	Intergene
25	rs989191	6	79,970,675	0.72	3.33	HMGN3 high mobility group nucleosomal binding domain 3
	rs3846741	6	80,198,421	0.83	4.78	Intergene
	rs2803183	6	80,235,582	0.87	5.54	Intergene
26	rs1753826	6	91,283,465	0.71	3.84	**MAP3K7 mitogen-activated protein kinase kinase kinase 7**
	rs806284	6	91,300,645	0.85	3.92	**MAP3K7 mitogen-activated protein kinase kinase kinase 7**
	rs1145735	6	91,335,141	1.19	3.16	**MAP3K7 mitogen-activated protein kinase kinase kinase 7**
27	rs2181069	6	155,896,762	1.16	3.26	Intergene
	rs1391655	6	156,092,837	1.34	3.82	Intergene
	rs9322535	6	156,158,886	0.85	3.52	Intergene
28	rs2191480	7	14,607,796	1.33	3.00	**DGKB diacylglycerol kinase, beta 90 kDa**[c]
	rs217560	7	14,684,842	1.17	3.69	**5′ flank DGKB diacylglycerol kinase, beta 90 kDa**[c]
	rs2191349	7	14,837,549	1.29	3.11	Intergene
29	rs10486606	7	28,956,707	1.19	4.07	CPVL carboxypeptidase, vitellogenic-like
	rs6978690	7	29,364,669	1.28	3.97	Intergene
	rs10499584	7	29,364,852	1.17	3.08	Intergene
	rs2391802	7	29,636,004	1.21	3.95	LOC441208
30	rs7789889	7	33,916,358	1.20	3.30	BMPER BMP-binding endothelial regulator precursor protein
	rs714588	7	34,466,466	1.19	3.78	**5′ flank GPR154 G protein-coupled receptor 154/NPS receptor**[c]
	rs1419794	7	34,474,245	0.84	3.54	**GPR154 G protein-coupled receptor 154/NPS receptor**[c]
	rs324960	7	34,575,631	1.15	3.97	**GPR154 G protein-coupled receptor 154/NPS receptor**[c]
31	rs321967	7	77,959,200	1.17	3.64	**AIP1**[c]
	rs2190662	7	77,966,715	1.18	4.14	**AIP1**[c]
	rs2190665	7	77,966,810	0.71	4.73	**AIP1**[c]
32	rs17071791	8	4,744,386	0.83	4.93	**Intron I CSMD1 CUB and Sushi multiple domains 1**[c]
	rs1377881	8	4,948,992	0.85	3.74	**5′ flank CSMD1 CUB and Sushi multiple domains 1**[c]
	rs7822993	8	5,187,543	0.85	3.48	**5′ flank CSMD1 CUB and Sushi multiple domains 1**[c]
	rs4146469	8	5,253,259	1.20	3.10	**5′ flank CSMD1 CUB and Sushi multiple domains 1**[c]
	rs10503302	8	5,269,281	1.29	3.17	**5′ flank CSMD1 CUB and Sushi multiple domains 1**[c]
33	rs10503688	8	20,635,206	1.34	3.12	Intergene
	rs1459328[a]	8	21,141,281	1.20	3.12	Intergene
	rs9314269	8	21,860,935	0.88	3.22	XPO7 exportin 7
34	rs1381113	8	28,202,619	1.18	3.18	Intergene
	rs524458	8	29,148,550	1.19	4.77	KIF13B kinesin family member 13B
	rs6558132	8	29,526,866	1.28	3.60	LOC392208similar to 60S ribosomal protein L17 (L23)
35	rs3935233	8	39,307,991	0.84	3.73	ADAM5 a disintegrin and metalloproteinase domain 5
	rs2980817	8	40,155,783	0.86	3.16	Intergene
	rs966169	8	40,156,816	1.21	3.58	Intergene
	rs724322	8	40,157,961	0.88	4.26	Intergene
	rs2980813	8	40,169,167	0.87	3.40	Intergene
36	rs723085	8	77,949,261	1.24	3.51	3′ flank ZFHX4 zinc finger homeodomain 4
	rs2128944	8	77,957,795	1.25	5.01	3′ flank ZFHX4 zinc finger homeodomain 4
	rs1545881	8	77,983,105	1.21	3.36	3′ flank ZFHX4 zinc finger homeodomain 4

TABLE I. (*Continued*)

#	SNP ID	Chr	Location	A/C	t value	Gene
37	rs10504751	8	83,374,045	0.82	3.69	LOC389674
	rs1404759	8	83,850,716	0.72	3.23	Intergene
	rs10504770	8	83,886,603	1.21	3.16	Intergene
38	rs572811	9	75,329,015	1.22	3.11	**5′ flank proprotein convertase subtilisin/kexin type 5 PCSK5**[c]
	rs10512042	9	75,348,913	1.33	4.23	**5′ flank proprotein convertase subtilisin/kexin type 5 PCSK5**[c]
	rs10521468	9	76,033,066	1.22	4.30	**proprotein convertase subtilisin/kexin type 5 PCSK5**[c]
	rs2270571	9	76,033,485	0.83	4.19	**proprotein convertase subtilisin/kexin type 5 PCSK5**[c]
39	rs7897412	10	82,792,026	1.26	3.20	Intergene
	rs7093354	10	82,800,051	1.26	3.90	Intergene
	rs7069222[a]	10	83,746,082	1.23	3.28	NRG3 neuregulin 3 NRG3 Pro-neuregulin-3 precursor
40	rs4298845	10	84,889,336	1.24	5.02	3′ flank NRG3 neuregulin 3
	rs10509466	10	84,890,641	1.40	3.97	Intergene
	rs987312	10	85,304,655	1.25	3.33	Intergene
	rs2350152	10	85,778,039	0.84	4.46	LOC439991
41	rs10505947	12	24,811,933	1.21	3.58	LOC387846 hypothetical LOC387846
	rs1489907	12	25,447,971	1.18	3.77	3′ flank of FLJ36004; likely ortholog of mouse Pas1 candidate 1
	rs7301836	12	25,537,419	0.83	3.11	FLJ36004; likely ortholog of mouse Pas1 candidate 1
	rs1049380	12	26,380,811	1.16	5.22	**3′ ITPR2 inositol 1,4,5-triphosphate receptor, type 2**
	rs4654	12	26,381,422	0.86	3.22	**3′ ITPR2 inositol 1,4,5-triphosphate receptor, type 2**
42	rs724026	12	112,555,413	1.23	3.49	Intergene
	rs3782422	12	112,722,445	0.80	4.79	RBM19 RNA binding motif protein 19
	rs2114865	12	113,034,425	0.82	3.22	Intergene
43	rs1927384	13	101,943,751	1.20	3.09	Intergene; 5′ flank of TPP2 tripeptidyl peptidase II
	rs279929	13	102,538,720	1.39	3.50	3′ flank SLC10A2 solute carrier family 10 (sodium/bile acid cotransporter family), member 2
	rs1529281	13	102,706,611	0.79	3.76	Intergene
	rs951348	13	102,788,188	0.86	3.10	Intergene
44	rs726449	13	107,949,490	0.88	3.19	**5′ flank MYR8 myosin heavy chain Myr 8**
	rs1033871	13	108,629,670	0.88	3.18	**MYR8 myosin heavy chain Myr 8**
	rs1328837	13	108,677,100	0.83	4.67	**3′ flank MYR8 myosin heavy chain Myr 8**
45	rs10498431	14	51,213,444	1.20	3.54	**C14orf31 FRMD6 FREM domain containing 6**
	rs9285578	14	51,222,351	0.85	3.62	**C14orf31 FRMD6 FREM domain containing 6**
	rs10498445	14	51,810,191	0.86	3.50	PTGDR prostaglandin D2 receptor
46	rs10518930	15	35,280,878	1.35	3.67	Intergene
	rs603575	15	35,462,954	0.85	3.44	Intergene
	rs10518950[a]	15	35,539,179	0.74	3.34	Intergene
	rs10518952	15	35,567,312	1.20	3.69	5′ flank LOC390576
	rs10520086[a]	15	35,861,390	1.24	3.21	Intergene
47	rs8049647	16	50,695,297	1.18	3.46	5′ flank LOC388276
	rs10521256	16	50,764,161	1.24	3.03	5′ flank LOC388277
	rs7206384	16	51,299,759	1.18	3.02	Intergene
48	rs40510	16	63,530,623	1.25	3.51	**3′ flank CDH11 cadherin 11, type 2**
	rs35164	16	63,532,702	1.19	4.28	**3′ flank CDH11 cadherin 11, type 2**
	rs35200	16	63,579,045	0.87	3.10	**CDH11 cadherin 11, type 2**
	rs10500508	16	63,679,379	1.19	4.03	**CDH11 cadherin 11, type 2**
49	rs10514542	16	81,115,185	0.88	3.21	**5′ flank cadherin 13**[c]
	rs192599	16	81,443,739	1.21	4.81	**cadherin 13**[c]
	rs7206473	16	81,458,377	0.73	3.91	**cadherin 13**[c]
50	rs10515112	17	51,240,116	0.76	3.07	3′ flank PCTP phosphatidylcholine transfer protein
	rs1879143	17	52,113,556	1.17	4.44	Intergene
	rs721427	17	52,519,635	0.85	3.13	**AKAP1 A kinase (PRKA) anchor protein 1**
	rs998113	17	52,520,226	1.32	3.08	**AKAP1 A kinase (PRKA) anchor protein 1**
	rs4793900	17	53,145,196	0.84	3.31	**3′ flank MSI2 musashi homolog 2 (*Drosophila*)**
	rs4793902	17	53,149,218	0.78	3.54	**3′ flank MSI2 musashi homolog 2 (*Drosophila*)**
51	rs10502332	18	5,703,232	1.18	4.49	LOC388459 TTM2 two transmembrane protein A
	rs341184	18	6,355,614	1.19	3.08	L3MBTL4 l(3)mbt-like 4
	rs971529	18	6,822,779	1.47	4.53	LOC388462/ARHGAP28 Rho GTPase activating protein 28

Shown are: cluster number, reference SNP (rs) identifier, chromosome number, chromosomal position (bp, NCBI), abuser/control ratio (A/C), t value for abuser/control ratio and nearby candidate genes, identified using criteria enumerated in Materials and Methods.

[a]SNPs for which allele frequencies in African-American versus European-American control samples lie among the most different sets of values.

[b]SNPs labeled as indicating "LRRTM4 leucine rich repeat transmembrane neuronal 4" since they lie near ESTs from brain that are likely to represent parts of this gene.

[c]Genes that are discussed in more detail in the text are indicated in boldface and genes that are supported by preliminary clustered positive data from additional samples (QRL, TD, CJ GRU and others in preparation).

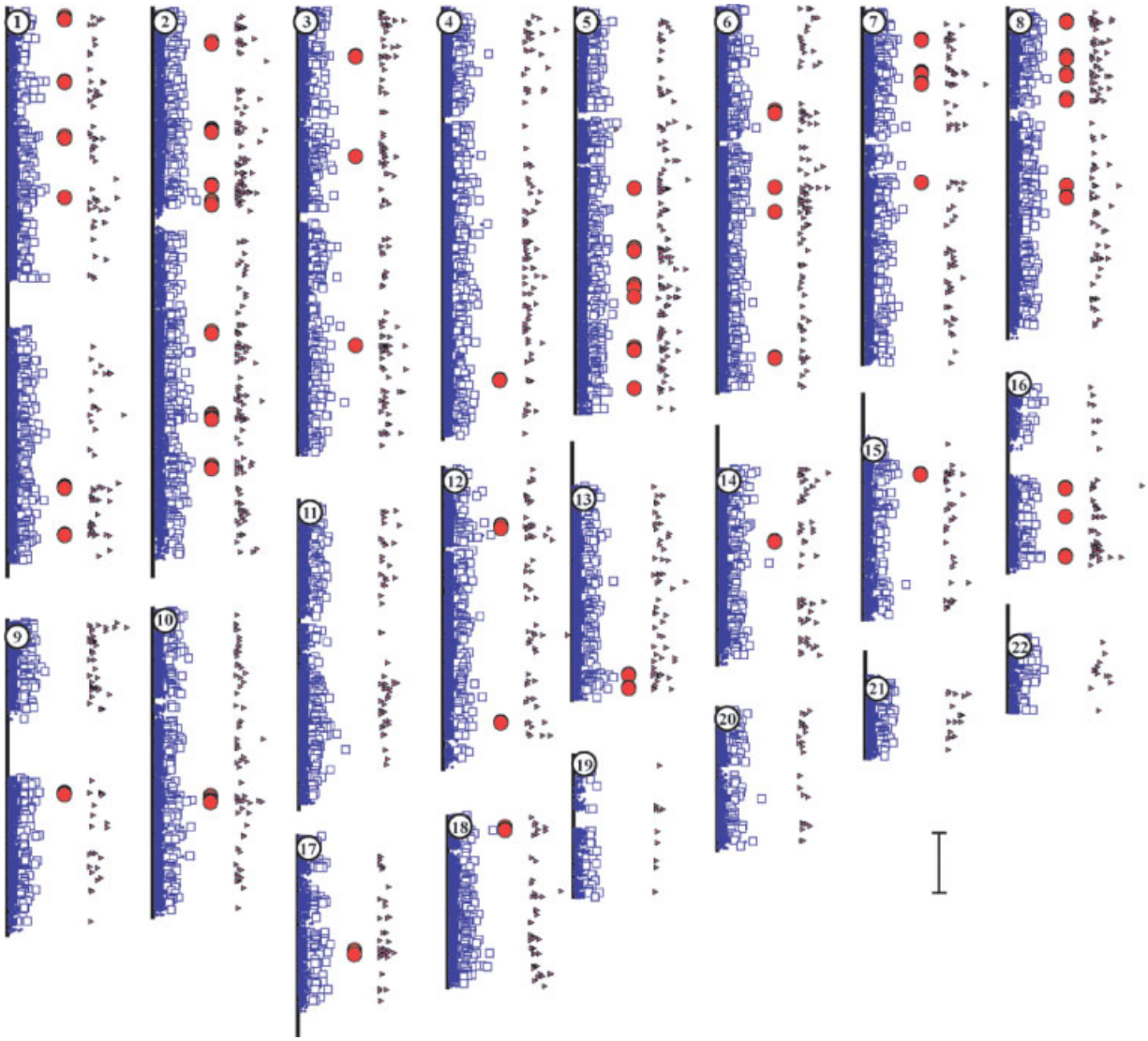

Fig. 1. Main axes: abuser/control ratios to the chromosomal position of each SNP for COGA alcohol-dependent and control individuals. The positions of the SNPs whose data yield outlier abuser/control values are indicated by larger symbols. Supplementary axis (right of main axis): SNPs for which abuser/control differences display t values ≥ 3. Red dots designate clustered positive SNPs that display outlier abuser/control and t values. Scale: chromosomal positions based on NCBI Map Viewer coordinates and supplemental data from NETAFFYX. The vertical bar represents 25 MB.

three positive SNPs met the same criterion ($P = 0.00034$). Thus when we performed 100,000 Monte Carlo simulation trials in each of which a random set of 667 SNPs was sampled, only 34 such trials involved 188 or more SNPs in such clusters.

We focus our subsequent analyses on these 51 clusters of candidate-positive SNPs (Table I). We identify candidate gene nominees for many of these 51 clusters of positive SNPs based on information from Mapviewer, HapMap, and Unigene and the criteria noted in Materials and Methods.

The clustered-positive results from this dataset can be compared with results from other association and linkage results for addictions. Fourteen of the two-SNP clusters and four of the three-SNP clusters from the present work lie within 1 Mb of at least one of the clustered positive results obtained from both NIDA European-American and NIDA African-American polysubstance abusers who report dependence on at least one illegal addictive substance ($P \leq 0.00001$ for both comparisons). These results provide additional support for positive SNP clusters 9, 11, 32, and 38 in the current work. Ten of the two-SNP clusters and five of the three-SNP clusters from a study of methamphetamine-dependent Japanese individuals versus controls also lie within 1 Mb of at least one of the current clusters ($P < 0.00001$), providing additional support for clusters 11, 20, 22, 31, and 49 from the current work. Of the 26 genes that are identified here by multiple-clustered positive SNPs, LRP1B, AIP1, CDH13, LRRTM4, CSMD1, PCSK5, CSMD2, GPR154, and DGKB also contain clustered positive results from at least one other addiction association genome-scanning sample. NMUR2 is also identified by clustered-positive results from other samples. This level of replication is especially remarkable since these convergences were sought for samples from different ethnic backgrounds and different addictions. Such a level of replication is consistent with false discovery rate calculations for the 51 loci, which range from 0.06 for joint false discovery rates for the eight positive SNP cluster to ca. 0.33 for the three positive SNP clusters. Such a level of replication is also consistent with simulation-based Monte Carlo P-values for each of these loci. Each of 100,000

trials selected a genomic segment that started with the beginning of a randomly selected annotated gene, continued 3′ for the same number of bases as that identified by the positive cluster and added an additional 1 Mb at either end of the segment. For each trial, the genomic segment was assessed to identify whether a cluster of positive SNPs with the same properties identified in the true dataset lay in the region. These studies appear to yield *P*-values that range from 0.001 to 0.03, uncorrected for the 51 multiple comparisons [Johnson et al., 2006].

Based on the number of pools assessed here and the pool-to-pool variability actually observed in these experiments (SEMs for four replicate arrays 0.03; mean ± SEMs for pool-to-pool variation 0.62 ± 0.02), we calculate 0.9 power to detect abuser/control allele frequency differences of 0.05.

DISCUSSION

The results of this study support the idea that array-based pooled association genome scanning approaches can identify chromosomal regions likely to contain allelic variants that differ in frequencies between alcohol-dependent and control individuals. This current identification of such frequency differences in alcohol-dependent individuals provides the first genome-wide association-based assessment for genomic loci likely to contain variants that contribute to dependence on alcohol. We discuss the strengths and possible limitations of these results, the ways in which they converge with results of previous association- and linkage-based studies and the classes of genes that they nominate to play roles in human vulnerabilities to alcohol dependence.

The reliability and validity of the current approach is supported by data that documents the reliability of clinical assessments made by multiple observers and the extent to which the markers that display nominally positive differences between abusers and controls cluster together in specific chromosomal regions in these samples. We and others have also provided extensive evidence for the reliability and validity of pooling approaches using related microarray-based assays [Butcher et al., 2004; Liu et al., 2005; Drgon et al., submitted]. Correlations between the current data and preliminary data obtained using the same samples and "10k" Affymetrix arrays were 0.98 for the overlapping SNPs that displayed outlier abuser/control values [CJ, QRL, TD, GRU et al., unpublished observations, 2004].

Modeling studies support significant power for the current methods and also support the likelihood of both false-positive and false-negative results. Power calculations support 0.9 power to detect 5% allele frequencies differences in the current experiments. Despite the relatively high marker density used in this report, however, there are still likely to be haplotype blocks that contain vulnerability-modifying alleles but do not contain three SNPs that are assessed in this report. Such blocks could thus provide false negative results in these studies. False positives are also likely, since we make many comparisons in this study. Simulation studies suggest a very low likelihood that all of the clustered positive results displayed here represent false positives. However, false positive results are still likely even among the clustered positive SNPs. False discovery rates lie between 0.06 and ca. 0.33 for the different clusters. Many, but not all, of these findings are supported by positive results from association genome scans of different addictions studied in different populations.

Since the COGA sample was not collected for association studies, it is possible that there might be occult stratification within these sample sets. We have examined this by comparing mean differences between arctangent transformed A/B allelic ratios between European- and African-American samples collected in Baltimore, Maryland [Liu et al., 2006] for the SNPs that are 5% outliers among the Affymetrix SNPs from the current dataset and with data for all Affymetrix SNPs. The average normalized allele frequency ethnic difference for the SNPs that displayed outlier alcoholic/control values was 0.154. The average normalized ethic differences for all of the SNPs represented in the current dataset was 0.147. There was thus no evidence for overall stratification.

These results support the possibility that careful evaluation of associations within unrelated members of samples collected for linkage may be possible with the relatively high marker densities provided by SNP methods, given the genomic controls that these high-SNP densities can also provide. These data and their convergence with prior results continue to provide support for the idea that common allelic variants contribute to human vulnerability to abuse of addictive substances. Finding a number of SNPs with substantial abuser-control differences near markers previously linked to alcoholism in this same dataset supports the idea that further fine mapping studies using association approaches in these samples, as well as others, might help to better define the specific genes, haplotypes and gene variants that contribute to previously observed linkage signals in these datasets.

When we assess the extent to which the SNPs that display outlier abuser/control values also display outlier t values, the observed results are found rarely by chance in simulation studies. When we examine the degree to which these nominally positive SNPs cluster together in groups of three or more on modest-sized chromosomal regions we also observe striking departures from chance values. Forty-two of the 51 clusters identified here contain positive SNPs from both *Xba* I and *Hind* III arrays. Twenty-two of the 51 clusters identified here receive at least some support from another linkage or association study; others also receive support from candidate gene studies (see below).

Assessing convergence of the current data with results of linkage analyses identifies 16 simple sequence length polymorphism (SSLP) markers that were previously linked to alcohol phenotypes with nominal statistical significance that lie within ±5 Mb of clustered positive results from the current study. Seven of these markers lie near linked markers from analyses of alcohol dependence in COGA pedigrees [Reich et al., 1998]; four from linkage analyses of alcohol dependence in Southwest Indians [Long et al., 1998] and five from linkage analyses of alcohol quantity/frequency phenotypes in data from the Framingham study [Bergen et al., 2003; Ma et al., 2003]. Each of these observations provide additional levels of support for the validity of the observed clustering.

Interesting candidate gene nominees lie near many of the clustered positive markers identified in this work. Cell-signaling molecule genes that lie near reproducibly positive SNPs (Table I) include those that signal within cells and between cells. Peptide signaling is implicated. Clustered-positive SNPs lie just in the 5′ flank of the GPR154 G protein-coupled receptor 154 that has been characterized as the receptor for neuropeptide S [Xu et al., 2004]. Several SNPs lie in 5′ and 3′ flanks of the AGTR1 angiotensin II receptor, type 1 gene. Clustered positive SNPs also flank and/or lie within enzymes that function to convert propeptides to biologically active peptides, including CPE carboxypeptidase E and proprotein convertase subtilisin/kexin type 5 (PCSK5).

Intracellular signaling with several different second messenger systems is implicated.

Phospholipid-signaling pathways could be altered by variations in several genes that lie near clustered-positive SNPs. Positive SNPs cluster in the 5′ flank and within the DGKB diacylglycerol kinase, beta 90 kDa gene, the gene that encodes

the ITPR2 inositol 1,4,5-triphosphate receptor, type 2 and the MAP3K7 mitogen-activated protein kinase kinase kinase 7 gene. Other phosphorylation patterns could well be altered by differences in the activities of the genes that encode the WW and PDZ domain containing BAIAP1/MAGI1 membrane-associated guanylate kinase and the anchor protein for AKAP1A kinase (PRKA).

Channels are implicated by these results. The KCNK2 potassium channel, subfamily K, member 2 is implicated by multiple positive SNPs.

Gene regulatory and/or developmental genes lie near reproducibly positive SNPs. The ephrin EFNA5 gene's 5′ flank contains positive SNPs that support roles for variations in this single transmembrane domain receptor protein kinase in addiction vulnerability. The DAB1-disabled homolog 1, DOCK2 dedicator of cytokinesis 2, CSMD1 CUB, and Sushi multiple domains 1, SESTD1 SEC14 and spectrin domains 1, ZNF533 zinc finger protein 533, and the MSH3 mutS homolog 3 (*E. coli*) genes each contain multiple clustered positive SNPs. The 3′ flank of the MSI2 musashi homolog 2 contains multiple positive SNPs. Each of these genes' products could alter brain developmental and/or adult form and function with consequences for addiction vulnerability.

The atrophin-1 interacting protein 1 (*AIP1*) gene is a disease-related gene that lies near clustered positive SNPs from the current dataset and reproducibly positive SNPs in studies of African- and European-American polysubstance abusers versus controls (see Liu et al. [2005]). This gene [Wood et al., 1998] is expressed largely in brain where it interacts with proteins including atrophin, the protein in which trinucleotide repeat expansions cause dentatorubral and pallidoluysian atrophy.

We have identified clustered positive SNPs near the genes that encode documented or suspected cell adhesion molecules and their possible ligands. These genes include the LRP1B low-density lipoprotein-related protein, cadherin 11 and cadherin 13 genes. Cadherin 13 is expressed in neurons and is abundant in interesting brain regions including amygdala. We have previously identified clustered-positive SNPs in the cadherin 13 gene in comparisons of methamphetamine abusers with controls [Drgon et al., submitted]. A number of SNPs are 3′ to the sequences currently annotated as the LRRTM4 leucine rich repeat transmembrane neuronal 4 gene. These SNPs lie near ESTs that derive from brain and seem likely to signal previously unelucidated more 3′ portions of this gene. These data add to previous nomination and or/confirmation of addiction-associated variants in cell adhesion molecules including neurexin 3 [Liu et al., 2005], NrCAM [Ishiguro et al., 2005], PTPRB [Ishiguro et al., submitted], the minor histocompatibility antigen HB-1 [Liu et al., 2005], multimerin 1 [Drgon et al., submitted], ADAM23 [Drgon et al., submitted], the FAT tumor suppressor homolog 3 and the Downs syndrome cell adhesion molecule [Drgon et al., submitted].

Clustered positive SNPs also lie near genes with other diverse cellular functions. DLAD DNase II-like acid DNase, MYR8 myosin heavy chain Myr 8, and the C14orf31 that encodes the FRMD6 FREM domain containing 6 protein each contain multiple clustered positives SNPs. In addition, clustered-positive SNPs also lie near genes that encode proteins of unknown function, including a number of hypothetical proteins (Table I).

While these data nominate interesting genes, it is only confirmation in multiple datasets in ongoing and future studies that will link each of them securely to addiction vulnerability. In preliminary results from higher density genome scanning studies from at least three additional samples, several of these genes receive substantial support (Table I, TD, QRL, CJ, GRU and others in preparation). Nevertheless, the current data provide support for loci nominated in prior SNP association and linkage-based studies and identify new chromosomal regions with clustered-positive SNPs and interesting genes. They provide a set of genomic markers in these 51 chromosomal regions that should be useful in subsequent studies of alcohol abusers. As we identify more and more of the allelic variants that contribute to vulnerability to abuse of alcohol and other substances, we will be better able to understand addictions themselves.

ACKNOWLEDGMENTS

We acknowledge passionate statistical discussions and help from Dr. Daniel Naiman, Department of Mathematical Sciences, Johns Hopkins University. For assistance in obtaining these datasets and samples, we are especially grateful the Genetics Analysis Workshop, NIAAA and COGA investigators, including PI: H. Begleiter; co-PIs L. Bierut, H. Edenberg, V. Hesselbrock, and B. Porjesz; 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 Downstate Medical Center (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 Southwest Foundation (L. Almasy) and Zhaoxia Ren as NIAAA staff collaborator. We acknowledge support for sample and data collection and storage from U10AA008401 (NIAAA and NIDA). COGA investigators especially acknowledge the fundamental scientific contributions of the late Theodore Reich, M.D., Co-Principal Investigator of COGA from its inception and a founder of modern psychiatric genetics. The authors acknowledge assistance from NIAAAA, the Genetics Analysis Workgroup and members of the Collaborative Study on the Genetics of Alcoholism.

REFERENCES

Bansal A. 2001. Trends in reporting of SNP associations. Lancet 358:2016.

Barcellos LF, Klitz W, Field LL, Tobias R, Bowcock AM, Wilson R, Nelson MP, Nagatomi J, Thomson G. 1997. Association mapping of disease loci, by use of a pooled DNA genomic screen. Am J Hum Genet 61:734–747.

Bergen AW, Yang XR, Bai Y, Beerman MB, Goldstein AM, Goldin LR. 2003. Genomic regions linked to alcohol consumption in the Framingham Heart Study. BMC Genet 4(Suppl 1):S101.

Bierut LJ, Saccone NL, Rice JP, Goate A, Foroud T, Edenberg H, Almasy L, Conneally PM, Crowe R, Hesselbrock V, et al. 2002. Defining alcohol-related phenotypes in humans. The Collaborative Study on the genetics of Alcoholism. Alcohol Res Health 26:208–213.

Butcher LM, Meaburn E, Liu L, Fernandes C, Hill L, Al-Chalabi A, Plomin R, Schalkwyk L, Craig IW. 2004. Genotyping pooled DNA on microarrays: A systematic genome screen of thousands of SNPs in large samples to detect QTLs for complex traits. Behav Genet 34:549–555.

Cadoret RJ, Troughton E, O'Gorman TW, Heywood E. 1986. An adoption study of genetic and environmental factors in drug abuse. Arch Gen Psychiatry 43:1131–1136.

Cadoret RJ, Yates WR, Troughton E, Woodworth G, Stewart MA. 1995. Adoption study demonstrating two genetic pathways to drug abuse. Arch Gen Psychiatry 52:42–52.

Cervino AC, Hill AV. 2000. Comparison of tests for association and linkage in incomplete families. Am J Hum Genet 67:120–132.

Drgon T, Lui OR, Johnson C, Walther D, Hishimoto A, Ujike H, Komiyama T, Harano M, Sekine Y, Inada T, Ozaki N, Iyo M, Iwata N, Yamada M, Sora I and Uhl GR. 2006. Addiction molecular genetics in Japanese methamphetamine-dependent individuals: Pooled association genome scanning identifies addiction vulnerability loci and genes. (submitted).

Dupont WD, Plummer WD Jr. 1990. Power and sample size calculations. A review and computer program. Control Clin Trials 11:116–128.

Germer S, Holland MJ, Higuchi R. 2000. High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. Genome Res 10:258–266.

Goldberg J, Henderson WG, Eisen SA, True W, Ramakrishnan V, Lyons MJ, Tsuang MT. 1993. A strategy for assembling samples of adult twin pairs in the United States. Stat Med 12:1693–1702.

Grove WM, Eckert ED, Heston L, Bouchard TJ Jr, Segal N, Lykken DT. 1990. Heritability of substance abuse and antisocial behavior: A study of monozygotic twins reared apart. Biol Psychiatry 27:1293–1304.

Gynther LM, Carey G, Gottesman II, Vogler GP. 1995. A twin study of non-alcohol substance abuse. Psychiatry Res 56:213–220.

Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, et al. 1999. Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. Nat Genet 22:164–167.

Ishiguro H, Liu QR, Gong JP, Hall FS, Ujike H, Morales M, Sakurai T, Grumet M, Uhl GR. 2006. NrCAM in addiction vulnerability: Positional cloning, drug-regulation, haplotype-specific expression, and altered drug reward in knockout mice. Neuropsychopharmacology 31:572–584.

Kaprio J, Hammar N, Koskenvuo M, Floderus-Myrhed B, Langinvainio H, Sarna S. 1982. Cigarette smoking and alcohol use in Finland and Sweden: A cross-national twin study. Int J Epidemiol 11:378–386.

Karkowski LM, Prescott CA, Kendler KS. 2000. Multivariate assessment of factors influencing illicit substance use in twins from female-female pairs. Am J Med Genet 96:665–670.

Kendler KS, Prescott CA. 1998. Cocaine use, abuse and dependence in a population-based sample of female twins. Br J Psychiatry 173:345–350.

Kendler KS, Karkowski LM, Corey LA, Prescott CA, Neale MC. 1999. Genetic and environmental risk factors in the aetiology of illicit drug initiation and subsequent misuse in women. Br J Psychiatry 175:351–356.

Liu QR, Drgon T, Johnson C, Walther D, Hess J and Uhl GR. 2006. Addiction molecular genetics: 639,401 SNP whole genome association reveals many "cell adhesion" gene variants. (submitted).

Liu QR, Drgon T, Walther D, Johnson C, Poleskaya O, Hess J, Uhl GR. 2005. Pooled association genome scanning: Validation and use to identify addiction vulnerability loci in two samples. Proc Natl Acad Sci USA 102:11864–11869.

Long JC, Knowler WC, Hanson RL, Robin RW, Urbanek M, Moore E, Bennett PH, Goldman D. 1998. Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an American Indian population. Am J Med Genet 81:216–221.

Ma JZ, Zhang D, Dupont RT, Dockter M, Elston RC, Li MD. 2003. Mapping susceptibility loci for alcohol consumption using number of grams of alcohol consumed per day as a phenotype measure. BMC Genet 4(Suppl 1):S104.

Maes HH, Woodard CE, Murrelle L, Meyer JM, Silberg JL, Hewitt JK, Rutter M, Simonoff E, Pickles A, Carbonneau R, et al. 1999. Tobacco, alcohol and drug use in eight- to sixteen-year-old twins: The Virginia Twin Study of Adolescent Behavioral Development. J Stud Alcohol 60:293–305.

Merikangas KR, Stolar M, Stevens DE, Goulet J, Preisig MA, Fenton B, Zhang H, O'Malley SS, Rounsaville BJ. 1998. Familial transmission of substance use disorders. Arch Gen Psychiatry 55:973–979.

Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, Foroud T, Hesselbrock V, Schuckit MA, Bucholz K, et al. 1998. Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet 81:207–215.

Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. Science 273:1516–1517.

Schork NJ, Nath SK, Fallin D, Chakravarti A. 2000. Linkage disequilibrium analysis of biallelic DNA markers, human quantitative trait loci, and threshold-defined case and control subjects. Am J Hum Genet 67:1208–1218.

Sham PC, Cherny SS, Purcell S, Hewitt JK. 2000. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. Am J Hum Genet 66:1616–1630.

Storey JD. 2002. A direct approach to false discovery rates. J R Stat Soc B 64:479–498.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc Natl Acad Sci USA 100:9440–9445.

Tsuang MT, Lyons MJ, Eisen SA, Goldberg J, True W, Lin N, Meyer JM, Toomey R, Faraone SV, Eaves L. 1996. Genetic influences on DSM-III-R drug abuse and dependence: A study of 3,372 twin pairs. Am J Med Genet 67:473–477.

Tsuang MT, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, True W, Lin N, Toomey R, Eaves L. 1998. Co-occurrence of abuse of different drugs in men: The role of drug-specific and shared vulnerabilities. Arch Gen Psychiatry 55:967–972.

Tsuang MT, Lyons MJ, Harley RM, Xian H, Eisen S, Goldberg J, True WR, Faraone SV. 1999. Genetic and environmental influences on transitions in drug use. Behav Genet 29:473–479.

Uhl GR. 1999. Molecular genetics of substance abuse vulnerability: A current approach. Neuropsychopharmacology 20:3–9.

Uhl GR. 2004. Molecular genetic underpinnings of human substance abuse vulnerability: Likely contributions to understanding addiction as a mnemonic process. Neuropharmacology 47(Suppl 1):140–147.

Uhl GR, Elmer GI, Labuda MC, Pickens RW. 1995. Genetic influences in drug abuse. In: Gloom FE, Kupfer DJ, editors. Psychopharmacology: The fourth generation of progress. New York: Raven Press. pp 1793–2783.

Uhl GR, Gold LH, Risch N. 1997. Genetic analyses of complex behavioral disorders. Proc Natl Acad Sci USA 94:2785–2786.

Uhl GR, Liu QR, Walther D, Hess J, Naiman D. 2001. Polysubstance abuse-vulnerability genes: Genome scans for association, using 1,004 subjects and 1,494 single-nucleotide polymorphisms. Am J Hum Genet 69:1290–1300.

Uhl GR, Liu QR, Naiman D. 2002. Substance abuse vulnerability loci: Converging genome scanning data. Trends Genet 18:420–425.

Wood JD, Yuan J, Margolis RL, Colomer V, Duan K, Kushi J, Kaminsky Z, Kleiderlein JJ, Sharp AH, Ross CA. 1998. Atrophin-1, the DRPLA gene product, interacts with two families of WW domain-containing proteins. Mol Cell Neurosci 11:149–160.

Woodward CE, Maes HH, Silberg JL, Meyer JM, Eaves LJ. 1996. Tobacco, alcohol and drug use in 8–16 year old twins. NIDA Res Monograph 162:309.

Xu YL, Reinscheid RK, Huitron-Resendiz S, Clark SD, Wang Z, Lin SH, Brucher FA, Zeng J, Ly NK, Henriksen SJ, et al. 2004. Neuropeptide S: A neuropeptide promoting arousal and anxiolytic-like effects. Neuron 43:487–497.