

Catechol-O-Methyl-Transferase Functional Polymorphism and Nicotine Dependence: an Evaluation of Nonreplicated Results

David T. Redden,¹ Peter G. Shields,² Leonard Epstein,³ E. Paul Wileyto,⁴ Stanislav O. Zakharkin,¹ David B. Allison,¹ and Caryn Lerman⁴

¹Department of Biostatistics, Section on Statistical Genetics, University of Alabama at Birmingham, Birmingham, Alabama;

²Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University, South Carolina, California;

³Department of Pediatrics, State University of New York at Buffalo, Buffalo, New York; and ⁴Department of Psychiatry, Transdisciplinary Tobacco Use Research Center, University of Pennsylvania, Philadelphia, Pennsylvania

Abstract

Review articles have focused attention on and cited possible reasons for the nonreplication of genetic association studies. Herein, we illustrate how one might work through these possible reasons to make a judgment about the most plausible reason(s) when faced with two or more studies which yield seemingly inconsistent results. In the first study, 342 treatment-seeking smokers were genotyped for the Val¹⁰⁸Met polymorphism in the functional catechol-O-methyltransferase (COMT) locus. Alleles coding Val at codon 108 are denoted as H and those coding Met are denoted as L. An association between presence of the "H" (high activity) allele and pretreatment level of nicotine dependence level using the Fagerstrom Test for Nicotine Dependence was detected ($P = 0.0072$), after controlling for baseline body mass index (BMI, kg/m²), depression symptoms, and age. To validate this initial

finding, 443 treatment-seeking smokers from an independent smoking cessation clinical trial were genotyped for the COMT polymorphism. Within the second study, no association between presence of the "H" allele and nicotine dependence was detected ($P = 0.6418$) after controlling for baseline BMI, depression symptoms, and age. We critically reviewed both studies with regard to often cited reasons for nonreplication, including type I error, population stratification, low statistical power, and imprecise measures of phenotype. Although in our opinion the failure to replicate the initial association in the second study is likely either the result of low statistical power to detect a small effect or effect heterogeneity, thorough analyses failed to definitively identify the reason for nonreplication. (Cancer Epidemiol Biomarkers Prev 2005;14(6):1384-9)

Introduction

Over the past decade, numerous research projects have reported associations between complex phenotypes [obesity, diabetes mellitus (types I and II), smoking persistence, etc.] and genetic polymorphisms in various regions of the human genome (1-3). A fundamental tenet of scientific studies is the need for replication. No single study is definitive, and failure to replicate a positive finding on a consistent basis would lessen the credibility of the reported association. Unfortunately, many reported associations have not been replicated in independent research. Nonreplication of these association findings is a concern and has caused some researchers to question the utility of association methodology in genetic association studies (4-6). The numerous reasons cited for the low replication rate of genetic association studies include spurious findings due to population stratification, lack of control of type I error rates, and insufficient statistical power within the replication sample (7).

Confounding due to population stratification is often emphasized in association studies. Simply stated, population stratification can create confounding leading to spurious results in genetic association studies. We define a spurious association as any association between allelic variation at a

marker locus and phenotypic variation that is due neither to the marker locus causing variation in the phenotype nor to being linked to a marker locus that causes variation in the phenotype. Confounding by population stratification can occur when both of two conditions are met (8-10). First, allele frequencies under investigation must vary among subpopulations. Second, disease prevalence or mean phenotypic value must vary among subpopulations. If researchers are unaware of the subpopulations, subjects can be unknowingly selected from differing subpopulations and a spurious association between genotype and phenotype can be created. Under such circumstances, the relationship between allele frequency and disease is confounded by subpopulation. Fortunately, statistical methods have been developed to identify when association studies may be influenced by population substructure and to adjust for the influence of population stratification in association studies (11-14). Even when population stratification is absent or has been controlled statistically, other causes that could lead to a false significance (inflated type I error rates due to multiple testing and influential observations) must be addressed. Within studies attempting to replicate reported associations, the issues of the appropriate effect size to investigate, statistical power to detect the genetic effect, and comparability of study populations and measures must be addressed.

This article presents two studies that used independent samples of treatment-seeking smokers to test for an association between a functional single nucleotide polymorphism (SNP) in the catechol-O-methyltransferase (COMT) gene and nicotine dependence. COMT is the primary enzyme involved in the degradation and inactivation of dopamine (15). The COMT gene has a SNP in exon 3, a G1947A transition, which results in the substitution of a methionine for valine at codon 108

Received 9/1/04; revised 1/10/05; accepted 2/10/05.

Grant support: National Cancer Institute, National Institutes on Drug Abuse grants P50 CA/DA 84718 and RO1 CA63562 (C. Lerman) and National Institutes of Diabetes, Digestive and Kidney Diseases grants K25 DK062817-01, R01DK056366, and P30DK056336.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: David T. Redden, Department of Biostatistics, RPHB 309D, 1530 3rd Avenue South, Birmingham, AL 35294-0022. Phone: 205-975-9165; Fax: 205-975-2540. E-mail: dredden@ms.soph.uab.edu

Copyright © 2005 American Association for Cancer Research.

(Val¹⁰⁸Met). This amino acid change affects the function of the transcribed protein 3- to 4-fold; the valine high activity (H) allele produces higher levels of enzyme activity and therefore decreased brain dopamine levels (16, 17). Because nicotine stimulates the secretion of dopamine into the neuronal synapse (18), we hypothesized that smokers who carried the high activity (H) allele and who therefore may have lower levels of dopamine, would experience a different level of reward from nicotine and possibly be more or less nicotine dependent.

The initial study of 342 individuals provided evidence in support of this hypothesis. However, the second study of 443 individuals failed to replicate the association between COMT and nicotine dependence detected within the first study. The analyses of this article document and illustrate a process to identify and examine possible reasons for the observed findings regarding the relationship between the COMT locus and nicotine dependence. Using the replication sample, this article illustrates the challenges of using a readily available independent sample to attempt confirmation of a genetic association. Furthermore, this article examines factors that may have contributed to nonreplication, including population stratification, influential observations, lack of power, and a type I error in the first study. Finally, this article presents recommendations regarding research designs to improve the reliability of findings.

Materials and Methods

Subjects and Procedures. Both the initial study and replication study included genetic information and baseline variables collected from smokers enrolled in smoking cessation clinical trials. For both studies, smokers responding to local media advertisements for free smoking cessation treatment and to physician referrals were screened for eligibility (study 1 recruited participants from Washington DC and Philadelphia and study 2 recruited participants from Washington, DC and Buffalo, NY). Eligible individuals were current cigarette smokers who were ages ≥ 18 years and had smoked at least 10 cigarettes/d for the prior 12 months. Exclusion criteria included pregnancy or lactation, uncontrolled hypertension, unstable angina, heart attack or stroke within the past 6 months, current treatment or recent diagnosis of cancer, drug or alcohol dependence, current diagnosis or history of a *Diagnostic and Statistical Manual Fourth Edition* psychiatric disorder, seizure disorder, and current use of bupropion or nicotine-containing products other than cigarettes. All participants completed self-report questionnaires during an in-person visit to the clinic, before the initiation of treatment.

To reduce potential bias due to racial admixture, analyses were limited to persons of European ancestry ($n = 342$ in study 1 and $n = 443$ in study 2). All participants provided a 40-mL blood sample for genotyping and completed a set of standardized self-report questionnaires.

Genotyping Methods. PCR for the COMT genotype was done as described (19). Briefly, genomic DNA (50 ng) was amplified using 20 pmol of each primer (5'-CTGTGGCTACT-CAGCTGTG-3' and 5'-CCTTTTCCAGTCTGACAA-3'), in Applied Biosystems buffer [10 mmol/L Tris-HCl (pH 8.3), 50 mmol/L KCl, 1.5 mmol/L MgCl₂], and Amplitaq DNA polymerase (1.0 unit; Roche Molecular Systems, Alameda, CA) with 2'-deoxynucleoside-3'-triphosphates (1.87 mmol/L; Invitrogen, Carlsbad, CA) in a 50- μ L reaction volume. The PCR reaction had an initial melting temperature of 95°C (5 minutes) followed by 35 cycles of melting (95°C, 1 minute), annealing (55°C, 1 minute), and extension (72°C, 1 minute). An extension period of 4 minutes at 72°C followed the final cycle. The PCR reactions were done using an Applied Biosystems GeneAmp PCR System 9700 (Foster City, CA). The 169-bp PCR product

was digested overnight at 37°C with Hsp92II (Promega, Madison, WI) and gave rise to fragments of 114, 23, and 32 bp for the high activity allele and 96, 23, 32, and 18 bp for the low activity allele. The digested PCR fragments were resolved by agarose gel electrophoresis (4% w/v, MetaPhor agarose, Cambrex Bio Science, Rockland, ME) and observed by ethidium bromide staining. The assay was validated by confirming polymorphic Mendelian inheritance patterns in 12 human family cell lines ($n = 133$ family members), encompassing three generations (data not shown; samples were obtained from National Institute of General Medical Sciences Human Genetic Mutant Cell Repository, Coriell Institute, Camden, NJ). Two independent investigators read the results of the genotyping study and 20% of the samples were repeated for quality control and their genotypes were confirmed. There were no discordant samples.

Forty-three SNPs were genotyped to assess possible population substructure. These 43 markers were spaced such that there were two to three SNPs per chromosome (three SNPs in the larger chromosomes and two in the smaller ones). None of these 43 SNPs were close to or in genes under study. These SNPs were not in coding regions. All SNPs have a minor allele frequency in people of European ancestry and African Americans of >0.1 . All SNPs were available as validated and inventoried assay kits from Applied Biosystems.

Measures

Outcome: Nicotine Dependence. Baseline nicotine dependence was assessed using the Fagerstrom Test for Nicotine Dependence. This scale is a six-item self-report measure of nicotine dependence derived from the Fagerstrom Tolerance Questionnaire (20). Scores can range from 1 to 10 with higher scores indicating greater nicotine dependence.

Covariates. Weight was measured at baseline using a balance beam scale that was calibrated daily. Self-reported weight and measured weight at baseline were highly correlated ($r = 0.97$, $P < 0.00001$). Gender, education, marital status, age, and ethnic ancestry of grandparents were assessed by self-report during the pretreatment assessment visit. The Center for Epidemiologic Studies Depression Scale was used to assess the severity of depression symptoms (21).

Explanatory Variable: COMT Genotype. Using methods described above, participants were classified as homozygous for the H (high activity) allele, heterozygous, or homozygous for the low activity (L) allele. Based on our hypotheses, we characterized participants according to the presence (H/H or H/L) or absence (L/L) of the high activity allele.

Statistical Methods. Statistical examination of the data from the two studies included the cross-tabulation of genotypes by study to ensure comparability of the prevalence of the COMT H allele in the two studies. Tests of Hardy-Weinberg equilibrium (HWE) were used to test whether the genotypic information conformed to the Hardy-Weinberg principles. Two-sample t tests assuming unequal variances were used to compare the two samples of smokers with regard to baseline nicotine dependence, age, depression symptoms, and self-reported body mass index (BMI). Analysis of covariance (ANCOVA) models were used to test for an association between the COMT locus and the baseline Fagerstrom Test for Nicotine Dependence. Specifically, a variable added-last test was conducted to determine if presence of the H allele was significantly associated with nicotine dependence in the presence of and controlling for baseline depression symptoms as measured by the Center for Epidemiologic Studies Depression Scale, baseline self-reported BMI (kg/m²), and age (22). The covariates were included in the models because previous research had indicated associations of these variables with nicotine dependence and smoking cessation (23, 24). All covariates were standardized to have a mean of zero and SD of

one to improve interpretability of the estimated variables. In addition to the analysis of the *COMT* locus, we examined hypotheses for the *DRD2 Taq1* locus and the *DAT VNTR* locus (results not included). For all models, there was only one phenotype under investigation, nicotine dependence. To account for the multiple models built to address different loci, Bonferroni corrections were used to maintain experiment wise error at or below 0.05. The effect size attributable to the presence of the H allele at the *COMT* locus was estimated by the difference in the coefficient of the determination (R^2) for the regression model containing baseline BMI, depression symptoms, age, and an indicator variable for the presence of the H allele and the coefficient of the determination (R^2) for the regression model containing only baseline BMI, depression symptoms, and age (i.e., semipartial R^2). Influence diagnostics, specifically Cook's D , were calculated to examine whether any individual observation exhibited undue influence upon the fitting of the ANCOVA models.

To examine for confounding due to population substructure, a method recommended by Pritchard and Rosenberg (13) was used. This method uses the sum of K single degree of freedom (df) Pearson χ^2 test statistics calculated from the cross-tabulation of case-control status and the frequency of alleles at K independent random markers. Evidence of substructure is indicated if the observed sum is an unlikely result from a χ^2 distribution with K degrees of freedom. Cases were defined as study participants with baseline Fagerstrom scores total greater than or equal to 6 points, indicating high levels of nicotine dependence (25). To assess false-positive report probability, the method proposed by Wacholder et al. (26) was employed.

All analyses were conducted using SAS 9.0 (Cary, NC).

Results

Characteristics of Study Participants within Study 1.

Characteristics of study participants within study 1 are presented in Table 1. The frequency of the H allele within study 1 is estimated to be 0.5. Within the initial association study, 91 individuals had the *COMT* HH genotype (26.6%), 160 individuals had the HL genotype (46.8%), and 91 individuals had the *COMT* LL genotype (26.6%). The participants of this study had a mean age of 46.7 years (SD = 10.9), smoked on average 24.8 cigarettes/d (SD = 9.1), scored on average 5.9 points on the Fagerstrom Test for Nicotine Dependence (SD = 2.1), had mean BMI of 27.5 kg/m² (SD = 5.7), and scored on average 11.6 points on the Center for Epidemiologic Studies Depression Scale (SD = 8.4). Males

Table 1. Comparison of characteristics between the two studies

Variable	Study 1 ($n = 342$), count (%)	Study 2 ($n = 443$), count (%)
<i>COMT</i> Genotype		
HH	91 (26.6)	145 (32.7)
HL	160 (46.8)	173 (39.1)
LL	91 (26.6)	125 (28.2)
Gender		
Male	155 (45.3)	231 (52.1)
Female	187 (54.7)	212 (47.9)
Nicotine dependence Genotype	Mean (SD)	Mean (SD)
HH	5.95 (2.09)	5.15 (2.01)
HL	6.06 (2.09)	5.55 (2.04)
LL	5.36 (2.00)	5.30 (2.00)
Age (y)	46.7 (10.9)	44.5 (11.3)
Cigarettes/d	24.8 (9.1)	22.8 (9.6)
BMI	27.5 (5.7)	25.87 (4.27)
Depression symptoms	11.6 (8.4)	12.55 (8.56)

accounted for 45.3% of the sample. A test of HWE provided no significant evidence ($\chi^2 = 1.42$, $df = 1$, $P = 0.23$) of deviations from the principles of Hardy-Weinberg.

Association of *COMT* Alleles with Nicotine Dependence within Study 1. Analysis of study 1 data found that the presence of the H allele was significantly associated with nicotine dependence ($P = 0.0072$), controlling for standardized baseline depression symptoms ($P = 0.0042$), self-reported BMI ($P = 0.0164$), and age ($P = 0.0078$). Details of the model are provided in Table 2. Before attempting to replicate this association within another sample, regression influence diagnostics were examined to ensure that the significance of the association could not be attributed to the influence of outliers or leverage points. Although two observations exhibited moderate influence relative to other observations within the analysis, removal of these two points actually increased the association between the presence of the H allele and nicotine dependence. Because confounding due to unrecognized population substructure has been discussed extensively within genetic association studies (7-9), 43 random markers were used to examine whether population substructure existed among the 342 subjects. Using the hypothesis test proposed by Pritchard and Rosenberg (13), no evidence of substructure was found ($\chi^2 = 39.87$, $df = 43$, $P = 0.61$). Based upon the observed P associated with the presence of the H allele and assuming a range of prior probability from 0.25 to 0.001 that the *COMT* locus affects nicotine dependence, false-positive report probability was calculated using the method of Wacholder et al. (26). If one assumes the prior probability of a true effect to be 0.25, 0.1, 0.01, or 0.001, the false-positive report probability is 0.05, 0.13, 0.63, or 0.95, respectively.

Characteristics of Study Participants within Study 2.

Characteristics of study participants within study 2 are presented in Table 1. The frequency of the H allele within the replication study is estimated to be 0.523. Within the replication study, 144 individuals had the *COMT* HH genotype (32.7%), 173 individuals had the HL genotype (39.2%), and 124 individuals had the *COMT* LL genotype (28.1%). The participants of this study had a mean age of 44.5 years (SD = 11.3), smoked on average 22.8 cigarettes/d (SD = 9.6), scored on average 5.4 points on the Fagerstrom Test for Nicotine Dependence (SD = 2.0), had mean BMI of 25.9 kg/m² (SD = 4.3), and scored on average 12.6 points on the Center for Epidemiologic Studies Depression Scale (SD = 8.56). Males accounted for 52.14% of the sample. A test of HWE provided significant evidence ($\chi^2 = 20.93$, $df = 1$, $P < 0.0001$) of deviations from the principles of Hardy-Weinberg. Examinations of genotype frequencies revealed that the sample contained more HH and LL genotypes and fewer HL genotypes than expected under the principles of HWE.

Association of *COMT* Alleles with Nicotine Dependence within Study 2. Given the lack of influential observations and lack of evidence of population substructure within the initial study, the association between the presence of the H allele and nicotine dependence was examined within an independent set of 443 individuals. The necessary outcomes, covariates, and genotypes on these individuals were readily available due to their participation in a parallel nicotine replacement study. Analysis of the study 2 data found the presence of the H allele was not significantly associated with nicotine dependence ($P = 0.6418$) in the presence of the standardized baseline depression symptoms ($P = 0.0132$), self-reported BMI ($P = 0.2956$), and age ($P = 0.0007$). Details of the model are provided in Table 2.

Given the failure to replicate the association, influence diagnostics were examined to ensure that the lack of significance could not be attributed to masking due to the influence of outliers or leverage points. Two observations

Table 2. Ordinary least squares estimates from the regression of Fagerstrom Test for Nicotine Dependence upon age, BMI, Center for Epidemiologic Studies Depression Scale, and presence of H allele

Predictor	Estimate	SE	T	P
Study 1				
Age	0.30093	0.11240	2.68	0.0078
BMI	0.23211	0.09626	2.41	0.0164
Depression symptoms	0.31920	0.11060	2.89	0.0042
COMT*	0.66827	0.24713	2.70	0.0072
MSE	4.04890			
Study 2				
Age	0.32139	0.09454	3.40	0.0007
BMI	0.11719	0.11191	1.05	0.2956
Depression symptoms	0.23855	0.09587	2.49	0.0132
COMT*	0.09916	0.21303	0.47	0.6418
MSE	3.95734			

*Represents an indicator variable equal to 1 if COMT genotype is either HH or HL and equal to 0 if COMT genotype is LL.

within the replication study exhibited moderate influence relative to other observations within the analysis, removal of these two points actually decreased the association between the presence of the H allele and nicotine dependence. Because confounding due to unrecognized population substructure can also mask associations within genetic association studies (27), the 43 random markers were again used to determine if substructure existed among the 443 subjects. Using the hypothesis test proposed by Pritchard and Rosenberg (13), no evidence of substructure was found ($\chi^2 = 40.51$, $df = 43$, $P = 0.58$).

Comparison of the Two Study Samples. To investigate potential reasons for the discrepancy of the results of the two analyses, comparisons of the genotype frequencies, allele frequencies, and means of age, BMI, depression symptoms, and nicotine dependence were conducted. Genotype frequencies for COMT did not differ between the two studies ($\chi^2 = 5.31$, $df = 2$, $P = 0.07$). The frequency of the H allele did not differ between the two studies ($\chi^2 = 0.25$, $df = 1$, $P = 0.62$). To examine whether the study groups differed with regard to the 43 random markers, a modified version of Pritchard and Rosenberg's test was used. Instead of the cross-tabulating case-control status and the frequency of alleles at K independent random markers, we cross-tabulated a study indicator and the frequency of alleles at K independent random markers and summed the K independent χ^2 statistics. Evidence of differing population structure between the two studies is indicated if the observed sum is an unlikely result from a χ^2 distribution with K degrees of freedom. However, this approach revealed no evidence of differing population structure ($\chi^2 = 46.4$, $df = 43$, $P = 0.34$).

Significant mean differences exist between the two samples with regard to nicotine dependence ($P = 0.0011$), age ($P = 0.0054$), and BMI ($P < 0.0001$). On average, the participants of the initial study were older, heavier, and more nicotine dependent than participants within the replication study. Results of other comparisons are contained in Table 1. To examine if these differences could account for the discrepancy in association results, an ANCOVA model on the pooled data was constructed.

Pooled Analysis. Given the different distribution of covariates between the two studies and the lack of population substructures within the data, a pooled ANCOVA analysis was constructed to adjust the analyses for differences in covariates. Regression analysis of the pooled samples indicated the presence of the H allele at the COMT locus was significantly associated with nicotine dependence ($P = 0.0345$) in the presence of the standardized baseline depression symptoms ($P = 0.0002$), self-reported BMI ($P = 0.0096$), age

($P < 0.0001$), and an indicator for initial study membership ($P = 0.0089$).

Pooling the data from the two studies allows for tests of effect heterogeneity (i.e., the effect of covariate adjustment between the two studies is different or the measured effect of the H allele at the COMT locus differed between studies). To further explore the pattern of results, ANCOVA models using the pooled data and including interactions between COMT and study indicator were developed. A significant association between COMT and nicotine dependence ($P = 0.0068$) and a nearly significant interaction between marker and study ($P = 0.0795$) were observed. There was no indication that the effects of age ($P = 0.8891$), BMI ($P = 0.5812$), and depression ($P = 0.4366$) upon baseline nicotine dependence differed between studies.

Statistical Power. After finding that population substructure and influential observations did not account for the results of the studies, we focused our attention upon statistical power. Specifically, we seek to estimate the power of the second study to replicate the statistically significant finding of the first study.

Arguably, there are two separate approaches that can be used to estimate statistical power within a replication study. The first approach, documented in Taylor and Muller (28), is to use the estimated residual variance and estimated effect size from the initial study in the calculation of the power. However, because both the estimated residual variance and estimated effect size are observed random variables, statistical power within the replication study becomes a random variable. Hence, the question becomes not what is the power of the replication study but what is the likely range of power for replication studies of fixed sample size based upon random estimates. Based upon the estimates of the initial study [residual variance (controlling for age, depression, and BMI) of 4.05 and effect size of 2%], power to replicate the initial findings with a fixed sample size of 443 ranges would likely range from as low as 20.9% to 100% (95% confidence interval). Ideally, this interval would have a much narrower range with the lower bound closer to 80%.

The second approach, documented in Taylor and Muller (29) is to a priori determine the minimum effect size of scientific relevance worth detecting and use only the estimated residual variance to estimate power in a replication study. Given nicotine dependence is a complex phenotype and variability in one locus is unlikely to explain a medium to large amount of the variability, an increase in R^2 of 1% might be a reasonable minimum effect size. However, because the residual variance is still estimated, statistical power for the replication study would still be a random variable. Based upon the estimates of the initial study [residual variance (controlling for age, depression, and BMI) of 4.05] and assuming a priori scientifically relevant effect size of 1%, power to replicate the initial findings with a fixed sample size of 443 would likely range from as low as 53% to 65% (95% confidence interval). Table 3 presents estimated power for a range for other possible effect sizes.

Table 3. 95% Confidence interval for power based upon estimates from the initial study

Effect size (partial R^2)	95% Confidence lower bound	95% Confidence upper bound
0.01	0.53	0.65
0.0125	0.61	0.74
0.015	0.69	0.81
0.0175	0.76	0.87
0.02	0.81	0.91
0.0225	0.86	0.94
0.025	0.90	0.96

Discussion

In the comparison of studies that report disparate results of significance, a series of questions naturally arise. Among these questions are how comparable are the studies in design and measurements, how comparable are the populations sampled, how likely is it that a significant report is actually a type I error, how likely is it that a nonsignificant study is actually a type II error, are the studies possibly confounded, and could there be effect modification? We address each of these questions below.

Both studies reported here were conducted by the same investigational team. Although designed to test different smoking interventions, both studies report the phenotypic outcome (baseline nicotine dependence) using the Fagerstrom Test for Nicotine Dependence before treatment. Genotyping for both studies was conducted by the same lab under the same quality control measures. Therefore, the phenotypic and genotypic measures between the two studies can be considered comparable. With regard to equivalence of samples, there were baseline differences between the samples. On average, the participants of the initial study were older, heavier, and more nicotine dependent. Because interactions between a study indicator and these covariates were not significant, it seems unlikely that these differences in covariate levels between studies explain the observed discrepancy in the observed tests of association between *COMT* and nicotine dependence. With regard to *COMT* genotype frequency and prevalence of the H allele, the two studies seem comparable. However, a test of HWE for the *COMT* locus was rejected for the replication study. Typically, rejection of HWE implies nonrandom mating or recent immigration. Given the reported results from the test of population substructure, these are unlikely interpretations. Given the quality control standards used and a review of genotyping results, we do not believe the rejection of HWE is due to genotyping errors. Another possible reason for the observed deviation from HWE is the violation of the assumption of random sampling from the population under investigation, which is required for the test of HWE to be valid. Both studies used nonrandom self-selecting samples defined by treatment-seeking smokers responding to local media advertisements. Other possible reasons for the observed rejection of HWE include genetic selection, assortative mating, or a type I error. However, we cannot definitively ascribe the apparent departure from HWE to any single cause.

With regard to the initial study, it is possible that the finding is a type I error. Within the study we did use type I error control methods (i.e., Bonferroni corrections). We examined specifically three hypothesis tests for three different loci (*COMT*, *DRD2 Taq1*, and *DAT VNTR*). Therefore, under the statistical assumptions of the hypothesis tests, there is an experiment-wise chance of no more than 5% that we have committed at least one type I error. As recommended by a reviewer, we estimated the false-positive report probability as detailed in Wacholder et al. (26). We would like to note that we agree with many of the criticisms presented by Thomas and Clayton (30) of the method. However, calculation of the false-positive report probability did not indicate that the results of the initial study were highly likely to be a false-positive result assuming a prior probability of 0.01.

We examined the power of our replication study to detect the association observed within the initial study. This calculation proved to be complex and must account for the facts that the effect size and variance are estimated from a previous study. As indicated in Table 3, if the true effect of the *COMT* locus was slightly less than the estimated effect in the initial study, then the method of Taylor and Muller (29) indicates that the replication study was likely underpowered. Statistical power within the replication study may have been adversely affected by both measurement error in the

phenotype as well as measurement error in the depression covariate. Therefore, we cannot rule out the possibility of a type II error within the replication sample. A future study designed to detect an effect size of partial R^2 equal to 1% based upon the information estimated from our initial study would require 850 individuals to be 95% confident that the power was at least 80%.

To address possible confounding due to population stratification, we measured 43 biallelic SNPs on the study participants in both studies and used the method of Pritchard and Rosenberg (13) to test for population substructure. Examination of evidence of substructures within the first study was warranted due to the possibility that the detected association between nicotine dependence and the *COMT* locus was spurious due to confounding by population substructure. Examination of evidence of substructures within the second study was warranted due to the possibility that the association between nicotine dependence and the *COMT* locus was masked due to confounding by population substructure. However, no evidence of population stratification was found. Although no evidence indicating population stratification was found, the possibility of confounding due to population substructure cannot be absolutely excluded. If more SNPs were available, our ability to detect possible population substructures would increase (13). However, given the lack of evidence of substructure provided by the analysis of the 43 available SNPs, we believe confounding due to population stratification studies is not a likely explanation for the observed discrepancy in the observed tests of association. Furthermore, examination of influential observations indicated that the different results from the studies were not attributable to outliers or leverage points.

Pooling the data from the two studies together into one ANCOVA model provided marginal evidence of effect heterogeneity between the two studies. Given that the participants of the two studies were recruited from different geographic locations, this marginal *P* may indicate that the observed associations between the high activity allele at the *COMT* locus and nicotine dependence is moderated by some unobserved environmental factor confounded between studies. Based upon the analyses done, we believe a likely explanation of our results is either a type II error in the replication study or possibly effect heterogeneity. Other possible explanations, which we believe to be less likely, include a type I error in the initial study or confounding due to undetected population stratification.

Conclusions

We report a positive association study between the high activity allele at the *COMT* locus and nicotine dependence within a Caucasian sample after controlling for age, BMI, and depression symptoms. Using a second independent study, we failed to replicate this association. A critical review of both studies examining the typically cited reasons for nonreplication in numerous review articles was done. Careful and thorough statistical review of the studies did not definitively identify a specific reason for nonreplication. Examination of the designs and results of the studies indicate a likely type II error within the second study or effect heterogeneity between the two studies. We cannot unequivocally exclude the type I error or confounding due to population stratification as other possible explanations. Using the replication sample, this article illustrates the challenges of using a readily available independent sample to attempt confirmation of a genetic association and highlights the need for research into the design and conduct of adequately powered replication studies. Our findings warrant future research on the association between the *COMT* locus polymorphism and nicotine dependence.

Acknowledgments

We thank Freddie Patterson, Angela Pinto, Susan Kucharski, and Vyga Kaufman for assistance with project implementation and Stephanie Restine and Drs. Wade Berrettini and Debra Leonard for their assistance with selection and analysis of SNPs to assess population stratification.

References

- Chagnon YC, Rankinen T, Snyder EE, et al. The human obesity gene map: the 2002 update. *Obes Res* 2003;11:313–43.
- Hirschhorn JN, Lohmueller K, Bryne E, et al. A comprehensive review of genetic association studies. *Genet Med* 2002;4:45–61.
- Lerman C, Shields PG, Wileyto EP, et al. Pharmacogenetic investigation of smoking cessation treatment. *Pharmacogenetics* 2002;12:627–34.
- Gambaro G, Anglani F, D'Angelo A. Association studies genetic polymorphisms and complex disease. *Lancet* 2002;355:308–11.
- Strohman R. Maneuvering in the complex path from genotype to phenotype. *Science* 2002;296:701–3.
- Holtzman NA. Putting the search for genes in perspective. *Int J Health Serv* 2001;31:445–61.
- Sullivan PF, Eaves LJ, Kendler KS, et al. Genetic case-control studies in neuropsychiatry. *Archives of Genetic Psychiatry* 2001;58:1015–24.
- Cardon LC, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;361:598–604.
- Wacholder S, Rothman N, Caporaso N. Population Stratification in epidemiological studies of common genetic variants and cancer: quantification of bias. *J Natl Cancer Inst* 2000;92:1151–8.
- Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;2:91–9.
- Allison DB, Neale MC. Joint tests of linkage and association for quantitative traits. *Theor Popul Biol* 2001;60:239–51.
- Satten GA, Flanders WD, Yang Q. Accounting for unmeasured population substructure in case control studies of genetic association studies using a novel latent-class structure. *Am J Hum Genet* 2001;68:466–77.
- Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999;65:220–8.
- Hoggart CJ, Esteban JP, Shriver MD, et al. Control of confounding of genetic association in stratified populations. *Am J Hum Genet* 2003;72:1492–504.
- Cooper JR, Bloom FE, Roth RH. The biochemical basis of neuropharmacology. Oxford: New York; 1996.
- Lachman HM, Papolos DF, Saito T, et al. Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders. *Pharmacogenetics* 1996;6:243–50.
- Lotta T, Vidgren J, Tigmann C, et al. Kinetics of human soluble and membrane-bound catechol-O-methyl-transferase: a revised mechanism and description of the thermolabile variant of the enzyme. *Biochemistry* 1995;4:4202–5.
- Pontieri FE, Tanda G, Orzi F, et al. Effects of nicotine on the nucleus accumbens and similarity to those of addictive drugs. *Nature* 1996;18:255–7.
- Thompson PA, Shields PG, Freudenheim JL, et al. Genetic polymorphisms in catechol-O-methyltransferase, menopausal status, and breast cancer risk. *Cancer Res* 1998;58:2107–10.
- Heatherton TF, Kozlowski LT, Frecker RC, et al. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Bri J Addict* 1991;86:1119–27.
- Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977;1:385–401.
- Muller KE, Fetterman BA. Regression and ANOVA: an integrated approach using SAS® software. Cary: SAS Institute; 2002.
- Lerman C, Kaufmann V, Rukstalis M, et al. Individualizing nicotine replacement therapy for the treatment of tobacco dependence: a randomized trial. *Ann Intern Med* 2004 Mar 16;140:426–33.
- Lerman C, Audrian J, Orleans CT, et al. Investigation of mechanisms linking depressed mood in nicotine dependence. *Addict Behav* 2004;21:9–19.
- Fagerstrom KO, Heatherton TF, Kozlowski LT. Nicotine addiction and its assessment. *Ear Nose Throat J* 1990;69:763–5.
- Wacholder S, Chanock S, Garcia-Closas M, et al. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004 Mar 17;96:434–42.
- Redden DT, Allison DB. Nonreplication in genetic association studies of obesity and diabetes research. *J Nutr* 2003;133:3323–6.
- Taylor DJ, Muller KE. Bias in linear model power and sample size calculation due to estimating noncentrality. *Communications in Statistics: Theory and Methods* 1996;25:1595–610.
- Taylor DJ, Muller KE. Computing confidence bounds for power and sample size of the general linear univariate model. *American Statistician* 1995;49:43–7.
- Thomas DC, Clayton DG. Betting odds and genetic associations. *J Natl Cancer Inst* 2004;96:421–3.