A Candidate Gene Approach Identifies the CHRNA5-A3-B4 Region as a Risk Factor for Age-Dependent Nicotine Addiction

Robert B. Weiss1*, Timothy B. Baker2, Dale S. Cannon3, Andrew von Niederhausern1, Diane M. Dunn1, Nori Matsunami1, Nanda A. Singh1, Lisa Baird1, Hilary Coon3, William M. McMahon3, Megan E. Piper2, Michael C. Fiore2, Mary Beth Scholand1, John E. Connett4, Richard E. Kanner5, Lorise C. Gahring5,6, Scott W. Rogers6,7, John R. Hoidal5, Mark F. Leppert1*

1 Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah, United States of America, 2 Department of Medicine, Center for Tobacco Research and Intervention, University of Wisconsin School of Medicine, Madison, Wisconsin, United States of America, 3 Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, Utah, United States of America, 4 Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota, United States of America, 5 Department of Internal Medicine, Division of Respiratory, Critical Care and Occupational Pulmonary Medicine, University of Utah School of Medicine, Salt Lake City, Utah, United States of America, 6 Geriatric Research Education and Clinical Center, Salt Lake City VA Medical Center, Salt Lake City, Utah, United States of America

Abstract

People who begin daily smoking at an early age are at greater risk of long-term nicotine addiction. We tested the hypothesis that associations between nicotinic acetylcholine receptor (nAChR) genetic variants and nicotine dependence assessed in adulthood will be stronger among smokers who began daily nicotine exposure during adolescence. We compared nicotine addiction—measured by the Fagerstrom Test of Nicotine Dependence—in three cohorts of long-term smokers recruited in Utah, Wisconsin, and by the NHLBI Lung Health Study, using a candidate-gene approach with the neuronal nAChR subunit genes. This SNP panel included common coding variants and haplotypes detected in eight α and three β nAChR subunit genes found in European American populations. In the 2,827 long-term smokers examined, common susceptibility and protective haplotypes at the CHRNA5-A3-B4 locus were associated with nicotine dependence severity (p = 2.0 × 10⁻⁵; odds ratio = 1.82; 95% confidence interval 1.39–2.39) in subjects who began daily smoking at or before the age of 16, an exposure period that results in a more severe form of adult nicotine dependence. A substantial shift in susceptibility versus protective diplotype frequency (AA versus BC = 17%, AA versus CC = 27%) was observed in the group that began smoking by age 16. This genetic effect was not observed in subjects who began daily nicotine use after the age of 16. These results establish a strong mechanistic link among early nicotine exposure, common CHRNA5-A3-B4 haplotypes, and adult nicotine addiction in three independent populations of European origins. The identification of an age-dependent susceptibility haplotype reinforces the importance of preventing early exposure to tobacco through public health policies.


Editor: Jonathan Flint, The Wellcome Trust Centre for Human Genetics, University of Oxford, United Kingdom

Received April 4, 2008; Accepted June 13, 2008; Published July 11, 2008

Copyright: © 2008 Weiss et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This research was supported, in part, by the National Institutes of Health: NIDA/NHLBI P01-HL72903, NHLBI N01-HR46014, NCI P50-CA84724 and NIDA P50-DA19706. The Lung Health Study (LHS) is conducted and supported by the NHLBI in collaboration with the LHS Study Investigators.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Bob.Weiss@genetics.utah.edu (RBW); Mark.Leppert@genetics.utah.edu (MFL)

Introduction

Nicotine addiction has profound clinical and public health consequences because it is associated with reduced ability to cease tobacco use [1,2], and tobacco use is the leading cause of preventable morbidity and mortality in developed countries [3]. Meta-analysis of numerous twin studies shows that both genes and environment play an important role in smoking-related behaviors [4]. Nicotine is the primary agent in tobacco smoke that leads to addiction, and while progress has been made in finding genes that contribute to nicotine addiction in humans [5–9], there is a great need for additional progress.

Both human and animal research support an a priori hypothesis that age of onset of daily smoking may influence the association of genetic risk variants with measures of nicotine dependence. In humans, early onset of smoking is associated with greater consumption of cigarettes in adulthood [10–15], a relative inability to quit smoking [10,16–18], and a more severe form of nicotine dependence [15,19–22]. Research in rodents shows that nicotine exposure during periadolescence induces long-lasting biochemical, anatomical, physiological, and behavioral changes that differ markedly from those seen with adult exposure [23–27].

The neuronal nicotinic acetylcholine receptor genes (nAChRs) are likely candidates for harboring functional variants contributing to nicotine addiction since these ligand-gated ion channels are the initial physiological targets of nicotine in the central and peripheral nervous system. They have also been implicated in nicotine addiction in animals where chronic nicotine exposure...
Author Summary

Tobacco use is a global health care problem, and persistent smoking takes an enormous toll on individual health. The onset of daily smoking in adolescence is related to chronic use and severe nicotine dependence in adulthood. Since nicotine is the key addictive chemical in tobacco, we tested the hypothesis that genetic variants within nicotinic acetylcholine receptors will influence the severity of addiction measured in adulthood. Using genomic resequencing to define the patterns of variation found in these candidate genes, we observed that common haplotypes in the CHRNA5-A3-B4 gene cluster are associated with adult nicotine addiction, specifically among those who began daily smoking before age 17. We show that in populations of European origins, one haplotype is a risk factor for dependence, one is protective, and one is neutral. These observations suggest that genetic determinants expressed during human adolescence contribute to the risk of lifetime addiction severity produced from early onset of cigarette use. Because disease risk from the adverse health effects of tobacco smoke is related to lifetime tobacco exposure, the finding that an age-dependent effect of these haplotypes has a strong influence on lifetime smoking behavior reinforces the public health significance of delaying smoking onset.

Results

Study Populations

Nicotine dependence was assessed with the Fagerstrom Test of Nicotine Dependence (FTND) because it predicts important dependence attributes such as the likelihood of relapse back to tobacco use and biochemical measures of nicotine self-administration [2,36,37]. The FTND assesses a pattern of heavy, compulsive smoking and has been used in other genetic association studies [5–9]. Genetic variants were assessed in 2,827 subjects from three European American cohorts with a mean age of 49.6 years (SD = 9.5), 1155 (41%) of whom were females (Table S1). All participants were either current or previous daily cigarette smokers; 222 (8%) had not smoked for at least 2 years prior to participation in the study. One cohort comprised participants in a study of genetic risk factors for nicotine dependence and chronic obstructive pulmonary disease (COPD) recruited in Salt Lake City, Utah (N = 486, UT). Another cohort was made up of participants in randomized trials of smoking cessation interventions recruited in Madison and Milwaukee, Wisconsin (N = 398, WI). A final cohort was drawn from the Lung Health Study (N = 1943, LHS), a multisite longitudinal study of COPD sponsored by the Division of Lung Disease of the National Heart, Lung and Blood Institute [38]. All ex-smokers were in the UT cohort and responded to smoking related assessments based upon their prior smoking patterns [39]. Participants began daily smoking at a mean age of 17.3 (SD = 4.1), smoked a mean of 28.5 cigarettes per day (CPD, SD = 13.9), smoked for a mean 30.7 years (SD = 9.5), and had a mean FTND score of 5.7 (SD = 2.2). Consistent with a history of chronic, heavy smoking, most subjects in the UT and LHS cohorts (N = 2302, 81% of the total sample) had mild to moderate chronic obstructive pulmonary disease (COPD) as determined by pulmonary function testing; lung function testing was not performed on the WI subjects.

Age of onset of daily smoking was dichotomized into early onset (onset of daily smoking at age 16 or younger) vs. late onset (onset of daily smoking at age 17 or older). Previous studies have shown that 16 vs. 17 is an appropriate age range to differentiate the impact of early from late nicotine exposure on dependence [40]. We used a dichotomous variable to represent age because the effect was hypothesized to be nonlinear in nature due to the fact that nicotine exerts distinct effects when administered during adolescence vs. adulthood [24]. We attempted to reflect this underlying causal model in our analyses by creating an age cut-score that reflected this nonlinear effect. Also, the age 16 vs. 17 dichotomy approximated a median split, which yielded near equivalent statistical power for tests within the two age samples; 46% of the sample were thus classified as early onset smokers. Research suggests that age of smoking onset can be reliably assessed with retrospective assessments [41]. To compare distinct levels of nicotine dependence, FTND scores were dichotomized into low (FTND = 0–4) and high (FTND = 6–10) dependence [7]. An FTND score of 6 or higher identifies subjects with high nicotine dependence [42], while a score of “5” has an ambiguous relation with dependence [42–44]. Of the initial sample of 2827, 751 (26%) were assigned to the low dependence condition, 1536 (55%) were assigned to the high dependence condition and 540 (19%) had an intermediate score of 5.

Consistent with previous reports, age of onset of daily smoking was inversely related to level of dependence in the present sample. Age of onset and FTND score were correlated, r(2135) = −0.18, P<0.001, and dichotomized age of onset and FTND were highly associated (χ² = 51.6, 1 df, P<0.001). In early onset smokers, 24.2% had low FTND scores compared with 38.7% of late onset smokers.
SNP and Haplotype Discovery

As an initial step, we conducted an exhaustive SNP discovery survey surrounding the neuronal nAChR coding regions in a small sample selected to represent the most extreme heavy and light dependent smokers. We used a genomic resequencing strategy to generate a dense set of variants in α-like nAChR subunits (CHRNA3, CHRNA5, CHRNA6, and CHRNA7) and β-like nAChR subunits (CHRNA2 and CHRNA10) in 144 smokers and 48 population-matched non-smokers. This survey identified 262 SNPs, including 38 non-synonymous, 35 synonymous, 37 UTR, 1 stop, and 1 frameshift (see Table S2 for location and allele frequency). The stop and frameshift alleles, both located in CHRNA6, were each observed as heterozygotes in single individuals, indicating that the depth of the resequencing survey reached the boundary of rare loss-of-function alleles. For a preliminary χ² analysis of association in the resequencing sample, we subdivided the 144 smokers into high (n = 72) and low dependence smokers. We used a genomic resequencing strategy to generate a dense set of variants in 10 nAChR subunits based on inference from the unphased resequencing data, and derived a minimal set of tagging SNPs (n = 87) for genotyping in the larger three-cohort sample.

To evaluate the significance of these CHRNA5–CHRNA3 single marker associations and the underlying haplotypes, five tagging SNPs capable of distinguishing these four haplotypes were assessed for association to FTND by age of daily smoking onset (daily smoking onset by vs. after age 16) in the combined UT-WI-LHS cohorts (total N = 2,037). Observed frequencies of Haplotypes A–D, respectively, by cohort were as follows: LHS, 39%, 37%, 19%, and 5%; UT, 36%, 38%, 20%, and 4%; and WI, 41%, 38%, 17%, and 4%. A test of all four haplotypes showed a significant omnibus association P value of 2.6 × 10⁻⁶ with high vs. low FTND score within the early onset group (Table 2), but not within the late onset group (P = 0.44). This omnibus haplotype test was supported by single marker association P values ranging from 1.1 × 10⁻⁵ to 1.7 × 10⁻⁴ (Figure 1). The main haplotype effect can be partitioned into two significant haplotype-specific associations (Table 2) with both a susceptibility effect for high dependence in the early onset group for haplotype H₄A (odds ratio (OR) 1.50, 95% CI 1.21 to 1.86, P = 1.3 × 10⁻³), and a protective effect for haplotype H₄C (OR 0.66, 95% CI 0.52 to 0.85, P = 1.1 × 10⁻³). Importantly, haplotype H₄ shows no significant association (OR 0.93, 95% CI 0.76 to 1.14, P = 0.50), suggesting that the two haplotype-specific associations are the result of distinct susceptibility (H₄A) and protective (H₄C) haplotype effects in the early onset group.

An interaction between haplotype status and age of onset in predicting dependence severity was also obtained via logistic
regression analyses (Table 3). A significant interaction effect ($P = 0.006$) was found, as well as a significant HA vs. HC contrast within the early onset condition ($P = 2.0 \times 10^{-5}$, OR = 1.82, 95% CI 1.39 to 2.39). HC was associated with reduced likelihood of severe nicotine dependence relative to HA, but only among those smoking daily at age 16 or younger. Gender was associated with age of onset in our cohorts (41% of females began daily smoking before age 17 compared with 49% of males). However, gender did not moderate the relation between dependence and haplotype status within age of onset groups and so was dropped from subsequent analyses. Based on these results for the entire sample, haplotype associations with the dichotomized FTND variable were tested by age of onset condition within each cohort separately (Table 4). There was a significant HA vs. HC contrast in the early onset group for each cohort ($P$s = 0.02–0.001, OR’s = 1.5–3.0), but this contrast was not significant in any cohort in the late onset condition. Thus, the relative risk/protective effect of haplotypes A and C was found only in early onset smokers in three separate cohorts.

Diplotype analyses also suggest that age-of-onset moderates the relation of these variants with nicotine dependence (Figure 2). Logistic regression analysis in the combined sample showed a significant interaction between the diplotype status (AA vs. BC) and age-of-onset ($P = 0.02$, OR = 0.43, 95% CI 0.86 to 0.21). Analyses in the early onset group found significant differences between diplotype AA and all other tested diplotypes ($P$ value...
range from 0.05 to 3.0 × 10⁻⁴. The contrast between diplotype AA vs. diplotype AC (P = 0.03, OR = 0.53, 95% CI 0.93 to 0.30) and diplotype AA vs. diplotype BC (P = 3.0 × 10⁻⁴, OR = 0.35, 95% CI 0.62 to 0.20) suggests a recessive effect of Haplotype A, and an incomplete dominance effect of Haplotype C (Figure 2A).

To ensure that the obtained results were not dependent upon the particular age of daily smoking onset or FTND dichotomy that was used, the effects for both phenotypes were investigated in more detail. First, the interaction between haplotype and age of onset as a continuous variable was tested based on the combined UT-WI-LHS cohorts and using FTND46 as the dependent variable. The HA vs. HC by daily age interaction effect was significant, p = 0.01, OR = 0.95 (95% CI = 0.91–0.99), but the HB vs. HC interaction effect was not significant, p = 0.24, OR = 0.98 (95% CI = 0.94–1.02). Next, subjects were divided into quartiles based on onset age of daily smoking and the haplotype effects within each quartile were tested separately. Table 5 shows the logistic regression analysis for the HA vs. HC effect was significant in the two younger quartiles (p = 0.0003 for 15–16 years and p = 0.01 for 16–17 years), but not in the two older quartiles (p = 0.43 for 17–18 years and p = 0.63 for >18 years). Figure 3 shows the relative haplotype frequencies across each age group quartile.

Prior research suggested that the FTND score of “5” constitutes an intermediate score that is ambiguous with regards to dependence level [42–44]. To explore this assumption we conducted a multinomial logistic regression with all subjects in which haplotype status was related to an FTND dependent variable that was split into three levels: “Low” (i.e., scores 0–4, Table 2. CHRNAS-A3-B4 Haplotype-Specific Association with Dichotomized FTND Scores in the UT-WI-LHS Cohorts.

<table>
<thead>
<tr>
<th>Haplotype Tags</th>
<th>Frequency FTND = 0–4 (n = 256)</th>
<th>Frequency FTND = 6–10 (n = 795)</th>
<th>Frequency FTND = 0–4 (n = 493)</th>
<th>Frequency FTND = 6–10 (n = 776)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA CCAGA</td>
<td>0.318</td>
<td>0.413</td>
<td>1.50 (1.21–1.86)</td>
<td>0.381</td>
</tr>
<tr>
<td>HB TCGGG</td>
<td>0.375</td>
<td>0.358</td>
<td>0.50 (0.76–1.14)</td>
<td>0.385</td>
</tr>
<tr>
<td>HC CTGAG</td>
<td>0.244</td>
<td>0.179</td>
<td>0.66 (0.52–0.85)</td>
<td>0.192</td>
</tr>
<tr>
<td>HD TCGAG</td>
<td>0.063</td>
<td>0.051</td>
<td>0.81 (0.53–1.23)</td>
<td>0.042</td>
</tr>
</tbody>
</table>

*The haplotype-tagging SNPs are rs680244, rs569207, rs16969968, rs578776, rs1051730, and the tag sequences are from the chr. 15 (+) strand. Data in bold indicate significant haplotype-specific tests (p<0.05); an omnibus haplotype test was significant in the early onset group (p = 0.00026) but not in the late onset group (p = 0.444).

To explore this assumption we conducted a multinomial logistic regression with all subjects in which haplotype status was related to an FTND dependent variable that was split into three levels: “Low” (i.e., scores 0–4, Table 3. Logistic Regression Analyses in which Low and High Nicotine Dependence (FTND = 0–4 vs. FTND = 6–10) Was the Dependent Variable and Haplotype C Was the Reference Condition.

<table>
<thead>
<tr>
<th>Haplotype Tags</th>
<th>Interaction OR (95% CI)</th>
<th>Early onset of daily smoking by 16 OR (95% CI)</th>
<th>Late onset of daily smoking after 16 OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA CCAGA</td>
<td>2210 0.006 1.65 (1.16–2.35)</td>
<td>988 2.0 × 10⁻⁴ 1.82 (1.39–2.39)</td>
<td>1222 0.38 1.11 (0.88–1.39)</td>
</tr>
<tr>
<td>HB TCGGG</td>
<td>2094 0.08 1.36 (0.96–1.94)</td>
<td>935 0.04 1.33 (1.02–1.74)</td>
<td>1159 0.83 0.98 (0.78–1.22)</td>
</tr>
</tbody>
</table>

*Interaction = interaction between haplotype and early vs. late onset of daily smoking. N = haplotype count and OR = odds ratio (95% confidence intervals in parentheses). Haplotype C counts by test are: Interaction, 1089; Early Onset, 511; and Late Onset, 578.

Prior research suggested that the FTND score of “5” constitutes an intermediate score that is ambiguous with regards to dependence level [42–44]. To explore this assumption we conducted a multinomial logistic regression with all subjects in which haplotype status was related to an FTND dependent variable that was split into three levels: “Low” (i.e., scores 0–4, Table 4. Logistic Regression Analyses by Cohort in which Low and High Nicotine Dependence (FTND = 0–4 vs. FTND = 6–10) Was the Dependent Variable and Haplotype C Was the Reference Condition.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Haplotype</th>
<th>Early onset of daily smoking by 16 OR (95% CI)</th>
<th>Late onset of daily smoking after 16 OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHS</td>
<td>HA</td>
<td>0.02 1.52 (1.07–2.14)</td>
<td>0.19 1.20 (0.91–1.58)</td>
</tr>
<tr>
<td>UT</td>
<td>HA</td>
<td>0.21 1.25 (0.88–1.78)</td>
<td>0.87 0.98 (0.74–1.29)</td>
</tr>
<tr>
<td>WI</td>
<td>HA</td>
<td>0.001 2.97 (1.53–5.76)</td>
<td>0.72 0.91 (0.56–1.48)</td>
</tr>
<tr>
<td>HB</td>
<td>0.27 1.42 (0.76–2.66)</td>
<td>0.88 0.96 (0.59–1.57)</td>
<td></td>
</tr>
<tr>
<td>HB</td>
<td>0.02 2.16 (1.16–4.00)</td>
<td>0.98 1.01 (0.49–2.08)</td>
<td></td>
</tr>
<tr>
<td>HB</td>
<td>0.13 1.58 (0.88–2.85)</td>
<td>0.94 1.03 (0.49–2.15)</td>
<td></td>
</tr>
</tbody>
</table>
This analysis showed significant haplotype effects only when individuals with Low vs. High FTND scores were contrasted with one another. That is, comparisons involving individuals with Intermediate scores, yielded no significant main effects or interaction effects. An inspection of haplotype distributions showed that individuals with mid-range levels of dependence (FTND of “5”) had haplotype distributions that were intermediate to those of subjects with Low and High scores. Therefore, as would be expected if haplotypes confer dependence vulnerability, when dependence levels became more extreme, so were the relative distributions of haplotypes.

However, we also report secondary analyses with individuals with the full range of FTND scores, and with different score cutpoints (i.e., FTND scores 0–4 vs. 5–10 and 0–5 vs. 6–10), in order to demonstrate that the reported pattern of results does not depend upon exclusion of mid-range scores. With both divisions (FTND scores 0–4 vs. 5–10 and 0–5 vs. 6–10), the Haplotype A vs. C interaction with age of onset was significant, as was the Haplotype A vs. C contrast within early onset but not late onset smokers (Table S4). Further, diplotype analyses indicated a significant AA vs. BC interaction with age of onset for both of the additional FTND divisions and a significant AA vs. BCcontrast within the early but not the late onset condition. Inspection of the relation between FTND deciles and the frequency of HA relative to HC in early onset smokers indicated nonlinearity, supporting the choice to relate genetic variants to dichotomous rather than continuous FTND scores. Finally, AA vs. BC diplotype differed significantly for mean FTND scores in early ($P=5\times10^{-5}$), but not late, onset smokers (Figure 2C and D).

The psychometric performance of the six-item FTND to assess physical dependence on tobacco smoking is well established, particularly items 1 and 4 which relate to heaviness of smoking [36,37]. The significance levels of analyses of the association

---

**Figure 2. Low Nicotine Dependence (%) and Mean FTND Scores by Diploype and Age of Onset of Daily Smoking in the Combined UT-WI-LHS Cohorts.** (A), (B) partitioning of diplotypes between FTND<5 and FTND>5 categories in early and late onset groups. The percentage of individuals in the FTND<5 category is shown by the dichotomous ‘FTND cut’ value (dashed line). (C), (D) mean FTND score of diplotypes by early and late onset; error bars indicate the S.E.M. Sample size by diploype within the early onset condition was: AA, 161; AB, 276; AC, 162; BB, 146; BC, 153; and CC, 34. Within the late onset condition, diplotype sample sizes were: AA, 203; AB, 384; AC, 186; BB, 160; BC, 190; and CC, 34. Diplotype counts are indicated by the width of each associated column within the plots.

doi:10.1371/journal.pgen.1000125.g002

---

N = 731), “Intermediate” (score 5, N = 540), and “High” (score>5, 1556). This analysis showed significant haplotype effects only when individuals with Low vs. High FTND scores were contrasted with one another. That is, comparisons involving individuals with Intermediate scores, yielded no significant main effects or interaction effects. An inspection of haplotype distributions showed that individuals with mid-range levels of dependence (FTND of “5”) had haplotype distributions that were intermediate to those of subjects with Low and High scores. Therefore, as would be expected if haplotypes confer dependence vulnerability, when dependence levels became more extreme, so were the relative distributions of haplotypes.
between individual FTND items and haplotypes are shown in Table 6 and positive scores on most items show a trend towards higher frequency of haplotype HA and a lower frequency of haplotype in the FTND score.

Table 6. Associations between Haplotypes and Individual FTND Items.

<table>
<thead>
<tr>
<th>FTND Item*</th>
<th>p-valuesb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inter</td>
</tr>
<tr>
<td>1. How soon after you wake up do you smoke your first cigarette?</td>
<td>0.56</td>
</tr>
<tr>
<td>2. Do you find it difficult to refrain from smoking in places where it is forbidden?</td>
<td>0.84</td>
</tr>
<tr>
<td>3. Which cigarette would you hate most to give up, first one in the morning?</td>
<td>0.93</td>
</tr>
<tr>
<td>4. How many cigarettes per day did/do you smoke?</td>
<td>0.18</td>
</tr>
<tr>
<td>5. Do you smoke more frequently during the first hours after waking than during the rest of the day?</td>
<td>0.28</td>
</tr>
<tr>
<td>6. Do you smoke if you are so ill that you are in bed most of the day?</td>
<td>0.25</td>
</tr>
</tbody>
</table>

*p-values for items 1 and 4 are from GLM tests in which the 4 levels of the item responses were taken to be an ordinal variable. For items 2, 3, 5 and 6, p-values are from logistic regression analyses.

**Inter** = the interaction between haplotype and age of onset, **All** = all subjects, **Early** = early onset smokers only, and **Late** = late onset smokers only. For GLM analyses, the p-value was obtained in a test across HA, HB, and HC. For logistic regression analyses, the p-value was for the HA vs. HC effect.

doi:10.1371/journal.pgen.1000125.t006
be involved with receptor trafficking. The local amino acid context surrounding the p.Asp398Asn variant also displays accelerated protein evolution (higher lineage-specific Ka/Ks) in primate lineages within a class of genes related to nervous system development [53]. In mouse brain, α3 is a widely distributed minor subunit within heteromeric brain nAChRs, and incorporation of an α3 subunit into brain nAChRs leads to changes in receptor-level function [54]. The association of HA and the p.Asp398Asn variant with a nicotine dependence phenotype in humans suggests further research, such as an engineered mouse model, to explore the functional role and developmental expression of these variants within the process leading to chronic nicotine dependence. Both rs16969968 and rs1051730, a synonymous α3 variant, are in nearly complete linkage disequilibrium and, along with the HA LD structure, occur at significantly higher frequencies in European-American populations based on HapMap allele frequencies.

In-depth haplotype analysis also revealed a protective effect of haplotype HC. Multiple non-coding variants are contained within the HC LD structure; therefore, potential functional variants may indirectly affect receptor function through developmental and/or cell-specific expression of α3, α3 or β4 levels. The risk versus protective effects of HA and HC reinforce an emerging theme in nicotine dependence. Both rs16969968 and rs1051730, a synonymous α3 variant, are in nearly complete linkage disequilibrium and, along with the HC LD structure, occur at significantly higher frequencies in European-American populations based on HapMap allele frequencies.

In our European-American population, we did not find evidence of significant associations (P<0.05 in early or late onset smokers) among the four CHRNA4 SNPs in common with our study (rs2273504, rs2236196, rs1044396, rs1044397; Table S3) and previous candidate gene association studies of nicotine dependence in Chinese men [5], and females of European-American and African-American descent [6]. A follow-up candidate gene study [9] to a genome-wide association design [8] with a definition of no dependence (FTND = 0) in controls and a FTND>3 for dependence in cases of European descent, identified significant associations within CHRNA3 and the CHRNA5-A3-B4 cluster. Although that study [9] and the current study employed tagging SNPs for similar LD bins, the current study did not find any association between nicotine dependence and CHRNA3 SNPs. However, similar LD bins within the CHRNA5-A3-B4 cluster showed a significant risk effect in both studies, including the nonsynonymous α3 rs16969968 SNP, even with the substantial difference in FTND criteria. Using similar criteria for low and high dependence as our study, a previous study [7] also reported a suggestive association (primary P=0.08) of the CHRNA5-A3-B4 cluster in a small sample population of 242 Israeli women. A recent follow-up candidate gene analysis using genome-wide association data from three European populations data suggested a risk effect of CHRNA5-A3-B4 locus for cigarettes per day regularly smoked [32].

Subsequent to submission of this article, a GWA study using smoking quantity (cigarettes per day, CPD) as a measure of nicotine dependence observed an association of rs1051730 with CPD at genome-wide significance levels in a large Icelandic population. That study used FTND item 4 as their measure of cigarettes per day. Table S7 shows the comparison of our study and the Icelandic study using cigarettes per day as the one, in common, phenotype and rs1051730 as the one, in common, genotype. The trend in the frequency of the rs1051730 T allele is generally similar, except for the 1–10 cigarettes per day group in the Icelandic study. Clearly, their population of smokers differs from the UT-WI-LHS smokers, which is understandable since our study preferentially enrolled long-term heavy smokers who had lung disease (UT-LHS) or sought cessation treatment (WI).

Two other GWA reports using lung cancer case-control designs, reported that rs1051730 exceeded genome-wide significance levels for association with lung cancer but having failed to measure significant associations with smoking behavior, concluded that the effect on lung cancer risk was independent of smoking behavior. Our analysis demonstrating that strong associations between CHRNA5-A3-B4 variants and nicotine dependence are seen only amongst smokers who began daily smoking relatively early in life, coupled with the detailed molecular definition of the CHRNA5-A3-B4 haplotype structures generated from resequencing, supports the hypothesis that the disease outcome effects of rs1051730, and other surrogate markers for Haplotype A, are mediated by nicotine addiction. In our opinion, the biological link between nicotinic receptor variants and smoking behavior is more plausible than a direct effect of these ion channels on lung cancer susceptibility. The association of rs1051730 with lung cancer may therefore be due to disease mortality related to long-term, persistent smoking caused by severe nicotine addiction [60,61].

In summary, we have demonstrated that major CHRNA5-A3-B4 haplotypes identify countervailing susceptibility (HA) and protective (HC) determinants for long-term nicotine dependence. A substantial shift in haplotype frequency (AA vs. CC = 10%) and diplotype frequency (AA vs. BC = 17%, AA vs. CC = 27%) is observed when age of exposure to nicotine is used to define an at-risk subgroup. Identifying this interaction of a common genetic risk factor with age of daily smoking onset among the complexity of factors that influence nicotine addiction indicates how genetics can augment public health approaches to the problem of smoking-related illness, because the risk is amenable to intervention. Identification of genetically high-risk individuals who would benefit from proactive interventions, such as adolescent education and cessation clinics, may result in a population with a lower rate of adult nicotine addiction.

Materials and Methods

Subjects
The UT cohort was made up of respondents to community advertising for persons who had smoked more than 100 cigarettes lifetime plus a subset of the originally recruited Utah LHS cohort who had follow-up phenotyping and biosample collection in Salt Lake City, UT; these Utah participants were excluded from the LHS cohort. Participants were not drawn from a psychiatric treatment population, and no medical or behavioral treatments were offered as part of the study. UT volunteers were not excluded simply because they had a lifetime diagnosis of psychosis or Bipolar Disorder, but they were excluded if their current mental status made it impossible for them to complete the questionnaires or interviews.

WI participants were recruited by media advertisements and had to be current smokers who were motivated to quit smoking, smoked more than 9 CPD, and produced a breath sample with carbon monoxide (CO) >9 ppm at baseline. Exclusion criteria included evidence of psychosis history based on the Prime-MD structured psychiatric interview [62], significant alcohol abuse based on the Michigan Alcoholism Screening Test [MAST] [63] and clinically significant depression symptoms based on the CES-D [64]. DNA extraction for the Wisconsin subjects was performed by the Wisconsin State Laboratory of Hygiene. Study procedures were approved by the institutional review boards at the University of Wisconsin and the University of Utah. All LHS participants had
COPD compared to 62% of the UT cohort. COPD was not assessed in the WI cohort, but given the younger age of those participants, it is likely that relatively few had developed COPD yet. All LHS and WI participants were smoking at the time they participated in the study, but 43% of the UT cohort had been tobacco abstinence for the previous two years. The LHS subjects were restricted to European Americans for whom DNA was available. De-identified demographic and phenotypic data was provided from the LHS limited access database by the NIH/NHLBI.

The cohorts differed in age, sex, age of onset of daily smoking, percent smoking daily by age 16, cigarettes smoked/day (CPD), and dichotomized Fagerstrom Test of Nicotine Dependence (FTND) score, but they did not differ in mean FTND (Table S1). FTND scores of 0–4 were designated low dependence and scores of 6–10 were designated high dependence. Omitting mid-range FTND scores eliminated 340 participants (19%) from further analyses. The number eliminated by cohort was: LHS, 392 (20%); UT, 73 (15%); and WI, 75 (19%). Thus, the final sample used for FTND association tests comprised 1513 LHS participants, 413 UT participants, and 323 WI participants (total N = 2287). Some participants (N = 692) in the LHS cohort were smoking at the time they participated in the study, likely that relatively few had developed COPD yet. All LHS and WI participants were assigned to the high dependence group, and those with scores 6–9 were assigned to the high dependence group, and those with scores of 4 or 5 (N = 392) were omitted.

Resequencing of nAChR Genes, SNP Selection, and Genotyping

DNA samples were obtained from all subjects. The resequencing sample comprised 144 smokers identified by a multiple predictor model with unit weighting as being either high or low on a broad range of dependence measures (e.g., relapse likelihood, withdrawal severity; N = 144) and an additional group of lifetime nonsmokers (N = 48). The University of California Santa Cruz genome browser was used to extract positions of proximal promoters and RefSeq exons for CHRNA2, CHRNA3, CHRNA4, CHRNA5, CHRNA6, CHRNA7 (unique 5’ only), CHRNA9, CHRNA10, CHRNA2, CHRNB3 and CHRNB4. Exons, genomic sequences conserved in mouse, and ~1 kb of 5’ and 3’ flanking regions were targeted by DNA sequencing, as previously described [63]. The resequencing scan surveyed a total range of 36,620 nts. in each of 192 individuals of Northern European descent; survey coordinates from NCBI Build 35 for genomic ranges surveyed with a Phrap quality score of >30 are: CHRNA2 [Chr. 8] 27374096::27375538, 27376211::27377643, 27380318::27381167, 27382561::27383658, 27384018::27384835, 27392219::27397230 nts.; CHRNA3-CHRNA3-CHRNB3 [Chr. 15] 7664811::7665440, 76651089::76658111, 76659985::76660542, 76665576::76666420, 76667214::76688294, 76668898::76670309, 76672128::76676199, 76680573::76681761, 76696115::76696969, 76697824::76698624, 76700082::76700676, 76703348::76704858, 76706170::76709554, 76710234::76710990, 76713688::76715398, 76720101::76721040 nts.; CHRNA4 [Chr. 20] 151353228::151353745, 151354953::151356224, 151356569::151357823, 151361153::151362332 nts.

Forward and reverse PCR primer sequences were chosen from the publicly available genomic sequence, and PCR amplification was carried out in 25 μl reaction volumes using standard techniques (primer sequences are available from the authors upon request). Amplification products and unincorporated nucleotides were removed using ExoSAP-IT (USB, Cleveland, Ohio). For sequencing, internal primers were used, and cycle sequencing was carried out using Applied Biosystems BigDye terminator chemistry. Cycling was done with an initial denaturation at 96°C for 30 sec.; followed by 45 cycles of 96°C for 10 sec., 50°C for 5 sec., 60°C for 4 min. Upon completion of sequencing, 20 μl of 62.5% EtOH/1M KOAc, pH 4.5 was added to each reaction and the sequence plates were centrifuged at 4,000 rpm at 4°C for 45 min. The samples were resuspended in 15 μl of formamide and electrophoresed on an ABI 3730 capillary instrument. Sequence trace files were evaluated using the Phred, Phrap and Consed programs, and potential variants were identified by using the PolyPhred program [66,67]. Single nucleotide polymorphisms (SNPs) were verified by manual evaluation of the individual forward and reverse sequence traces. In addition, all sequence assemblies were manually scanned for insertions, deletions and polymorphic positions undetected by PolyPhred. Tag SNPs were selected using the lSelect algorithm [68] with an r^2 cutoff value of 0.64 on all SNPs with a minor allele frequency >5%.

Genotyping was carried out using one of two different methods: an oligonucleotide ligation, PCR assay for simultaneous genotyping of 48 single nucleotide polymorphisms (SNPs) in a multiplexed, 384-well plate method, SNPlex (Applied Biosystems), and a single-step homogeneous SNP genotyping using a 5’-nuclease assay, TaqMan (Applied Biosystems). The SNPlex assay, pooling 48 different SNPs into a single genotyping assay, was applied to DNA samples processed in 384-well plate format through an automated protocol which included: SNP-specific oligonucleotide ligation, PCR amplification, immobilization, hybridization, elution, and separation and fluorescent detection on an ABI 3730xl capillary instrument. For all SNPs genotyped by SNPlex, the call rate on the 932 individuals in the UT and WI cohort was 99.26% after 35 analyses were used to test the association of SNPs with dichotomized FTND, as well as the Cochran-Armitage trend test and the Cochran-Armitage trend test using the computer program, PLINK [69]. Empirical significance levels of allelic tests were evaluated by phenotype-genotype permutation testing using the PLINK adaptive mode. The logistic regression analyses reported in Tables 3–4 and Figure 2 were computed using SYSTAT 10.2 (Richmond, CA: SYSTAT Software Inc.). Dummy coding was used for haplotype and sex, while ordinal coding was used for age of onset of daily smoking. Haplotype estimation and individual assignment were carried out on genotypic data using fastPHASE [70], and independently...
evaluated using the EM algorithm implemented in SNPHAP (http://www-gene.cimr.cam.ac.uk/clayton/software). Haplotype-based association analyses for omnibus and haplotype-specific tests were carried out using PLINK. To evaluate potential population stratification effects, the UT and WI cohorts were analyzed for population admixture using STRUCTURE [71] on 94 non-related loci and an assumed population of 2. No significant admixture was observed, supporting the European American self-report. Also, cohorts did not differ in CHRNA5-A3-B4 haplotype frequency ($\chi^2, 6, N = 5654 = 8.2, P = 0.22$, see Table 2), suggesting they were drawn from the same population.

Supporting information

Table S1  Demographics and smoking history by cohort.
Found at: doi:10.1371/journal.pgen.1000125.s001 (0.01 MB PDF)

Table S2  Single nucleotide polymorphism discovery by genomic resequencing in 9-like nAChR ($\alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_9,$ and $\alpha_{10}$) and B-like nAChR ($\beta_2, \beta_3,$ and $\beta_4$) subunit genes. Found at: doi:10.1371/journal.pgen.1000125.s002 (0.05 MB PDF)

Table S3  Association of nicotinic receptor variants using dichotomized FTND Scores within age of onset of daily smoking in the combined UT and WI cohorts. Found at: doi:10.1371/journal.pgen.1000125.s003 (0.23 MB DOC)

Table S4  Results of logistic regression analyses in which Haplotype C was the reference condition and two alternative criteria for low and high nicotine dependence were used (0–4 vs. 5–10 and 0–5 vs. 6–10). Found at: doi:10.1371/journal.pgen.1000125.s004 (0.01 MB PDF)

References