Published in final edited form as:

Behav Genet. 2011 September ; 41(5): 680–690. doi:10.1007/s10519-011-9476-z.

A Twin Association Study of Nicotine Dependence with Markers in the *CHRNA3* and *CHRNA5* Genes

Hermine H. Maes

Department of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, P.O. Box 980003, Richmond, VA 23298-0003, USA

Massey Cancer Center, Virginia Commonwealth University, Richmond, VA, USA

Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

Michael C. Neale

Department of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, P.O. Box 980003, Richmond, VA 23298-0003, USA

Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

Xiangning Chen

Department of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, P.O. Box 980003, Richmond, VA 23298-0003, USA

Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

Jingchun Chen

Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

Carol A. Prescott

Department of Psychology, University of Southern California, Los Angeles, CA, USA

Kenneth S. Kendler

Department of Human and Molecular Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, P.O. Box 980003, Richmond, VA 23298-0003, USA

Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

Abstract

Twin and family studies have provided overwhelming evidence for the genetic basis of individual differences in tobacco initiation (TI), regular smoking (RS) and nicotine dependence (ND). However, only a few genes have been reliably associated with ND. We used a finite mixture distribution model to examine the significance and effect size of the association of previously identified and replicated specific variants in the *CHRNA5* and *CHRNA3* receptor genes with ND, against the background of genetic and environmental risk factors for ND. We hypothesize that additional phenotypic information in relatives who have not been genotyped can be used to

increase the power of detecting the genetic variant. The nicotine measures were assessed by personal interview in female, male and opposite sex twin pairs ($N = 4,153$) from the population-based Virginia Twin Registry. Three SNPs in the *CHRNA5* and *CHRNA3* receptor genes, previously shown to be significantly associated with ND in this sample, were replicated in the augmented analyses; they accounted for less than one percent of the genetic variance in liability to ND, which is estimated to be over 50% of the phenotypic variance. The significance of these effects was increased by adding twins with phenotype but without genotype data, but gains are limited and variable. The SNPs associated with ND did not show a significant association with either TI or RS and appear to be specific to the addictive stage of ND, characterized by current smoking and smoking a large amount of cigarettes per day. Furthermore, these SNPs did not appear to be associated with the remaining items comprising the FTND scale. This study confirmed a significant contribution of the *CHRNA* receptor on different forms of tobacco dependence. However, the genetic variant only accounted for little of the total genetic variance for liability to ND. Including phenotypic data on ungenotyped relatives can improve the statistical power to detect the effects of genetic variants when they contribute to individual differences in the phenotype.

Keywords

Nicotine; Dependence; Association; Twin studies; Smoking; Receptors

Introduction

Smoking is a serious public health problem. Briefly, tobacco smoking is associated with increased morbidity, mortality, and personal and public cost (US Department of Health and Human Services 1989; World Health Organization 1997). In the US, cigarettes are responsible for 30% of all cancer deaths and 21% of deaths from cardiovascular disease (US Department of Health and Human Services 1989). Half of individuals beginning to smoke in adolescence will die from a cigarette related cause (World Health Organization 1997).

Twin studies have consistently shown a significant genetic component to the liability to smoking initiation and nicotine dependence. A recent review showed that the heritability estimates for TI and for ND range from 40 to 70% (Maes and Neale 2009) with shared environmental influences more pronounced in adolescence than adulthood. Furthermore, there is evidence for genetic and environmental correlation between TI and ND (Maes et al. 2004) suggesting that at least partly the same genes contribute to liability to TI and ND.

Linkage and association studies have identified specific regions on the genome that may harbor susceptibility genes, but few loci have been replicated (Straub et al. 1999; Wessel et al. 2010). Two notable exceptions are the $\alpha 5$ and $\alpha 3$ subunits of the nicotinic acetylcholine receptors (*CHRNA5* and *CHRNA3*). A non-synonymous SNP (rs16969968) was first identified to be associated with ND in a candidate gene study (Saccone et al. 2007) and independently replicated using fine mapping (Bierut et al. 2008) and genome wide association (Berrettini et al. 2008). Chen et al. (2009) also found a significant association between rs16969968 (as well as a highly correlated SNP rs1051730) when testing seven SNPs in the *CHRNA5* and *CHRNA3* genes. Polymorphisms in these genes were the only ones to show genomewide significance in their association with number of cigarettes smoked in three meta-analysis of GWAS studies involving over 100,000 individuals (Tobacco Genetics Consortium 2010; Thorgeirsson et al. 2010; Liu et al. 2010). More recently, at least four statistically independent loci for ND have been identified by dense coverage of the nicotinic receptor subunit genes (Saccone et al. 2009) and confirmed in meta-analysis (Saccone et al. 2010). Three of the four loci are captured by the SNPs in the

Chen et al. (2009) study. Number of cigarettes smoked is a good indicator of ND, and the main contributing factor to symptom count scores for ND using the FTND scale (Fagerström and Schneider 1989; Heatherton et al. 1991).

The number of studies of relatives (sib-pair, twin & family studies) for which not only phenotypic but also genotypic data are available is rapidly increasing. However, in most cases not all relatives will be phenotyped and genotyped. Visscher and Duffy (2006) noted that the statistical power of association studies of quantitative traits can be increased by including ungenotyped relatives with phenotypes in the analyses. Furthermore, little power is lost when using relatives rather than unrelated individuals in GWAS studies (Visscher et al. 2008). Statistical models for data on relatives have been extended to include measured genotypes (van den Oord and Snieder 2002). However most applications are limited to analyzing relatives which have been both phenotyped and genotyped, as the models are conditioned on measured genotypes.

In this report, we incorporate association tests within the traditional twin modeling framework, using a mixture method to estimate the effects of measured genes even when not all phenotyped relatives are genotyped. This approach allows us to quantify the contributions of specific variants as well as background genetic and environmental contributions to variation in ND. In addition, we test whether including phenotypic information from ungenotyped twins influences the precision and magnitude of the estimates of the allelic effect. This twin-association model is applied to data on a large sample of adult twins from the Virginia Twin Registry.

Methods

Subjects

Participants in the present investigation were drawn from two longitudinal studies conducted in a similar manner by the same research group, combined in the Virginia Adult Twin Study of Psychiatric and Substance Use Disorders (VATSPSUD). Both investigations were reviewed by ethical review boards and all participants provided written informed consent (or verbal consent for telephone interviews) before participation. Each sample was ascertained from the population-based Virginia Twin Registry, now the Mid-Atlantic Twin Registry. The first study was of female–female twin pairs (FF) and the second of male-male and male-female twin pairs (MMMF). All twins were Caucasian. These studies are described elsewhere (Kendler and Prescott 2006). Zygosity determination was based on questionnaire responses which were validated against DNA polymorphisms (Kendler and Prescott 2006). Phenotypic data from the third interview wave of the FF study (1992–1995) on 1,846 individuals were used in this report (88% of the wave one sample). The mean age of the female twins was 35.1 years (SD 7.5). These were combined with data from the second interview wave of the MMMF study (1994–1998), which included 4,959 twins (82.6% of the wave one sample) with a mean age of 37.0 (SD 9.1) years. Genotyping data was generated by three projects, each selecting a subset of individuals from the VATSPSUD studies (see Chen et al. 2009 for details). The first of the three panels is the Virginia Study of Nicotine Dependence (VAND) including never smokers, individual twins with low ND and with high ND (defined as FTQ score ≥ 7). The second panel is the Virginia Study of Anxiety and Depression (VAANX) including only regular smokers ($N = 815$ for first two panels). The third panel ($N = 1,128$) included almost all regular smokers for whom DNA was available and who were not included in the other two panels. As a result, individuals for whom genotypes are available are not a random sample of the VATSPSUD subjects. The panel used in this report included an additional 378 individuals who were co-twins of either of the two previous panels resulting in a total of 2,321 genotyped individuals.

Measures

The phenotypic data analyzed here were collected as part of a 1–3 h personal interview. The interviews for both the FF and MMMF studies were highly homologous. Buccal swaps were collected for DNA typing.

Phenotyping—In the MMMF study, all common forms of tobacco self-administration (cigarettes, cigars, pipe tobacco, chewing tobacco, and snuff) were assessed, whereas FF study participants were only asked about cigarettes. To equilibrate these data, we were forced to assume that the 1,846 FF study participants had an extremely low prevalence of non-cigarette forms of tobacco use. This assumption was strongly supported by data from the MMMF study in which none of 1,195 women from opposite sex twin pairs reported the use of non-cigarette forms of tobacco use.

For the purposes of this report, we focused on five tobacco related variables. *Tobacco initiation* (TI) was coded 0/1 and defined according to the responses to the questions, “Have you ever smoked cigarettes?” and the follow-up question “Not even once?”. *Regular smoking* (RS) was coded 0/1 and defined as the use of an average of at least seven cigarettes per week for a minimum of 4 weeks. Individuals who met criteria for RS were given a modified version of the Fagerström Tolerance Questionnaire (FTQ). The Fagerström Test for Nicotine Dependence (FTND) score was calculated during the period of lifetime maximal cigarette use (Fagerström 1978; Fagerström and Schneider 1989; Heatherton et al. 1991). Adjustments of scoring of non-cigarette tobacco use are described in detail elsewhere (Maes et al. 2004). In brief, most of the FTQ items adapted readily to all forms of tobacco use, except two which required modification (number of cigarettes and inhale). Both the symptom count of the six items (FTND) as well as dichotomous variable created based on having a FTND score greater or equal to 7 (ND) were used. *Current smoking* (CS) was coded 0/1 and defined according to the responses to the question “Do you currently smoke cigarettes?”.

Genotyping—DNA was extracted from buccal swaps and genotyping was performed with the Taqman genotyping method. A detailed description can be found in Chen et al. (2009). Markers were selected to cover the CHRNA5, CHRNA3 and CHRNB4 genes based on the Caucasian dataset of the HapMap project and positive association reports in the literature, which resulted in seven SNPs covering the 55 kb genomic distance: rs16969968, rs1051730, rs684513, rs578776, rs2869546, rs6495308 and rs8192475. Marker characteristics, including gene, sample allele frequency and HWE, are reported in Chen et al. (2009). A recent study found evidence for at least four statistically independent loci in the CHRNA5/CHRNA3/CHRNB4 gene cluster (Saccone et al. 2009). The seven SNPs which were genotyped for the Chen et al. paper tag Locus 1 with rs16969968 and rs1051730, Locus 2 with rs578776, and Locus 4 with rs8192475, but not Locus 3.

Statistical analyses

Structural equation modeling was used to model jointly the effects of quantitative trait loci (QTL) and background additive genetic, shared and within-family environment factors (Neale and Cardon 1992). Between-family (or shared) environmental effects make family members relatively more similar, whereas within-family (or specific) environmental factors are unique to individuals within a family and contribute to differences between family members.

The allelic effects of SNPs were specified as deviations from the population mean μ . An additive or dominant model can be used for the allelic effect. If we assume that the alleles at a particular locus are A and a, the additive deviation is aS and the dominance deviation is

dS , then the expected mean for AA homozygotes is $\mu + aS$, for aa homozygotes is $\mu - aS$, and for Aa heterozygotes is $\mu + dS$. MZ twin pairs have identical genotypes and thus belong to one of three classes (AA-AA, Aa-Aa or aa-aa). DZ twin pairs can have any of the 3×3 combinations of the three genotypes, resulting in nine classes. For MZ pairs where one twin is genotyped (which is typical for MZ pairs), the same genotype class can be assigned to the ungenotyped twin. For DZ pairs where one twin is genotyped, a probability is assigned of belonging to each of the three possible classes given the class of the genotyped twin, using the allele frequencies. For the ungenotyped pairs of twins, we can use information from the genotyped twins and the allele frequencies to assign probabilities of membership in any of the 3 (for MZ) or 9 (for DZ) possible genotype combinations of sibpairs.

SNP allele frequencies were computed using the total sample of genotyped individuals. Assuming Hardy–Weinberg equilibrium, the expected proportions in each of the genotypic categories of twin pairs can be calculated. If we assume that the alleles at a particular locus are A and a with probabilities p and q respectively, then MZ twin pairs belong to one of three classes with the following proportions: AA-AA with a probability of p^2 , Aa-Aa with a probability of $2pq$ and aa-aa with a probability of q^2 . DZ pairs belong to one of nine classes. The corresponding probabilities for DZ twin (or sibling) pairs (Lynch and Walsh 1998) are listed in Table 1.

Using a mixture distribution approach, the likelihood of the data is given by:

$$\sum_g^m w_j \{L|\text{genotype}=g\}$$

where the likelihood L , conditional on having genotype j , is weighted by the probability that the pair have genotype pairing g , summed over the $j = 1 \dots m$ genotypes pairings that are possible for the pair. Note that only the predicted means differ between the pair genotype combinations; residual genetic variance is assumed to be equal for all nine pairwise combinations. Individual means are modeled as a grand mean μ , an additive deviation aS and a dominance deviation dS , as shown in Table 1. In the case of MZ twins with one twin genotyped, this results in assigning genotypes of the genotyped twin to their ungenotyped cotwin. However, when neither member of an MZ pair was genotyped, the likelihood is simply the weighted sum of the likelihood of the three pairwise genotype combinations (AA-AA, Aa-Aa, aa-aa), where the weights are a function of the allele frequencies. The same applies in the case of DZ pairs concordant for not being genotyped, except that the likelihood is the weighted sum of all nine possible pair types. Scripts for the analyses were written for classic Mx (Neale et al. 2002) and OpenMx (Boker et al. 2010) and both are available upon request.

To evaluate the gain in statistical power by adding phenotypic data for ungenotyped twins, we compared the results under five different scenarios. We will use the following schematic in Table 2 to clarify the various combinations. Traditional association analyses use only individuals who have been both phenotyped and genotyped (GP). However, if both members of a twin pair are GP, typically only one twin, selected at random, is included in the analyses (Chen et al. 2009). In our analyses we included both twins as we model the residual variances and covariances within and across twins, which appropriately accounts for the non-independence of the twin data, while modeling the effect of the SNP on the means. Genotyped twins who lack phenotypic data (G), either as a result of missing values, do not contribute to the analyses, although their genotype data may be used to estimate allele frequencies in the population. Missing phenotype data may occur for either random or non-random reasons. For example, data may be missing on ND due to non-initiation. The reverse

situation, in which twins who were phenotyped but not genotyped (P) is the focus of potentially added information, due to the availability of genotyped co-twins and/or allele frequencies. This group is further subdivided in four groups (see Table 3) according to zygosity and availability of genotyped co-twin, to evaluate the relative contributions of the different pair combinations. Note that for one of the four groups, namely the discordantly genotyped MZ group, the genotypes of the ungenotyped MZ cotwins can be assigned with high probability.

Results

Descriptive statistics

The twin sample contained 6,805 individuals of whom 55.3% were male and 44.7% were female. The mean age of the sample was 36.2 (SD 8.6) years with a range of 20.4–59.5 years. Overall, individuals in the sample reported: lifetime SI 78.2%, lifetime RS 54.2%, and 19.8% met criteria for lifetime ND. Those who reported that they were regular smokers were given the Fagerström Questionnaire, from which the FTQ score and FTND score were calculated.

Univariate analysis of ND

Five scenarios, presented in Table 3, were used to evaluate the increase of power resulting from adding phenotypes for ungenotyped twins, and assigning or estimating their genotypes based on allele frequencies, and genotype of co-twin when available. In all cases, twins for whom both phenotype and genotype data are available are included (GP) regardless of whether they constitute a complete twin pair or not. This corresponds to traditional association analyses, except that in such analyses typically only one twin per pair is included. This scenario includes 2,321 individuals.

In the second scenario, genotypes for MZ twins whose cotwin was genotyped (P1) are estimated—which is equivalent to assigning the genotypes as they are known without error, assuming that the measured genotypes are correct, increasing the sample size to 2,610 individuals. Alternatively, genotypes for DZ twins whose cotwin was genotyped (P2), regardless of whether the co-twin was phenotyped, are estimated based on the probabilities of the allele frequencies in the three categories corresponding to the genotype of the other twin, resulting in 2,815 individuals for scenario III. In scenario IV, information from the discordantly genotyped pairs—those where one member of a pair is genotyped and the other is not—is used regardless of zygosity, thus combining scenarios II and III, resulting in 3,104 individuals with phenotype-genotype data.

In the most complete scenario (V, $N = 4,132$), genotypes for MZ and DZ twins where neither twin was genotyped are estimated based on the allele frequencies, resulting in three possibilities for MZ (P3) and nine possibilities for DZ twins (P4). The latter scenario uses all available phenotypic data. Note that when phenotypic data are not available (combinations 4, 8 and 12 in Table 2), even when genotypic data are available (G), those twins are not contributing to the analyses (except for the estimate of the allele frequencies) and were therefore omitted from Table 3.

A breakdown of the sample size by availability of genotype and phenotype data and by zygosity and twin order is presented in Table 4. Note that the phenotype of FTND is only available for regular smokers with non-regular smokers having missing values for FTND thereby reducing the overall sample sizes.

Variance decomposition

The results of the augmented univariate twin analyses for FTND showed that additive genetic factors accounted for slightly over half of the variance in liability to FTND (~55%, CI 37–63%), whereas the contribution of shared environmental factors was small (3%, CI 0–17%). Unique environmental factors made up the remaining part of the variance (42%, CI 37–47%). Of the seven SNPs tested, three reached significance at the 0.05 level (rs16969968 and rs1051730, which were previously identified in association analyses, and marker rs2869546 which is novel). Note that markers rs16969968 and rs1051730 are highly correlated ($D = 0.98$) and as such likely represent the same association signal. Even though these three SNPs reached genome-wide significance in previous analyses, including one on the same sample, they each account for less than one percent of the variance in liability to ND (0.35%, CI 0.03–0.99% for rs16969968, 0.30%, CI 0.02–0.99% for rs1051730 and 0.27%, CI 0.01–0.86% for rs2869546), or close to 1% of the genetic variance.

By comparing models that included the effect of the measured genotype with those in which it is excluded, we can evaluate whether the variance accounted for by the SNP forms part of the residual additive genetic variance or either of the environmental variance sources, as the estimates of the background additive genetic and shared environmental variance are not entirely independent and derive from the relative magnitude of the MZ and DZ covariance. The relative difference in unstandardized variance components is presented in Fig. 1 and Table 5, as well as the Chi-squares and P values associated with estimating the allelic effects of the respective SNPs. For the three significant SNPs, the variance accounted for by the SNP is mostly picked up by the additive genetic variance when the SNP is not modeled, although a small proportion of it is picked up by the shared environmental variance component, possibly as a consequence of population stratification or between-family effects of unknown origin. The difference in variance accounted for by specific environmental factors is very small, suggesting its estimation is not biased.

Significance of allelic effect

In comparing the different scenarios, including individual twins who were phenotyped but not genotyped and using information of the genotyped co-twin or allele frequencies, minor differences in the significance level of the SNP were noted between the scenarios for non-significant SNPs, as shown in Fig. 2. For the two alleles of largest effect—and previously identified—SNPs, adding phenotypic information of discordantly genotyped DZ twins slightly increased the significance level. However, adding this information for discordantly genotyped MZ twins surprisingly decreased the significance level, resulting in a lower significance level than just using twins who are both phenotyped and genotyped. For rs2869546, the significance level went up with adding ungenotyped but phenotyped individuals regardless of zygosity, resulting in a pattern of increasing Chi-squares with increasing sample size.

Association with other smoking related phenotypes

We further evaluated whether these replicated SNPs associated with FTND also conferred risk for other smoking related phenotypes. Of interest, given the correlated liabilities of TI, RS and ND, was whether SNPs in the CHRNA receptor gene also contribute to variability in TI and RS. We also included CS, a dichotomous ND measure (based on FTND 7) and the individual items comprising the FTND scale: number of cigarettes smoked (Cigs), how soon after waking do you smoke your first cigarette (Wake), refrain from smoking where it is prohibited (Refrain), smoking when ill in bed (Ill), smoking more in the morning (Morning), first cigarette of day most satisfying (First).

The chi-square test statistics for testing the allelic effect are presented in Fig. 3. Note that for these analyses, we used the mixture approach and thus included all available phenotypes. As a result, sample sizes varied by phenotype, as more individuals had a non-missing value for TI and RS than for FTND which was only obtained in individuals who have smoked regularly. The three SNPs identified for FTND did not contribute significantly to variability in TI or RS, although some of the other SNPs appeared to be associated but their allelic effect was not statistically significant. The same three SNPs did contribute to risk for CS but to a lesser degree. Not surprisingly, the two items contributing the most information to the FTND score, Cigs and Wake, show interesting results. The two previously replicated SNPs seem to be most strongly associated with Cigs, while the third SNP (rs2869546) appear most strongly associated to Wake item, as well as with smoking when ill. No significant associations were found for Refrain, Morning and First.

Discussion

We attempted to simultaneously model the effect of a measured gene and background genetic and environmental contributions to ND and related phenotypes. To our knowledge, it is the first time that phenotypes of ungenotyped twins are included in such an approach using a mixture distribution. We evaluated whether this approach influenced the precision and power of the allelic effect in a population sample of Virginian twins.

Our analyses confirmed the significance of the association of two previously identified SNPs (rs16969968 and rs1051730, P values <0.01) with FTND and suggested one additional marker rs2869546 (P value <0.05). Each of these markers, tested individually, contributed less than 1% of the total variance, or around one percent of the genetic variance. As predicted, the amount of variance accounted for by specific environmental factors remained unchanged with or without estimating the allelic effect. The background genetic variance decreased slightly as expected as a result of including the measured gene, as did the shared environmental variance, albeit to a smaller extent. The latter may be due to the effects of population stratification, which is confounded with the effects of the shared environment in the classical twin study.

For the two replicated markers (rs16969968 and rs1051730) the statistical significance of the allelic effect increased when including phenotypes of ungenotyped DZ twins whose co-twin was genotyped. However, the significance of the effect decreased when including phenotypes of ungenotyped MZ twins with a genotyped co-twin. Likely, this could be due to the non-random sampling of part of the sample for the purposes of identifying loci related to ND, which would have selected the more highly dependent member of a twin pair for genotyping. For the third SNP (rs2869546), the statistical significance appeared to increase gradually as additional phenotypic data were included. For the remaining SNPs which were not significantly associated with ND, no clear trend of the significance of the allelic effect was observed, which is expected if the null hypothesis of no association is correct. As the selection of twins to be genotyped was not random in this sample, we cannot draw firm conclusions about the possible increase in power of detecting an allelic effect as a result of adding phenotyped relatives.

We also evaluated whether the seven markers in the *CHRNA5* and *CHRNA3* genes were significantly associated with other smoking related phenotypes. Of particular interest is whether the allelic effects are specific to ND (as measured by the FTND score) or whether they also contribute to the genetic variance of TI and RS. We did not find evidence for a significant association with the earlier stages of smoking. These results are consistent with the negative findings of the ENGAGE consortium for genomewide evidence for age at initiation (Thorgerirsson et al. 2010). In contrast, two studies reported an association between

independent SNPs in the CHRNA5/CHRNA3/CHRNA4 cluster and with age of initiation (Schlaepfer et al. 2007; Weiss et al. 2008). However, the SNPs associated with ND in this study also appeared to confer risk for CS, which is of course associated with FTND. When analyzing the specific symptoms of FTND, the two markers (rs16969968 and rs1051730) were primarily associated with number of cigarettes smoked daily, while the third marker (rs2869546) showed a moderate association with time to first cigarette after waking. These results are consistent with the fact that the two replicated markers were established in studies based on the number of cigarettes smoked (Tobacco Genetics Consortium 2010). These results are inconsistent with a common factor model for FTND which would predict, within statistical limits, that the association with risk factors for FTND should be equivalent to the degree to which the individual criterion loads on the common factor.

In summary, this study showed a significant contribution of the CHRNA receptor on different forms of tobacco dependence (FTND and CS). However, the genetic variants only accounted for a very small proportion of the total genetic variance for liability to ND. Furthermore, these replicated variants were not significantly associated with TI and RS. We also established that including phenotypic data on ungenotyped co-twins can improve the statistical power to detect the effects of genetic variants when they contribute to individual differences in the phenotype.

Limitations

This study should be interpreted in the context of five potential limitations. First, our sample was entirely Caucasian and we do not know whether a similar pattern would hold in other ethnic groups. Second, despite the large initial sample, power to detect genetic and environmental factors specific to ND (as opposed to common for TI and ND) is limited. However, the current results are consistent with previous studies in suggesting that there exist genetic factors specific to ND.

Third, twin resemblance for ND was predicted by frequency of adult contact which could be a violation of the equal environment assumption. However, that would assume that frequent contact 'causes' resemblance for ND which may be less likely than twins with dissimilar smoking habits choosing to have less close contact. The validity of the equal environment assumption was supported for most psychiatric disorders including substance use (Kendler and Gardner 1998).

Fourth, the subset of individuals who were genotyped was not a random sample of all twins participating in VATSPSUD. DNA was largely collected only in wave 2 of the MMMF study and wave 4 of the FF study, and therefore was not available on twins who did not participate in those waves or refused to donate DNA. Furthermore, since the aim of the previous genotyping efforts was to locate genes affecting ND, samples selected for high versus low ND were used to maximize power. As such, the gain in power from adding genotypic information for ungenotyped co-twins was limited. Also, only one of the previously genotyped samples included individuals who had initiated smoking but not progressed to regular smoking, thus leading to unequal frequencies of SI, RS and ND in the non-genotyped versus the genotyped sample (see Table 6) and a higher representation of individuals with higher FTND scores genotyped (see Fig. 4).

Fifth, the analyses presented in this paper did not include sex and age effects on the phenotype, as the focus was on the incorporation of measured genotype effects. Such extensions to the models, as well as more advanced multivariate modeling of the data, will be implemented in future applications.

Acknowledgments

This work was funded by National Institutes of Health (DA018673, CA085739, DA024304, DA024413, DA027070, DA025109, DA022989) and Virginia Tobacco Settlement Foundation grant 8520012). The first author is supported by the Massey Cancer Center

References

- 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; 13:368–373. [PubMed: 18227835]
- 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]
- Boker S, Neale MC, Maes HH, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J. OpenMx: an open source extended structural equation modeling framework. *Psychometrika*. 2010; 76:306–317.
- Chen X, Chen J, Williamson VS, An SS, Hetta 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:926–933. [PubMed: 19132693]
- Fagerström KO. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addict Behav*. 1978; 11:331–335.
- Fagerström KO, Schneider NG. Measuring nicotine dependence: a review of the Fagerström tolerance questionnaire. *J Behav Med*. 1989; 12:159–182. [PubMed: 2668531]
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. *British J Addic*. 1991; 86:1119–1127.
- Kendler KS, Gardner CO. Twin studies of adult psychiatric and substance dependence disorders: are they biased by differences in the environmental experiences of monozygotic and dizygotic twins in childhood and adolescence. *Psychol Med*. 1998; 28:825–833. [PubMed: 9723139]
- Kendler, KS.; Prescott, CA. Genes, environment, and psycho-pathology: understanding the causes of psychiatric and substance use disorders. Guilford Press; New York: 2006.
- Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, Berrettini W, Knouff CW, Yuan X, Waeber G, Vollenweider P, Preisig M, Wareham NJ, Zhao JH, Loos RJ, Barroso I, Khaw KT, Grundy S, Barter P, Mahley R, Kesaniemi A, McPherson R, Vincent JB, Strauss J, Kennedy JL, Farmer A, McGuffin P, Day R, Matthews K, Bakke P, Gulsvik A, Lucae S, Ising M, Brueckl T, Horstmann S, Wichmann HE, Rawal R, Dahmen N, Lamina C, Polasek O, Zgaga L, Huffman J, Campbell S, Kooner J, Chambers JC, Burnett MS, Devaney JM, Pichard AD, Kent KM, Satler L, Lindsay JM, Waksman R, Epstein S, Wilson JF, Wild SH, Campbell H, Vitart V, Reilly MP, Li M, Qu L, Wilensky R, Matthai W, Hakonarson HH, Rader DJ, Franke A, Wittig M, Schäfer A, Uda M, Terracciano A, Xiao X, Busonero F, Scheet P, Schlessinger D, Rujescu D, Abecasis GR, Grabe HJ, Teumer A, Petersmann A, John U, Rudan I, Hayward C, Wright AF, Kolcic I, Wright BJ, Thompson JR, Balmforth AJ, Hall AS, Samani NJ, Anderson CA, Ahmad T, Mathew CG, Parkes M, Satsangi J, Caulfield M, Munroe PB, Farrall M, Dominiczak A, Worthington J, Thomson W, Eyre S, Barton A, Wellcome Trust Case Control Consortium. Mooser V, Francks C, Marchini J. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet*. 2010; 42(5):436–440. [PubMed: 20418889]
- Lynch, M.; Walsh, B. Genetics and analysis of quantitative traits. Sinauer; Sunderland: 1998.
- Maes, HH.; Neale, MC. Genetic modeling of tobacco use behavior and trajectories. In: Swan, GE.; Baker, TB.; Chassin, L.; Conti, DV.; Lerman, C.; Perkins, KA., editors. NCI Tobacco Control Monograph Series 20: phenotypes and endophenotypes: foundations for genetic studies of nicotine

use and dependence. US Department of Health and Human Services; National Institutes of Health; Bethesda: 2009.

- Maes HH, Sullivan PF, Bulik CM, Neale MC, Prescott CA, Eaves LJ, Kendler KS. A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychol Med*. 2004; 34:1251–1261. [PubMed: 15697051]
- Neale, MC.; Cardon, LR. Methodology for genetic studies of twins and families. Kluwer Academic Publishers BV; Dordrecht; 1992.
- Neale, MC.; Boker, SM.; Xie, G.; Maes, HH. Mx: statistical modeling. 6th edn.. Department of Psychiatry; Richmond, VA: 2002. VCU Box 90012623298
- 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; 16:36–49. [PubMed: 17135278]
- 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: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) pii 1001053.
- 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*. 2007; 63(11):1039–1046. [PubMed: 18163978]
- Straub RE, Sullivan PF, Ma Y, Myakishev MV, Harris-Kerr C, Wormley B, Kadambi B, Sadek H, Silverman MA, Webb BT, Neale MC, Bulik CM, Joyce PR, Kendler KS. Susceptibility genes for nicotine dependence: a genome scan and followup in an independent sample suggest that regions on chromosomes 2, 4, 10, 16, 17 and 18 merit further study. *Mol Psychiatry*. 1999; 4:129–144. [PubMed: 10208445]
- Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Sulem P, Rafnar T, Esko T, Walter S, Gieger C, Rawal R, Mangino M, Prokopenko I, Mägi R, Keskitalo K, Gudjonsdottir IH, Gretarsdottir S, Stefansson H, Thompson JR, Aulchenko YS, Nelis M, Aben KK, den Heijer M, Dirksen A, Ashraf H, Soranzo N, Valdes AM, Steves C, Uitterlinden AG, Hofman A, Tönjes A, Kovacs P, Hottenga JJ, Willemsen G, Vogelzangs N, Döring A, Dahmen N, Nitz B, Pergadia ML, Saez B, De Diego V, Lezcano V, Garcia-Prats MD, Ripatti S, Perola M, Kettunen J, Hartikainen AL, Pouta A, Laitinen J, Isohanni M, Huei-Yi S, Allen M, Krestyaninova M, Hall AS, Jones GT, van Rij AM, Mueller T, Dieplinger B, Haltmayer M, Jonsson S, Matthiasson SE, Oskarsson H, Tyrfingsson T, Kiemeny LA, Mayordomo JI, Lindholt JS, Pedersen JH, Franklin WA, Wolf H, Montgomery GW, Heath AC, Martin NG, Madden PA, Giegling I, Rujescu D, Järvelin MR, Salomaa V, Stumvoll M, Spector TD, Wichmann HE, Metspalu A, Samani NJ, Penninx BW, Oostra BA, Boomsma DI, Tiemeier H, van Duijn CM, Kaprio J, Gulcher JR, ENGAGE Consortium. McCarthy MI, Peltonen L, Thorsteinsdottir U, Stefansson K. Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet*. 2010; 42(5): 448–453. [PubMed: 20418888]
- Tobacco Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010; 42(5):441–447. [PubMed: 20418890]

- US Department of Health and Human Services. Reducing the health consequences of smoking: 25 years of progress. Department of Health and Human Services; Public Health Service; Centers for Disease Control; Center for Chronic Disease Prevention and Health Promotion; Office on Smoking and Health; Rockville: 1989.
- van den Oord EJCG, Snieder H. Including measured genotypes in statistical models to study the interplay of multiple factors affecting complex traits. *Behav Genet.* 2002; 32:1–21. [PubMed: 11958539]
- Visscher PM, Duffy DL. The value of relatives with phenotypes but missing genotypes in association studies for quantitative traits. *Genet Epidemiol.* 2006; 30:30–36. [PubMed: 16355405]
- Visscher PM, Toby Andrew T, Nyholt DR. Genome-wide association studies of quantitative traits with related individuals: little (power) lost but much to be gained. *Eur J Human Genet.* 2008; 16:387–390. [PubMed: 18183040]
- 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]
- Wessel J, McDonald SM, Hinds DA, Stokowski RP, Javitz HS, Kennemer M, Krasnow R, Dirks W, Hardin J, Pitts SJ, Michel M, Jack L, Ballinger DG, McClure JB, Swan GE, Bergen AW. Resequencing of nicotinic acetylcholine receptor genes and association of common and rare variants with the Fagerström test for nicotine dependence. *Neuropsychopharmacology.* 2010; 35:2392–2402. [PubMed: 20736995]
- World Health Organization. Tobacco or Health: A Global Status Report. World Health Organization; Geneva: 1997.

Fig. 1.
Difference in variance components between modeling with or without allelic effects

Fig. 2.
Chi-squares associated with allelic effect of SNPs by scenario

Fig. 3.
Chi-squares for association of SNPs with smoking related phenotypes

Fig. 4.
Distribution of FTND scores in genotyped and non-genotyped samples

Table 1

Expected proportions and means in genotypic categories of twin pairs

Genotype		Expected proportion		Expected mean	
T1	T2	MZ	DZ	T1	T2
AA	AA	p^2	$p^4 + p^3q + (pq)^2/4$	$\mu + aS$	$\mu + aS$
AA	Aa	0	$p^3q + (pq)^2/4$	$\mu + aS$	$\mu + dS$
AA	aa	0	$(pq)^2/4$	$\mu + aS$	$\mu - aS$
Aa	AA	0	$p^3q + (pq)^2/4$	$\mu + dS$	$\mu + aS$
Aa	Aa	$2pq$	$p^3q + 3(pq)^2 + pq^3$	$\mu + dS$	$\mu + dS$
Aa	aa	0	$pq^3 + (pq)^2/4$	$\mu + dS$	$\mu - aS$
aa	AA	0	$(pq)^2/4$	$\mu - aS$	$\mu + aS$
aa	Aa	0	$pq^3 + (pq)^2/4$	$\mu - aS$	$\mu + dS$
aa	aa	q^2	$q^4 + pq^3 + (pq)^2/4$	$\mu - aS$	$\mu - aS$

A, a: alleles, p : probability of A, q : probability of a, μ : grand mean, aS : additive deviation, dS : dominance deviation

Table 2
Possible combinations of availability (genotyped/phenotyped) of twin data by zygosity (Color table online)

Combination	Twin Data Availability		Twin Zygosity			
	Genotyped	Phenotyped	MZ		DZ	
			twin 1	twin 2	twin 1	twin 2
1	both	both	GP	GP	GP	GP
2		twin 1	GP	G	GP	G
3		twin 2	G	GP	G	GP
4		neither	G	G	G	G
5	one	both	GP	P	GP	P
6		twin 1	GP		GP	
7		twin 2	G	P	G	P
8		neither	G		G	
9	neither	both	P	P	P	P
10		twin 1	P		P	
11		twin 2		P		P
12		neither				
Legend	GP	genotyped+phenotyped	G	genotyped	P	phenotyped

MZ monozygotic, DZ dizygotic

Table 3

Five scenarios (I–V) to evaluate power by adding information from ungenotyped twins (Color table online)

	I				II				III				IV				V			
	MZ		DZ		MZ		DZ		MZ		DZ		MZ		DZ		MZ		DZ	
	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
GG	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP
	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP
	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP	GP
Gt1	G	GP	G	GP	G	GP	G	GP	G	GP	G	GP	G	GP	G	GP	G	GP	G	GP
	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P
	GP		GP		GP		GP		GP		GP		GP		GP		GP		GP	
Pt1	G	P	G	P	G	P	G	P	G	P	G	P	G	P	G	P	G	P	G	P
	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P
	GP		GP		GP		GP		GP		GP		GP		GP		GP		GP	
Pt2	G	P	G	P	G	P	G	P	G	P	G	P	G	P	G	P	G	P	G	P
	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P	GP	P
	GP		GP		GP		GP		GP		GP		GP		GP		GP		GP	
NG	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
	P		P		P		P		P		P		P		P		P		P	
	P		P		P		P		P		P		P		P		P		P	
Pt2	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P	P
Legend	N=2321				N=2610				N=2815				N=3104				N=4153			
	G				+ assigned genotypes (MZ)				+ assigned genotypes (MZ)				+ mixture of 3 genotypes (MZ)				+ mixture of 3 genotypes (MZ)			
	P				+ mixture of 3 genotypes (DZ)				+ mixture of 3 genotypes (DZ)				+ mixture of 9 genotypes (DZ)				+ mixture of 9 genotypes (DZ)			

MZ monozygotic, DZ dizygotic, t1 twin 1, t2 twin 2, GG both members of twin pair genotyped, PP both members of twin pair phenotyped, Pt1 twin 1 phenotyped, Pt2 twin 2 phenotyped, Gt1 twin 1 genotyped (patterns are similar if twin 2 but not twin 1 was genotyped), GP genotyped + phenotyped, NG neither member of twin pair genotyped

Table 4
Sample sizes for combinations of genotyped and phenotyped twins (Color table online)

Twin Data Availability		MZ twins			DZ twins		
Genotyped	Phenotyped	fams	t1	t2	fams	t1	t2
both	both	109	109	109	158	158	158
	twin 1	4	4		54	54	
	twin 2	24		24	38		38
	neither	15			13		
one	both	267	267	267	440	440	440
	twin 1	227	227		733	733	
	twin 2	22		22	54		54
	neither	226			206		
neither	both	103	103	103	100	100	100
	twin 1	98	98		253	253	
	twin 2	93		93	199		199
	neither	467			647		
	totals	1655		1333	2895		2528
							5727

Twin pairs are classified as concordant for being genotyped, discordant or concordant for not being genotyped. Similar classifications are made to distinguish pairs who are concordant for being phenotyped, discordant or concordant for not being phenotyped

MZ monozygotic, *DZ* dizygotic, *t1* twin 1, *t2* twin 2, *fams* families

Table 5

Difference in variance components and Chi-square associated with allelic effects

	Δa	Δc	Δe	Δg	ΔV	χ^2	P
rs684513	-0.0013	-0.0018	0.0007	0.0028	0.0003	2.030	0.154
rs16969968	-0.0062	-0.0031	0.0007	0.0107	0.0021	7.999	0.005
rs578776	0.0023	0	0.0013	0.0002	0.0038	0.169	0.681
rs1051730	-0.0049	-0.0023	-0.0017	0.0093	0.0006	6.915	0.009
rs2869546	-0.0082	-0.0005	0.0013	0.0084	0.0001	6.405	0.011
rs6495308	-0.0009	-0.0017	0.0004	0.0029	0.0008	2.065	0.151
rs8192475	-0.0011	-0.0011	0.0005	0.0025	0.0009	1.708	0.191

Δa : difference in estimate of additive genetic variance; Δc : difference in estimate of shared environmental variance; Δe : difference in estimate of unique environmental variance; Δg : difference in estimate of variance due to the SNP; ΔV : difference in estimate of total variance; χ^2 : difference in chi-square between model including the effect of the SNP versus model not including the effect of the SNP; P: P value associated with likelihood ratio test

Table 6

Frequency (and percentage) of response by genotype status for key stages of smoking

	<u>NO</u>		<u>YES</u>	
	Non-genotyped	Genotyped	Non-genotyped	Genotyped
Initiation	1,294 (28)	397 (13)	3,345 (72)	2,607 (87)
Regular smoking	1,601 (46)	279 (11)	1,863 (54)	2,329 (89)
ND (dichotomous)	1,416 (77)	1,668 (72)	416 (23)	653 (28)