Published in final edited form as:

Behav Genet. 2012 July ; 42(4): 636–646. doi:10.1007/s10519-012-9529-y.

The *CHRNA5/A3/B4* Gene Cluster and Tobacco, Alcohol, Cannabis, Inhalants and Other Substance Use Initiation: Replication and New Findings Using Mixture Analyses

Gitta H. Lubke,

Department of Psychology, University of Notre Dame, 118 Haggard Hall, Notre Dame, IN 46556, USA

Sarah H. Stephens,

University of Maryland School of Medicine, Boulder, CO, USA

Jeffrey M. Lessem,

Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA

John K. Hewitt, and

Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA. Department of Psychology and Neuroscience, University of Colorado, Boulder, CO, USA

Marissa A. Ehringer

Institute for Behavioral Genetics, University of Colorado, Boulder, CO, USA. Department of Integrative Physiology, University of Colorado, Boulder, CO, USA

Gitta H. Lubke: glubke@nd.edu

Abstract

Multiple studies have provided evidence for genetic associations between single nucleotide polymorphisms (SNPs) located on the *CHRNA5/A3/B4* gene cluster and various phenotypes related to Nicotine Dependence (Greenbaum et al. 2009). Only a few studies have investigated other substances of abuse. The current study has two aims, (1) to extend previous findings by focusing on associations between the *CHRNA5/A3/B4* gene cluster and age of initiation of several different substances, and (2) to investigate heterogeneity in age of initiation across the different substances. All analyses were conducted with a subset of the Add Health study with available genetic data. The first aim was met by modeling onset of tobacco, alcohol, cannabis, inhalants, and other substance use using survival mixture analysis (SMA). Ten SNPs in *CHRNA5/A3/B4* were used to predict phenotypic differences in the risk of onset, and differences between users and non-users. The survival models aim at investigating differences in the risk of initiation across the 5–18 age range for each phenotype separately. Significant or marginally significant genetic effects were found for all phenotypes. The genetic effects were mainly related to the risk of initiation and to a lesser extent to discriminating between users and non-users. To address the second goal, the survival analyses were complemented by a latent class analysis that modeled all phenotypes jointly. One of the ten SNPs was found to predict differences between the early and late onset classes. Taken together, our study provides evidence for a general role of the *CHRNA5/A3/B4* gene cluster in substance use initiation that is not limited to nicotine and alcohol.

Keywords

Survival mixture analysis; Nicotine initiation; Substance use; *CHRNA5/B3/A4* cluster

Introduction

Genetic association studies commonly involve fitting a simple statistical model that relates genetic markers to a phenotype. This approach is especially useful when investigating a large number of markers. If the focus is narrowed to a small region of the genome with only a few markers more complex models of the phenotype can be specified that permit the simultaneous analysis of multiple markers. This paper presents two complementary sets of mixture analyses. First, survival mixture models are used to investigate potential heterogeneity in self-reported first use of tobacco, alcohol, cannabis, inhalants, and other substances for each phenotype separately. Second, all phenotypes are analyzed simultaneously in a latent class analysis in order to discriminate between potentially different patterns of initiation across substances. In both sets of analyses, ten single nucleotide polymorphisms (SNPs) in the *CHRNA5/A3/B4* gene cluster are used to predict different aspects of heterogeneity in substance use initiation. The *CHRNA5/A3/B4* gene cluster is a well-replicated region that has been associated with drug behaviors in over thirty different studies.

Nicotinic receptors

SNPs in the *CHRNA5/A3/B4* region have been related to several different nicotine and other substance use behaviors (genetic studies reviewed below, see also Greenbaum and Lerer 2009). These genes encode for $\alpha 5$, $\alpha 3$, $\beta 4$ subunits which contribute to neuronal acetylcholine receptors (nAChRs). nAChRs are ligand-gated ion channels composed of five subunits, which may combine in a variety of configurations and are specifically expressed in different brain regions. Different combinations of subunits lead to differences in receptor pharmacology. For example, $\alpha 4\beta 2^*$ receptors (* indicates some other receptor subunit) are most widely expressed in the brain, show high nicotine affinity, and slow desensitization (Gotti et al. 2009). In brief, nicotine directly binds to nAChRs, which initiates a receptor activation-desensitization cycle. Activation of nAChRs has been shown to activate dopamine release, establishing their role in the dopaminergic reward system (Schlaepfer et al. 2008). Alcohol is thought to moderate the function of nAChRs, which might explain at least partially the comorbidity of nicotine and alcohol dependence (Chatterjee 2010). Receptors containing the $\alpha 5$, $\alpha 3$, $\beta 4$ subunits are expressed in limited regions of the brain (Gotti et al. 2009), but recent studies have provided strong evidence for their role in mediating drug responses.

In particular, studies of null mutant ("knock-out", KO) mice, which lack the $\alpha 3$ and $\beta 4$ subunits have shown differential behavioral responses to nicotine. Heterozygous $\alpha 3$ animals (i.e. mice which carry one copy of the gene) are resistant to nicotine-induced seizures and hypolocomotion, suggesting a possible role of this gene in central responses to nicotine. $\beta 4$ KO mice also show reduced sensitivity to the effects of nicotine on seizure induction and locomotion (Salas et al. 2004a), but otherwise appear normal and do not show any symptoms of nicotine withdrawal (Salas et al. 2004b). In addition, the $\alpha 3$ and $\beta 4$ -containing receptors in the habenula-interpeduncular nucleus have recently emerged as potential targets for smoking cessation strategies (see review by (De Biasi and Salas 2008). For example, although the commonly prescribed treatment for smoking cessation, Varenicline, is a partial agonist of $\alpha 4\beta 2^*$ receptors, it is also a full agonist of $\alpha 7$ and a partial agonist of $\alpha 3\beta 4^*$ receptors, which have been shown to display higher efficacy (maximal release) than $\alpha 4\beta 2^*$ receptors (Mihalak et al. 2006). Moreover, the antidepressant Bupropion, which

alters brain reward circuits influenced by nicotine and is used as an anti-tobacco agent, has been shown to act as a nAChR antagonist, including the $\alpha 3\beta 4^*$ subtypes (Fryer and Lukas 1999; Slemmer et al. 2000; Cheverud 2001, 2007; Mansvelder et al. 2007). The role of $\alpha 3\beta 4^*$ nAChRs in the habenula and interpeduncular nucleus has been studied using an alkaloid, ibogaine, and its derivative, 18-Methoxycoronaridine (MC-18) (Glick et al. 2001). Importantly, in rodents, MC-18 leads to decreased self-administration not only of nicotine and alcohol, but also of morphine, cocaine and methamphetamine (Maisonneuve and Glick 2003). Thus, there is evidence from animal models that these particular subunits, and their corresponding genes, are likely to be involved in multiple substance-related behaviors, including behaviors related to substances other than nicotine and alcohol.

Human genetic studies of the *CHRNA5/A3/B4* region

The main human genetic evidence comes from studies investigating nicotine dependence (ND) or ND related phenotypes including first use of tobacco. Some evidence has been reported also for alcohol initiation (Schlaepfer et al. 2008), alcohol dependence (Chen et al. 2009; Sherva et al. 2010), level of response to alcohol and other substance use. Converging evidence has been established for rs16969968 in the *CHRNA5* gene. This non-synonymous SNP has been associated with ND (Saccone et al. 2009a), ND severity/heavy smoking (Weiss et al. 2008, Stevens et al. 2008), habitual smoking (Bierut et al. 2008) and early onset of smoking (Weiss et al. 2008). It has also been shown to affect receptor function, where the risk allele was associated with decreased response to a nicotine agonist (Bierut et al. 2008). Several recent papers have begun to disentangle the multiple distinct genetic signals that are present in this region, which may be associated with different functions, and perhaps slightly different endophenotypes (Saccone et al. 2009a, b; Saccone et al. 2010). SNP rs1051730, present on many of the GWA platforms, is in near perfect linkage disequilibrium (LD, as measured by r^2) with rs16969968, and has emerged from several GWA studies of nicotine behaviors and lung cancer (Amos et al. 2008; Thorgeirsson et al. 2008, 2010). It has also been examined in meta-analyses where association with tobacco behaviors reaches genome-wide significance levels for two independent loci (Saccone et al. 2010; Thorgeirsson et al. 2010). More recently, SNPs in this region have been associated with externalizing behaviors, supporting a more general role of these genes that is not specific to nicotine (Stephens et al. 2011). Our aim is to extend these findings by focusing on a more fine-grained definition of different substance initiation phenotypes and their relative association with specific genetic signals in the *CHRNA5/A3/B4* gene cluster.

The current study

We investigated age of onset of nicotine and alcohol use and their association with SNPs in the *CHRNA5/A3/B4* gene cluster. Since especially mouse models indicated involvement of subunits in a broader range of substance use behaviors, we investigated three additional substances or substance groups in an attempt to explore the specificity of the *CHRNA5/A3/B4* gene cluster in substance use initiation. The three additional phenotypes were age of initiation of cannabis, inhalants, and “other substances”. “Other substances” included LSD, PCP, ecstasy, and amphetamines.

The current study focuses specifically on potential heterogeneity of the phenotype. Since we consider age of onset of different substances, and it is likely that the general population from which the sample was drawn contains (at least) two subpopulations, namely subjects who will never initiate use, and subjects who have already initiated or who might initiate use in the future. Therefore, two separate sets of mixture analyses were conducted, namely (1) a survival mixture analysis (SMA) to investigate potential heterogeneity regarding the risk of first use of tobacco, alcohol, cannabis, inhalants, and other substances (i.e., the category including LSD, PCP, amphetamines, and ecstasy), and (2) a latent class analysis (LCA)

conducted on subjects with reported first use to examine differential patterns of first use of the same five (groups of) substances. The two sets of analyses aimed at answering different but complementary questions. The SMA addressed the risk of initiation for the five phenotypes separately. Survival mixture models permit differentiation between subjects who never use (i.e., “long-term survivors”), and subjects who have initiated use or are likely to initiate use in the future. We compared survival models with or without a long-term survivor class, and with increasing constraints on how risk is modeled (no constraint, proportional hazard, and constant hazard). The SNPs are integrated as predictors of the risk of first use, and in case of the 2-class models, also as predictors of the difference between users and long-term survivors. The LCA, on the other hand, specifically focuses on differences in the pattern of initiation of tobacco, alcohol, cannabis, inhalants, and other substance use. The latent classes may for instance differentiate between subjects with an early versus a late onset of these substances. Alternatively, latent classes may also group subjects with a different pattern of initiation. For example, it may be possible to observe a class with an early age of initiation of alcohol and tobacco and a late first use of cannabis, versus a class with a reverse pattern of initiation. In the LCA models, SNPs were used to predict differences between classes.

Data were used from the genetic pairs subsample of the National Longitudinal Study of Adolescent Health (Add Health), which is a longitudinal study of a nationally representative sample of adolescents in the United States (Harris 2009).

We analyzed self-reported first use of tobacco, alcohol, cannabis, inhalants, and other substances obtained during the first wave when the subjects were on average between 16 and 17 years old. The sample was limited to Caucasians due to allele frequency differences across different ethnic groups (Schlaepfer et al. 2008), and to subjects with available genetic data. Subjects had been previously typed for ten SNPs on *CHRNA5/A3/B4* including rs16969968 and nine SNPs that were examined by Schlaepfer et al. (2008) in two samples, a Colorado sample and the National Youth Survey sample, that were independent of the Add Health sample used here. Hence, the current study was designed to replicate previous results, and to extend findings by focusing on population heterogeneity in the phenotype. Using a subset of a population data set provides a wide dispersion of the phenotype measures, and therefore affords the possibility of capturing potential heterogeneity in the phenotype, and of modeling the association of SNPs with subgroup differences. A second aim was to obtain more detailed information concerning the specificity of the *CHRNA5/A3/B4* gene cluster with respect to the risk of initiating a variety of different substances, in addition to detecting SNPs that differentiate between users and non-users.

Methods

Subjects

The behavioral data used in the present study are part of the National Longitudinal Study of Adolescent Health (Add Health). Add Health is a longitudinal study of a nationally representative sample of adolescents that were in grades 7–12 in the United States during the 1994–95 school year. The current analyses utilized data from respondents in the sibling-pairs subsample for which phenotypic and DNA were available. Details about the Add Health data are provided at <http://www.cpc.unc.edu/projects/addhealth> and in Harris et al. 2006. Study protocols were approved by Institutional Review Boards (NC and CO, USA), and subjects provided written informed consent.

For all analyses, we used subjects with available genetic data. A reduction of the sample size was due to the fact that mixture models are estimated conditional on covariates, hence only

subjects with complete genetic data on all ten SNPs are analyzed. Depending on the phenotype, sample sizes were approximately $N = 775$ in the survival models.

Sex was evenly distributed in the sample. Mean age at the time the phenotypic data were collected was 16.7 (SD = 1.4). Sex was included as covariate in all models to account for potential differences in age of onset.

The items used to measure initiation of using tobacco, cannabis, cocaine, and inhalants (such as glue or solvents) were worded as “How old were you when you... for the first time?”. First use of alcohol was measured with the item “Think about the first time you had a drink of beer, wine, or liquor when you were not with your parents or other adults in your family. How old were you then?”, and first use of other substances was measured with the item “How old were you when you first tried any other type of illegal drug, such as LSD, PCP, ecstasy, mushrooms, speed, ice, heroin, or pills, without a doctor’s prescription?” (see <http://www.cpc.unc.edu/projects/addhealth/codebooks/wave1>). The prevalence of the five phenotypes (reported age of first use vs. reported zero use) was tobacco = 0.46, alcohol = 0.43, cannabis = 0.26, inhalants = 0.06, and other substances = 0.09. Since the distribution of age of onset was skewed, medians are more informative than means. Median age of first use was tobacco = 13, alcohol = 14, cannabis = 15, inhalants = 13, and other substance = 15. Rank correlations between the reported age of first use of the different substances based on pair-wise complete data ranged between 0.38 and 0.59.

Selection of SNPs and genotyping

Ten SNPs in *CHRNA5/A3/B4* were genotyped using Taq-Man® assays for allelic discrimination (Applied Biosystems) per manufacturer’s instructions using an ABI PRISM® 7900 instrument. SNPs were selected based on validation status, minor allele frequency greater than 0.10, and location in the gene to be approximately evenly distributed. The SNPs used in this study were selected for a previous study (Schlaepfer et al. 2008), and the selection was made prior to availability of much linkage disequilibrium (LD) data on HapMap (www.hapmap.org), but a comparison to other studies (Saccone et al. 2010) suggests the three major LD signals (locus 1, locus2, and locus 3) in the region are captured by the ten selected SNPs. The selected SNPs (in order along the physical map of the gene cluster) were rs684513, rs680244, rs16969968, rs514743, rs11637630, rs8040868, rs8023462, rs1948, rs1316971, and rs11634351. Two of these SNPs (rs16969968 and rs8040868) tag the locus 1 signal defined by Saccone et al. (2009a, b, 2010). One SNP (rs11637630) tags locus 2, and two SNPs (rs680244 and rs514743) tag locus 3. The other SNPs appear to be separate LD signals from the three identified major loci. LD for five of the SNPs is available on Hapmap, with r^2 ranging between 0.286 and 0.802, shown in Fig. 1. The correlation between SNPs in the sample selected for the mixture analyses was generally much lower, with r^2 not exceeding 0.22, hence the ten SNPs could be included in the analyses simultaneously without the risk of instability of parameter estimates due to multicollinearity. Minor allele frequencies in the sample ranged between 0.235 and 0.402. SNPs were scored 0, 1, and 2 indicating the number of minor alleles.

Analytical methods

SMA models and LCA models were fitted using the software program Mplus version 6.1 for linux (Muthén and Muthén 1998/2010).

SMA models—As mentioned above, data from individuals who were on average 16.7 years old at Wave 1 collection were analyzed. Using data from this age group increased the reliability of the self-reported age of first use, but it also introduces the possibility that individuals might not yet have initiated use. For these potential future users, the age of first

use is missing. This type of missing data is known as right censoring, and is explicitly accounted for in the likelihood function (Vermunt 1996). SMA models add the additional benefit of discriminating between long-term survivors (non-users) and individuals who have reported first use or might use in the future. This is similar to two-part random effect models where one class is used to group subjects with zero scores. Two-part random effects models are designed for symptom endorsement or severity scores, which can be modeled as a function of time. Our analyses focus on age of onset, which does not vary over time.

More specifically, survival analysis is designed to model the risk of onset, also known as hazard. The hazard can be modeled for continuous or for discrete time. Due to the character of the Add Health data (reported year of first use), discrete time survival models are more adequate. Discrete time survival models have been described in detail in Vermunt (1996), and Muthén and Masyn (2005). In essence, discrete time survival models serve to investigate the predictors of the risk of initiating a behavior. In this study, eight discrete age intervals were defined to correspond to reported first year of use and to ensure a sufficient number of subjects within each group: 5–10, 10–12, 13, 14, 15, 16, 17, 18.

Within each of the intervals, an individual can have one of three different states: (1) the individual has not started to use yet and is still at risk, (2) the individual has started to use, and (3) has reported use previously and is not at risk of initiation anymore. The hazard rate is computed simply by considering the subjects who report use as a fraction of all subjects who are still at risk. For example, if $N = 100$, and in the third interval 60 subjects have already previously reported use, then only 40 subjects are still at risk. If in the third interval ten subjects report first use, the hazard rate would be $1/4 = 0.25$.

We compared three survival models that place different constraints on how the hazard rate is modeled. Model 1 (M1) is unconstrained. This means that the hazard can be different in each time interval, and that the hazard is regressed on the ten SNPs. In this model, the SNPs have time interval-specific regression weights. This model is included as a lenient baseline model. Model 2 (M2) constrains the hazard to be proportional. This model permits the hazard to increase and/or decrease over time, however, covariate effects follow the proportionality constraint such that only one regression coefficient is estimated for each SNP. Model 3 (M3), finally, constrains the hazard to be constant over time. Again, there is only one regression coefficient estimated per covariate. The substantive meaning of the constant hazard is that for the observed time span (i.e. adolescence) the risk of onset does not change substantially. Although due to availability there might be a peak risk at a younger age for some substances, it might be possible that the risk of initiating use of other substances is fairly constant during adolescence.

All three models are also fitted as a mixture model with two classes. One of the classes is fitted as in the above single class models whereas the other class serves to group subjects who are “long-term survivors”, that is, subjects who never use. To fit the 2-class models, we created so-called “training variables”, which are used to aid in the distinction between classes. Training variables are binary variables that indicate whether or not class membership is known. For subjects who have initiated use it is known that they cannot belong to the non-user, long-term survivor class. For these subjects the training variable is zero, reflecting that class membership does not need to be estimated. All other subjects might still engage in substance use in the future, and hence it needs to be estimated from the data whether they are more likely to belong to the users or the non-users class. For these subjects the training variable equals one, indicating that class membership needs to be estimated. Potential future users and true long-term survivors are distinguished during model fitting based on similarity of patterns of the covariates. If a covariate of an individual without reported first use is similar to a reported first user then this individual will have a

higher probability of belonging to the user class. It should be noted though that given the small expected differences between non-users and future users with respect to our covariates, it is likely that at least a proportion of future users will have high probabilities of belonging to the long-term survivor class. A simple comparison of class proportions and prevalence rates of reported first use was used to evaluate this issue.

The survival mixture analysis provides an estimation of the risk to initiate substance use across the observed time. Depending on which model fits best, the risk is more likely constant, proportional, or unconstrained. Furthermore, the analysis provides information regarding the relation between the SNPs and the risk over time, and the relation between the SNPs and the differences between users and long-term survivors.

The five phenotypes, first use of tobacco, alcohol, cannabis, inhalants, and other substances, were analyzed separately. Due to the correlations between the five phenotypes, when fitting survival models for one of the phenotypes, we condition on initiation of the other four phenotypes. Therefore, effects of SNPs predicting risk of onset are specific for the phenotype that is investigated.

Sex was also included as a covariate in all models to account for potential gender differences in age of onset. The models are labeled as follows. Depending on whether a single or 2-class model is fitted, C1 or C2 is appended to M1–M3, leading to model labels M1C1, M1C2, M2C1, etc. (see Table 1).

LCA models—The second set of analyses utilized LCA models. The LCA models were fitted to self-reported age of onset of tobacco, alcohol, cannabis, inhalants, and other substance. The models serve to investigate whether different groups exist in the sample regarding the *pattern* of drug use initiation. It is possible that groups have the same pattern of initiation and only differ with respect to mean age of onset (e.g., young age of onset vs. older age). It is also possible that groups differ with respect to initial drug preference such as early use of alcohol and tobacco but late onset of cannabis versus late onset of alcohol and tobacco but early onset of cannabis. How subjects are grouped in LCA is driven by the data. In the current analysis, only subjects who had reported use at some age were used in order to focus on differences between users, and to avoid bias due to right censoring. Right censoring is a form of missingness that is only taken into account in survival models. When fitting the latent class models, the ten SNPs on *CHRNA5/A3/B4* were used to predict differences between classes. Latent class models are fitted with an increasing number of classes. The five phenotypes were analyzed jointly.

Multiple testing—To correct for multiple testing of SNPs in linkage disequilibrium, we used an implementation of methods described by Li and Ji (2005). Similar to Nyholt (2004), their approach is based on Cheverud (2001), and involves first estimating the “effective number” of independent tests using a spectral decomposition of the correlation matrix of the SNPs, followed by an adjustment of the correlated tests using the effective number of independent tests. In our case, this number was estimated to equal 5, hence the significance level is $0.05/5 = 0.01$, whereas marginal significance is $0.1/5 = 0.02$. Note that this correction does not correct for testing multiple phenotypes, or fitting multiple models (see also “Limitations” in the “Discussion” section).

Results

Survival mixture analysis

The size of the sample containing subjects with or without first use and available genetic data was $N \approx 775$ for the five phenotypes. The results of the model comparisons are

presented first together with a conceptual interpretation of the best fitting model for each phenotype in terms of risk of first use over time. This is followed by the results concerning genetic effects.

Model comparisons—The fit of the survival models was compared separately for each of the phenotypes using information criteria (Akaike's Information Criterion, AIC, the standard Bayesian Information Criterion, BIC, and an "sample-size adjusted" BIC which corrects for sample size in a milder way than the standard BIC). Note that the BIC has the strongest tendency to favor parsimonious models compared to AIC and saBIC. The survival models differed significantly with respect to model parsimony. For instance, constraining the hazard to be proportional and constant reduces the number of estimated parameters from 126 (M1C1) to 24 (M2C1) and 17 (M3C1), respectively. Lubke and Neale (2006, 2008) have shown that when comparing mixture models more parsimonious models can be preferred incorrectly by the BIC in case of smaller sample sizes, although Nylund et al. (2007) show that the BIC performs well in general mixture models. No specific simulation results are available for survival mixture models with training data.

Importantly, it is common that mixture analyses do not necessarily result in a single "best-fitting" model. The different information criteria can point to different "best-fitting" models, and for instance BICs for models with two and three classes may only differ very slightly. Whenever this occurs, we consider the results from the different indicated models (Lubke 2010). The results of the model comparison are summarized in Table 2.

All analyses were conducted using data from all family members. Non-independence of observations was accounted for using a sandwich-type estimated in conjunction with family id as a clustering variable.

Although all models converged properly, the models with unconstrained hazard (models M1C1 and M1C2) required fixing numerous parameter estimates during estimation. This was mainly due to small proportions of reported first use during one or more of the time intervals, and to lack of variability in the number of minor alleles of one or more of the SNPs at a given time interval. Recall that in unconstrained models, SNP effects are estimated specific for each time-interval. Results of the unconstrained models are therefore not presented in Table 2.

For tobacco, alcohol, and cannabis, the survival mixture Model M2C2 was the best fitting model according to all fit indices. The M2C2 model discriminates between long-term survivors (i.e., zero risk of onset), and subjects who reported an age of first use or might engage in use in the future. The class proportion of the users class were similar for tobacco and alcohol (0.45 and 0.49, respectively), and smaller in the case of cannabis (0.26). In the case of tobacco and alcohol, the class proportions were somewhat higher than the reported first use, showing that this class contains a small proportion subjects who, based on their higher similarity in the covariates with the users, are considered to be potential future users. The class proportion of cannabis reflects the reported first use. There were no sex differences between the classes in either one of these three phenotypes.

The three information criteria indicated two different models for inhalants and other substances, namely the M2C2 model and the M3C1 model with constant hazard. In case of first use of other substances AIC was lowest for the M2C2 model whereas the two different BIC's pointed to the more parsimonious M3c1 model. For other substances the pattern was similar although the saBIS was also lowest for the M2C2 model.

Genetic effects—The evidence for a role of the *CHRNA5/B4/A3* cluster for the five phenotypes is presented in Table 3. Several different SNPs predicted the risk to initiate substance use. Note that with a single exception all p values <0.1 were observed for risk of onset within the user class. All SNP effects predicting differences between subjects with and without reported first use had p values >0.1 with the exception of rs1948 which had a p value of .037 in the 2-class proportional hazard model for other substances.

Surprisingly, the weakest results were found for tobacco. Furthermore, our study showed a marginal effect of rs514743 with the risk of first use of alcohol (p value 0.011). This SNP was associated with age of initiation of alcohol and tobacco in a different population sample (Schlaepfer et al. 2008). Rs8040868, which had a marginal effect in Schlaepfer's study on first use of tobacco, marginally predicted risk of onset of cannabis (p value 0.011). Both marginal effects were in the predicted direction based on previous work, where the more rare variant was associated with higher risk. The risk of first use of inhalants was significantly predicted by rs680244 and rs11637630 (p values 0.004 and 0.002). Rs8023462 and rs11634357 were marginally significant. Rs11637630 significantly predicted risk of onset for other substances.

Latent class analysis

Model comparisons—Since the risk of first use showed a differential pattern across substances, a latent class analysis focusing on the users was conducted. The sample size in the LCA was $N=486$. The goal of this complementary analysis was to identify subgroups that may differ with respect to the pattern of initiation across the different substances.

Latent class models for 2–5 classes were fitted to first use of alcohol, tobacco, cannabis, inhalants, and other substances. The 3-class model had the lowest BIC, whereas the 4-class model had the lowest AIC and saBIC, which has a less severe correction for sample size than the standard BIC. Importantly, in both models, the classes were ordered with respect to onset. This means that the subjects in the Add Health sample can best be grouped with respect to an early onset class, one or two medium onset classes, and a late onset class. The 5-class model had one empty class and resembled the 4-class model with respect to the non-empty classes. The 5-class model is therefore not further considered.

The early onset class of the 3-class model contained 14.6% of the subjects, followed by 39.9 and 45.5% for the medium and late onset classes. Class proportions in the 4-class model were 11.5, 16.33, 32.9 and 39.3%. The fit indices of the models, and the estimated mean age of first use in the 3-class model are presented in Tables 4 and 5, respectively.

As can be seen in Table 5, the classes were strictly ordered with regard to mean age of onset for all five phenotypes. The same ordering was observed in the 4-class model. Furthermore, inhalants were consistently used prior to other substances. There were significant sex differences between classes, with boys having higher probabilities of belonging to earlier onset classes.

Genetic effect in the three and four class models—In the 3-class model, several of the estimated regressions of class differences on the ten SNPs had p values <0.1 , however only rs11637630 was marginally significant after correction for multiple testing (early vs. medium onset, p value 0.011). In the 4-class model rs8040868 differentiated early and medium onset from late onset (p values 0.007 and 0.017). This SNP had a p value of 0.027 in the 3-class model (early vs. late onset). Rs11637630 significantly predicted differences between early and later onset (p value 0.006). Rs684513 was marginally significant in differentiating early from late onset (p value 0.017).

The direction of all effects indicated that a larger number of the minor allele was more likely in the earlier onset group. Regarding rs8040868, note that according to Hapmap LD data from the merged Phase I, II, and III CEU sample (Utah residents with ancestry from northern and western Europe) this SNP is highly correlated with rs16969968 ($r^2 = 0.802$), which has been associated with severity of smoking and/or early onset of smoking and alcohol use in several studies (Saccone et al. 2009a, b; Weiss et al. 2008; Bierut et al. 2008; Sherva et al. 2009; Chen et al. 2009; Stevens et al. 2008). However, in our subsample from the Add Health data, the r^2 value between these two SNPs is only 0.52, which may explain the lack of association with rs16969968.

Discussion

The mixture analyses in this study afforded a detailed analysis of the phenotypes in conjunction with an equally detailed investigation of genetic effects. Nine of the ten investigated SNPs were previously analyzed by Schlaepfer et al. (2008) in two samples that were independent of the current sample, thereby providing ample opportunity to replicate and extend previous findings. In addition, SNP rs16969968 was investigated, which had been replicated in several different studies using various different substance-related phenotypes, and which has been shown to change receptor function (Bierut et al. 2008). Although our results concerning onset of smoking are weak, the main body of results from this study is consistent with previous findings regarding the role of the *CHRNA5/B4/A3* gene cluster. The allelic effects reported by Schlaepfer et al. (2008) were in the same direction as the corresponding effects in our analyses. Importantly, our work shows that the importance of this gene cluster is not limited to tobacco and alcohol initiation, but that it extends to the onset of use of various different substances.

In the survival analysis, ten SNPs on *CHRNA5/B4/A3* were used to predict the risk to engage in using tobacco, alcohol, cannabis, inhalants, and/or other substance between ages 5 and 19. Several SNPs were significant or marginally significant predicting onset of alcohol, cannabis, inhalants, and “other substances” after correcting for multiple testing. This result indicates a general role of the *CHRNA5/B4/A3* gene cluster that is not limited to nicotine dependence (ND) and ND related phenotypes. The latent class analysis, which modeled first use of the different substances jointly, showed a clear effect of rs11637630 and rs8040868 discriminating between early and later onset of first use. The effect of rs8040868 corroborates previous findings.

The survival mixture analysis provided interesting results not only with respect to the role of the *CHRNA5/B4/A3* gene cluster but also with respect to the structure of the risk to engage in the use of different substances during adolescence. For the phenotypes with a higher reported prevalence of first use (tobacco, alcohol, and cannabis) the best fitting models were 2-class models, which discriminate between reported (or potential future) users and individuals who have not reported first use and are considered to be unlikely users in the future. Likewise, the data indicated that the risk to engage in using these substances is proportional rather than constant during adolescence. For the phenotypes inhalants and other substances, the 2-class proportional hazard model and a survival model constraining the risk of using inhalants or other substances (a group including LSD, PCP, amphetamines, and ecstasy) to be constant across the age range emerged as the best fitting model for these phenotypes. The lack of clearly detecting two classes might be due to the much lower prevalence of reported first use in the Add Health sample.

Since the survival mixture analysis showed a different structure of risk of initiation across substances, a latent class analysis focusing on the users only was also examined. Importantly, the latent class analysis distinguished groups that were clearly ordered with

respect to age of onset. Furthermore, within the early onset class, inhalants had the lowest mean age of first use, whereas for instance “other substances” were initiated at a significantly later mean age. This pattern of first use is likely related to accessibility of different substances, which has been shown to be an important predictor of substance use (Gillespie et al. 2011).

Limitations

Regarding the reliability of the results concerning individual SNPs, some caution is warranted. In samples of limited size, small changes in observed allele frequencies due to sampling fluctuation can substantially affect results (Duncan and Keller 2011). Furthermore, multi-collinearity can render individual effects unreliable, although importantly, the joint prediction of correlated predictors is asymptotically correct. All analyses presented in the results included all family members, and the models were fitted using a sandwich-type estimator with family id as the clustering variable. We conducted additional analyses with only one subject per family. The additional analyses did not change our main conclusions, although the decrease in sample size rendered some of the significant effects non-significant. Also, not all of the individual SNP effects coincided.

The issue of multiple testing equally deserves consideration. We fitted a number of different models, and the effects of the ten SNPs were tested simultaneously in each model. Our correction for multiple testing was based on testing ten correlated SNPs, and did not take into account the different fitted models. The exact testing burden for an analysis that consists of fitting multiple different models has not been studied in detail, but is going to be addressed in the future.

Notwithstanding these limitations, the general pattern of our results indicates that the importance of the *CHRNA5/A3/B4* gene cluster is not limited to ND and ND related phenotypes. It is more broadly involved in the first use of various different substances, where it seems to have a role particularly in early onset of substance use behaviors.

Acknowledgments

The research of the first author was supported through grant DA018673 by NIDA. ME was supported by AA015336, AA107889, DA026901. SHS was supported by AA007464. JKH and JML were supported by HD031921. This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (<http://www.cpc.unc.edu/addhealth>).

References

- 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. Genomewide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nature Genetics*. 2008; 40:616–622.10.1038/ng.109 [PubMed: 18385676]
- Bierut LJ, Stitzel JA, Wang JC, Hinrichs AL, Gruzza 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; 165:1163–1171. [PubMed: 18519524]

- Chatterjee S, Bartlett SE. Neuronal nicotinic acetylcholine receptors as pharmacotherapeutic targets for the treatment of alcohol use disorders. *CNS Neurol Disord Drug Targets*. 2010; 9(1):60–76. Review. [PubMed: 20201817]
- Chen X, Chen J, Williamson VS, An SS, Hettema JM, Aggen SH, et al. Variants in nicotinic acetylcholine receptors alpha5 and alpha3 increase risks to nicotine dependence. *Am J Med Genet B Neuropsychiatr Genet*. 2009 (e-pub ahead of print).
- Cheverud JM. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity*. 2001; 87:52–58. [PubMed: 11678987]
- De Biasi M, Salas R. Influence of neuronal nicotinic receptors over nicotine addiction and withdrawal. *Exp Biol Med (May-wood)*. 2008; 233(8):917–929.
- Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am J Psychiatry*. 2011; 168(10):1041–1049. [PubMed: 21890791]
- Fryer JD, Lukas RJ. Noncompetitive functional inhibition at diverse, human nicotinic acetylcholine receptor subtypes by bupropion, phencyclidine, and ibogaine. *J Pharmacol Exp Ther*. 1999; 288(1): 88–92. [PubMed: 9862757]
- Gillespie NA, Lubke GH, Gardner CO, Neale MC, Kendler KS. Two-part random effects growth modeling to identify risks associated with alcohol and cannabis initiation, initial average use and changes in drug consumption in a sample of adult, male twins. *Drug Alcohol Depend*. 2011 (Epub ahead of print).
- Glick SD, Maisonneuve IM, et al. Mechanisms of action of ibogaine: relevance to putative therapeutic effects and development of a safer iboga alkaloid congener. *Alkaloids Chem Biol*. 2001; 56:39–53. [PubMed: 11705115]
- Gotti C, Clementi F, Fornari A, Gaimarri A, Guiducci S, Manfredi I, Moretti M, Pedrazzi P, Pucci L, Zoli M. Structural and functional diversity of native brain neuronal nicotinic receptors. *Biochem Pharmacol*. 2009; 78(7):703–711. [PubMed: 19481063]
- Greenbaum L, Lerer B. Differential contribution of genetic variation in multiple brain nicotinic cholinergic receptors to nicotine dependence: recent progress and emerging open questions. *Mol Psychiatry*. 2009; 14:912–945. [PubMed: 19564872]
- Greenbaum L, Rigbi A, Teltsh O, Lerer B. Role of genetic variants in the CHRNA5-CHRNA3-CHRNA4 cluster in nicotine dependence risk: importance of gene-environment interplay. *Mol Psychiatry*. 2009; 14:828–830. [PubMed: 19696770]
- Harris, KM. The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002; Wave IV, 2007–2009 [machine-readable data file and documentation]. Carolina Population Center, University of North Carolina; Chapel Hill: 2009.
- Harris KM, Halpern CT, Smolen A, Haberstick BC. The national longitudinal study of adolescent health (Add Health) twin data. *Twin Res Hum Gen*. 2006; 9:977–988.
- Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*. 2005; 95:221–227. [PubMed: 16077740]
- Lubke, GH. Latent variable mixture modeling. In: Hancock, GR.; Mueller, RO., editors. *The reviewer's guide to quantitative methods in the social sciences*. Routledge; New York: 2010. p. 283–297.
- Lubke GH, Neale MC. Distinguishing between latent classes and continuous factors: resolution by maximum likelihood? *Multivar Behav Res*. 2006; 41:499–532.
- Lubke GH, Neale MC. Distinguishing between latent classes and continuous factors with categorical outcomes: class invariance of parameters of factor mixture models. *Multivar Behav Res*. 2008; 43:592–620.
- Maisonneuve IM, Glick SD. Anti-addictive actions of an iboga alkaloid congener: a novel mechanism for a novel treatment. *Pharmacol Biochem Behav*. 2003; 75(3):607–618. [PubMed: 12895678]
- Mansvelder HD, Fagen ZM, et al. Bupropion inhibits the cellular effects of nicotine in the ventral tegmental area. *Biochem Pharmacol*. 2007; 74(8):1283–1291. [PubMed: 17868653]
- Mihalak KB, Carroll FI, et al. Varenicline is a partial agonist at alpha4beta2 and a full agonist at alpha7 neuronal nicotinic receptors. *Mol Pharmacol*. 2006; 70(3):801–805. [PubMed: 16766716]

- Muthén B, Masyn K. Discrete-time survival mixture analysis. *J Educational Behavioral Stat.* 2005; 30:27–28.
- Muthén, LK.; Muthén, BO. *Mplus User's Guide*. 6. Los Angeles, CA: Muthén and Muthén; 1998/2010.
- Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet.* 2004; 74:765–769. [PubMed: 14997420]
- Nylund KL, Asparouhov T, Muthen B. Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. *Struct Equ Model.* 2007; 14:535–569.
- Saccone NL, Wang JC, Breslau N, Johnson EO, Hatsukami D, Saccone SF, Grucza RA, Sun L, Duan W, Budde J, et al. The CHRNA5-CHRNA3-CHRNA4 nicotinic receptor subunit gene cluster affects risk for nicotine dependence in African-Americans and in European-Americans. *Cancer Res.* 2009a; 69:6848–6856. [PubMed: 19706762]
- 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.* 2009b; 150B(4):453–466. [PubMed: 19259974]
- Saccone NL, Culverhouse RC, Schwantes-An TH, Cannon DS, Chen X, Cichon S, Giegling I, Han S, Han Y, Keskitalo-Vuokko K, Kong X, Landi MT, Ma JZ, Short SE, Stephens SH, Stevens VL, Sun L, Wang Y, Wenzlaff AS, Aggen SH, Breslau N, Broderick P, Chatterjee N, Chen J, Heath AC, Heliövaara M, Hoft NR, Hunter DJ, Jensen MK, Martin NG, Montgomery GW, Niu T, Payne TJ, Peltonen L, Pergadia ML, Rice JP, Sherva R, Spitz MR, Sun J, Wang JC, Weiss RB, Wheeler W, Witt SH, Yang BZ, Caporaso NE, Ehringer MA, Eisen T, Gapstur SM, Gelernter J, Houlston R, Kaprio J, Kendler KS, Kraft P, Leppert MF, Li MD, Madden PA, Nöthen MM, Pillai S, Rietschel M, Rujescu D, Schwartz A, Amos CI, Bierut LJ. Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.* 2010; 6(8):e1001053. [PubMed: 20700436]
- Salas R, Cook KD, et al. The alpha3 and beta4 nicotinic acetylcholine receptor subunits are necessary for nicotine-induced seizures and hypolocomotion in mice. *Neuropharmacology.* 2004a; 47(3): 401–407. [PubMed: 15275829]
- Salas R, Pieri F, et al. Decreased signs of nicotine withdrawal in mice null for the beta4 nicotinic acetylcholine receptor subunit. *J Neurosci.* 2004b; 24(45):10035–10039. [PubMed: 15537871]
- Schlaepfer IR, Hoft NR, Ehringer MA. The genetic components of alcohol and nicotine co-addiction: from genes to behavior. *Curr Drug Abuse Rev.* 2008a; 2008(1):124–134. [PubMed: 19492010]
- Schlaepfer IR, Hoft NR, Collins AC, Corley RP, Hewitt JK, Hopfer CJ, Lessem JM, McQueen MB, Rhee SH, Ehringer MA. The CHRNA5/A3/B4 gene cluster variability as an important determinant of early alcohol and tobacco initiation in young adults. *Biol Psychiatry.* 2008b; 63:1039–1046. [PubMed: 18163978]
- Sherva R, Rice JP, Neuman RJ, Rochberg N, Saccone NL, Bierut LJ. Associations and interactions between SNPs in the alcohol metabolizing genes and alcoholism phenotypes in European Americans. *Alcohol Clin Exp Res.* 2009; 33:848–857. [PubMed: 19298322]
- Sherva R, Kranzler HR, Yu Y, Logue MW, Poling J, Arias AJ, Anton RF, Oslin D, Farrer LA, Gelernter J. Variation in nicotinic acetylcholine receptor genes is associated with multiple substance dependence phenotypes. *Neuropsychopharmacology.* 2010; 35:1921–1931. [PubMed: 20485328]
- Stemmer JE, Martin BR, et al. Bupropion is a nicotinic antagonist. *J Pharmacol Exp Ther.* 2000; 295(1):321–327. [PubMed: 10991997]
- Stephens SH, Hoft NR, Schlaepfer IR, Young SE, Corley RC, McQueen MB, Hopfer C, Crowley T, Stallings M, Hewitt J, Ehringer MA. Externalizing behaviors are associated with SNPs in the CHRNA5/CHRNA3/CHRNA4 gene cluster. *Behav Genet.* 2011
- Stevens VL, Bierut LJ, Talbot JT, Wang JC, Sun J, Hinrichs AL, et al. Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev.* 2008; 17:3517–3525. [PubMed: 19029397]

- Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP, Manolescu A, Thorleifsson G, Stefansson H, Ingason A, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature*. 2008; 452:638–642. [PubMed: 18385739]
- Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, et al. ENGAGE Consortium . Sequence variants at CHRNA3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet*. 2010; 42:448–453. [PubMed: 20418888]
- Vermunt, JK. Log-linear event history analysis: a general approach with missing data, unobserved heterogeneity, and latent variables. Tilburg University Press; Tilburg: 1996.
- Weiss RB, Baker TB, Cannon DS, von Niederhausen A, Dunn DM, Matsunami N, et al. A candidate gene approach identifies the CHRNA5-A3-B4 region as a risk factor for age-dependent nicotine addiction. *PLoS Genet*. 2008; 4:e1000125. [PubMed: 18618000]

Fig. 1. Heatmap showing Linkage Disequilibrium (r^2) in Caucasians of the 10 SNPs on *CHRNA5/A3/B4* used in this study

Table 1

Fitted models in the survival mixture analysis

M1C1	Single class, unconstrained hazard, one estimated coefficient for each SNP for each time interval
M2C1	Single class, proportional hazard, the estimated coefficient for each SNP is the same for all time intervals
M3C1	Single class, constant hazard, the estimated coefficient for each SNP is the same for all time intervals
M1C2	Two classes (users and non-users), unconstrained hazard in the users class, one estimated coefficient for each SNP for each time interval, SNPs also predict differences between classes
M2C2	Two classes (users and non-users), proportional hazard in the users class, the estimated coefficient for each SNP is the same for all time intervals, SNPs also predict differences between classes
M3C2	Two classes (users and non-users), constant hazard in the users class, the estimated coefficient for each SNP is the same for all time intervals, SNPs also predict differences between classes

Table 2

Fit indices of the survival mixture analysis

Model	Estimated parameters	AIC	BIC	saBIC
<i>Tobacco</i>				
M2C1	24	2091.588	2203.287	2127.076
M2C2	42	1990.900	2186.375	2053.004
M3C1	17	2141.551	2220.672	2166.689
M3C2	36	2114.038	2281.587	2167.270
<i>Alcohol</i>				
M2C1	24	2070.743	2182.566	2106.354
M2C2	42	1959.369	2155.059	2021.689
M3C1	17	2173.648	2252.856	2198.873
M3C2	36	<i>2124.816</i>	<i>2292.550</i>	<i>2178.233</i>
<i>Cannabis</i>				
M2C1	24	1309.454	1421.123	1344.911
M2C2	42	1204.181	1399.602	1266.232
M3C1	17	1422.662	1501.760	1447.777
M3C2	36	1416.546	1584.049	1469.732
<i>Inhalants</i>				
M2C1	24	397.771	509.347	433.136
M2C2	42	392.322	587.579	454.210
M3C1	17	400.751	479.783	425.800
M3C2	36	<i>375.520</i>	<i>542.884</i>	<i>428.567</i>
<i>Other (LSD, Amphetamines, XTC,...)</i>				
M2C1	24	526.388	638.118	561.907
M2C2	42	492.968	688.496	555.126
M3C1	17	549.205	628.347	574.364
M3C2	36	<i>529.725</i>	<i>697.321</i>	<i>583.004</i>

Note Models M1–M3 differ with respect to the hazard function. M1 = unconstrained (not presented), M2 = proportional, M3 = constant (for details, see Table 1). AIC stands for Akaike Information Criterion, BIC stands for Bayesian Information Criterion, saBIC has an additional sample size adjustment. Information criteria for the best fitting model(s) are presented in bold. Smaller values indicate better fit. Note that AIC, BIC, and saBIC do not necessarily favor the same model. Fit indices in italic indicate that one or more parameter estimates had to be fixed during estimation to avoid singularity of the information matrix. This can be indicative of model misspecification or an overly high model complexity given the information contained in the observed data

Table 3*P* values <0.1 for SNPs predicting risk of onset in the survival mixture analysis

Risk of onset tobacco	rs8040868	0.079
Risk of onset alcohol	rs680244	0.094
	rs514743	0.011
	rs11637630	0.071
	rs8040868	0.056
Risk of onset cannabis	rs684513	0.032
	rs8040868	0.011
Risk of onset inhalants	rs680244	0.004
	rs11637630	0.002
	rs8023462	0.012
	rs11634357	0.019
Risk of onset other substances	rs684513	0.060
	rs514743	0.022
	rs11637630	0.005

Note *P* values in bold are significant (*p* value <0.01) or marginally significant (*p* value <0.02) after correcting for multiple testing

Table 4

Fit indices latent class analysis models

Model	Estimated parameters	AIC	BIC	saBIC
2-class	33	4444.231	4582.376	4477.635
3-class	56	4279.686	4514.114	4336.373
4-class	79	4190.366	4521.077	4270.335
5-class	102	<i>4236.366</i>	<i>4663.360</i>	<i>4339.617</i>

Note AIC stands for Akaike Information Criterion, BIC stands for Bayesian Information Criterion, saBIC has an additional sample size adjustment. Information criteria for the best fitting model are presented in bold. Smaller values indicate better fit. Note that AIC, BIC, and saBIC do not necessarily indicate the same model

Table 5

Estimated mean age of onset and standard errors in the 3-class model

Phenotype	Early onset class	Medium onset class	Later onset class
Tobacco	9.897	12.377	14.161
Alcohol	11.369	12.993	14.994
Cannabis	11.221	13.496	15.536
Inhalants	8.526	12.896	14.673
Other substances	11.986	14.223	16.035

Note Estimated average age of onset is presented in the first column for each class, standard errors of these estimates are in italic in the second column