Association of the Tryptophan Hydroxylase Gene With Smoking Initiation But Not Progression to Nicotine Dependence

Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia

We investigated the association between two markers in the seventh intron of the tryptophan hydroxylase gene (TPH C218A and C779A) in a population-based case control study of 780 genotyped subjects. As in prior studies, the two markers were in strong linkage disequilibrium. The phenotypes we studied were smoking initiation and progression to nicotine dependence. Allele, genotype, and estimated haplotype frequencies for each marker were highly significantly different for smoking initiation (P < 0.0004 for each comparison) and were nonsignificant for progression to nicotine dependence. An empirical test suggested that the positive results were unlikely to have resulted from population stratification. Our results are similar to those of Lerman et al. [2001: Am J Med Genet (Neuropsychiatr Genet) 105:000–000] in associating these TPH markers with a construct related to smoking initiation but dissimilar in the variable implicated. If these results replicate in other samples, the serotonergic system may be involved in the etiology of smoking initiation given the rate-limiting role of TPH in the biosynthesis of serotonin.

INTRODUCTION

Cigarette smoking is a major public health problem [U.S. Department of Health and Human Services, 1989; World Health Organization, 1997]. Adoption of the smoking habit consists of several phases, including initiating smoking and the subsequent development of nicotine dependence [U.S. Department of Health and Human Services 1989; Sullivan and Kendler 1999]. There are substantial individual differences in each of these phases: although approximately 75% of U.S. adults have initiated smoking (i.e., smoked at least one cigarette) [Andreski and Breslau, 1993; Anthony et al., 1994; Centers for Disease Control, 1998], only about 20% develop nicotine dependence [Anthony et al., 1994; Breslau et al., 1994; Kendler et al., 1999].

A substantial body of evidence suggests that genetic effects are etiologically important for both smoking initiation and progression to nicotine dependence. In a recent review of the genetic epidemiology of smoking, we estimated that additive genetic effects accounted for 56% of the variance in liability to smoking initiation and 67% of the variance for progression to nicotine dependence [Sullivan and Kendler, 1999]. Attempting to elucidate the genetic basis of various smoking phenotypes is an increasingly active area of study.

Recently, Lerman et al. [2001] informed us that they had found a significant association of age at onset of smoking with an intronic variant in the tryptophan hydroxylase gene (TPH C779A) in a multiple regression model (P = 0.009). The hypothesis that genetic variation in TPH is related to smoking behavior is plausible. TPH catalyzes the rate-limiting step in serotonin biosynthesis and a component of cigarette smoke irreversibly inhibits the activity of TPH in vitro [Minami et al., 1993]. Some studies have found associations between TPH polymorphisms with impulsivity, hostility, and/or suicidal behavior [Nielsen et al., 1994, 1998; Mann et al., 1997; New et al., 1998; Manuck et al., 1999], which, except for the latter, may be elevated in smokers. However, our linkage study of nicotine dependence did not suggest that TPH was a positional candidate gene [Straub et al., 1999]. Moreover,

KEY WORDS: nicotine; smoking; dependence; initiation; tryptophan hydroxylase; serotonin; genetic association
although alterations in serotonergic function have been implicated in smoking, serotonergic antidepressants are not particularly effective in smoking cessation [Hughes et al., 2000] and there has been greater emphasis in the literature on mesolimbic dopaminergic pathways [Dani and Heinemann, 1996; Koob, 1996].

TPH C218A and C779A are both located in intron 7 of the TPH gene and are in strong linkage disequilibrium [Nielsen et al., 1997; Kunugi et al., 1999; Marshall et al., 1999]. The C779A polymorphism lies in a polypropylidine consensus sequence immediately upstream of the 3’ acceptor splice site, and transversions to purines have been found to cause hemophilia and other genetic diseases. However, sequencing of TPH cDNA generated from transformed lymphoblast mRNA from homozygotes at either C779A or C218A revealed no exon skipping or alternative splicing [Nielsen et al., 1997].

The C218A SNP is located in a possible GATA transcription factor–binding site, but there have been no reports of the relative expression levels of the different TPH alleles. A report associating C779A with CSF 5-HIAA levels [Nielsen et al., 1994] was not replicated [Mann et al., 1997; Nielsen et al., 1998].

We report here results from a population-based association study of smoking initiation and progression to nicotine dependence for two polymorphisms located in intron 7 of the TPH gene.

MATERIALS AND METHODS

Subjects

Our sample was selected from two population-based twin studies described elsewhere [Kendler and Prescott, 1999]; no subject participated in our prior linkage study [Straub et al., 1999]. We studied twins who were of European ancestry, a randomly selected member of a dizygotic pair, and had an available DNA sample. All subjects were unrelated and provided informed consent.

Smoking data were collected in a similar manner in each twin study [Kendler et al., 1999] and included basic smoking history, the Fagerstrom Tolerance Questionnaire (FTQ) [Fagerstrom, 1978], and the severity of withdrawal symptoms. The widely used FTQ is an eight-item scale (range, 0–11) that assesses the degree of dependence on nicotine. Scores of seven or more are consistent with nicotine dependence [Fagerstrom and Schneider, 1989]. The time frame for the FTQ was the subject’s lifetime period of maximum cigarette use. We defined three groups: lifetime nonsmokers who reported never having smoked an entire cigarette; regular smokers with a low degree of nicotine dependence (defined as an FTQ score of 0–3 in the lifetime period of maximum use) despite having smoked regularly for at least 5 years; and lifetime regular smokers with a high degree of nicotine dependence (FTQ score of 7–11 in the lifetime period of maximum use). The process of becoming dependent on nicotine includes at least two critical steps [U.S. Department of Health and Human Services 1989; Kendler et al., 1999]: beginning to smoke regularly and the development of nicotine dependence contingent upon beginning to smoke regularly. Thus, we defined two dichotomous comparisons: smoking initiation (never smoked vs. regular smokers with low and high FTQ scores) and progression to nicotine dependence (regular smokers with low vs. high FTQ scores).

DNA Extraction

DNA was obtained during direct interview by having the subjects collect buccal epithelial cells using standard cytology brushes (Fisher Scientific) [Richards et al., 1993]. DNA preparation was done from four brushes per subject using the InstaGene Purification Matrix (Biorad Laboratories).

TPH Markers

In Genbank sequence AC005728, SNP C218A is at position 103,820 and SNP C779A is at position 103,259. The use of fluorescence polarization (FP), template-directed dye-terminator incorporation (TDI) to analyze single-base-extension reaction products as a method of genotyping has been described in detail [Chen and Kwok, 1999; Chen et al., 1999]. We have optimized conditions and our modified protocol follows.

Step 1: PCR amplification. For SNP C218A, the PCR primers were VCU-P1316 (GGGACTCAAAACCAATGGAAT) and VCU-P1317 (CTGGTTACTT). For SNP C719A, the TDI primer was VCU-P1258 (CCTAAACTTCTGGTTACTT). For SNP C729A, the PCR primers were VCU-P1260 (ACCGTTGCAGTTTTTGAAC) and VCU-P1367 (GAGTGGCCAAGGTTTTGAAC). PCR amplification was performed in a total volume of 10 ml, containing 100 nM of each primer, 10 mM dNTPs, 0.075 U/ml PE-ABI AmpliTaq Gold and 5 ng/ml of single-strand binding protein (USB). Cycling conditions were 10 min at 95°C, then 1 min–2 min–1 min cycles at the following temperatures: 94°C for 30 min and the enzymes were heat-inactivated at 95°C for 15 min. The resulting mixture was kept at 4°C for 1 h.

Step 2: Primer and dNTP degradation. Six μl of a cocktail containing 0.08 μl/ml shrimp alkaline phosphatase (Roche), 0.05 μl/ml E. coli exonuclease I (USB) in shrimp alkaline phosphatase buffer (14 mM Tris-HCl, pH 8.0, 50 mM KCl, 2.1 mM MgCl2, 25 μM dNTPs, 0.075 U/μl PE-ABI AmpliTaq Gold and 5 ng/μl of single-strand binding protein (USB). Cycling conditions were 10 min at 95°C, then 1 min–2 min–1 min cycles at the following temperatures: 94°C for 30 min and the enzymes were heat-inactivated at 95°C for 15 min. The resulting mixture was kept at 4°C and used in the step 3 FP-TDI assay without further quantitation or characterization.

Step 3: Genotyping by the FP-TDI assay. For SNP C218A, the TDI primer was VCU-P1260 (TTACATTCCCTATGCTCAGAATAAGCAGCTA); for SNP C719A, the TDI primer was VCU-P1258 (CTTAAGAGGAGGAGTGTTATTCTGAA). To the step 2 product was added 6 μl of TDI reaction mixture, which contained TDI buffer (24 mM Tris-HCl, pH 9.0, 36 mM KCl, 5 mM MgCl2, 1.4 mM NaCl, 2.2% glycerol),
0.27 μM TDI primer, 34 nM each of R6G and Rox dye-labeled ddNTPs, 34 nM of the other two ddNTPs (unlabeled), and 0.025 U/μl Thermosequenase (USB). The reaction mixtures were incubated at 95°C for 2 min, followed by 35 cycles of 94°C for 15 sec and 57°C for 30 sec. At the end of the reaction, the samples were held at 20°C for 10 min.

**Step 4: Fluorescence polarization reading.** Immediately after completion of step 3, FP measurements were made using the Analyst fluorescence reader (LJL Biosystems, Sunnyvale, CA). The supplied software calculates the degree of polarization, expressed by the unit mP, which is written directly to a text file on the Analyst. We created custom spreadsheets and macros in Microsoft Excel that automatically plot the mP values to allow for visual inspection and quality control and converts mP values to allele numbers.

**Assessment of Population Stratification**

Population stratification is a potential threat to the validity of case control association studies. We used the stratification chi-square statistic [Pritchard and Rosenberg, 1999] to test for population stratification empirically. All cases and controls were genotyped at 16 unlinked microsatellite markers (heterozygosity mean, 0.73; range, 0.62–0.85) to assess the presence of population stratification.

**Statistical Analysis**

We used the Pearson chi-square test to compare the allele and genotype distributions for the two TPH markers for groups defined by smoking initiation and progression to nicotine dependence. We estimated statistical power [Schlesselman, 1982] for allelic comparisons assuming a type I error rate of 0.05 and genetic homogeneity. The allele frequencies of the two TPH markers were approximately 0.30. Given the allele frequencies and the sample size for the smoking initiation comparison, there was 80% power to detect the effect of a predisposing allele if its frequency in cases was ~ 9% greater than controls (relative risk ~ 1.5). For progression to nicotine dependence, the power calculations were only slightly different despite the smaller sample sizes.

Haplotype frequencies in the three groups of subjects were estimated by maximum likelihood using an approach described by Sham [1998] and implemented with Mx [Neale et al., 1999]. We performed planned comparisons to test two hypotheses, namely, the estimated TPH haplotypes of regular smokers with low and high FTQ scores did not differ (a test of progression to nicotine dependence), and lifetime nonsmokers did not differ from the two groups of regular smokers (a test of smoking initiation). Initially, a saturated model was fit with all three groups having separate haplotype frequencies. The first hypothesis was tested by equating haplotype frequencies for the low- and high-nicotine-dependent groups while allowing the nonsmokers to differ (model A). The difference of fit between this model and the saturated model is asymptotically distributed as chi-square with degrees of freedom equal to the difference in the number of parameters equated. The second hypothesis was tested by constraining all three groups to have the same haplotype frequencies and comparing the fit of this model to that of model A.

**RESULTS**

As we reported previously [Silverman et al., 2000], genotypes from 16 unlinked microsatellite markers suggested no significant evidence of population stratification across groups defined by smoking initiation (stratification chi-square = 134.2; df = 117; \( P = 0.13 \)) and for groups defined by progression to nicotine dependence (stratification chi-square = 110.1; df = 107; \( P = 0.40 \)).

Table I shows comparisons of genotype and allelic frequencies for the two biallelic TPH markers for the smoking phenotypes of initiation and progression to nicotine dependence. For C218A, there were highly significant differences in both genotype and allele frequencies for smoking initiation (i.e., lifetime non-smokers vs. regular smokers with low and high FTQ scores). However, genotype and allele frequencies for C218A were similar for progression to nicotine dependence (i.e., regular smokers with low FTQ scores vs. those with high FTQ scores). The results for the second TPH marker (C779A) were quite consistent with C218A. As in other samples [Nielsen et al., 1997; Kunugi et al., 1999; Marshall et al., 1999], these two markers were in strong linkage disequilibrium (\( D^2 = 0.97 \)) [Devlin and Risch, 1995].

As expected, there were highly significant differences for smoking initiation for estimated TPH C218A–C779A haplotypes. Estimates of haplotype frequencies were marginally significant for progression to nicotine dependence (\( P = 0.042 \)), but the absolute differences were slight and of dubious meaning as several cells had low expectations.

To place these findings in an experimental context, we have genotyped polymorphic markers in 22 genes which we believed to be defensible candidate genes for smoking behavior. Results from three genes were significant at the 0.05 level; two results were in the 0.01–0.05 range (HPRT and DRD5) and the TPH results show the strongest association with any smoking phenotype.

We next attempted to gain further understanding of the association between smoking initiation and these TPH polymorphisms by analyzing additional phenotypic data available on these subjects. We chose to analyze C779A because of the greater numbers of available genotypes. As shown in Table II, there was no significant effect of genotype (controlling for group and gender) on any of the smoking history variables (including age at onset of regular smoking), lifetime diagnoses of major depression and alcoholism, and on novelty seeking scores.

**DISCUSSION**

In this large (\( n = 780 \)) population-based association study of smoking behavior and two markers in intron 7 of...
of the tryptophan hydroxylase gene (TPH), we found significantly different allele, genotype, and estimated haplotype frequencies for smoking initiation (comparing lifetime nonsmokers with regular smokers). Each of these comparisons were highly significant for smoking initiation and none was compellingly significant for progression to nicotine dependence (regular smokers with low vs. high FTQ scores).

False positive results are a known problem in association studies (i.e., case control studies with genetic markers as risk factors) [Crowe, 1993; Kidd, 1993; Gambaro et al., 2000] and arise from three basic processes. First, false positive results could arise by chance particularly if a large number of genes were studied. Although TPH was among 22 genes we studied, the probability of such a strong association ($P = 0.0001$) arising by chance alone is small (Bonferroni the probability of such a strong association ($P = 0.0001$)) and, consequently, most significant associations will be false positives.

In conclusion, consistent with the findings of Lerman et al. [2001], we have found a highly significant association between smoking initiation and two markers located in intron 7 of TPH. It will be important to see whether positive results are found in independent replication studies.

### Table I. Associations of Two TPH Markers With Smoking Phenotypes

<table>
<thead>
<tr>
<th>TPH Marker</th>
<th>Smoking initiation</th>
<th>Progression to nicotine dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>TPH C218A</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genotype frequencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>0.473</td>
<td>0.334</td>
</tr>
<tr>
<td>C/A</td>
<td>0.430</td>
<td>0.471</td>
</tr>
<tr>
<td>A/A</td>
<td>0.097</td>
<td>0.195</td>
</tr>
<tr>
<td>Allele frequencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.688</td>
<td>0.570</td>
</tr>
<tr>
<td>A</td>
<td>0.311</td>
<td>0.430</td>
</tr>
<tr>
<td><strong>TPH C779A</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genotype frequencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C/C</td>
<td>0.476</td>
<td>0.345</td>
</tr>
<tr>
<td>C/A</td>
<td>0.448</td>
<td>0.489</td>
</tr>
<tr>
<td>A/A</td>
<td>0.076</td>
<td>0.166</td>
</tr>
<tr>
<td>Allele frequencies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0.700</td>
<td>0.590</td>
</tr>
<tr>
<td>A</td>
<td>0.300</td>
<td>0.410</td>
</tr>
<tr>
<td><strong>C218A–C779A haplotypes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-C</td>
<td>0.685</td>
<td>0.563</td>
</tr>
<tr>
<td>C-A</td>
<td>0.010</td>
<td>0.022</td>
</tr>
<tr>
<td>A-C</td>
<td>0.005</td>
<td>0.000</td>
</tr>
<tr>
<td>A-A</td>
<td>0.300</td>
<td>0.416</td>
</tr>
<tr>
<td>TPH C779A Genotype</td>
<td>Smoking phenotype</td>
<td>Smoking (low FTQ score)</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Effect of genotype controlling for gender</td>
<td>Initiation</td>
<td>Progression</td>
</tr>
<tr>
<td>CC</td>
<td>1.8 (3.4)</td>
<td>7.3 (3.1)</td>
</tr>
<tr>
<td>CA</td>
<td>1.9 (1.1)</td>
<td>1.8 (1.0)</td>
</tr>
<tr>
<td>AA</td>
<td>1.9 (1.1)</td>
<td>1.8 (1.0)</td>
</tr>
</tbody>
</table>

*The data shown are mean (SD) or proportion by smoking classification and by genotype. The statistical comparisons in the rightmost two columns are from multivariate regression (continuous data) and logistic regression (proportional data) controlling for the effects of smoking classification and gender. All comparisons were not significant (P > 0.10).

The data are for the lifetime period of maximum cigarette use and are undefined for lifetime nonsmokers. FTQ means Fagerstrom Tolerance Questionnaire. Cessation depression and cessation anxiety identify subjects who became very depressed or very anxious in the context of an attempt to quit smoking.

These data were derived for all smoking classification groups. Life-time diagnoses of DSM-III-R major depression and alcohol dependence were determined via structured personal interview.

This study is supported by DA-10228 to K.S.K. and MH-01458 to M.C.N.

**ACKNOWLEDGMENTS**

We thank Dr. Caryn Lerman, Georgetown University, for informing us of their findings in advance of publication. Supported by DA-10228 to K.S.K. and MH-01458 to M.C.N.

**REFERENCES**


