

## Haplotypes at the OPRM1 Locus Are Associated With Susceptibility to Substance Dependence in European-Americans

Xingguang Luo,<sup>1,2</sup> Henry R. Kranzler,<sup>3</sup> Hongyu Zhao,<sup>4</sup> and Joel Gelernter<sup>1,2\*</sup>

<sup>1</sup>Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut

<sup>2</sup>VA Connecticut Healthcare System, West Haven Campus, Connecticut

<sup>3</sup>Department of Psychiatry, University of Connecticut School of Medicine, Alcohol Research Center, Farmington, Connecticut

<sup>4</sup>Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut

Our objective was to investigate the relationship between the gene encoding the  $\mu$ -opioid receptor (OPRM1) and susceptibility to substance dependence in European-American (EA) and African-American (AA) subjects. Eight single nucleotide polymorphisms (SNPs) at the OPRM1 locus, i.e., –2044C/A, –1793T/A, –1699insT, –1469T/C, –1320A/G, –111C/T, +17C/T (Ala6Val), and +118A/G (Asn40Asp) were genotyped in 676 subjects: 318 EA subjects and 124 AA subjects with substance dependence, and 179 EA normal controls, and 55 AA normal controls. Affection status was defined by each unique combination of alcohol, cocaine, and opioid dependence and analysis of association examined in relation to the possible combinations. We used a newly implemented permutation method to evaluate statistical significance. In EAs, a significant difference in haplotype frequency distributions was found between controls and “alcohol + opioid” dependent patients ( $P = 0.0036$ ). This finding is also supported by logistic regression analysis and a simulation method. The frequencies of allele –2044A and haplotypes

including –2044A are higher in these patients than in controls. In AAs, no allele, haplotype, or genotype frequencies were significantly different between cases and controls. There were highly significant differences in the allele, haplotype, and genotype frequencies between EA and AA controls. Four of the variants [–1793T/A, –1699insT, –1320A/G, and –111C/T] are in virtually complete linkage disequilibrium (LD) to compose a sequence pattern, which does not associate with any of the seven categories of substance dependence. In EAs, allele –2044A and haplotypes that include –2044A are the susceptibility allele and haplotypes for substance dependence. These findings suggest that OPRM1 may play a role in the pathophysiology of substance dependence and this role is population- and diagnosis-specific. © 2003 Wiley-Liss, Inc.

**KEY WORDS:** substance dependence; OPRM1 gene;  $\mu$ -opioid receptor (MOR); haplotype; allele

Grant sponsor: NIH; Grant numbers: MH01387, DA12849, DA12690, AA03510, AA11330, AA12870, AA13736, RR06192, GM59507; Grant sponsor: US Department of Veterans Affairs; the VA Medical Research Program; the VA Connecticut-Massachusetts Mental Illness Research, Education, and Clinical Center (MIRECC); the REAP (Research Enhancement Award Program); and the Yale-VA Alcohol Research Center.

\*Correspondence to: Dr. Joel Gelernter, M.D., Yale University School of Medicine, VA Psychiatry 116A2, 950 Campbell Avenue, West Haven, CT 06516. E-mail: joel.gelernter@yale.edu

Received 16 October 2002; Accepted 30 January 2003

DOI 10.1002/ajmg.b.20034

### INTRODUCTION

Twin, family, and adoption studies have indicated that susceptibility to substance dependence, including opioid dependence, cocaine dependence, and alcohol dependence, is genetically influenced, with strong influences from environmental factors as well [Mirin et al., 1984; Cadoret et al., 1986; Goodwin, 1989; Grove et al., 1990; Pickens et al., 1991; Rounsaville et al., 1991; McGue et al., 1992; Cadoret et al., 1995]. Since these disorders are heritable, it should be possible to identify specific genetic loci that influence risk. The candidate gene strategy may be used to help identify certain genes that play important roles in the etiology of substance dependence.

The  $\mu$ -opioid receptor (MOR) (which is encoded by genetic locus OPRM1) is a primary site of action for many endogenous opioid peptides, including  $\beta$ -endorphin and enkephalin, and also for the most commonly used and abused opioids, including morphine, heroin, fentanyl, and methadone [Basbaum and Fields, 1984; Pasternak, 1993; Zadina et al., 1997]. In contrast to the selective action of opioids on the  $\mu$  receptor, some non-opioid drugs and abused substances with other primary sites of action (including alcohol and cocaine) exert some of their ultimate effects through interaction with the opioid receptor [Kreek, 1996; Herz, 1997]. Some drugs of abuse can cause a euphoric effect through rapidly activating the MOR. The euphoric effect confers reinforcing or rewarding effects of the drug, which constitute a key psychomotor mechanism of development of addiction [Bond et al., 1998].

Animal models have provided evidence for a role of the opioid system in the effects of alcohol. Alcohol-preferring rats show a significantly higher density of MORs in the ventral tegmental area (and related limbic brain regions), in comparison to alcohol avoiding (ANA) rats [De Waele et al., 1995]. In humans,  $\beta$ -endorphin immunoreactivity in blood, plasma, or CSF in alcoholics is less than that in controls [Inder et al., 1998]. Subjects from families with a high density of alcohol dependence show reduced endogenous opioid activity in hypothalamus [Wand et al., 1998]. MOR agonists and antagonists have been shown to affect alcohol consumption. Naltrexone, a MOR antagonist, is one of only two medications approved in the US to treat alcohol dependence [O'Malley et al., 1992; Volpicelli et al., 1992]. A series of clinical and experimental studies have suggested possible involvement of MOR-related mechanisms in the reward, tolerance, and withdrawal processes of alcohol dependence [Kreek, 1996; Herz, 1997]. Genetic variation at loci coding for opioid-system proteins, including OPRM1, might, therefore, be expected to affect risk for drug and alcohol dependence, and possibly other addictive behaviors as well, and to be involved in the pathophysiology of substance dependence.

There is also considerable evidence for opioid effects on cocaine self-administration. Animal studies have shown that pre-treatment with an opioid agonist (e.g., methadone) enhances cocaine's reinforcing properties and increases cocaine self-administration [Mello et al., 1990; Corrigan and Coen, 1991]. Conversely, pre-treatment with an opioid antagonist (e.g., naltrexone) typically results in reduced reinforcement and attenuates cocaine self-administration [De Vry et al., 1989; Houdi et al., 1989; Bilsky et al., 1992]. A study in living human subjects with PET has found a significant upregulation of MOR binding in cocaine-dependent men and an association between this upregulation and cocaine craving [Zubieta et al., 1996]. The opioid-dependent human subjects who were treated with naltrexone used significantly less cocaine than those maintained on methadone [Kosten et al., 1989].

Many genetic variants at the OPRM1 locus have been identified and their relationships to drug dependence, alcohol dependence, and related phenotypes have been studied. Our research group previously examined the

association of a polymorphic (CA)<sub>n</sub> repeat at OPRM1 locus to alcohol or drug dependence in 320 Caucasian and 108 AA substance-dependent or control subjects, with suggestion of a modest association ( $P = 0.03$ ), observed between OPRM1 alleles and substance (alcohol, cocaine, or opioid) dependence [Kranzler et al., 1998]. Other studies [e.g., Bergen et al., 1997; Berrettini et al., 1997] have considered putatively functional OPRM1 variants, with mixed results. Polymorphisms commonly studied include +17C/T, located in position +17 in exon 1, leading to Ala6Val in the N-terminus; and +118A/G, located in position +118 in exon 1, leading to Asn40Asp. This Asn40Asp substitution can differentially affect receptor activity and affinity for endogenous ligand  $\beta$ -endorphin. The MOR Asp40 isoform binds  $\beta$ -endorphin approximately three times more potently than the more common Asn40 variant [Bond et al., 1998]. Also,  $\beta$ -endorphin increases the agonist-induced activation of G protein-coupled potassium channels of the variant Asp40 three times more than that of the common Asn40 [Bond et al., 1998]. Thus, this substitution can alter both binding and signal transduction. Subjects expressing Asp40 have a greater cortisol response to opioid receptor blockade with naloxone and a lower agonist effect of morphine-6-glucuronide [Höllt, 2002; Lötsch et al., 2002; Wand et al., 2002; Hernandez-Avila et al., 2003].

Several studies have reported positive associations of +17C/T or +118A/G with substance dependence. A study on Hispanics found that the common allele +118A was present in a significantly higher proportion of opioid-dependent subjects than non-opioid-dependent subjects [Bond et al., 1998], a result that we could not replicate [Gelernter et al., 1999]. Another research group genotyped the +118A/G polymorphism in 105 alcoholics (80% Caucasian) and 122 normal controls, and reported an association between the +118A allele and alcohol dependence [Town et al., 1999]. A most recent study on 297 Caucasian controls and 179 Caucasian patients with substance dependence, who abused alcohol, cigarette, cocaine, marijuana, and/or other illegal drugs, found that there was a significant association between substance dependence and +118A [Schinka et al., 2002]. A previous study on +118A/G and another variant, C1031G, in 200 Chinese heroin-dependent and 97 control subjects showed a significant association for both allele +118G and +1031G polymorphisms and opioid dependence [Szeto et al., 2001]. In addition, as noted above, an OPRM1 (CA)<sub>n</sub> polymorphism has also been studied. This polymorphism would presumably not affect function directly, but could show association with phenotype due to linkage disequilibrium (LD) with a functional polymorphism (we previously reported that it is in LD with the exon 1 polymorphisms [Gelernter et al., 1999]). The positive findings discussed above (albeit in limited samples) suggest that OPRM1 may play a role in the pathophysiology of substance dependence.

Hoehe et al. [2000] genotyped 250 cases and controls by multiplex sequence comparison. In the subgroup of 172 African-Americans (AAs), including 137 opioid/cocaine dependent patients and 35 normal controls, a

total of 43 biallelic variants of OPRM1 were identified by sequencing, 40 of which were substitution SNPs and three of which were insertion/deletion (Ins/Del) variants. Of these 43 variants, 24 variants were in the 5'-regulatory region, four in the 5'-UTR, one in the 3'-UTR, eight in intronic regions, and six in the coding region (of which five affect predicted protein sequence). Although the 5'-regulatory region is non-coding, some variants in this region could affect transcription regulatory motifs. Alteration of 5'-untranslated mRNA sequences could readily explain different levels of MOR mRNA stability or even translational efficacy, and could contribute to the expression of differing levels of this protein in different individuals or cell types.

Hoehe et al. [2000] reported that the sequence variant pattern [-1793T/A, -1699insT, -1320A/G, -111C/T, and +17C/T], which is predominantly composed of a specific set of changes in putative transcription regulatory motifs, positively associated with opioid or cocaine dependence in AAs. The sample for that analysis included 158 affected subjects and 51 controls; data from 137 cases and 35 controls were shown in the original paper.

In the present study, five variants in the 5' putative regulatory region, including four substitution SNPs, i.e., -2044C/A, -1793T/A, -1469T/C, and -1320A/G, and one insertion variant -1699T, one SNP -111C/T in the 5'-UTR, and two SNPs in exon 1, i.e., +17C/T and +118A/G were genotyped. These variants were chosen because they showed the greatest potential to distinguish between substance dependent and control subjects in previous studies, including that recent study by Hoehe et al. [2000]. The relationships between the alleles and the haplotypes of these eight variants and the susceptibility to substance dependence including opioid dependence, cocaine dependence, and alcohol dependence were investigated.

## MATERIALS AND METHODS

### Subjects

We examined allele and haplotype frequencies in 497 European-Americans (EAs) and 179 AAs. Of this number, 318 EAs and 124 AAs met lifetime DSM-III-R criteria [American Psychiatric Association, 1987] for alcohol, cocaine, or opioid dependence or a combination of these disorders. These subjects were predominantly male (76.0%). Diagnoses were made using the Structured Clinical Interview for DSM-III-R (SCID) [Spitzer et al., 1992], the computerized Diagnostic Interview Schedule for DSM-III-R (C-DIS-R) [Blouin et al., 1988], or a checklist comprised of DSM-III-R symptoms. The 234 control subjects, which included 179 EAs and 55 AAs, were screened using the SCID, the C-DIS-R, or the Schedule for Affective Disorders and Schizophrenia [Spitzer and Endicott, 1975] to exclude alcohol or drug dependence. A majority of the control subjects were male (55.5%).

Subjects were recruited at the University of Connecticut Health Center or the VA Connecticut Healthcare System, West Haven Campus. All subjects gave written informed consent before participating in the study,

which was approved by the Institutional Review Board at the relevant institutions. The present data set includes many of those subjects for whom we previously reported allele frequencies for the OPRM1 STR [Kranzler et al., 1998] and the two exon 1 polymorphisms (Ala6Val and Asn40Asp, or +17C/T and +118A/G) [Gelernter et al., 1999].

### Genotyping

Genomic DNA was extracted from peripheral blood by standard methods. The eight SNPs were genotyped by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) techniques using five pair of primers and eight different restriction enzymes, as described in Table I. Mismatched bases were inserted to create artificial restriction sites for genotyping (Table I). PCR was performed in a final volume of 15  $\mu$ l for each system. From the 15  $\mu$ l reaction product in the first system, 5  $\mu$ l were digested in 10 U HhaI (all restriction enzymes purchased from New England Biolabs, Beverly, MA), and 5  $\mu$ l digested in 10 U DraI, to yield genotypes for two SNPs from a single PCR reaction. The forward primer (5'-end) was used to genotype the -2044C/A variant. The reverse primer (3'-end) was used to genotype the -1793T/A variant. The second PCR system also serves to genotype two SNPs from a single reaction. From the 15- $\mu$ l reaction product in the third PCR system, 10  $\mu$ l of the reaction mixture were digested in 10 U Sau96I to yield genotypes for the -1320A/G variant. The fourth PCR system also served to genotype only one SNP. In the fifth PCR system, +17C/T and +118A/G were genotyped from a single PCR reaction as described previously [Gelernter et al., 1999].

Approximately 8% of all genotypes were repeated at random for quality control, with complete agreement.

### Statistical Analysis

For discussion of haplotype frequency, we notate the haplotype in the 5'- to 3'-direction, i.e., -2044C/A, -1793T/A, -1699insT, -1469T/C, -1320A/G, -111C/T, +17C/T (Ala6Val), and +118A/G (Asn40Asp). The "1" notation corresponds to the following eight common alleles: [-2044C, -1793T, -1699 non-T, -1469T, -1320A, -111C, +17C, and +118A]. The "2" notation corresponds to the following eight variant alleles: [-2044A, -1793A, -1699T, -1469C, -1320G, -111T, +17T, and +118G].

**Haplotype frequency estimation.** The most likely haplotype pair for each genotype in each individual was estimated and the haplotype frequency distributions were obtained with an expectation-maximum (EM) algorithm, using MLOCUS software [Long et al., 1995; Peterson et al., 1999; Long, personal communication]. We observed six different eight-variant haplotypes in EAs and seven different eight-variant haplotypes in AAs (see Table II). The genotype for each sample and the genotype frequency distributions were then derived from the haplotype pair and the haplotype frequency distributions. We note that the haplotype pairs for most individuals can be uniquely identified, i.e., there is at

TABLE I. Five Pairs of Primer Sets and Their Corresponding Polymerase Chain Reaction (PCR) Reaction Conditions

System no.	Primer sequences	Product size (bp)	Annealing temperature	Restriction buffers	Restriction enzymes	Variants
1F	5'-TTTAAACACTGGAACTGTAGTTTCAGcG-3'	307	65°C → 55°C(TD)	D	HhaI	-2044C/A
R	5'-CAGGATGACAGTCTTTCATTAATTAA-3'				DraI	-1793T/A
2F	5'-ACAAAGCTGATTTATAAAATGTT-3'	278	65°C → 55°C(TD)	D	HincII	-1699InsT
R	5'-CTATTAATCATAGGCTATTTTgTC-3'				TaqAI	-1469T/C
3F	5'-ACAAAGCTGATTTATAAAATGGTT-3'	427	55°C → 45°C(TD)	PC-2	Sau96I	-1320A/G
R	5'-CATAAACATTGAAAATACATGgCC-3'					
4F	5'-CGCAGAGGAGATGTCAGAT-3'	110	60°C	PC-2	PvuII	-111C/T
R	5'-TCGCCGCTGCCACAGTCCaGCT-3'					
5F	5'-CCGTCAGTACCATGGACAGCGTG-3'	154	60°C	PC-2	BanI	+17C/T
R	5'-GTTCCGACCCGATGGTCCGACAGaT-3'				DpnII	+118A/G

F: forward primer; R: reverse primer. Small caps in primer sequences represent mismatched bases that create artificial restriction sites for genotyping; TD: "touchdown" protocol. The annealing temperature varied according to a "touchdown" protocol in which the initial annealing temperature was lowered by 1°C each cycle for 10 cycles, and then 25 additional cycles were run with the lowered annealing temperature; 1 × buffer D (3.0 mM MgCl<sub>2</sub>, 50 mM KCl, and 10 mM Tris-HCl, pH 8.4), 1 × buffer PC-2 (Ab Peptides, Inc., St. Louis, MO). All restriction enzymes were purchased from New England Biolabs (Beverly, MA).

TABLE II. Distributions of Eight-Variant Haplotype Frequencies in EA and AA Subjects

Haplotype	Controls		Opioid		Op + Coc		Cocaine		Coc + Alc		Alcohol		Alc + Op <sup>b</sup>		Op + Coc + Alc		Cases (total) <sup>a</sup>	
	f	N	f	N	f	N	f	N	f	N	f	N	f	N	f	N	f	N
EA	0.838	300	0.850	17	0.875	28	0.788	41	0.886	117	0.869	245	0.806	29	0.805	66	0.854	543
11111111	0.137	49	0.100	2	0.125	4	0.154	8	0.098	13	0.106	30	0.056	2	0.159	13	0.113	72
11111112	0.014	5	0.000	0	0.000	0	0.019	1	0.000	0	0.000	0	0.028	1	0.000	0	0.003	2
11121111	0.003	1	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0
12212221	0.000	0	0.000	0	0.000	0	0.000	0	0.008	1	0.000	0	0.000	0	0.000	0	0.002	1
21111111	0.008	3	0.050	1	0.000	0	0.038	2	0.008	1	0.025	7	0.111	4	0.037	3	0.028	18
Total	1.000	358	1.000	20	1.000	32	1.000	52	1.000	132	1.000	282	1.000	36	1.000	82	1.000	636
AA	0.827	91	0.800	8	0.846	22	0.704	38	0.767	66	0.813	13	0.750	6	0.771	37	0.766	190
11111112	0.027	3	0.000	0	0.000	0	0.037	2	0.023	2	0.000	0	0.000	0	0.000	0	0.016	4
11111121	0.055	6	0.100	1	0.115	3	0.130	7	0.105	9	0.125	2	0.125	1	0.167	8	0.125	31
11121111	0.000	0	0.000	0	0.000	0	0.037	2	0.012	1	0.000	0	0.000	0	0.021	1	0.016	4
12212221	0.082	9	0.000	0	0.038	1	0.093	5	0.070	6	0.063	1	0.125	1	0.042	2	0.065	16
12212222	0.000	0	0.100	1	0.000	0	0.000	0	0.012	1	0.000	0	0.000	0	0.000	0	0.008	2
21111111	0.009	1	0.000	0	0.000	0	0.000	0	0.012	1	0.000	0	0.000	0	0.000	0	0.004	1
Total	1.000	110	1.000	10	1.000	26	1.000	54	1.000	86	1.000	16	1.000	8	1.000	48	1.000	248

EA: European-American; AA: African-American; Op: opioid dependence; Coc: cocaine dependence; Alc: alcohol dependence; N: number of chromosomes; In EAs, in the most significant logistic regression equation (diagnosis ~ age + sex + haplotype),  $P^b = 0.005$  for the contribution of covariant "haplotype with -2044A" to dependent variable "alcohol + opioid." In EAs, when comparing the eight-variant haplotype frequency distribution in each case group against controls with CLUMP,  $P^b = 0.0207$  for total case group,  $P^b = 0.0036$  for "alcohol + opioid" subgroups.  $P = 1 \times 10^{-8}$ , when the haplotype frequency distribution was compared between EA and AA controls. In each eight-variant haplotype, 1 and 2 are identical with the reference allele and the variant allele of each corresponding variant, respectively. For the eight variants specified by positions see Table III.

most one heterozygous site among the studied polymorphisms (when the four-variant pattern is considered as one marker). Therefore, the impact of uncertainty in haplotype inference on association studies is minimal. We observed eleven different genotypes in EAs and eight genotypes in AAs (data not shown).

Several variants in complete LD can compose a sequence variant pattern. As described in detail in the results below, we found that the four variants, i.e., [-1793T/A, -1699insT, -1320A/G, and -111C/T] were in virtually complete LD to compose a sequence pattern that produced, in effect, a single marker. Therefore, these eight SNPs can be represented as five markers.

All phases of genotypes based on five-marker haplotypes (considering the four-variant pattern as one marker) can be unambiguously inferred with the EM algorithm; the majority of the individuals (99.9% in EAs, 92.5% in AAs) have at most one heterozygous allele of these five markers at each chromosome (see Table II).

**Evaluation of the potential for confounding of age and sex on the association between haplotypes and phenotypes.** Differences in age and sex between cases and controls may sometimes obscure an alternate explanation for association. We, therefore, tested whether possible associations between haplotypes and phenotypes were robust after excluding possible confounding of age and sex. Such confounding could be attributable to age and sex related penetrance differences, and could increase either type I or type II error. Such tests were performed before analysis of the relationship between haplotypes and diagnoses, because it is impossible to incorporate age and sex in the simulation test for association, given the complexity of the method of analysis.

The data were analyzed using logistic regression analysis in SPSS 11.0. In the regression models, age, sex, and the haplotypes served as the independent variables (i.e., covariates) and the diagnosis was the dependent variable. The multinomial haplotypes were separated into five binary covariates before being entered into the regression models, i.e., haplotype with/without -2044A, haplotype with/without -1793A, haplotype with/without -1469C, haplotype with/without +17T, and haplotype with/without +118G. The dependent variable is also binary, i.e., "control" = 0 and "affected" = 1, in which "affected" is any one of the seven case diagnoses (enumerated below).

**Comparison of haplotype, allele, and genotype frequencies: 2 × 2 contingency table comparison:  $\chi^2$ -test, Fisher's exact test.** Comparison of haplotype frequency was defined as the primary analysis; other analyses on allele and genotype frequency were considered exploratory. Comparisons were made based on the total case group: seven different, mutually exclusive diagnostic subgroups (opioid, opioid + cocaine, cocaine, cocaine + alcohol, alcohol, alcohol + opioid, and alcohol + opioid + cocaine dependence) versus normal controls. These diagnostic subgroups were set to identify the source of any statistically significant difference observed between the total case group and controls. Contingency table (2 × 2) comparisons were performed using the  $\chi^2$ -test in S-PLUS or the two-tail Fisher's exact

test online [Langsrud, 1999]. Empirical  $P$  values for 2 ×  $k$  ( $k > 2$ ) contingency table comparisons were computed via permutation by CLUMP software (T1 statistic, 1,000 simulations) [Sham and Curtis, 1995]. The adjusted empirical  $P$  values for  $m \times k$  ( $m, k > 2$ ) contingency table comparisons were computed by a simulation method described below.

**Bonferroni correction.** When multiple statistical tests on the same data were performed, a Bonferroni-corrected  $\alpha$  level ( $\approx 0.05/n$ , where  $n$  is the number of tests) was selected as the primary threshold for statistical significance. We obtained this exact Bonferroni-corrected  $\alpha$  value via SISA [Uitenbroek, 1997]. To avoid multiple tests that could occur in our  $m \times k$  contingency table comparisons, we developed a program to run Monte Carlo simulation test described below.

**Monte Carlo simulation test.** In our  $m \times k$  tables, which include several haplotypes in a normal control group and seven mutually exclusive diagnosis groups, some expected cell counts are smaller than five, which may nullify the asymptotic sampling distribution of the Pearson  $\chi^2$  statistic and, therefore, the associated test of significance. The large number of haplotypes increases the degrees of freedom, which would reduce the power to detect a potentially small difference in haplotype frequency between cases and controls. Traditionally, two empirical methods are used for handling these kinds of statistical problems: (1) grouping together the rare haplotypes such that no expected cell count is less than five. This method may result in a loss of power; (2) comparing each haplotype in turn against the rest. This method would create many 2 × 2 tables and an adjustment such as the Bonferroni correction for multiple comparisons is always required, but it is also always difficult to ascertain the exact number of independent tests to understand and clarify the Bonferroni-corrected significance level. Both methods could reduce the power to detect a true association when the number of haplotypes or alleles is large. To avoid this loss of information, we performed Monte Carlo simulations to estimate the significance level. With this test, we are no longer limited to statistics of known asymptotic distributions. In the EAs and AAs, if we take each marker as a special form of a haplotype, i.e., a one-marker haplotype, the five markers (i.e., eight variants) can consist of up to 31 kinds of different haplotypes ( $31 = C_5^1 + C_5^2 + C_5^3 + C_5^4 + C_5^5$ , in which  $C$  denotes the combinatorial approach, with formula  $C_m^n = m! / [(m-n)!n!]$ ). Of these 31 kinds of haplotypes, only the one-marker haplotypes and the eight-variant haplotypes were listed in Tables II and III, respectively. Thus, for Table II, we could perform  $\chi^2$ -tests on 217 ( $= 31 \times 7$ ) 2 ×  $k$  contingency tables, each of which compares the frequency distributions of one kind of haplotype between controls and any one of the seven diagnostic groups and get 217  $P$ -values, where  $k \geq 2$  is the number of different chromosomes in each kind of haplotype. The minimum of these 217  $P$ -values is designated as  $P_0$ .

Then, according to the principle of Monte Carlo simulation tests, we randomly generate 10,000 simulated contingency tables constrained to have the same marginal totals as the original data, under the null

TABLE III. Distributions of Allele Frequencies of Eight Variants of OPRM1 in EA and AA Subjects

Marker	Position	Controls		Opioid		Op+Coc		Cocaine		Coc+Alc		Alcoholism		Alc+Op		Op+Coc+Alc		Cases (total)	
		f	N	f	N	f	N	f	N	f	N	f	N	f	N	f	N	f	N
EA I	-2044C	0.992	355	0.950	19	1.000	32	0.962	50	0.992	131	0.975	275	0.889 <sup>b</sup>	32	0.963	79	0.972 <sup>a</sup>	618
	A	0.008	3	0.050	1	0.000	0	0.038	2	0.008	1	0.025	7	0.111	4	0.037	3	0.028	18
	-1793T	1.000	358	1.000	20	1.000	32	1.000	52	0.992	131	1.000	282	1.000	36	1.000	82	0.998	635
II	A	0.000	0	0.000	0	0.000	0	0.000	0	0.008	1	0.000	0	0.000	0	0.000	0	0.002	1
	-1469T	0.997	357	1.000	20	1.000	32	1.000	52	1.000	132	1.000	282	1.000	36	1.000	82	1.000	636
	C	0.003	1	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0	0.000	0
IV	+17C	0.986	353	1.000	20	1.000	32	0.981	51	0.992	131	1.000	282	0.972	35	1.000	82	0.995	633
	T	0.014	5	0.000	0	0.000	0	0.019	1	0.008	1	0.000	0	0.028	1	0.000	0	0.005	3
	+118 A	0.863	309	0.900	18	0.875	28	0.846	44	0.902	119	0.894	252	0.944	34	0.841	69	0.887	564
V	G	0.137	49	0.100	2	0.125	4	0.154	8	0.098	13	0.106	30	0.056	2	0.159	13	0.113	72
	-2044C	0.991	109	1.000	10	1.000	26	1.000	54	0.988	85	1.000	16	1.000	8	1.000	48	0.996	247
	A	0.009	1	0.000	0	0.000	0	0.000	0	0.012	1	0.000	0	0.000	0	0.000	0	0.004	1
II	-1793T	0.918	101	0.900	9	0.962	25	0.907	49	0.919	79	0.937	15	0.875	7	0.958	46	0.927	230
	A	0.082	9	0.100	1	0.038	1	0.093	5	0.081	7	0.063	1	0.125	1	0.042	2	0.073	18
	-1469T	1.000	110	1.000	10	1.000	26	0.963	52	0.988	85	1.000	16	1.000	8	0.979	47	0.984	244
III	C	0.000	0	0.000	0	0.000	0	0.037	2	0.012	1	0.000	0	0.000	0	0.021	1	0.016	4
	+17C	0.864	95	0.800	8	0.846	22	0.778	42	0.814	70	0.812	13	0.750	6	0.792	38	0.802	199
	T	0.136	15	0.200	2	0.154	4	0.222	12	0.186	16	0.188	3	0.250	2	0.208	10	0.198	49
V	+118A	0.973	107	0.900	9	1.000	26	0.963	52	0.965	83	1.000	16	1.000	8	1.000	48	0.976	242
	G	0.027	3	0.100	1	0.000	0	0.037	2	0.035	3	0.000	0	0.000	0	0.000	0	0.024	6

EA: European-American; AA: African-American; Op: opioid dependence; Coc: cocaine dependence; Alc: alcohol dependence; f: frequencies; N: number of chromosomes. I-V, markers; II, complete LD marker pattern, i.e., [-1793T/A, -1699insI, -1320A/G, -111C/T], of which only -1793T/A is listed in this table;  $P^a = 0.039$ ,  $P^b = 0.0017$ , when compared with controls, respectively; when comparing the allele frequencies between EA and AA controls,  $P_{-1793} = P_{-1699} = P_{-1320} = P_{-111} = 2 \times 10^{-6}$ ,  $P_{+17} = 9 \times 10^{-7}$ , and  $P_{+118} = 8 \times 10^{-4}$ , respectively.

hypothesis that cases and controls have the same haplotype frequencies. We calculated the minimal  $P$ -value ( $P'_0$ ) of each of these 10,000 simulated tables using the same procedure as the original data set. The statistical significance of the observed association can, therefore, be evaluated on the basis of how often the observed minimal  $P$ -value is smaller than the calculated minimal  $P$ -values when applied to random subsets of the original data. If  $r$  is specified as the number of times that  $P'_0$  is less than or equal to  $P_0$  in the 10,000 simulations, then the adjusted empirical  $P$ -value is just  $(r + 1)/10,001$  [North et al., 2002].

We conducted the above procedures using an R program (version 1.5.0, available from <http://www.r-project.org>); code is available from us upon request.

## RESULTS

The common allele and the variant type allele of each of the SNPs are shown in Table III. Among the eight SNPs studied, we found four SNPs, i.e.,  $-1793T/A$ ,  $-1699insT$ , and  $-1320A/G$ , which are located in the 5'-regulatory region, and  $-111C/T$  in the 5'-UTR that are in virtually complete LD to compose a sequence pattern like a biallelic marker. This characteristic pattern results in only two common haplotypes, 1111 and 2222, so these eight SNPs can be taken as five markers (notated as I–V, see Table III).

Age, sex, and haplotype served as covariates and the diagnosis served as the dependent variable; in the different logistic regression models (seven models in EAs and seven models in AAs), all  $P$  values for the contributions of haplotypes to phenotypes are  $>0.05$ , with the exception of  $P = 0.005$  for the contribution of haplotype with  $-2044A$  to diagnosis of "alcohol + opioid" in EAs. This unique association remains significant even

after Bonferroni correction ( $\alpha = 0.0073 \approx 0.05/7$ , with 7 being the number of subgroups), which indicates that this association was robust after excluding the potential confounding of age and sex.

The five markers (i.e., eight variants) can generate at most 31 different n-marker haplotypes ( $n = 1-5$ ). All 31 haplotypes were investigated in the present study, in order to identify any subsets of markers that may be useful for risk assessment. The eight-variant haplotype frequencies are shown in Table II, one-variant haplotype (i.e., allele) frequencies are shown in Table III, and the sequence four-variant pattern haplotype frequencies are shown in Table IV. Comparing the frequencies of each of these 31 n-marker haplotypes between controls and any one of the seven case groups, we found some nominal differences in EAs; exact  $P$  values less than 0.05 are shown in Table V. In EAs, there is a significant difference in the eight-variant haplotype frequency distribution between controls and the total case group ( $P = 0.0207 < 0.05$ , see Table II). Among the diagnostic subgroups, there is a significant difference in the eight-variant haplotype frequency distribution only between controls and the "alcohol + opioid" group ( $P = 0.0036$ , Table II), which remains significant even after Bonferroni correction ( $\alpha = 0.05/7 = 0.0071$ ). There are nominal differences between controls and the "alcohol + opioid" dependent group in the frequency distributions of all of the n-marker haplotypes that contain the marker  $-2044C/A$  (minimal  $P = 0.0013$ , see Table V). However, because 217 ( $= 31 \times 7$ ) comparisons were performed, a threshold of 0.00023 ( $\approx 0.05/217$ ) should be used to identify statistical significance. No  $P$  values in Table V reach this corrected threshold. However, it should be noted that the Bonferroni correction is only an estimate and that it may yield too low an  $\alpha$  level, making it overly conservative. To provide a valid test of the association,

TABLE IV. Distributions of Haplotype Frequencies of Four-Variant Pattern of OPRM1 in Our EA, AA Subjects, and AA Subjects in the Study by Hoehe et al. [2000]

Pattern haplotype	Present study				Hoehe et al. [2000]			
	Controls		Cases (total)		Controls <sup>a</sup>		Op/Coc <sup>b</sup>	
	f	N	f	N	f	N	f	N
<b>AA</b>								
*11*11**	0.918	101	0.927	230	0.986	69	0.901	247
*12*11**	0.000	0	0.000	0	0.000	0	0.004	1
*12*22**	0.000	0	0.000	0	0.000	0	0.007	2
*22*22**	0.082	9	0.073	18	0.014	1	0.088	24
Total	1.000	110	1.000	248	1.000	70	1.000	274
<b>EA</b>								
*11*11**	1.000	358	0.998	635				
*22*22**	0.000	0	0.002	1				
Total	1.000	358	1.000	636				

EA: European-American; AA: African-American; Op: opioid dependence; Coc: cocaine dependence; Alc: alcohol dependence; f: frequencies; N: number of chromosomes.  $P^a = 0.091$ , when the haplotype frequencies were compared with our AA controls;  $P^b = 0.147$  (T1) or 0.020 (T2), when the overall haplotype frequency were compared with controls in the study by Hoehe et al. [2000] using CLUMP.  $P^b = 0.024$ , when the haplotypes "1222 + 2222" frequency was compared with controls in the study by Hoehe et al. [2000] using Fisher's exact test. When comparing the distribution of pattern frequencies between our EA and AA controls,  $P = 2 \times 10^{-6}$ . In each haplotype above, "1" designates the common allele, "2" designates the variant allele, and "\*" designates any allele in the corresponding position as Table II. This four-variant pattern is defined by  $[-1793T/A, -1699insT, -1320A/G, -111C/T]$ .

TABLE V. Different n-Marker Haplotype Frequency Distributions Between Controls and “Alcohol + Opioid” Dependent Subgroups in EAs

Haplotypes <sup>a</sup>	Exact <i>P</i> -values <sup>b</sup>
I	0.0017
I–II	0.0017
I–III	0.0032
I–IV	0.0017
I–V	0.0013
I–II–III	0.0032
I–II–IV	0.0017
I–II–V	0.0013
I–III–IV	0.0028
I–III–V	0.0024
I–IV–V	0.0022
I–II–III–IV	0.0028
I–II–III–V	0.0024
I–II–IV–V	0.0022
I–III–IV–V	0.0036
I–II–III–IV–V	0.0036

n-Marker: n is 1–5. Markers I–V, same as Table III. (Note: “II,” as noted elsewhere, refers to a four-marker pattern.)

<sup>a</sup>There were a total of 31 haplotypes ( $= C_5^1 + C_5^2 + C_5^3 + C_5^4 + C_5^5$ ) compared, but only those 16 with the *P* values less than 0.05 listed.

<sup>b</sup>The exact *P* values were calculated by Fisher’s exact test or computed via permutation by CLUMP software (T1 statistic, 10,000 simulations). The empirical *P* value adjusted by Monte Carlo simulation test is 0.0250 for overall haplotype frequency distribution comparison.

we performed 10,000 simulations to obtain a more exact empirical *P* value for the comparisons of the frequencies of the 31 haplotypes between controls and all case groups. The empirical *P* value adjusted by Monte Carlo simulation test is 0.0250 in EAs. In contrast, among AAs, there is no significant difference in the haplotype frequency distributions between controls and any one of the case subgroups (Table II). The empirical *P* value adjusted by the Monte Carlo simulation test is 0.3332 in AAs.

Distributions of allele frequencies of eight variants of OPRM1 in EA and AA subjects are shown in Table III. Compared with controls, the frequency of allele –2044A is nominally higher in the total case group in EAs ( $P = 0.039$ ), but it does not reach Bonferroni-corrected statistical significance ( $\alpha = 0.05/5 = 0.01$ ). Among the seven diagnostic subgroups, only the frequency of the –2044A allele in EA patients with alcohol + opioid dependence was nominally higher than in controls ( $P = 0.0017$ ), but it did not reach Bonferroni-corrected statistical significance, either [ $\alpha = 0.05/(5 \times 7) = 0.0015$ ].

The comparative distributions of haplotype frequencies of this four-variant pattern found in the present study and in a recent study by Hoehe et al. [2000] are shown in Table IV. The variant haplotype of this pattern was reported by Hoehe et al. [2000] to be positively associated with opioid/cocaine dependence, but the pattern in the present study is not associated with any of the seven substance dependence classifications.

Distributions of eight-variant genotype frequencies in EAs and AAs were also examined (data not shown). All observed genotype frequencies in all subgroups were in Hardy-Weinberg equilibrium. When compared with controls, the only nominal difference in the genotype frequency distribution was found in the “alcohol +

opioid” subgroup in EAs, when using either the CLUMP program ( $P = 0.0410$ ) or logistic regression analysis (diagnosis as the dependent variable, age, sex, and genotypes as the independent variables;  $P = 0.0150$  for contribution of genotype with –2044A to diagnosis of “alcohol + opioid”). We did not find any suggestion of a genotype effect distinct from haplotype effects.

When we compared the distribution of eight-variant haplotype frequencies or genotype frequencies between EA and AA controls, we observed a highly significant difference ( $P = 1 \times 10^{-8}$  and  $P = 3 \times 10^{-8}$ , respectively). When we compared the allele frequencies between EA and AA controls, there were significantly lower allele frequencies of –1793A, –1699T, –1320G, –111T, and +17T and a significantly higher allele frequency of +118G in EA ( $P_{-1793} = P_{-1699} = P_{-1320} = P_{-111} = 2 \times 10^{-6}$ ,  $P_{+17} = 9 \times 10^{-7}$ , and  $P_{+118} = 8 \times 10^{-4}$ , respectively). Comparison of the distribution of sequence pattern frequencies between the EA and AA controls revealed a highly significant difference ( $P = 2 \times 10^{-6}$ ). There was also a trend level difference in the distribution of the haplotype frequencies of this pattern in the AA normal controls between the present study and the recent study by Hoehe et al. [2000] ( $P = 0.091$ ). Therefore, it is critical to match cases and controls in terms of ethnic background in genetic association studies, and association results will be strengthened from analyses that are robust to population stratification, for example, family-based association studies and genomic control methods.

As seen in Table II, we also found that +17C/T is in strong LD with this above four-variant characteristic pattern. When this four-variant pattern is in variant type (i.e., 2222), the +17C/T is in variant type in all subjects also, but when +17C/T is in variant type, not all the four-variant patterns are in variant type. Within the heterozygous haplotypes, the other two variants, i.e., –2044A and –1469C, and this four-variant pattern (i.e., –1793A + –1699T + –1320G + –111T) are complementary, i.e., they were never observed on the same chromosome.

## DISCUSSION

We demonstrated highly significant differences in the total distribution of allele, haplotype, and genotype frequencies of the eight variants or the four-variant pattern between EA and AA controls ( $P = 10^{-3} - 10^{-8}$ ), consistent with our previous report that considered only two of these variants in a smaller sample [Gelernter et al., 1999]. The alleles of –1793A, –1699T, –1320G, –111T, and +17T are more common in AA controls, but the allele +118G is more common in EA controls. Among the eight-variant haplotype frequency distributions, in both populations studied, the haplotype 11111111 was the most common, regardless of phenotype. The haplotype 12212222 was not observed in EA subjects, but was seen rarely in AA subjects. Only two haplotypes of the four-variant pattern, i.e., 1111 and 2222, were observed in both EA and AA subjects, but 2222 is rare (<1%) in EA subjects and common (>5%) in AA subjects. However, four haplotypes 1111, 1211, 1222, and 2222 were

observed in AA subjects in the study by Hoehe et al. [2000]. These differences between EAs and AAs indicate that the allele, haplotype, and genotype frequency distributions strongly associate with ethnicity, suggesting that the LD between the markers studied and a putative disease susceptibility variant at OPRM1 locus may differ by population.

The interpretation of the significant result is based on the assumption that the control group is fairly representative for the cases, i.e., that it is sampled from a similar population. We believe that this is a reasonable assumption. The cases and controls were sampled in the same geographic area, and indeed, for most sub-phenotypes we studied, no differences in allele or haplotype frequencies were observed, which further supports the comparability of the samples.

Analysis of the eight-variant haplotypes shows that, in the EA population, there is a significant difference in the haplotype frequency distribution between the total case group and the controls ( $P = 0.0207$ , see Table II). This difference is attributable to the "alcohol + opioid" dependent diagnostic subgroup ( $P = 0.0036$ , significant after Bonferroni correction). This finding is robust after excluding the potential confounding from age and sex structures ( $P = 0.005$ , logistic regression analysis, significant after Bonferroni correction), suggesting that there may be a true association between the haplotypes at the OPRM1 locus and substance dependence. Further analyses of allele, n-marker haplotype, and eight-variant genotype frequency distributions show that this difference is compatible with the interpretation that it results from allele  $-2044A$  ( $P = 0.0017$ , see Table III), haplotypes with marker  $-2044C/A$  ( $P = 0.0013-0.0037$ , see Table V) and genotypes with  $-2044A$  ( $P = 0.0150$ , data not shown) in the "alcohol + opioid" subgroup, although these differences all are nominal and do not reach the Bonferroni-corrected statistical significance threshold.

In the present study, some markers are dependent on one another, for example, the marker IV (+17C/T) is in strong LD with marker II (four-variant pattern); the 31 n-marker haplotypes overlap one another and are dependent on one another too, for example, the n-marker haplotypes are subsets of, not parallel to, (n-1)-marker haplotypes. Therefore, simple Bonferroni correction may be overly conservative to our data analysis and may result in the loss of information. To preserve the information, we used a Monte Carlo simulation test to adjust an empirical  $P$  value for the 31 n-marker haplotype analysis. We obtained an adjusted empirical  $P$  value of 0.0250, which is more exact (and more accurate) than that predicted by Bonferroni correction. In addition, with 10,000 simulations, the standard error for our estimate of statistical significance level is 0.0016 when the true significance level is 0.0250. Therefore, these data support an association between haplotypes with  $-2044A$ , including allele  $-2044A$  itself, and susceptibility to substance dependence. A putative functional variant at the OPRM1 locus that causes a direct effect on the development of substance dependence may thus cosegregate with allele  $-2044A$  or haplotypes with  $-2044A$ .

In the AA population, there were no differences in allele, haplotype, or genotype frequency distributions between cases and controls (see Tables II and III). The adjusted empirical  $P$  value for the 31 n-marker haplotype analysis is 0.3332, via 10,000 simulations. These negative findings indicate that there might be no biological association or true genetic linkage between any of these markers and substance dependence in this AA population. One possible explanation for the lack of association is that a putative functional OPRM1 polymorphism directly responsible for a phenotypic effect (if it is not  $-2044A$  itself) may not cosegregate with these markers in the AA population. Thus, the role of OPRM1 in substance dependence, or at least the specific associated allele(s), is population-specific. We cannot exclude the possibility that the negative findings in AAs may occur by chance due to insufficient statistical power, because our AA sample size is small (55 controls; 62 cases).

+17C/T is a functional variant. Some haplotypes with marker IV (+17C/T) were found to have associations with substance dependence in EAs (see Table V). However, we observed that adding marker IV into any haplotypes, for example, I-IV ( $P = 0.0017$ ), did not increase the statistical significance over that observed with marker I ( $-2044C/A$ ) alone ( $P = 0.0017$ , see Table V). We also failed to find a statistically significant frequency difference when comparing allele +17T alone between cases and controls either in EAs or AAs (Table III). These findings indicate that marker +17C/T does not contribute to the statistical difference in haplotype frequencies. That is, +17C/T may not really associate with substance dependence (or may not be in sufficient LD with the putative disease susceptibility allele in our populations to be detected). However, since the variant allele +17T is rare in our EAs (1.4% in controls, 0.5% in cases), this negative finding may be due to inadequate statistical power. A previous study on US Caucasian population samples ( $n = 100$ ), Finnish sample ( $n = 324$ ), and American Indian population sample ( $n = 367$ ) also found that +17T was rare (1.5%) and failed to find a significant frequency difference between patients with drug abuse or alcohol dependence and controls [Bergen et al., 1997]. In our previous study (in a sample that is partially overlapping with the present sample), we also found that +17T is rare in the EA population (0.6%) and failed to find an association between this polymorphism and substance dependence [Gelernter et al., 1999]. Although +17T is more common in AAs (17.9% in total) (Table III), we failed to find significant allele, haplotype, or genotype frequency differences at this marker between cases and controls in the AA population, which is consistent with our previous study [Gelernter et al., 1999] and the study by Hoehe et al. [2000].

+118A/G is another functional variant, but +118A/G also does not significantly add information to haplotypes listed in Table V. We also did not find a statistically significant difference at allele +118G between cases and controls among either EAs or AAs (Table III). This extends our earlier report [Gelernter et al., 1999] through an increase in sample size in EAs. Since the variant

allele +118G is rare in AAs (2.5% in total subjects), we cannot exclude inadequate statistical power as an explanation for the lack of association. In the study by Hoehe et al. [2000], +118G is also rare and does not associate with substance dependence in the AA population (4.0% in cases, 4.3% in controls), consistent with the findings in the present study. The +118G allele is more common in our EAs (12.2% in total). This negative finding in EAs is consistent with several previous studies. A previous study on US Caucasian population samples, a Finnish sample, and an American Indian population sample found that there was no difference in +118A/G between patients with drug abuse or alcohol dependence and controls [Bergen et al., 1997]. Another study of AAs, Caucasians, and native North Americans identified five different SNPs including +17C/T and +118A/G in the coding region of the opioid receptor gene, but did not find the +118A/G variant allele was present in a significantly different proportion of non-opioid-dependent subjects when compared to opioid-dependent subjects [Bond et al., 1998]. A study of German individuals failed to detect any significant association between +118A/G and alcohol dependence [Sander et al., 1998]. We previously found no association between +118A/G alleles and substance dependence in populations of EA, AA, and Hispanics [Gelernter et al., 1999]. An investigation of two German samples, including a large case-control sample and family-controlled samples of heroin-dependent and alcohol-dependent subjects found no support for the hypothesis that the +118A/G polymorphism is a particular risk factor for either opioid or alcohol dependence [Franke et al., 2001]. In summary, a number of studies, including the present study, suggest that the susceptibility allele for substance dependence is not in LD with marker +118A/G.

However, as described in Introduction, several studies reported positive associations of the common allele +118A or the variant allele +118G with substance dependence in Hispanic, Caucasian, and Chinese populations. Because the sample size of the study on Hispanics [Bond et al., 1998] was small ( $n = 58$  opioid-dependent cases vs.  $n = 9$  controls) and our previous study [Gelernter et al., 1999] did not confirm this result in a somewhat larger Hispanic subsample, we conclude that the positive finding in Hispanics [Bond et al., 1998] probably represents a Type I error. In the study on the mixed samples of Caucasians and non-Caucasians, the positive finding could be due to population stratification effects [Town et al., 1999]. A most recent study increased the Caucasian sample size of this study by Town et al. [1999] and found that there was a significant association between substance dependence and the common allele, +118A ( $P = 0.004$ ) [Schinka et al., 2002]. Due to the strong associations in that study by Schinka et al. [2002] between sex and age and diagnosis (cases vs. controls: male, 97.8 vs. 48.1%; mean age, 46.9 vs. 72.8 years), these could be confounding factors interfering with the association between diagnosis and genotype. It is in some cases necessary to test whether the association between diagnosis and allele is robust after excluding the confounding from sex and age using stratification analysis or multivariate regression analysis. Alterna-

tively, the difference between the study by Schinka et al. [2002] and the present study could be explained by phenotypic heterogeneity. Their diagnostic subtypes included dependence on alcohol, cigarettes, cocaine, marijuana, and/or other illegal drugs, but we included dependence on alcohol, cocaine, and/or opiates in the present study. In contrast, the study on a Chinese sample found the variant allele (+118G) to be a risk factor for heroin dependence [Szeto et al., 2001]. If this finding is correct, the difference between the study on Chinese sample and the present study could be explained by population stratification effects as described above. But it is still difficult to understand why substance dependence in different populations [Chinese: Szeto et al., 2001; Caucasians: Schinka et al., 2002] associates to different alleles, i.e., +118A and +118G, which express different functional isoforms, Asn40 and Asp40, with opposite phenotypic effects. In summary, in view of the inconsistent data, we conclude that any effect of +118A/G on susceptibility to substance dependence remains to be established.

Interestingly, we found that the four sequence variants [−1793T/A, −1699insT, −1320A/G, and −111C/T] were in virtually complete LD and cosegregated in our EAs and AAs. This four-variant pattern segregates like a biallelic marker with only two haplotypes 1111 and 2222, i.e., [−1793T, −1699 non-T, −1320A, −111C] and [−1793A, −1699T, −1320G, −111T]. The frequency of the variant type haplotype 2222 is extremely low in EA subjects (0% in controls and 0.2% in cases), but more common in AA subjects (8.2% in controls and 7.3% in cases). We cannot find any evidence of association between this sequence pattern and any of the seven subtypes of substance dependence in either population (see Tables III and IV). A similar LD pattern in AAs was also reported by Hoehe et al. [2000], in which an association with opioid/cocaine dependence in AAs was reported. However, when we re-analyzed their data using CLUMP (1,000 simulations) to directly compare the 52 haplotype frequency distributions between cases and controls, we obtained an exact  $P$  value of 0.522 ( $T_1$  statistic). In the original analyses in that study by Hoehe et al. [2000], they divided the haplotypes into two groups with cluster analysis, and then ran simulations on the clustering process and got an adjusted  $P$  value of 0.011. Actually, the haplotype clustering process is performed essentially through collapsing some columns together. Therefore, their simulation-adjusted  $P$  value might not be stringent. Grouping these four haplotypes into two clusters, i.e., 1111 + 1211 vs. 1222 + 2222, we obtain a  $P$  value of 0.024 (cases vs. controls, see Table IV), but this  $P$  value is non-significant after correction due to multiple tests ( $\alpha = 0.05/10 = 0.005$ , in which 10 is the number of possible combinations, i.e.,  $C_4^1 + C_4^2$ , to classify four haplotypes into two categories). Thus, the positive result may have resulted in part from insufficiently corrected multiple statistical tests. Additionally, the frequency of the 2222 pattern in the AA cases in that study by Hoehe et al. [2000] is similar to that in the present study (8.8 vs. 7.3%), but the frequency of that pattern in their AA controls is lower than that in the present study (1.4 vs. 8.2%) (see Table IV), which

indicates that their nominally positive association may result from the relatively lower frequency of the haplotype in their controls, not from the relatively higher frequency in their cases. Possible explanations for the low haplotype frequency in their controls might be due to chance sampling, given the small number of controls ( $n = 35$ ). Data for the other 21 AA cases and 16 AA controls in that study by Hoehe et al. [2000] were not available from the original paper, so it is difficult to compare them with our findings. Though the small size of the AA control sample was, clearly, a limitation for the present study as well ( $n = 55$ ), we preliminarily conclude that this four-variant sequence pattern might not play a major role in the susceptibility to AA substance dependence (see Table IV).

Additionally, we demonstrated that in a range of samples of different population origin and phenotype, four of the variants always occur together ( $[-1793T/A, -1699insT, -1320A/G]$  and the 5'-UTR  $[-111C/T]$ ); there might be exceptions if studied in a larger sample; however, our cumulative sample is already quite large. Almost all of the information is, therefore, obtainable by studying only one of these variants. Therefore, future studies should focus on a single one of these four.

In summary, we conclude that allele  $-2044A$  (and haplotypes that include  $-2044A$ ) may be a susceptibility allele (or susceptibility haplotypes) for the combination of alcohol and opioid dependence in EAs. This association suggests a population- and diagnosis-specific role of OPRM1 in the pathophysiology of substance dependence. We propose that study of possibly functional correlates of this variant is warranted.

## ACKNOWLEDGMENTS

A. Lacobelle provided excellent technical assistance.

## REFERENCES

- American Psychiatric Association. 1987. Diagnostic and statistical manual of mental disorders, third edition, revised. Washington, DC: American Psychiatric Press.
- Basbaum AI, Fields HL. 1984. Endogenous pain control systems: Brainstem spinal pathways and endorphin circuitry. *Ann Rev Neurosci* 7:309–338.
- Bergen AW, Kokoszka J, Peterson R, Long JC, Virkkunen M, Linnoila M, Goldman D. 1997. Mu opioid receptor gene variants: Lack of association with alcohol dependence. *Mol Psychiatry* 6:490–494.
- Berrettini WH, Hoehe MR, Ferraro TN, Demaria PA, Gotthel D. 1997. Human mu opioid receptor gene polymorphisms and vulnerability to substance abuse. *Addict Biol* 2:303–308.
- Bilsky E, Montegut M, Delong C, Reid L. 1992. Opioidergic modulation of cocaine conditioned place preferences. *Life Sci* 50:85–90.
- Blouin AG, Perez EL, Blouin JH. 1988. Computerized administration of the diagnostic interview schedule. *Psychiatry Res* 23:335–344.
- Bond C, LaForge KS, Tian M, Melia D, Zhang S, Borg L, Gong J, Schluger J, Strong JA, Leal SM, et al. 1998. Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: Possible implications for opiate addiction. *Proc Natl Acad Sci USA* 16:9608–9613.
- Cadoret TJ, 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 TJ, Yates WR, Troughton E, Woodworth G, Stewart MA. 1995. Adoption study demonstrating two genetic pathways to drug abuse. *Arch Gen Psychiatry* 52:42–52.
- Corrigall W, Coen K. 1991. Opiate antagonists reduce cocaine but not nicotine self-administration. *Psychopharmacology* 104:167–170.
- De Vry J, Donselaar I, Van Ree J. 1989. Food deprivation and acquisition of intravenous cocaine self-administration in rats: Effect of naltrexone and haloperidol. *J Pharmacol Exp Ther* 251:735–740.
- De Waele JP, Kiiianmaa K, Gianoulakis C. 1995. Distribution of the  $\mu$  and  $\delta$  opioid binding sites in the brain of the alcohol-preferring AA and alcohol-avoiding ANA lines of rats. *J Pharmacol Exp Ther* 275:518–527.
- Franke P, Wang T, Nothen MM, Knapp M, Neidt H, Albrecht S, Jahnes E, Propping P, Maier W. 2001. Nonreplication of association between mu-opioid-receptor gene (OPRM1) A118G polymorphism and substance dependence. *Am J Med Genet* 1:114–119.
- Gelernter J, Kranzler HR, Cubells J. 1999. Genetics of two mu opioid receptor gene (OPRM1) exon I polymorphisms: Population studies, and allele frequencies in alcohol- and drug-dependent subjects. *Mol Psychiatry* 5:476–483.
- Goodwin FK. 1989. From the alcohol, drug abuse, and mental health administration. *JAMA* 19:2654.
- Grove WM, Eckert ED, Heston L, Bouchard TJ, Segal N, Lykken DT. 1990. Heritability of substance abuse and antisocial behavior: A study of monozygotic twins reared apart. *Biol Psychiatry* 27:1293–1304.
- Höhl V. 2002. A polymorphism (A118G) in the  $\mu$ -opioid receptor gene affects the response to morphine-6-glucuronide in humans. *Pharmacogenetics* 12:1–2.
- Hernandez-Avila CA, Wand GS, Luo XG, Gelernter J, Kranzler HR. 2003. Association between the cortisol response to opioid blockade and alleles of the *Asn40Asp* polymorphism at the  $\mu$ -opioid receptor locus (OPRM1). *Am J Med Genet (Neuropsychiatr Genet)* 118B:60–65.
- Herz A. 1997. Endogenous opioid systems and alcohol addiction. *Psychopharmacology* 129:99–111.
- Hoehe MR, Kopke K, Wendel B, Rohde K, Flachmeier C, Kidd KK, Berrettini WH, Church GM. 2000. Sequence variability and candidate gene analysis in complex disease: Association of mu opioid receptor gene variation with substance dependence. *Hum Mol Genet* 19:2895–2908.
- Houdi A, Bardo M, Van Loon G. 1989. Opioid mediation of cocaine-induced hyperactivity and reinforcement. *Brain Res* 497:195–198.
- Inder WJ, Livesey JH, Donald RA. 1998. Peripheral plasma levels of beta-endorphin in alcoholics and highly trained athletes and the relationship to a measure of central opioid tone. *Horm Metab Res* 30:523–525.
- Kosten T, Kleber H, Morgan C. 1989. Role of opioid antagonists in treating intravenous cocaine abuse. *Life Sci* 44:887–892.
- Kranzler HR, Gelernter J, O'Malley S, Hernandez-Avila CA, Kaufman D. 1998. Association of alcohol or other drug dependence with alleles of the mu opioid receptor gene (OPRM1). *Alcohol Clin Exp Res* 6:1359–1362.
- Kreek MJ. 1996. Opiates, opioids, and addiction. *Mol Psychiatry* 1:232–254.
- Langsrud Ø. 1999. Fisher's Exact Test. Modified August 3, 1999, from the World Wide Web <http://www.matforsk.no/ola/fisher.htm>.
- Long JC, Williams RC, Urbanek M. 1995. An E-M algorithm and testing strategy for multiple locus haplotypes. *Am J Hum Genet* 56:799–810.
- Lötsch J, Skarke C, Grösch S, Darimont J, Schmidt H, Geisslinger G. 2002. The polymorphism A118G of the human mu-opioid receptor gene decreases the pupil constrictory effect of morphine-6-glucuronide but not that of morphine. *Pharmacogenetics* 12:3–9.
- McGue M, Pickens RW, Svikis DS. 1992. Sex and age effects on the inheritance of alcohol problems: A twin study. *J Abnorm Psychol* 101:3–17.
- Mello N, Mendelson M, Luksa B, Lukas S. 1990. Buprenorphine and NTX effects on cocaine self-administration by Rhesus monkeys. *J Pharmacol Exp Ther* 254:926–939.
- Mirin SM, Weiss RD, Sollogub A, Michael J. 1984. Substance abuse and psychopathology. Washington DC: American Psychiatric Association Press. pp 75–106.
- North BV, Curtis D, Sham PC. 2002. A note on the calculation of empirical P values from Monte Carlo procedures. *Am J Hum Genet* 71:439–441.
- O'Malley SS, Jaffe AJ, Chang G, Schottenfeld RS, Meyer RE, Rounsaville B. 1992. Naltrexone and coping skills therapy for alcohol dependence. A controlled study. *Arch Gen Psychiatry* 49:881–887.
- Pasternak GW. 1993. Pharmacological mechanisms of opioid analgesics. *Clin Neuropharmacol* 1:1–18.

- Peterson R, Goldman D, Long JC. 1999. Nucleotide sequence diversity in non-coding regions of ALDH2 as revealed by restriction enzymes and SSCP. *Hum Genet* 104:177–187.
- Pickens RW, Svikis DS, McGue M, Lykken DT, Heston LL, Clayton PL. 1991. Heterogeneity in the inheritance of alcoholism. *Arch Gen Psychiatry* 48:19–28.
- Rounsaville BJ, Kosten TR, Weissman MM, Prusoff B, Pauls D, Anton SF, Merikangas K. 1991. Psychiatric disorders in relatives of probands with opiate addiction. *Arch Gen Psychiatry* 46:33–42.
- Sander T, Gscheidel N, Wendel B, Samochowiec J, Smolka M, Rommelspacher H, Schmidt LG, Hoehe MR. 1998. Human  $\mu$  opioid receptor variation and alcohol dependence. *Alcohol Clin Exp Res* 22:2108–2110.
- Schinka JA, Town T, Abdullah L, Crawford FC, Ordorica PI, Francis E, Hughes P, Graves AB, Mortimer JA, Mullan M. 2002. A functional polymorphism within the mu-opioid receptor gene and risk for abuse of alcohol and other substances. *Mol Psychiatry* 7:224–228.
- Sham PC, Curtis D. 1995. Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. *Ann Hum Genet* 59:97–105.
- Spitzer RL, Endicott J. 1975. Schedule for affective disorders and schizophrenia: Lifetime version. New York: New York Biometrics Research Division, New York State Psychiatric Institute.
- Spitzer RL, Williams JBW, Gibbon M, First MB. 1992. The structured clinical interview for DSM-III-R (SCID), I: History, rationale, and description. *Arch Gen Psychiatry* 49:624–629.
- Szeto CY, Tang NL, Lee DT, Stadlin A. 2001. Association between mu opioid receptor gene polymorphisms and Chinese heroin addicts. *Neuroreport* 6:1103–1106.
- Town T, Abdullah L, Crawford F, Schinka J, Ordorica PI, Francis E, Hughes P, Duara R, Mullan M. 1999. Association of a functional mu-opioid receptor allele (+118A) with alcohol dependency. *Am J Med Genet* 5:458–461.
- Uitenbroek DG. 1997. SISA Bonferroni. Southampton: D.G. Uitenbroek. Retrieved January 01, 2002, from the World Wide Web <http://home.clara.net/sisa/bonfer.htm>.
- Volpicelli JR, Alterman AI, Hayashida M, O'Brien CP. 1992. Naltrexone in the treatment of alcohol dependence. *Arch Gen Psychiatry* 49:876–880.
- Wand GS, Mangold D, El Deiry S, McCaul ME, Hoover D. 1998. Family history of alcoholism and hypothalamic opioidergic activity. *Arch Gen Psychiatry* 55:1114–1119.
- Wand GS, McCaul M, Yang X, Reynolds J, Gotjen D, Lee S, Ali A. 2002. The mu-opioid receptor gene polymorphism (A118G) alters HPA axis activation induced by opioid receptor blockade. *Neuropsychopharmacology* 1:106–114.
- Zadina JE, Hackler L, Ge LJ, Kastin AJ. 1997. A potent and selective endogenous agonist for the mu-opiate receptor. *Nature* 6624:499–502.
- Zubieta JK, Gorelick DA, Stauffer R, Ravert HT, Dannals RF, Frost JJ. 1996. Increased  $\mu$  opioid receptor binding detected by PET in cocaine-dependent men is associated with cocaine craving. *Nat Med* 2:1225–1229.