Published in final edited form as:

Addict Biol. 2010 July ; 15(3): 346–357. doi:10.1111/j.1369-1600.2010.00220.x.

Incorporating age at onset of smoking into genetic models for nicotine dependence: Evidence for interaction with multiple genes

Richard A. Grucza¹, Eric O. Johnson⁴, Robert F. Krueger^{1,2}, Naomi Breslau⁵, Nancy L. Saccone³, Li-Shiun Chen¹, Jaime Derringer², Arpana Agrawal¹, Micheal Lynskey¹, and Laura J. Bierut¹

¹ Department of Psychiatry, Washington University, Saint Louis, MO

² Department of Psychology, Washington University, Saint Louis, MO

³ Department of Genetics, Washington University, Saint Louis, MO

⁴ RTI International, Research Triangle Park, NC

⁵ Department of Epidemiology, Michigan State University, East Lansing, MI

Abstract

Nicotine dependence is moderately heritable, but identified genetic associations explain only modest portions of this heritability. We analyzed 3,369 SNPs from 349 candidate genes, and investigated whether incorporation of SNP-by-environment interaction into association analyses might bolster gene discovery efforts and prediction of nicotine dependence. Specifically, we incorporated the interaction between allele count and age-at-onset of regular smoking (AOS) into association analyses of nicotine dependence. Subjects were from the Collaborative Genetic Study of Nicotine Dependence, and included 797 cases ascertained for Fagerström nicotine dependence, and 811 non-nicotine dependent smokers as controls, all of European descent. Compared with main-effect models, SNP x AOS interaction models resulted in higher numbers of nominally significant tests, increased predictive utility at individual SNPs, and higher predictive utility in a multi-locus model. Some SNPs previously documented in main-effect analyses exhibited improved fits in the joint-analysis, including rs16969968 from CHRNA5 and rs2314379 from MAP3K4. CHRNA5 exhibited larger effects in later-onset smokers, in contrast with a previous report that suggested the opposite interaction (Weiss et al, PLOS Genetics, 4: e1000125, 2008). However, a number of SNPs that did not emerge in main-effect analyses were among the strongest findings in the interaction analyses. These include SNPs located in GRIN2B ($p=1.5 \times 10^{-5}$), which encodes a subunit of the NMDA receptor channel, a key molecule in mediating age-dependent synaptic plasticity. Incorporation of logically chosen interaction parameters, such as AOS, into genetic models of substance-use disorders may increase the degree of explained phenotypic variation, and constitutes a promising avenue for gene-discovery.

Keywords

addiction; SNP; age-at-onset; interaction; environment; nicotine dependence

INTRODUCTION

Progress in understanding the genetics of nicotine dependence has advanced rapidly in recent years, with numerous replications of the association between non-synonymous polymorphism on the nicotinic receptor gene CHRNA5 (rs16969968 and highly correlated

proxy SNPs) and nicotine dependence (Amos *et al.* 2008; Berrettini *et al.* 2008; Bierut *et al.* 2008; Grucza *et al.* 2008; Thorgeirsson *et al.* 2008; Weiss *et al.* 2008; Caporaso *et al.* 2009; Chen *et al.* 2009; Saccone *et al.* 2009; Wang *et al.* 2009). Additional promising findings have been identified through both genome-wide and large-scale candidate gene approaches (Bierut *et al.* 2007; Saccone *et al.* 2007; Caporaso *et al.* 2009). However, for both nicotine dependence in particular, and complex diseases in general, a paradox has emerged that discoveries to date explain only small portions of the variance for such phenotypes (Orr and Chanock 2008; Goldstein 2009; Hirschhorn 2009; Kraft and Hunter 2009). Factors invoked to explain the “missing heritability” include polygenic effects (very large numbers of polymorphisms having very small effects), undiscovered rare variants, copy-number variants that have yet to be systematically analyzed, and interactions among genes, or between genes and environmental factors (Maher 2008; Phillips 2008; Goldstein 2009). Investigation of these potential sources of undiscovered phenotypic variation can lead to novel genetic discoveries, provide new insights into biological pathways involved in disease, and potentially provide clues to environmental interventions capable of mitigating genetic risk.

One of the key risk factors for addiction in general, as well as for nicotine dependence in particular, is early onset of substance use, as assessed by retrospectively reported age at first use or age at onset of regular use (Robins and Przybeck 1985; Breslau *et al.* 1993; Breslau and Peterson 1996; Grant and Dawson 1997; Grant *et al.* 2006; Grucza *et al.* 2008; Agrawal *et al.* 2009). Recently, advances in developmental neurobiology have pointed to adolescence as a potential critical period of vulnerability for the development of substance use disorders (Chambers *et al.* 2003; Crews *et al.* 2007). The development of adolescent animal models for nicotine dependence has enabled researchers to show that both neurobiological changes and long-term behavioral sequelae differ for adolescent as compared to adult nicotine exposure (Trauth *et al.* 2000; Trauth *et al.* 2001; Slotkin 2002; Adriani and Laviola 2004; Slotkin *et al.* 2008; Kota *et al.* 2009; Slotkin and Seidler 2009). Hence, the molecular mechanisms and neuroadaptive processes involved in adolescent exposure and adult exposure may differ considerably. The ramifications of these animal models for human genetic analysis are clear: If the etiology for the development of nicotine dependence differs by developmental stage, then the age at which smoking is initiated could be a critical moderator of SNP-associations for nicotine dependence.

In this paper, we examine genotype-phenotype associations with nicotine dependence, and investigate whether modeling statistical interactions between SNP and age-at-onset of regular smoking (AOS) can increase the amount of variation explained by genetic polymorphisms. Specifically, we posit that early and late onset smokers are likely to differ in terms of etiology for development of nicotine dependence, and that this will result in differences in the relative influence of polymorphisms as a function of AOS. We take a large-scale candidate gene approach (i.e., 3,369 SNPs from 348 genes) using data from the Collaborative Study for the Genetics of Nicotine Dependence (COGEND), a case-control study of the genetics of nicotine dependence among smokers. Because we are interested in the full explanatory power of the model (as opposed to the interaction term alone), we utilize a two degree-of-freedom test that jointly evaluates both SNP main effects and potential SNP x AOS interactions (Kraft *et al.* 2007).

Gene-environment interactions, including interactions with age-at-onset of smoking, have been the focus of several recent candidate gene analyses (Weiss *et al.* 2008; Chen *et al.* 2009; Schmid *et al.* 2009). However, we are unaware of multi-gene analyses of gene-environment interactions for smoking; i.e., analyses in which many polymorphisms are tested in interactive models with putative environmental moderators. Such exploratory approaches may prove to be fruitful for both enhancing our understanding of the etiology of nicotine dependence, as well as for gene discovery and validation (Kraft *et al.* 2007).

METHODS

Sample Ascertainment and Description

The Collaborative Genetic Study of Nicotine Dependence (COGEND), recruited participants from the St. Louis, Missouri and Detroit, Michigan, metropolitan areas. Both cases and controls were required to have smoked at least 100 cigarettes in their lifetime, a frequently used threshold to define being a smoker (e.g. CDC 2005). In addition to this criterion, subjects were ascertained based on their Fagerström Test for Nicotine Dependence (FTND) scores for their period of heaviest smoking (Heatherton *et al.* 1991). Cases were required to score 4 or higher, while controls were required to report never having symptoms of dependence (lifetime FTND = 0). Cases and controls were identified through a screening process in which telephone interviews on 54,644 individuals across the two sites were conducted (screening response rate= 74%). Among screened individuals, subjects were ineligible if they did not meet the requirements listed above, or were not within the targeted age range for the study (25–44 years). Among the 5,010 individuals meeting eligibility requirements, 3,137 were interviewed (response rate = 63%), and 2,949 donated a blood sample for genetic testing (response rate = 94%). The initial wave of genotyping was limited to European-Americans ascertained prior to summer 2005, who comprise the 789 cases and 811 controls analyzed here. Procedures were approved by Institutional Review Boards of participating institutions, and were carried out in accord with the Helsinki Declaration.

Measures

All subjects were personally interviewed using the Semi-Structured Assessment for Nicotine Dependence (SSAND) which was developed specifically for COGEND and is modeled after the widely use Semi-structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz *et al.* 1994; Hesselbrock *et al.* 1999). The tobacco section is poly-diagnostic, and includes the assessment of nicotine dependence using both the FTND, and DSM-IV criteria. Alcohol and drug use, abuse, and dependence, and other psychiatric disorder sections are identical to those in the SSAGA.

In addition to case-control status, variables used in these analyses included age at first full cigarette (AFC), age at onset of regular smoking (AOS), age at first full drink (AFD) and age at onset of regular drinking (AOD). All of these variables were retrospectively assessed as single items. Regular smoking was defined as having smoked at least once a week for a period of at least two months. Regular drinking was defined as drinking at least once a month for a period of 6 months or more. Individuals who had never initiated regular drinking were excluded from the AOD analysis. To reduce the influence of extreme values on age at onset variables in regression estimates, AFC, AOS, AFD and AOD were subjected to a “Winsorization” procedure (Tukey 1960). Winsorization involves selecting an equal number of observations at both tails of a distribution, and reassigning their values to be identical to the next-highest or next-lowest value in the distribution. For example, subjects reporting AOS <8 (n=7) were assigned an AOS value of 8, those with AOS over 30 (n=13) were assigned a value of 30, and so on. This transformation is appropriate for variables that have approximately normal distribution, but exhibit kurtosis due to small numbers of outliers (Fernandez *et al.* 2002; Shete *et al.* 2004). We selected Winsorization thresholds that reduced positive kurtosis without inducing negative kurtosis. This corresponded to highest and lowest 1–2% of the sample on each variable. Finally, all of these covariates were recoded so that the zero-value would reflect the sample median on the raw variable (AOS=16), and higher values would reflect risk, or earlier onset, rather than protection. We designate these recoded variables using a prefixed “r”. Hence, rAOS=16-AOS, after Winsorization.

Gene and SNP Selection

The analyses presented here utilize the candidate gene panel presented in Saccone et al (Saccone *et al.* 2007), in addition to a number of “fine mapping” SNPs from follow-up genotyping of the nicotinic acetylcholinergic receptor (CHRN) gene family. The initial candidate gene panel consisted of SNPs from genes selected based on known neurobiological and metabolic pathways, previous reports on the genetics of nicotine dependence, and previous reports on the genetics of other psychiatric disorders (see Saccone *et al.* 2007) for more details on gene and SNP selection). An additional 107 SNPs from 16 CHRN genes were genotyped in a separate experiment to follow-up and fine-map initial discoveries in this gene family (Bierut *et al.* 2007; Saccone *et al.* 2007; Saccone *et al.* 2009).

Genotyping and Quality Control

For the initial candidate gene study, SNPs were chosen to cover physical regions as uniformly as possible, while denser coverage was used for SNPs within exons and near candidate gene promoter regions. Quality control measures are described elsewhere (Saccone *et al.* 2007; Saccone *et al.* 2009) and include reliability of genotype-calls, Hardy-Weinberg equilibrium > 0.01 , and 95% genotyping call-rate. Because of the additional statistical power-requirements for gene-environment analyses, we eliminated SNPs from analysis if they exhibited a minor allele frequency (MAF) of 4% or less. This resulted in a total of 3369 SNPs tested in this report. These SNPs cover a total of 348 genes.

Association Analyses

Our overall purpose is not solely to detect interactions, but rather to understand the genotype-phenotype association in relation to onset age of smoking. As such, it is the joint-significance of the main effect and the interaction terms that is of interest. Hence, for each SNP, we allow for the possibility of statistical interaction between age-at-onset of smoking (and other covariates) and allele count. This involves evaluation of the non-genetic base model, the main-effect model and the interaction model:

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \beta_1(\text{age}) + \beta_2(\text{sex}) + \beta_3(\text{rAOS}) \quad (\text{Eq. 1})$$

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \beta_1(\text{age}) + \beta_2(\text{sex}) + \beta_3(\text{rAOS}) + \beta_4 G \quad (\text{Eq. 2})$$

$$\ln\left(\frac{P}{1-P}\right) = \alpha + \beta_1(\text{age}) + \beta_2(\text{sex}) + \beta_3(\text{rAOS}) + \beta_4^* G + \beta_5(G \times \text{rAOS}) \quad (\text{Eq. 3})$$

where α is the intercept, age is age at interview, sex is coded as 0 or 1, rAOS is age-at-onset of regular smoking (reversed and recoded as described), G is genotype for a given SNP, presumed to be log-additive. The final term in Equation 3 represents the product or interaction term. Note that in Equation 3, the term β_4^* designates the genetic main effect when rAOS=0, which corresponds to AOS=16, the sample median.

Because we are interested in the full explanatory power of the interaction model (Equation 3) as compared to the main-effect model (Equation 2), we compare each model to the non-genetic base mode (Equation 1). When the main effect model is compared to the base model using a likelihood ratio Chi-square test with one degree of freedom, the significance of the

overall main effect (β_4) is evaluated. To evaluate the joint significance of the main effect and interaction, the interaction model (Equation 3) is compared with the base model (Equation 1) using a two-degree of freedom likelihood ratio Chi-square test. This approach has been formally explicated by Kraft and colleagues, who showed it to be comparable in power to other approaches for detecting G x E, with little loss of power in detecting main effect association where G x E is non-significant (Kraft *et al.* 2007). To facilitate evaluation of the goodness-of-fit for both main effect and interaction models, we present Akaike's information criterion (AIC) for all genetic analyses. AIC is a log-likelihood based measure of model fit that penalizes for model complexity (Akaike 1974). Lower AIC values are indicative of better-fitting models; here we present the reduction in AIC relative to the non-genetic base model (Equation 1).

In addition to rAOS, we incorporated several other variables to examine the specificity of smoking onset age and to examine whether changes in model fits might be due to modeling artifacts. These included permuted rAOS, to simulate the null hypothesis with respect to the interaction – that is, individual rAOS values were randomly re-assigned to other individuals in the data set. This is labeled “pAOS.” Other moderators, that are correlated with rAOS but which were not hypothesized *a priori* to have moderating effects were age at first cigarette (rAFC), age at first drink (rAFD) and age at onset of drinking (rAOD). Strong results for these alternative moderators might indicate that a more general behavioral factor is involved in the interaction, as opposed to tobacco use.

Comparison of Models using p-value Thresholds and False Discovery Rates

As a simple metric of whether or not the incorporation of SNP x rAOS or other interaction terms into the genetic models would result in improved model fits, we tallied the number of SNPs that met traditional significance thresholds (<0.05 , <0.01 and <0.001) in the main effect and interaction models. The p-values were those derived from the main effect model (equation 2) and the joint p-value for the main effect and interaction terms from the interaction model (Equation 3). Overall false-discovery rates (π_0 parameters) were also computed for each set of analyses using the program QVALUE (Storey 2002). The π_0 parameter is an estimate of the total proportion of false-positives in the complete distribution of p-values if the null hypothesis is rejected for all tests.

Concordance Rates

The logistic regression concordance rate is a measure of concordance between predicted and observed risk between all possible case-control pairs. Pairs are considered to be concordant if the predicted risk for the case is higher than the predicted risk for the control, and discordant if the opposite is true. We employed “leave-one-out” cross-validation as implemented in SAS Proc Logistic (SASInstituteInc 2002–03) to compute predicted risk scores and concordance rates. In this procedure, the risk for each observation was predicted from a model parameterized on all other observations. This approach reduces inflated concordance rates that result from over-fitting.

Multi-Locus Model Building

In order to evaluate whether or not modeling of rAOS interactions would improve the predictive utility of multiple SNPs, we constructed multi-locus genetic models to predict case-control status from either : 1.) multiple SNPs; or 2.) multiple SNPs paired with their respective SNP x rAOS interaction terms. For the former model, the top 30 SNPs were entered into the model, and a backward elimination procedure was applied until all remaining SNPs in the model were independently associated with $p < 0.15$ (a commonly used threshold for multiple regression model building). A similar procedure was done for the model that included SNP x rAOS interaction except in this case, interactions were only

entered into the initial model if they were significant in single-locus analyses, and SNPs were eliminated until all remaining SNPs exhibited either a main effect or interaction with $p < 0.15$.

RESULTS

Description of sample

The analysis sample contains 797 cases and 811 controls of European descent. The average age at interview for nicotine dependent cases was 35.7 years ($SD=5.5$), the average age for controls was 36.9 years ($SD=5.3$); this was a small, but significant difference ($p < 0.001$). There were significantly more women among controls than cases (70% vs. 56%, $p < 0.001$). As such, age and sex were included as covariates in all analyses. To test for the possibility of population sub-structure not accounted for by self-reported race/ethnicity, STRUCTURE analyses (Pritchard *et al.* 2000) were conducted; no evidence of confounding substructure was detected (Bierut *et al.* 2007). This was subsequently confirmed using principal components analysis in Eigenstrat (Li and Yu 2008). Mean AOS was 15.4 years for cases ($SD=3.6$) and 18.0 years for controls ($SD=3.6$); a highly significant difference (ANOVA $t=13.55$, $p < 0.001$).

Number of Nominally Significant SNPs for Various Models

For all 3369 SNPs, we used logistic regression to predict case-status using the main effect model (Equation 2) and rAOS interaction model (Equation 3), and tallied the number of nominally significant loci using three traditional thresholds ($p < 0.05$, $p < 0.01$ and $p < 0.001$) – results of these counts are summarized in Table 1. This count is intended as a simple metric of the relative utility of each model, and does not reflect adjustment for multiple testing or linkage disequilibrium. In interaction models, the p-values reflect joint effects of SNP and SNP x rAOS interactions (2 degree of freedom test), whereas for the main-effect models, tallies reflect 1-degree of freedom tests for the effect of SNP only. For each threshold, substantially more loci met the p-value criterion under the interaction model than under the main-effect model, with 40–60% more “hits” in each case (e.g., 448 SNPs with $p < 0.05$ in the interaction model, vs. 268 in the main-effect model). The increased number of low p-values may stem in part from overfitting, resulting from the additional degree of freedom to the model. However, the addition of alternative interaction terms to the model (Equation 3), including randomly reassigned rAOS (permuted AOS, designated pAOS; i.e., SNP x pAOS), age-at-first cigarette (SNP x rAFC), age-at-onset of drinking (SNP x rAOD) and age-at-first drink (SNP x rAFD) yielded fewer nominally significant p-values than the SNP x rAOS model and, in many cases, the main-effect model. (The age-at-onset variables rAFC, rAOD, and rAFC correlate with rAOS with $R=0.67$, $R=0.26$ and $R=0.18$, respectively). False-discovery rate analysis π_0 parameters, derived from each of the complete p-value distributions, also suggest that incorporation of the SNP x rAOS interaction yields more overall findings than the main effect or alternative interaction models (Table 1).

Strongest Associations in Main Effect and Interaction Models

Table 2 summarizes the results from the 30 lowest p-values from the SNP main-effect model (Equation 2). These findings correspond to p-values of 0.0021 and lower. An R^2 -binning procedure (Carlson *et al.* 2004) was used to identify bins of highly correlated SNPs (all R^2 of 0.80 or higher), and only the top results from each bin are listed. Results are similar to those presented in Saccone et al (Saccone *et al.* 2007), but some discrepancies were expected because of differences in model specification, analysis sample composition, and SNP inclusion criteria. They are presented here for comparison with the interaction model results, and will not be discussed in detail except in that context.

Results from the SNP x rAOS interaction model are summarized in Table 3; the top 30 findings correspond to p-values less than or equal to 0.001. These p-values correspond to the 2-degree of freedom for the joint-significance of main effect and interaction terms. Genes and chromosomal positions are also listed for these top findings, and p-values from the main-effect model (Equation 2) for these loci are listed. A number of the genes enumerated in Table 3 correspond to findings that were identified in the initial candidate gene analysis of these data (Saccone *et al.* 2007); these include the nicotinic receptor gene cluster on chromosome 15 (CHRNA5-A3-B4) and its flanking regions (e.g., IREB2), DBH on chromosome 9, and KCNJ6 on chromosome 21. However, a number of new findings emerge. The most prominent new finding involves GRIN2B, for which two SNPs with rather low inter-correlation ($R^2=0.06$) yielded p-values less than 10^{-4} in the interaction model. Other examples of genes containing SNPs that were highly significant in the interaction model, even though they did not reach $p<0.05$ in the main effect model are DBI, ADCY8, VAPA and GABRR1. Two genes, DBH and KCNB1 were among the top findings in the interaction model, despite the fact that their p-values in the main-effect model were actually lower, suggesting that these SNPs exhibit robust main effects across all values of rAOS. Finally, several SNPs exhibited relatively low p-values in the main effect model but yielded improved fits under the interaction model, suggesting statistically significant (or near-significant) interactions; these included SNPs from CHRNA5, MAP3K4, KCNJ6.

As a measure of the overall predictive power, the logistic regression concordance rate (C) was computed for each of the top findings in Tables 2 and 3 – the reported ΔC parameter is the difference in concordance rates between the genetic model (equations 2 or 3 for Tables 2 and 3, respectively) and the base model (equation 1). A “leave-one-out” cross-validation approach was employed to reduce overfitting (i.e., risk for each observation is predicted from model parameterized on all other observations). For the main effect models, ΔC values ranged from 0.18 to 0.50 with a median of 0.28%, whereas for the interaction model ΔC values ranged from 0.18 to 0.72, with a median value of 0.40%.

Table 4 provides parameter estimates for the top findings derived from the interaction model. In addition to presenting parameter estimates from regression models that treat AOS as a continuous variable, we also present results of post-hoc analyses in which the sample was divided based on the median value for AOS (age 16). This latter set of analyses was largely conducted for illustrative purposes; i.e., to facilitate examination of differences in odds ratios between earlier and later onset smokers. The sample split with AOS of 16 or lower comprised 562 cases and 280 controls while the split with AOS of 17 or higher comprised 219 cases and 500 controls. In addition to being the sample median, this split-value corresponds to the dichotimization threshold utilized in a previous analysis of SNP x AOS interactions (Weiss *et al.* 2008). Main effect odds ratios were computed separately for the two sample halves. This approach, unlike the primary analyses, makes no assumptions about the scale of the putative interaction variable (i.e., does not assume that genetic effect sizes scale multiplicatively with AOS). Parameter estimates indicate that some interactions are synergistic; i.e., the main effect estimate at the sample median for environmental risk (onset of smoking at age 16) is in the same direction as the interaction odds ratio. In other words, SNP effects are stronger in earlier onset smokers. This was the case for GRIN2B (rs890), MAP3K4 and AGPAT4 (which is in LD with MAP3K4). In the cases of CHRNA5, KCNJ6 and IREB2 (in the flanking region of CHRNA5), the main effect and interaction odds ratios were in opposite directions (SNP effects stronger in later-onset smokers). Other cases of significant interactions were ambiguous, i.e., the main effect odds ratio was not significant at the median value for AOS; including GRIN2B (rs17760877), DBI, ADCY8, VAPA and GABRR1.

Multi-Locus Models

The top SNP findings summarized in Tables 2 and 3 were used in a multi-locus model to predict nicotine dependence. The initial model included main effects for all 30 top SNPs. A backwards elimination procedure was applied to eliminate all SNPs that were not significant in the multiple regression at $p < 0.15$. This was done to eliminate loci that are moderately correlated with nearby loci due to linkage disequilibrium. Thirteen loci remained in the model after the elimination procedure. For the interaction model, both main effect and interaction terms from the top 30 SNPs were initially entered, except in cases where interaction terms were not significant in the single-locus model (DBH and KCNB1). After eliminating redundant SNPs through the backwards elimination procedure, 13 SNPs remained. Predicted logistic regression risk scores and concordance rates, employing the leave-one-out cross-validation procedure were computed for each of these multi-locus models, compared with the non-genetic base model (Equation 1). Results of these parameters for both models, as well as covariates-only (non-genetic) model are presented in Table 5. Incorporation of the top genetic main effects into the model containing age, sex and the rAOS improves cross-validated concordance from 73.7% to 76.6%, and changes the predicted risk difference between cases and controls by 0.06 standard-deviations, relative to the non-genetic model (Table 5, second column, difference between rows 3 and 2). However, incorporation of top findings from the interaction-model, improves the cross-validated concordance value to 78.1%; and changes the predicted risk score difference between cases and controls by 0.09 SD, relative to the non-genetic base model. Hence, the incorporation of rAOS interactions into the genetic analyses results in a 50% relative increase in the predicted risk due genetic variables.

DISCUSSION

In this work, we demonstrate that incorporation of age-at-onset of smoking into a model that allows for moderation of genetic effects in predicting nicotine dependence results in improved SNP discovery using a variety of metrics, compared with a model that did not allow for such moderation. This effect seems to be specific to age-at-onset of regular smoking, which was selected based on the results of experimental studies in animal models (Trauth *et al.* 2000; Trauth *et al.* 2001; Rodd-Henricks *et al.* 2002; Rodd-Henricks *et al.* 2002; Slotkin 2002; Adriani *et al.* 2004; Slotkin *et al.* 2008; Slotkin *et al.* 2009); other age-at-onset variables did not exhibit similar effects, despite being correlated with rAOS. Finally, in a multi-locus prediction, the interaction model exhibited greater predictive power than did the main-effect model with a comparable number of loci.

Like other complex diseases with moderate heritability, substance use disorders are likely to be influenced by numerous common polymorphisms, each with only a small to moderate influence on the overall liability to the disorder (Orr *et al.* 2008). However, it is clear that there are also numerous environmental contributions to liability. Distinct genetic pathways may influence biological responses to specific environmental risk factors. Put simply, any given polymorphism may be important in some risk environments but not in others. It logically follows that specification of key environmental variables may aid in identifying genes that are otherwise masked under heterogeneity of the risk environment.

While concerns about multiple testing apply in this exploratory work, there are several discoveries that warrant further discussion. The first and third top findings in this paper involved two weakly correlated SNPs from the GRIN2B gene; these were not detected under the main-effect model in the initial candidate gene study. This gene encodes subunits of the N-methyl-D-aspartate (NMDA) receptor channel, which appears to be involved in mediating neural-plasticity leading to drug dependence (Kelley 2004; Hyman *et al.* 2006). Notably, the expression and subunit composition of NMDA receptors change markedly over the course

of development, and NMDA-mediated neuroplasticity may be heightened in adolescence (Carpenter-Hyland and Chandler 2007; Lau and Zukin 2007). GRIN2B SNPs have been implicated in correlated phenotypes such as ADHD and alcohol-related traits (Wernicke *et al.* 2003; Dorval *et al.* 2007) and one very recent report provides strong evidence for an association between SNPs in this gene and smoking initiation (Vink *et al.* 2009). The results here suggest that GRIN2B polymorphisms may be important, age-dependent factors in the development of nicotine dependence. Based on both compelling biological evidence and association studies, the SNPs identified here and correlated polymorphisms warrant more detailed studies in this and other samples.

Another noteworthy finding is that the effect of the non-synonymous coding SNP rs16969968 in CHRNA5 appears to be more important in later onset smokers. This contrasts with the findings of Weiss and colleagues, who suggested that haplotypes involving this polymorphism were more important among early onset smokers (Weiss *et al.* 2008). There are a number of differences in phenotype definitions, sampling approaches, and comorbidity between our sample and that of Weiss and colleagues; however there is no obvious reason for the clear discrepancy between findings. It is noteworthy that a significant interaction appeared in both samples, but further examination in larger samples and across populations will be required to discern the functional significance of this interaction, if any. The confidence intervals for this interaction parameter are close to 1.0 (0.89, 0.99), whereas interaction parameter estimates for some other SNPs were more robust (e.g., GRIN2B and ADCY8). Differences in sampling, secular trends in environmental contributions to nicotine dependence, as well as random fluctuation could all contribute to discrepancies in estimates of effect sizes.

In summary, the incorporation of AOS into genetic models of nicotine dependence seems to be a promising avenue to discover novel SNPs that may be important in the context of different environments. The age at which a person begins smoking is a complex phenomenon that is correlated with a number of behaviors, and so it would be naive to treat AOS as a simple environmental exposure in interpreting these findings. Nonetheless, it seems reasonable to pursue some of the present findings in replication attempts, and to extend this approach to genome-wide association studies.

Acknowledgments

This work was supported by NIH R21 DA026612 (RAG), R01DA25888, K02DA021237, P01 CA089392 (LJB), and DA02691 (NLS).

References

- Adriani W, Laviola G. Windows of vulnerability to psychopathology and therapeutic strategy in the adolescent rodent model. *Behav Pharmacol.* 2004; 15(5–6):341–52. [PubMed: 15343057]
- Agrawal A, Sartor CE, Lynskey MT, Grant JD, Pergadia ML, Grucza R, Bucholz KK, Nelson EC, Madden PA, Martin NG, Heath AC. Evidence for an Interaction Between Age at First Drink and Genetic Influences on DSM-IV Alcohol Dependence Symptoms. *Alcohol Clin Exp Res.* 2009; 33(12):2047–56. [PubMed: 19764935]
- Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control.* 1974; 19(6):716–723.
- Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet.* 2008; 40(5):616–22. [PubMed: 18385676]

- Berrettini W, Yuan X, Tozzi F, Song K, Francks C, Chilcoat H, Waterworth D, Muglia P, Mooser V. Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry*. 2008; 13(4):368–73. [PubMed: 18227835]
- Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, Swan GE, Rutter J, Bertelsen S, Fox L, Fugman D, Goate AM, Hinrichs AL, Konvicka K, Martin NG, Montgomery GW, Saccone NL, Saccone SF, Wang JC, Chase GA, Rice JP, Ballinger DG. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet*. 2007; 16(1):24–35. [PubMed: 17158188]
- Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Grucza RA, Xuei X, Saccone NL, Saccone SF, Bertelsen S, Fox L, Horton WJ, Breslau N, Budde J, Cloninger CR, Dick DM, Foroud T, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Kuperman S, Madden PA, Mayo K, Nurnberger J Jr, Pomerleau O, Porjesz B, Reyes O, Schuckit M, Swan G, Tischfield JA, Edenberg HJ, Rice JP, Goate AM. Variants in nicotinic receptors and risk for nicotine dependence. *Am J Psychiatry*. 2008; 165(9):1163–71. [PubMed: 18519524]
- Breslau N, Fenn N, Peterson EL. Early smoking initiation and nicotine dependence in a cohort of young adults. *Drug Alcohol Depend*. 1993; 33(2):129–37. [PubMed: 8261877]
- Breslau N, Peterson EL. Smoking cessation in young adults: age at initiation of cigarette smoking and other suspected influences. *Am J Public Health*. 1996; 86(2):214–20. [PubMed: 8633738]
- Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol*. 1994; 55(2):149–58. [PubMed: 8189735]
- Caporaso N, Gu F, Chatterjee N, Sheng-Chih J, Yu K, Yeager M, Chen C, Jacobs K, Wheeler W, Landi MT, Ziegler RG, Hunter DJ, Chanock S, Hankinson S, Kraft P, Bergen AW. Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS ONE*. 2009; 4(2):e4653. [PubMed: 19247474]
- Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet*. 2004; 74(1):106–20. [PubMed: 14681826]
- Carpenter-Hyland EP, Chandler LJ. Adaptive plasticity of NMDA receptors and dendritic spines: implications for enhanced vulnerability of the adolescent brain to alcohol addiction. *Pharmacol Biochem Behav*. 2007; 86(2):200–8. [PubMed: 17291572]
- CDC. Cigarette Smoking Among Adults --- United States, 2003. *MMWR Weekly*. 2005; 54(20):509–513.
- Chambers RA, Taylor JR, Potenza MN. Developmental neurocircuitry of motivation in adolescence: a critical period of addiction vulnerability. *Am J Psychiatry*. 2003; 160(6):1041–52. [PubMed: 12777258]
- Chen L, Johnson EO, Breslau N, Hatsukami D, Saccone NL, Grucza RA, Wang JC, Hinrichs AL, Fox L, Goate AM, Rice JP, Bierut LJ. Interplay of genetic risk factors and parent monitoring in risk for nicotine dependence. *Addiction*. 2009; 104:1731–1740. [PubMed: 20871796]
- Chen X, Chen J, Williamson VS, An SS, Hettema JM, Aggen SH, Neale MC, Kendler KS. Variants in nicotinic acetylcholine receptors alpha5 and alpha3 increase risks to nicotine dependence. *Am J Med Genet B Neuropsychiatr Genet*. 2009; 150B(7):926–33. [PubMed: 19132693]
- Crews F, He J, Hodge C. Adolescent cortical development: a critical period of vulnerability for addiction. *Pharmacol Biochem Behav*. 2007; 86(2):189–99. [PubMed: 17222895]
- Dorval KM, Wigg KG, Crosbie J, Tannock R, Kennedy JL, Ickowicz A, Pathare T, Malone M, Schachar R, Barr CL. Association of the glutamate receptor subunit gene GRIN2B with attention-deficit/hyperactivity disorder. *Genes Brain Behav*. 2007; 6(5):444–52. [PubMed: 17010153]
- Fernandez JR, Etzel C, Beasley TM, Shete S, Amos CI, Allison DB. Improving the power of sib pair quantitative trait loci detection by phenotype winsorization. *Hum Hered*. 2002; 53(2):59–67. [PubMed: 12037405]
- Goldstein DB. Common genetic variation and human traits. *N Engl J Med*. 2009; 360(17):1696–8. [PubMed: 19369660]

- Grant BF, Dawson DA. Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: results from the National Longitudinal Alcohol Epidemiologic Survey. *J Subst Abuse*. 1997; 9:103–10. [PubMed: 9494942]
- Grant JD, Scherrer JF, Lynskey MT, Lyons MJ, Eisen SA, Tsuang MT, True WR, Bucholz KK. Adolescent alcohol use is a risk factor for adult alcohol and drug dependence: evidence from a twin design. *Psychol Med*. 2006; 36(1):109–18. [PubMed: 16194286]
- Grucza RA, Norberg K, Bucholz KK, Bierut LJ. Correspondence between secular changes in alcohol dependence and age of drinking onset among women in the United States. *Alcohol Clin Exp Res*. 2008; 32(8):1493–501. [PubMed: 18564104]
- Grucza RA, Wang JC, Stitzel JA, Hinrichs AL, Saccone SF, Saccone NL, Bucholz KK, Cloninger CR, Neuman RJ, Budde JP, Fox L, Bertelsen S, Kramer J, Hesselbrock V, Tischfield J, Nurnberger II Jr, Almasy L, Porjesz B, Kuperman S, Schuckit MA, Edenberg HJ, Rice JP, Goate AM, Bierut LJ. A risk allele for nicotine dependence in CHRNA5 is a protective allele for cocaine dependence. *Biol Psychiatry*. 2008; 64(11):922–9. [PubMed: 18519132]
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict*. 1991; 86(9):1119–27. [PubMed: 1932883]
- Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA--a comparison with the SCAN. *Addiction*. 1999; 94(9):1361–70. [PubMed: 10615721]
- Hirschhorn JN. Genomewide association studies--illuminating biologic pathways. *N Engl J Med*. 2009; 360(17):1699–701. [PubMed: 19369661]
- Hyman SE, Malenka RC, Nestler EJ. Neural mechanisms of addiction: the role of reward-related learning and memory. *Annu Rev Neurosci*. 2006; 29:565–98. [PubMed: 16776597]
- Kelley AE. Memory and addiction: shared neural circuitry and molecular mechanisms. *Neuron*. 2004; 44(1):161–79. [PubMed: 15450168]
- Kota D, Robinson SE, Imad Damaj M. Enhanced nicotine reward in adulthood after exposure to nicotine during early adolescence in mice. *Biochem Pharmacol*. 2009; 78(7):873–9. [PubMed: 19576867]
- Kraft P, Hunter DJ. Genetic risk prediction--are we there yet? *N Engl J Med*. 2009; 360(17):1701–3. [PubMed: 19369656]
- Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered*. 2007; 63(2):111–9. [PubMed: 17283440]
- Lau CG, Zukin RS. NMDA receptor trafficking in synaptic plasticity and neuropsychiatric disorders. *Nat Rev Neurosci*. 2007; 8(6):413–26. [PubMed: 17514195]
- Li Q, Yu K. Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. *Genet Epidemiol*. 2008; 32(3):215–26. [PubMed: 18161052]
- Maher B. Personal genomes: The case of the missing heritability. *Nature*. 2008; 456(7218):18–21. [PubMed: 18987709]
- Orr N, Chanock S. Common genetic variation and human disease. *Adv Genet*. 2008; 62:1–32. [PubMed: 19010252]
- Phillips PC. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*. 2008; 9(11):855–67. [PubMed: 18852697]
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155(2):945–59. [PubMed: 10835412]
- Robins LN, Przybeck TR. Age of onset of drug use as a factor in drug and other disorders. *NIDA Res Monogr*. 1985; 56:178–92. [PubMed: 3929100]
- Rodd-Henricks ZA, Bell RL, Kuc KA, Murphy JM, McBride WJ, Lumeng L, Li TK. Effects of ethanol exposure on subsequent acquisition and extinction of ethanol self-administration and expression of alcohol-seeking behavior in adult alcohol-preferring (P) rats: I. Periadolescent exposure. *Alcohol Clin Exp Res*. 2002; 26(11):1632–41. [PubMed: 12436051]
- Rodd-Henricks ZA, Bell RL, Kuc KA, Murphy JM, McBride WJ, Lumeng L, Li TK. Effects of ethanol exposure on subsequent acquisition and extinction of ethanol self-administration and

- expression of alcohol-seeking behavior in adult alcohol-preferring (P) rats: II. Adult exposure. *Alcohol Clin Exp Res*. 2002; 26(11):1642–52. [PubMed: 12436052]
- Saccone NL, Saccone SF, Hinrichs AL, Stitzel JA, Duan W, Pergadia ML, Agrawal A, Breslau N, Grucza RA, Hatsukami D, Johnson EO, Madden PA, Swan GE, Wang JC, Goate AM, Rice JP, Bierut LJ. Multiple distinct risk loci for nicotine dependence identified by dense coverage of the complete family of nicotinic receptor subunit (CHRN) genes. *Am J Med Genet B Neuropsychiatr Genet*. 2009; 150B(4):453–66. [PubMed: 19259974]
- Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau O, Swan GE, Goate AM, Rutter J, Bertelsen S, Fox L, Fugman D, Martin NG, Montgomery GW, Wang JC, Ballinger DG, Rice JP, Bierut LJ. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet*. 2007; 16(1):36–49. [PubMed: 17135278]
- SAS Institute Inc. SAS Version 9.1. Cary, NC; USA: 2002–03.
- Schmid B, Blomeyer D, Becker K, Treutlein J, Zimmermann US, Buchmann AF, Schmidt MH, Esser G, Banaschewski T, Rietschel M, Laucht M. The interaction between the dopamine transporter gene and age at onset in relation to tobacco and alcohol use among 19-year-olds. *Addict Biol*. 2009; 14(4):489–99. [PubMed: 19740369]
- Shete S, Beasley TM, Etzel CJ, Fernandez JR, Chen J, Allison DB, Amos CI. Effect of winsorization on power and type 1 error of variance components and related methods of QTL detection. *Behav Genet*. 2004; 34(2):153–9. [PubMed: 14755180]
- Slotkin TA. Nicotine and the adolescent brain: insights from an animal model. *Neurotoxicol Teratol*. 2002; 24(3):369–84. [PubMed: 12009492]
- Slotkin TA, Bodwell BE, Ryde IT, Seidler FJ. Adolescent nicotine treatment changes the response of acetylcholine systems to subsequent nicotine administration in adulthood. *Brain Res Bull*. 2008; 76(1–2):152–65. [PubMed: 18395624]
- Slotkin TA, Seidler FJ. Nicotine exposure in adolescence alters the response of serotonin systems to nicotine administered subsequently in adulthood. *Dev Neurosci*. 2009; 31(1–2):58–70. [PubMed: 19372687]
- Storey JD. A direct approach to false discovery rates. *J Roy Statistical Society, Series B*. 2002; 64:479–498.
- Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, Stacey SN, Bergthorsson JT, Thorlacius S, Gudmundsson J, Jonsson T, Jakobsdottir M, Saemundsdottir J, Olafsdottir O, Gudmundsson LJ, Bjornsdottir G, Kristjansson K, Skuladottir H, Isaksson HJ, Gudbjartsson T, Jones GT, Mueller T, Gottsater A, Flex A, Aben KK, de Vegt F, Mulders PF, Isla D, Vidal MJ, Asin L, Saez B, Murillo L, Blondal T, Kolbeinnsson H, Stefansson JG, Hansdottir I, Runarsdottir V, Pola R, Lindblad B, van Rijn AM, Dieplinger B, Haltmayer M, Mayordomo JI, Kiemeny LA, Matthiasson SE, Oskarsson H, Tyrfinsson T, Gudbjartsson DF, Gulcher JR, Jonsson S, Thorsteinsdottir U, Kong A, Stefansson K. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008; 452(7187):638–42. [PubMed: 18385739]
- Trauth JA, Seidler FJ, Ali SF, Slotkin TA. Adolescent nicotine exposure produces immediate and long-term changes in CNS noradrenergic and dopaminergic function. *Brain Res*. 2001; 892(2):269–80. [PubMed: 11172774]
- Trauth JA, Seidler FJ, Slotkin TA. An animal model of adolescent nicotine exposure: effects on gene expression and macromolecular constituents in rat brain regions. *Brain Res*. 2000; 867(1–2):29–39. [PubMed: 10837795]
- Tukey, JW. A survey of sampling from contaminated distributions. In: Olkin, I.; Ghurye, SG.; Hoefding, W.; Madow, WG.; Mann, HB., editors. *Contributions to Probability and Statistics Essays in Honor of Harold Hotelling*. Stanford, CA: Stanford Univ. Press; 1960. p. 448–485.
- Vink JM, Smit AB, de Geus EJ, Sullivan P, Willemsen G, Hottenga JJ, Smit JH, Hoogendijk WJ, Zitman FG, Peltonen L, Kaprio J, Pedersen NL, Magnusson PK, Spector TD, Kyvik KO, Morley KI, Heath AC, Martin NG, Westendorp RG, Slagboom PE, Tiemeier H, Hofman A, Uitterlinden AG, Aulchenko YS, Amin N, van Duijn C, Penninx BW, Boomsma DI. Genome-wide association study of smoking initiation and current smoking. *Am J Hum Genet*. 2009; 84(3):367–79. [PubMed: 19268276]

- Wang JC, Cruchaga C, Saccone NL, Bertelsen S, Liu P, Budde JP, Duan W, Fox L, Grucza RA, Kern J, Mayo K, Reyes O, Rice J, Saccone SF, Spiegel N, Steinbach JH, Stitzel JA, Anderson MW, You M, Stevens VL, Bierut LJ, Goate AM. Risk for nicotine dependence and lung cancer is conferred by mRNA expression levels and amino acid change in CHRNA5. *Hum Mol Genet.* 2009; 14(5): 501–510.
- Weiss RB, Baker TB, Cannon DS, von Niederhausern A, Dunn DM, Matsunami N, Singh NA, Baird L, Coon H, McMahon WM, Piper ME, Fiore MC, Scholand MB, Connett JE, Kanner RE, Gahring LC, Rogers SW, Hoidal JR, Leppert MF. A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genet.* 2008; 4(7):e1000125. [PubMed: 18618000]
- Wernicke C, Samochowiec J, Schmidt LG, Winterer G, Smolka M, Kucharska-Mazur J, Horodnicki J, Gallinat J, Rommelspacher H. Polymorphisms in the N-methyl-D-aspartate receptor 1 and 2B subunits are associated with alcoholism-related traits. *Biol Psychiatry.* 2003; 54(9):922–8. [PubMed: 14573320]

Table 1
Number of SNPs meeting various significance thresholds under main effect and interaction models, and overall false-discovery rates.

	df	p<0.05	p<0.01	p<0.001	π_0^I
SNP + rAOS	1	268	104	20	86%
SNP + rAOS + SNP x rAOS	2	448	142	29	76%
SNP + nAOS + SNP x pAOS	2	296	84	14	95%
SNP + rAFC + SNP x rAFC	2	329	105	15	82%
SNP + rAOD + SNP x rAOD	2	268	58	7	83%
SNP + rAFD + SNP x rAFD	2	332	107	8	93%

Note:

I Overall false discovery-rate (π_0) derived from QVALUE. This parameter estimates the percentage of false-positive tests if the null hypothesis were rejected for all tests (e.g., $\pi_0 = 1$ if the null hypothesis is true for all SNPs).

Table 2

Summary of Top 30 findings from SNP main effect model – top findings from each of 19 R² bins are presented.

Rank	Bin	SNP	GENE	Chr	Pos (Mb)	Minor Allele	Freq.	P	OR (95% CI)	Δ AIC ³	Δ C(%) ⁴
1	1	rs1488	MAP3K4	6	161.46	G	0.36	0.00027	1.35 (1.15, 1.59)	-11.3	0.30
2	2	rs3025382	DBH	9	135.49	A	0.11	0.00033	0.64 (0.50, 0.82)	-10.9	0.28
3	3	rs17483548	IREB2 ¹	15	76.52	A	0.34	0.00035	1.35 (1.14, 1.59)	-10.8	0.21
4	4	rs13433293	KCNB1	20	47.43	T	0.18	0.00037	0.69 (0.57, 0.85)	-10.7	0.38
5	5	rs578776	CHRNA5	15	76.68	T	0.28	0.00038	0.73 (0.61, 0.87)	-10.7	0.35
7	6	rs16969968	CHRNA5	15	76.67	A	0.35	0.00044	1.34 (1.14, 1.58)	-10.5	0.50
11	7	rs6517442	KCNJ6	21	38.21	C	0.31	0.00063	1.32 (1.12, 1.56)	-9.7	0.26
12	8	rs1206549	CLTCL1	22	17.60	A	0.15	0.00064	0.69 (0.56, 0.86)	-9.7	0.24
15	9	rs9217	ZBTB4 ²	17	7.30	G	0.35	0.00069	0.75 (0.64, 0.89)	-9.6	0.32
16	10	rs9764	NPY1R	4	164.46	C	0.27	0.00074	1.34 (1.13, 1.60)	-9.6	0.36
19	11	rs2314379	MAP3K4	6	161.39	T	0.43	0.00093	0.77 (0.66, 0.90)	-9.0	0.30
20	12	rs2236196	CHRNA4	20	61.45	G	0.26	0.00093	1.35 (1.13, 1.61)	-9.0	0.18
22	13	rs17228571	SCG5	15	30.75	T	0.37	0.00109	1.31 (1.11, 1.54)	-8.8	0.26
23	14	rs4559074	MAP3K4	6	161.39	G	0.05	0.00131	1.85 (1.26, 2.72)	-8.7	0.26
26	15	rs7164594	LOC123688 ¹	15	76.59	T	0.22	0.00150	0.74 (0.61, 0.89)	-8.1	0.20
27	16	rs17732878	CHRNBI	17	7.30	C	0.21	0.00173	0.73 (0.61, 0.89)	-7.8	0.24
29	17	rs637137	CHRNA5	15	76.66	A	0.22	0.00184	0.74 (0.62, 0.90)	-7.8	0.29
31	19	rs1378650	CHRM2	7	136.36	T	0.46	0.00209	1.29 (1.10, 1.52)	-7.7	0.24

Notes: Italics indicate SNPs that are in moderate LD ($R^2=0.5$ to 0.8) with more highly ranked SNPs.

¹ Proximal to CHRNA5-A3-B4 Cluster.

² Proximal to CHRNBI.

³ Akaike's information criterion compared with base model (Equation 1); higher values indicate better model fit.

⁴ Improvement in logistic regression concordance rate (cross-validated) on introducing SNP into the model.

Table 3

Summary of Top 30 findings from SNP x rAOS Interaction model – Top findings from each of 16 R² bins are presented

Rank	LD Bin	SNP	Gene	Chr	Position (Mb)	Minor Allele	Freq.	P-Main (Main Effect Model)	P-Joint (SNP, SNP x rAOS)	ΔAIC^3 (Main Effect Model)	ΔAIC (SNP, SNP x rAOS)	$\Delta C(\%)^4$
1	1	rs17760877	GRIN2B	12	13.71	C	0.11	0.238236	0.000015	+0.6	-18.3	0.72
2	2	rs11693629	DBI	2	119.84	T	0.29	0.171475	0.000077	+0.1	-14.9	0.31
3	3	rs890	GRIN2B	12	13.61	G	0.49	0.081236	0.000086	-1.0	-14.7	0.57
4	4	rs16969968	CHRNA5	15	76.67	A	0.35	0.000440	0.000109	-10.4	-14.2	0.79
5	5	rs2314379	MAP3K4	6	161.39	T	0.43	0.000929	0.000142	-9.0	-13.7	0.20
7	6	rs10505566	ADCY8	8	131.95	C	0.21	0.331764	0.000173	+1.1	-13.3	0.28
9	7	rs17483548	IREB2 ¹	15	76.52	A	0.34	0.000355	0.000202	-10.8	-13.0	0.51
10	8	rs2289948	DBI	2	119.85	G	0.20	0.067328	0.000233	-1.3	-12.7	0.41
18	9	rs6517442	KCNJ6	21	38.21	C	0.31	0.000628	0.000413	-9.7	-11.6	0.26
19	10	rs2277091	AGPAT4	6	161.49	A	0.42	0.004145	0.000458	-6.2	-11.4	0.48
23	11	rs3025382	DBH	9	135.49	A	0.11	0.000326	0.000662	-10.9	-10.6	0.22
25	12	rs29202	VAPA	18	9.92	C	0.05	0.766735	0.000717	-8.2	-10.6	0.18
26	13	rs13433293	KCNB1	20	47.43	T	0.18	0.000368	0.000735	-10.7	-10.4	0.45
28	14	rs452667	GABRR1	6	89.96	T	0.19	0.575845	0.000785	-9.6	-10.3	0.22
29	15	rs1488	MAP3K4	6	161.46	G	0.36	0.000268	0.000849	-1.7	-10.3	0.39
30	16	rs12545452	ADCY8	8	131.85	T	0.36	0.582631	0.001001	-11.3	-10.1	0.48

Note: Italics indicate SNPs that are in moderate LD ($R^2=0.5$ to 0.8) with more highly ranked SNPs P-Joint is derived from the interaction model analysis (Equation 3). P-Main is the p-value for the same SNP when analyzed without the interaction term (Equation 2).

¹ Proximal to CHRNA5-A3-B4 Cluster.

³ Akaike's information criterion compared with base model (Equation 1); lower values indicate better model fit.

⁴ Improvement in logistic regression concordance rate (cross-validated) on introducing SNP and SNP x rAOS into the model.

Table 4

Parameter estimates for top findings from SNP x rAOS Interaction Model, (left two columns) and main effect models computed on sub-samples stratified by AOS (median-split; right two columns).

Rank	LD Bin	Rs Number	Gene	Interaction Model (Full Sample)		Main Effect ORs (Sample split by AOS)	
				Main Effect OR (95% CI) [/]	Interaction OR, Per Yr. (95% CI)	Main effect OR for AOS <16 Sub-Sample	Main effect OR for AOS ≥16 Sub-Sample
1	1	rs17760877	GRIN2B	0.88 (0.67, 1.14)	1.25 (1.12, 1.41) ****	1.12 (0.81, 1.53)	0.54 (0.35, 0.85) **
2	2	rs11693629	DBI	0.93 (0.78, 1.10)	1.13 (1.06, 1.20) ***	1.10 (0.88, 1.37)	0.74 (0.57, 0.96) *
3	3	rs890	GRIN2B	0.83 (0.70, 0.97) *	0.90 (0.86, 0.95) ****	0.67 (0.54, 0.82) ****	1.18 (0.94, 1.49)
4	4	rs16969968	CHRNA5	1.29 (1.09, 1.51) **	0.94 (0.89, 0.99) *	1.16 (0.94, 1.43)	1.55 (1.21, 1.97) ***
5	5	rs2314379	MAP3K4	0.74 (0.63, 0.87) ***	0.94 (0.89, 0.99) *	0.68 (0.55, 0.84) ***	0.90 (0.71, 1.13)
7	6	rs10505566	ADCY8	0.93 (0.77, 1.14)	1.15 (1.07, 1.23) ***	1.15 (0.89, 1.48)	0.68 (0.49, 0.93) *
9	7	rs17483548	IREB2	1.31 (1.11, 1.54) **	0.94 (0.89, 0.99) *	1.17 (0.95, 1.46) ****	1.60 (1.24, 2.05) ****
10	8	rs2289948	DBI	0.87 (0.72, 1.06)	1.14 (1.05, 1.22) ***	1.01 (0.79, 1.29)	0.70 (0.51, 0.94) *
18	9	rs6517442	KCNJ6	1.29 (1.09, 1.52) **	0.95 (0.90, 1.00)	1.18 (0.95, 1.47)	1.53 (1.20, 1.95) ***
19	10	rs2277091	AGPAT4	0.76 (0.65, 0.89) ***	0.93 (0.88, 0.98) **	0.71 (0.58, 0.89) **	0.93 (0.73, 1.18)
23	11	rs3025382	DBH	0.65 (0.51, 0.84) ***	1.06 (0.97, 1.15)	0.66 (0.48, 0.91) *	0.54 (0.36, 0.80) **
25	12	rs29202	VAPA	0.90 (0.65, 1.23)	0.82 (0.75, 0.90) ****	0.60 (0.40, 0.90) *	1.74 (1.03, 2.93) *
26	13	rs13433293	KCNB1	0.70 (0.57, 0.86) ***	1.04 (0.97, 1.12)	0.65 (0.49, 0.84) **	0.70 (0.51, 0.96) *
28	14	rs452667	GABRR1	1.00 (0.81, 1.23)	1.14 (1.06, 1.23) ***	1.13 (0.86, 1.48)	0.73 (0.54, 0.99) *
29	15	rs1488	MAP3K4	1.37 (1.16, 1.62) ***	1.02 (0.97, 1.08)	1.36 (1.09, 1.69) **	1.24 (0.97, 1.57)
30	16	rs12545452	ADCY8	0.99 (0.84, 1.16)	0.91 (0.87, 0.96) ***	0.92 (0.75, 1.14)	1.17 (0.93, 1.48)

Note:

[/] Estimated odds ratio for the genetic effect when rAOS=0 (or AOS=16). Italics indicate SNPs that are in moderate LD ($R^2=0.5$ to 0.8) with more highly ranked SNPs. P-values for parameter estimates indicated as follows:

* p<0.05,

** p<0.01,

*** p<0.001,

100000>d

*

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 5
Logistic Regression Concordance Values from Multi-Locus main effect and interaction models.

	AIC ³	Difference in Predicted Risk Score (Cases – Controls)		Concordance Rate	
		Standard	Cross-Validated	Standard	Cross-Validated
Age + Sex	-----	0.032 (0.004)	0.031 (0.004)	60.3%	58.8%
Age + Sex + rAOS	0	0.156 (0.009)	0.154 (0.009)	74.1%	73.7%
Age + Sex + rAOS + SNPs ¹	-100.6	0.224 (0.011)	0.215 (0.011)	77.7%	76.6%
Age + Sex + rAOS + (SNPs + rAOS*SNPs) ²	-143.2	0.261 (0.011)	0.243 (0.011)	80.0%	78.1%

¹ Include 13 SNPs remaining in model after backward elimination procedure (rs3025382, rs13433293, rs578776, rs17405217, rs6517442, rs1206549, rs9764, rs2314379, rs2236196, rs17228571, rs4559074, rs17732878, rs2656052).

² Includes 13 (SNP + SNP x rAOS) terms (rs17760877, rs890, rs2314379, rs10505566, rs17483548, rs6756192, rs6517442, rs2036527, rs10519203, rs29202, rs17483686, rs452667, rs12545452) terms and 3 additional SNP main-effect terms (rs3025382, rs13433293, rs452667) remaining after backward elimination procedure.

³ Akaike's information criterion compared with base model (Equation 1); lower values indicate better model fit.