Association Study of CHRM2 Polymorphisms with Substance-Use Pathology and Personality Traits

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Robert M. Kirkpatrick

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Adviser: Matt McGue

January 2011

Acknowledgements

The author extends his special thanks to Matt McGue, William G. Iacono, Michael B. Miller, and Brian M. Hicks.

Abstract

A growing literature exists on the association of CHRM2 (the gene coding for cholinergic muscarinic receptor 2) with alcohol dependence (Wang et al., 2004), with the broader phenotype of substance dependence in general (Luo et al., 2005; Dick, Agrawal, et al., 2007), and with the still-broader phenotype of externalizing psychopathology (Dick et al., 2008). Additionally, one study has found association between CHRM2 and the personality traits of Agreeableness and Conscientiousness; its authors suggest that personality may substantially mediate the heritability of substance dependence (Luo et al., 2007). Guided by the relevant literature, which is reviewed at length, the present study investigates: (1) the association of CHRM2 and risk of substance use disorders; (2) the association of CHRM2 with personality characteristics; and, conditional on replication of these main effects, (3) the extent to which personality mediates CHRM2's influence on substance use disorders. We use data from genotyped participants in two longitudinal studies, one of twins and their families and one of adopted siblings and their families. We use Raw Maximum Likelihood in Mx (see Neale et al., 2003) to examine the association of 4 CHRM2 SNPs with personality traits and with composite measures of disordered substance use, while taking into account the phenotypic covariance in different types of families. Our results provide no clear evidence of association of CHRM2 polymorphisms with broad personality traits or substance-abuse pathology. However, the validity of our results is considerably limited by the non-multivariate-normal distribution of the substance-use pathology variable, incompleteness of the available data, use of self-reported ethnicity instead of genomically-determined ancestry, and sparse coverage of the CHRM2 gene.

Table of Contents

List of Tables

Background

Despite requiring environmental exposure to some substance of abuse or addiction, it is clear that substance-use pathology has an appreciable hereditary component (Rose, Broms, Korhonen, Dick, & Kaprio, 2009; Dick, Prescott, & McGue, 2009). Personality traits, like substance-use pathology, are also partly heritable—reviews (e.g., Jang & Yamagata, 2009; Bouchard & McGue, 2003) indicate that that heritability estimates range from 0.30 to 0.60, with most between 0.40 and 0.50. Investigators are currently searching for specific genetic variants underlying individual variation in behavioral traits. Some existing genetic and neurobiological evidence, to be reviewed subsequently, implicates the involvement of cholinergic pathways in both substance-use pathology and personality.

Genome-wide linkage studies have implicated a disease-risk locus for alcohol dependence on chromosome 7q (Reich et al., 1998; Foroud et al., 2000). CHRM2, located on chromosome 7q, codes for the muscarinic acetylcholine receptor-2, a G-protein-coupled receptor typically functioning as an inhibitory pre-synaptic autoreceptor (Peralta et al., 1987; Zhuo, Fryer, & Jacoby, 2001; Sofuoglu & Mooney, 2009). CHRM2 serves as a plausible candidate gene for substance-use disorders in light of emerging evidence of the role of cholinergic systems in connection with dependence on alcohol and other drugs. CHRM2 codes the protein of one of the muscarinic receptors; it goes without saying that the *other* class of cholinergic receptor, the nicotinic receptors, is intimately involved in the abuse and dependence of its eponymous agonist (Markou, 2008, provides a recent review of the neurobiology of nicotine dependence). However,

recent literature suggests involvement of the nicotinic receptors in the addiction to other drugs, including alcohol, opiates, and stimulants (Rahman, Lopez-Hernandez, Corrigall, & Papke, 2008), and has characterized acetylcholine's involvement in the addiction to stimulants (Sofuoglu & Mooney, 2009), particularly cocaine (Williams & Adinoff, 2008, provide a thorough review). Striatal acetylcholine interacts with the dopaminergic systems underlying the reward mechanisms common to all drugs of abuse: activation of certain muscarinic subtypes modulates striatal dopamine release (Zhang, Yamada, Gomeza, Basile, & Wess, 2002), and in turn, striatal acetylcholine release is upregulated by activation of the D1 dopamine receptor and downregulated by activation of the D2 receptor (reviewed in Williams & Adinoff). In addition to the reinforcing effects of drug abuse, Hoebel and colleagues have reported that striatal acetylcholine is involved in the aversive effects of drug withdrawal, whereby it may be important for the development of drug addiction. Specifically, they observed high striatal acetylcholine levels relative to dopamine levels during antagonist-induced withdrawal in rats treated with morphine (Rada, Pothos, Mark, & Hoebel, 1991), nicotine (Rada, Jensen, & Hoebel, 2001), ethanol (Rada, Jensen, Lewis, & Hoebel, 2004, inducing withdrawal via blockade of endogenous opioid receptors thought to mediate the reward response to ethanol), and benzodiazepine (Rada & Hoebel, 2005). Hoebel, Avena, and Rada (2007) theorize that the balance of dopamine with acetylcholine is important to approach and avoidance motivation in general. They propose that whereas dopamine largely produces approach-motivation, acetylcholine produces avoidance-motivation, in the form of satiety when dopamine levels are high, and in the form of an aversive state of anxiety, depression, or withdrawal

when dopamine levels are low[1]. Finally, the involvement of cholinergic systems in learning and cognition helps to explain the more direct, genetic evidence connecting the CHRM2 gene with alcohol dependence, which we discuss in the subsequent paragraphs.

CHRM2 variants have exhibited allelic association with substance-use pathology in previous studies. The first allelic association study of CHRM2 and substance-use pathology (Wang et al., 2004) was motivated by an earlier finding from a genome-wide linkage scan for the theta- and delta-band event-related oscillations underlying the P300 event-related potential (ERP) from a visual oddball task, by Jones et al. (2004; also see Jones et al., 2006). The P300 has been of interest in alcoholism research since the seminal study by Begleiter, Porjesz, Bihari, and Kissin (1984), wherein P300 amplitude reduction distinguished adolescent boys with a family history of alcoholism from controls. The linkage scan by Jones et al. identified a significant (LOD = 3.5) linkage signal for frontal-electrode, target-case theta oscillation between the markers D7S1837 and D7S509—the locus of CHRM2. The observed linkage signal for the theta-band oscillation is especially interesting because of this oscillation's cognitive correlates (e.g., Doppelmayr, Klimesch, Schwaiger, Auinger, & Winkler, 1998; Klimesch et al., 2001) and because of known involvement of cholinergic systems in cognition (Everitt & Robbins, 1997; Baxter & Chiba, 1999). The M2 receptors, specifically, likely play a role in learning and memory by inhibiting long-term potentiation (Calabresi, Centonze, Gubellini, Pisani, & Bernardi, 1998). Long-term potentiation, in turn, may be one of the mechanisms necessary for the development of addiction (Wolf, 2003).

---

[1]The details of striatal dopamine-acetylcholine interaction are actually more complex, though they do not concern us here; see Williams and Adinoff (2008) and Cragg (2006) for review.

Using a sample of alcohol-dependent probands and their families from COGA (the Collaborative Study on the Genetics of Alcoholism), Wang et al. (2004) conducted a linkage analysis for alcohol dependence and observed a peak signal on chromosome 7 some distance away from CHRM2. However, they followed up with a Pedigree Disequilibrium Test using 11 CHRM2 single-nucleotide polymorphisms (SNPs) and observed highly significant association for three SNPs, two in the 5th intron and one in the 4th intron (all $p \leq 0.007$), with a few nominally significant results for other SNPs. Haplotype analysis indicated that the most common haplotype of the three 4th-intron SNPs had a protective effect against risk of alcohol dependence. Interestingly, the complement of this protective haplotype was over-transmitted to alcohol-dependence cases with comorbid major depression. Also, a rare haplotype was over-transmitted to alcohol-dependence-only cases.

An independent team of researchers (Luo et al., 2005) obtained results consistent with those of Wang et al. in a case-control association study of alcohol dependence and drug dependence (defined as dependence on either cocaine or opioids), observing nominally significant association of both alleles and genotypes, at six CHRM2 SNPs, with both of the substance-dependence phenotypes, but only in their African-American subsample. Luo et al. conclude that their "regression analysis demonstrated that [certain] alleles, genotypes, haplotypes and diplotypes at the CHRM2 locus affected risk for [alcohol dependence and drug dependence]" (p. 2424). Their results certainly provide suggestive, if not clear, evidence of an association between CHRM2 variation and

substance dependence (though their post-hoc stepwise regression analyses using haplotypes and diplotypes is not easy to interpret).

COGA investigators re-examined CHRM2's association with alcohol dependence in a study by Dick, Agrawal, et al. (2007), using a sample and genetic analyses comparable to those used by Wang et al. (2004). However, Dick, Agrawal, et al. (2007) used more genetic markers (27 CHRM2 SNPs) and compared results for alcohol-dependence cases *with* comorbid illicit-drug dependence to those for alcohol-dependence cases *without* comorbid illicit-drug dependence (drug dependence diagnoses in the sample included dependence on marijuana, cocaine, opioids, stimulants, and sedatives). Among cases *with* comorbid drug dependence, these authors observed nominally significant association for nine SNPs. However, they did not correct for multiple significance testing. Assuming 27 independent significance tests (which is overly conservative, as it ignores linkage disequilibrium among the SNPs) and a study-wide alpha of 0.05, a Bonferroni correction would leave only one of these nine results significant (PDT "sum" statistic for rs978437, $p = 0.001$). The statistics for rs324640 and rs324650, which were implicated by Wang et al. (2004), were not nominally significant in this study, though they approached nominal significance (i.e., $p < 0.10$). This discrepancy is presumably due to sampling error and the fact that different criteria for alcohol dependence were used in these two studies (DSM-IIIR and Feighner versus DSM-IV). Nonetheless, the most interesting finding from this study is quite clear: all of the (nominally) significant results were obtained with alcohol-dependence cases *with* comorbid illicit-drug dependence, and no SNP showed significant association with

alcohol-dependence among cases *without* illicit-drug dependence. Additionally, cases with comorbid illicit-drug dependence differed from cases without in severity of pathological alcohol consumption (e.g., age of onset, largest number of drinks in 24-hour period), level of novelty-seeking, and level of other comorbid psychopathology. Notably, COGA investigators had previously discovered that the observed association of GABRA2 with alcohol dependence in their sample was entirely due to those alcohol-dependence cases with comorbid drug dependence (Agrawal et al., 2006). Dick, Agrawal, et al. interpret their findings as evidence that alcohol dependence with comorbid drug dependence is a more severe manifestation of the disorder, under greater genetic influence.

Different substance use disorders are comorbid with one another, and with conduct disorder and with antisocial behavior in adulthood (Krueger & Markon, 2006), which together fall under the general label of externalizing psychopathology. Evidence from twin studies suggests that comorbid risk for substance use pathology is largely genetic (Young, Stallings, Corley, Krauter, & Hewitt, 2006), and that comorbid risk for substance use pathology and behavioral disinhibition is best modeled as stemming from a single, highly heritable externalizing factor (Krueger et al., 2002). Dick et al. (2008) attempted to identify a gene associated with general risk for externalizing psychopathology, using a sample of COGA participants. They used 27 SNPs in and around CHRM2 in an association analysis of the following phenotypes: DSM-IIIR symptom counts for antisocial personality disorder, conduct disorder (in childhood), and drug dependence; DSM-IV symptom counts for alcohol dependence; two personality

traits, Novelty-Seeking and Sensation-Seeking; and scores on the first principal component (which accounted for over half the phenotypic variance) extracted from the foregoing. All phenotypes were nominally associated with at least one SNP, with the principal-component scores exhibiting strongest association ($p < 0.01$ for six SNPs), though there was no correction for multiple significance testing. The strongest evidence of association came from those SNPs in an LD block spanning portions of the 3rd and 4th introns. Dick et al. also conducted a genome-wide linkage scan, hypothesizing that a gene associated with general risk for externalizing psychopathology would manifest a strong linkage signal for the principal-component scores at its locus, and modest linkage signals for the specific phenotypes in the same vicinity. Such a locus was identified on the long arm of chromosome 7, with a peak lod value of 1.57, for principal-component scores. Notably, analysis of phenotypes residualized for the first principal component yielded evidence neither of linkage nor association.

First and foremost, the report by Dick et al. (2008) provides plausible evidence that CHRM2 is a gene contributing to general risk for externalizing psychopathology. Aside from this, it is suggestive of a contribution by CHRM2 to individual variation in personality—both Novelty Seeking and Sensation Seeking were significantly ($p < 0.01$) associated with two SNPs each. Personality traits related to behavioral disinhibition—such as Novelty Seeking (Young et al., 2000) and (reversed) Constraint (Conway, Swendsen, Rounsaville, & Merikangas, 2002; Elkins, King, McGue, & Iacono, 2006)—are strongly correlated with externalizing psychopathology and form an integral part of the externalizing spectrum (Krueger et al., 2002).

The first report of association between CHRM2 and personality of which we are aware is due to Luo et al. (2007), which has largely motivated the present study. These authors analyzed the association of five personality variables (as measured by the NEO-FFI; Costa & McCrae, 1997) and six SNPs within and flanking CHRM2, using a mixed sample (N = 514) of both substance-dependence cases and healthy controls, and both European-Americans and African-Americans. Age, disease (alcohol or drug dependence) affection, ancestry, and sex were entered into the analyses as covariates. Again, the results of Luo et al. are not easily interpreted due to their use of stepwise MANCOVA and ANCOVA (conducted subsequent to exploratory *t*-tests and ANOVAs), and the fact that the five personality variables exhibited highly significant pairwise intercorrelation (all $p < 0.001$), but they report significant ($p < 0.01$) association results for the traits of Agreeableness and Conscientiousness. Luo et al. speculate that "personality traits might substantially underlie the heritable component of [substance dependence]" (p. 1557).

The dominant framework in trait-based theories of personality is the Five-Factor Model, which, as its name implies, posits that there exist five fundamental traits on which individuals differ with respect to personality. One alternative framework is that of Tellegen, which posits a three-factor model of personality. Tellegen's model is particularly attractive due to the thorough, exploratory, iterative method by which evolved both the theory and its operationalization, the Multidimensional Personality Questionnaire (Tellegen, 1982; Tellegen & Waller, 2008). The MPQ consists of eleven primary scales and three orthogonal factor scales—Positive Emotionality (PEM), Negative Emotionality (NEM), and Constraint (CON)—such that each factor scale is a

weighted linear composite of the primary scales. Each primary scale has a clear major loading onto one of the factors (except for one primary scale, Absorption). Replication of CHRM2's association with personality variables measured with a different personality instrument would constitute "constructive replication" (Lykken, 1968) and provide robust evidence that CHRM2 polymorphisms truly account for variance in personality traits and not merely variance specific to a particular personality measure.

Of course, "constructive replication" requires that the replication involve the same theoretical constructs as the original finding. Evidence concerning the structural relationship between the scales of the MPQ and NEO reveals their considerable overlapping variance. Data from two moderately large samples (Tellegen & Waller, 2008; Church, 1994) reveal substantial correlations between MPQ and NEO scales. These data enable investigators to formulate reasonable predictions of what results would be obtained in an association study of CHRM2 with MPQ data, conditional on a true association with personality traits. For example, NEO Conscientiousness and MPQ (factor) Constraint are substantially correlated ($r = 0.46$ in Church, 1994), so one might reasonably hypothesize that CHRM2 would have significant association with MPQ Constraint. Still more compelling is the excellent paper by Markon, Krueger, and Watson (2005), in which was reported a hierarchical factor analysis applied both to a meta-analytic dataset and a new sample of undergraduates measured with four different personality questionnaires. Markon et al. conclude from this analysis that there exists a hierarchical structure of personality traits, so that a five-, four-, three-, or two-factor solution is defensible, depending on the degree of abstraction from the observed data.

Because Markon et al. used orthogonal rotation in their exploratory factor analysis, the correlations between personality scales can be estimated by the summed products of their loadings on each factor.

Somewhat complicating matters is the fact that Markon et al. (2005) and Tellegen and Waller (2008) do not report such correlations or common-factor loadings using MPQ factor scales, and that Church (1994), who does, found that the two NEO scales for which Luo et al. (2007) obtained significant allelic-association results have their strongest correlations with MPQ *primary* scales rather than factor scales. NEO Agreeableness correlated -0.48 with MPQ Aggression, compared with -0.40 for MPQ NEM, the MPQ factor on which Aggression has its primary loading. Similarly, NEO Conscientiousness correlated 0.56 with MPQ Control and 0.48 with MPQ Achievement, compared with 0.46 with MPQ CON and 0.31 with MPQ PEM. These correlations, combined with the relative lack of information concerning the structural relationships of MPQ *factors* with NEO Agreeableness and Conscientiousness, suggest that an attempt to replicate Luo et al. (2007) using MPQ data would best be accomplished via allelic association analysis using MPQ primary scales. Using this approach, we would use MPQ Aggression as analog for NEO Agreeableness, and MPQ Control as analog for NEO Conscientiousness, because there is remarkable consistency between the reports by Church, Tellegen and Waller, and Markon et al. in identifying these as the strongest correlates of the NEO scales of interest. Church's data are summarized above. Tellegen and Waller (2008, p. 281) obtained a correlation of -0.50 between NEO Agreeableness and MPQ Aggression, whereas no other MPQ primary scales had a correlation greater than 0.30 in magnitude with

Agreeableness. The two correlations between NEO Conscientiousness and MPQ scales that exceeded 0.30 in magnitude were with MPQ Achievement (0.42) and MPQ Control (0.52). From the meta-analytic exploratory factor models reported by Markon et al. (2005, which used the MPQ, NEO, and three other personality instruments), MPQ Aggression was the MPQ primary scale exhibiting the highest correlation (as estimated from the reported factor loadings) with NEO Agreeableness, in three of the four such models. The exception was the two-factor model, which by necessity accounted for less common variance than any the others, and which Markon et al. interpret to be at the greatest level of abstract remoteness from the observed data. Here, MPQ Stress Reactivity had the highest such correlation. On the other hand, the situation is less clear for NEO Conscientiousness, which had its largest estimated MPQ-scale correlation with Control in the four- and five-factor models, with Aggression in the three-factor model, and with Stress Reactivity in the two-factor model. While not very helpful to our present purpose, this pattern neatly coincides with the hierarchical interpretation Markon et al (p. 148) offer.

However, we decided to analyze the three MPQ factor scales, for the following reasons. First, we decided to include all three MPQ factor scales in the analysis, because Church (1994) reports at least one correlation greater than 0.30 in magnitude between each of them and one of the NEO factors of interest (see preceding paragraph). Church also reports that NEO Agreeableness and Conscientiousness both have non-trivial correlations with MPQ primary scales that have their primary loadings on different factors. For example, NEO Conscientiousness correlated 0.48 with MPQ Achievement

(primary loading on PEM) and 0.56 with MPQ Control (primary loading on CON). Second, as opposed to an analysis of eleven intercorrelated primary scales, Type I error correction would be easier in an analysis of three scales that are mostly uncorrelated with each other (the factor scales were constructed with orthogonal rotation, and correlate pairwise approximately $r = 0.1$ in absolute magnitude in the present sample). Third, the factor scales represent more general traits than the primary scales. All MPQ items contribute to the factor scores, so the factors are measured with less measurement error than the primaries, and an analysis using the factors represents, at least indirectly, an analysis using all of the primaries. Finally, significant or suggestive results for a factor can always be followed up with analysis of its primary scales, and we judged the results of such follow-ups to be easier to interpret than follow-ups of factors suggested by the primaries.

Besides the observed CHRM2 association with Agreeableness and Conscientiousness due to Luo et al. (2007), Dick et al. (2008) found significant association between CHRM2 SNPs and the personality variables of Novelty-Seeking and Sensation-Seeking. However, it is not completely clear what might constitute constructive replication of this result using MPQ data. In the meta-analytic exploratory factor analyses by Markon et al. (2005), Novelty-Seeking exhibits nontrivial loadings on more than one common factor in all four models examined, leaving its structural relationships with MPQ primary scales rather ambiguous, and suggesting that it is a somewhat heterogeneous construct. From the Big-Five standpoint, reversed Conscientiousness appears to be its closest equivalent, but it has also shown association

with Extraversion, Openness, and reversed Agreeableness (DeYoung & Gray, 2009). We are not aware of any studies examining the structural relationships between Sensation-Seeking and MPQ variables, though some studies (e.g., Zuckerman, Kuhlman, Joireman, Teta, & Kraft, 1993) indicate that reversed Agreeableness and reversed Conscientiousness are its closest Big-Five equivalents. Others (reviewed by DeYoung & Gray, 2009) observe nontrivial correlations between Sensation-Seeking and Extraversion, Openness, and reversed Agreeableness, highlighting the theoretical similarity between Sensation-Seeking and Novelty-Seeking.

On the other hand, Dick et al. (2008) do not emphasize the specific personality traits of Novelty-Seeking and Sensation-Seeking, but upon the more general trait of externalizing psychopathology. Their study highlights the strengths of using a composite and *dimensional,* rather than categorical, model of externalizing pathology, and likewise, we employ a dimensional, quasi-continuous model of substance-use pathology, representing variance common to multiple indicators thereof. The present study investigates: (1) the association of CHRM2 and risk of substance-use pathology; (2) the association of CHRM2 with personality characteristics; and, conditional on replication of these main effects, (3) the extent to which personality mediates CHRM2's influence on substance-use pathology.

Method

Sample

Tissue samples were collected from participants from the Minnesota Twin Family Study ("MTFS;" Iacono, Carlson, Taylor, Elkins, & McGue, 1999; Iacono & McGue, 2002), and the Sibling Interaction and Behavior Study ("SIBS;" McGue et al., 2007). We (Kirkpatrick, McGue, & Iacono, 2009) have previously described MTFS as "a longitudinal study of a community-based sample of same-sex twins (N = 3779, including five sets of triplets), born between 1972 and 1994 in the State of Minnesota, and their parents," and SIBS as "an adoption study of sibling pairs (N = 1232) and their parents [comprising a community-based sample of] families where both siblings are adopted, where both are biologically related to the parents, or where one is adopted and one is biologically related." Among these participants, 2257 families were represented in the present analysis, comprising six distinct "family types":

1) monozygotic-twin families (3970 individuals in 1033 families),
2) dizygotic-twin families (2322 in 612),
3) biological-sibling families (679 in 173),
4) adopted-sibling families (682 in 239),
5) "mixed" families (357 in 92), and
6) "residual singletons" (108 individuals).

The sixth type, "residual singletons," comprises stepparents, the third members of sets of triplets, and individuals who were the only members of their families to provide a tissue sample.

The sample is overwhelmingly Caucasian (90.72%), but it includes a distinct East-Asian minority (4.58%), most of whom are SIBS offspring adopted from South

Korea. The remainder of the sample (3.93%), whose ethnic background was coded as "Other," consists of individuals who were neither Asian nor Caucasian, or were of mixed or unknown ancestry.

SNP Selection & Genotyping

Participants were genotyped on seven SNPs in and around CHRM2 (out of 35 SNPs and variable-number tandem repeats total). These seven SNPs were selected using the Tagger feature of Haploview (version 3.xx; Barrett, Fry, Maller, & Daly, 2005) using release 21 data from www.hapmap.org , with the objectives of adequately representing the gene's variation conditional on its linkage-disequilibrium structure, while also including SNPs showing association with phenotypes of interest in prior research. The SNPs were genotyped with Sequenom MassArray [Sequenom, Inc., San Diego, CA]. A total of thirty-five markers were typed, seven of which were in or near CHRM2.

Some participants had provided multiple tissue samples that had all been genotyped; the genotypes for these individuals were combined during the quality-control procedure applied to the genotypic datafile. Also, only one twin per MZ-twin pair was genotyped, so these genotypes were assigned to the co-twin of the pair. The datafile was additionally screened for data-entry errors, poor-quality tissue samples, recalcitrant markers, and Mendelian errors, using a program written in the R programming language.

Three CHRM2 SNPs were excluded from analysis due to excessive missing data (rs17496259 and rs11982108) or failure of Hardy-Weinberg equilibrium (rs2113545). The following four SNPs were entered into our analysis: (1) rs10228878, an upstream, flanking SNP; (2) rs2350780, located in the 3rd intron; (3) rs324640, located in the 3rd

intron; and (4) rs8191992, located in the downstream untranslated region (3'UTR) of CHRM2. The alleles and minor allele frequencies for these SNPs are presented in Table II. We identify the "minor allele" as the less-common allele in the sample when all available genotypes are considered, irrespective of ethnicity or founder status. We also calculated allele frequencies separately for Asian and Caucasian founders. This separate calculation by ethnic group is informative because allele frequencies can differ markedly in different populations, and calculating only from founders provides a better estimate of the allele frequencies in the respective populations because it can reasonably be assumed that founders are sampled independently.

Phenotypes

Substance-use pathology. We used a composite measure of substance-use pathology that we (Kirkpatrick, McGue, & Iacono, 2009) have described previously as:

> scores on the first principal component extracted from measures of nicotine dependence (e.g., symptom count, per-day frequency of use), heavy alcohol consumption (e.g., frequency of drinking, largest number of drinks in one day), alcohol abuse and dependence (e.g., withdrawal and tolerance, interference with social and occupational functioning), illicit drug abuse and dependence (e.g., symptom count, number of different drugs used), and frequency of illicit drug use.

Because the frequency distribution of substance-use pathology scores deviated substantially from normality in our sample, we subjected these scores to a $\log_e(x + 1)$ transformation, and re-standardized them with respect to the post-transformation mean and standard deviation of the sample.

Personality assessment. Personality was measured using a 198-item version of the Multidimensional Personality Questionnaire (Tellegen, 1982; Tellegen & Waller, 2008), with items scored on a four-point Likert scale. We analyzed the three higher-order MPQ factors—Positive Emotionality, Negative Emotionality, and Constraint—for association with CHRM2 SNPs. Personality data for parents was obtained at the intake assessment of MTFS and SIBS. We also analyzed intake personality data for those offspring who were sixteen years of age or older at the time, which includes the seventeen-year-old MTFS cohort and some of the older siblings from SIBS. For the eleven-year-old-cohort twins, we used their personality data collected at the second MTFS follow-up assessment, when their average age was seventeen. For SIBS offspring younger than sixteen at intake, we used their most recent personality assessment (the first SIBS follow-up), provided that they were sixteen or older at that time. The full MPQ is not given to MTFS and SIBS participants younger than sixteen, because the item content of some of the scales lack face validity for use among children and young adolescents (for example, questions about one's typical behavior in the workplace). Prior to analysis, we rescaled the three factor scores to the T-score metric (M = 50, SD = 10).

Type I error correction

Conventionally, when conducting multiple significance tests, researchers will correct the per-comparison Type I error rate (α_{pc}) so that the study-wide Type I error (α_{sw}) remains fixed at the desired level. Because we are conducting an association study with four phenotypes and four genetic polymorphisms, we are conducting sixteen significance tests. A standard Sidak correction of the Type I error ($\alpha_{pc} = 1 - (1 - \alpha_{sw})^{1/M}$, for M

statistical tests) would be straightforward but will be overly conservative in many cases, because it assumes that the significance tests are independent. Both Nyholt (2004) and Li and Ji (2005) have attempted to tackle the multiple-testing problem for analyses using SNPs that are in linkage disequilibrium (LD). Nyholt's approach calculates a Sidak correction for the "effective" number of independent statistical tests (M_{eff}) as calculated from the eigen decomposition of the correlation matrix of the variables: $M_{eff} = 1 + (M - 1)(1 - \frac{var(\lambda_{obs})}{M})$, where M is the number of variables and $var(\lambda_{obs})$ is the variance of the eigenvalues of the observed correlation matrix. When the variables are SNPs, "correlation," of course, refers to LD. When M_{eff} is calculated, Type I error is corrected with the formula $\alpha_{pc} = 1 - (1 - \alpha_{sw})^{1/M_{eff}}$. The Li-Ji approach is similar to Nyholt's, except that it always computes an integer value for M_{eff}.

The four SNPs entered into our analysis are not in tight LD (Nyholt M_{eff} = 3.48; Li-Ji M_{eff} = 3). The intercorrelation of the four phenotypes is negligible (Nyholt M_{eff}= 3.98; Li-Ji M_{eff}= 4). Both methods yielded α_{PC} = 0.004, within rounding error.

Power Analysis

Prior to allelic association analysis, we conducted a power-analysis in Quanto (Gauderman & Morrison, 2006) for detecting an allele with an additive effect on a quantitative trait. With 2000 independent observations (slightly fewer than the number of families in our sample), and a Type I error of 0.004, we would have 94.6% power to detect an allele that accounts for 1% of the variance in the trait.

Association Analysis—Raw Maximum Likelihood in Mx

The individuals in our sample are clustered within families, and therefore obviously do not provide independently sampled observations. Our sample presents an unusual methodological challenge, in that the covariance of observations within each family will depend upon what type of family it is. Our sample comprises six distinct types of family, described above. Raw maximum likelihood in Mx (Neale, Boker, Xie, & Maes, 2003) enables estimation of SNP effects from an association analysis while accounting for the within-family covariance structure, conditional on family-type. It also provides an attractive method of handling missing data (see below).

For each family *i* of family-type *k,* represent scores for a given phenotype as a 1 × 4 vector,

$\mathbf{x_i} = [x_{i1} \quad x_{i2} \quad x_{iM} \quad x_{iF}]$, where:

x_{i1} = score for offspring 1,

x_{i2} = score for offspring 2,

x_{iM} = score for mother,

x_{iF} = score for father.

The joint distribution of phenotype scores for family-type *k* is modeled as the multivariate normal distribution of four variables with mean $\boldsymbol{\mu_k}$, a 1 × 4 vector, and variance $\boldsymbol{\Sigma_k}$, a 4 × 4 symmetric matrix, such that

$\boldsymbol{\mu_k} = [\mu_{1k} \quad \mu_{2k} \quad \mu_{Mk} \quad \mu_{Fk}] = [E(x_{1k}) \quad E(x_{2k}) \quad E(x_{Mk}) \quad E(x_{Fk})]$

and

$$\boldsymbol{\Sigma_k} = \begin{bmatrix} var(x_{1k}) & cov(x_{1k}, x_{2k}) & cov(x_{1k}, x_{Mk}) & cov(x_{1k}, x_{Fk}) \\ cov(x_{2k}, x_{1k}) & var(x_{2k}) & cov(x_{2k}, x_{Mk}) & cov(x_{2k}, x_{Fk}) \\ cov(x_{Mk}, x_{1k}) & cov(x_{Mk}, x_{2k}) & var(x_{Mk}) & cov(x_{Mk}, x_{Fk}) \\ cov(x_{Fk}, x_{1k}) & cov(x_{Fk}, x_{2k}) & cov(x_{Fk}, x_{Mk}) & var(x_{Fk}) \end{bmatrix} = (\mathbf{L_k L_k^T}) ,$$

where $\mathbf{L_k}$ is a lower-triangular matrix and $\mathbf{L_k^T}$ is its transpose.

The maximum-likelihood estimates of the parameters within $\boldsymbol{\mu_k}$ and $\mathbf{L_k}$, given the data from all *n* vectors of family data in family-type *k*, will be those yielding the global maximum of the function,

$$L(\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k} \mid \mathbf{x_1} \ldots \mathbf{x_n}) = \prod_{i=1}^{n} \frac{\exp\left(-\frac{1}{2}[\mathbf{x}_i - \boldsymbol{\mu_k}]\, \boldsymbol{\Sigma_k^{-1}} [\mathbf{x}_i - \boldsymbol{\mu_k}]^{\mathbf{T}}\right)}{(2\pi)^2 |\boldsymbol{\Sigma_k}|^{1/2}} ,$$

which function Mx numerically optimizes in the following alternate form,

$$-2\log L(\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k} \mid \mathbf{x_1} \ldots \mathbf{x_n}) = \sum_{i=1}^{n} [-4\log(2\pi) + \log|\boldsymbol{\Sigma_k}| + (\mathbf{x_i} - \boldsymbol{\mu_k})\, \boldsymbol{\Sigma_k^{-1}} (\mathbf{x_i} - \boldsymbol{\mu_k})^{\mathbf{T}}] .$$

Note that the software actually optimizes the elements of $\mathbf{L_k}$, and thus indirectly estimates the parameters of $\boldsymbol{\Sigma_k}$. Defining $\boldsymbol{\Sigma_k}$ as the product of a lower-triangular matrix and its transpose ensures that $\boldsymbol{\Sigma_k}$ remain in the parameter space for covariance matrices during numerical optimization (i.e., that it remains positive definite). Obviously, in the special case of family-type 6, the residual singletons, each family contains exactly one observation, so the distribution of phenotype scores is modeled as a (univariate) normal distribution with scalar mean and variance parameters.

Family data vectors with some missing observations on the phenotype variable nonetheless provide information about a subset of the parameters being estimated. One of the attractive features of maximum-likelihood methods is that they do not require complete data for every unit of observation, and can "work around" missing values.

Conceptually, those data vectors which have missing values on a common set of variables are used by Mx in a function for a subspace of the likelihood hypersurface; it is as though the variables with missing values have been integrated out of the function for that particular subset of observations (see Schafer & Graham, 2002), which is straightforward under a multivariate-normal likelihood model. For example, if family-type *k* consists of fifty families providing data for all four members, and ten families with missing data for the mother, Mx optimizes the following fit function,

$$-2\log L(\boldsymbol{\mu}_{\mathbf{k}}, \boldsymbol{\Sigma}_{\mathbf{k}} \mid \mathbf{x_1} \ldots \mathbf{x_{60}}) =$$

$$\sum_{i=1}^{50} \left(-4\log(2\pi) + \log|\boldsymbol{\Sigma}_{\mathbf{k}}| + (\mathbf{x_i} - \boldsymbol{\mu}_{\mathbf{k}})\boldsymbol{\Sigma}_{\mathbf{k}}^{-1}(\mathbf{x_i} - \boldsymbol{\mu}_{\mathbf{k}})^{\mathbf{T}}\right) +$$

$$\sum_{i=51}^{60} \left(-3\log(2\pi) + \log|\boldsymbol{\Sigma}_{\mathbf{k}'}| + (\mathbf{x_i} - \boldsymbol{\mu}_{\mathbf{k}'})\boldsymbol{\Sigma}_{\mathbf{k}'}^{-1}(\mathbf{x_i} - \boldsymbol{\mu}_{\mathbf{k}'})^{\mathbf{T}}\right)$$

where $\boldsymbol{\mu}_{\mathbf{k}'} = [\mu_{1k'} \quad \mu_{2k'} \quad \mu_{Fk'}]$ and

$\boldsymbol{\Sigma}_{\mathbf{k}'} = \begin{bmatrix} var(x_{1k'}) & cov(x_{1k'}, x_{2k'}) & cov(x_{1k'}, x_{Fk'}) \\ cov(x_{2k'}, x_{1k'}) & var(x_{2k'}) & cov(x_{2k'}, x_{Fk'}) \\ cov(x_{Fk'}, x_{1k'}) & cov(x_{Fk'}, x_{2k'}) & var(x_{Fk'}) \end{bmatrix}$, and the parameters within $\boldsymbol{\mu}_{\mathbf{k}'}$ and $\boldsymbol{\Sigma}_{\mathbf{k}'}$ are equated to their counterparts in $\boldsymbol{\mu}_{\mathbf{k}}$ and $\boldsymbol{\Sigma}_{\mathbf{k}}$. This "work-around" for missing phenotypic data will result in parameter estimates that are not biased by the missing-data mechanism, provided that the data are missing at random (Schafer & Graham, 2002; Rubin, 1976).

In twin families (family-types 1 and 2), the assignment of the labels "Offspring 1" and "Offspring 2" mentioned above are basically arbitrary, though in actual practice, the labels are used to identify the first-delivered and second-delivered twin (during birth of the twin pair), respectively. The labels "Offspring 1" and "Offspring 2" likewise serve to

identify the older sibling and younger sibling, respectively, in family-types 3 and 4. In family-type 5, "Offspring 1" always denotes the adopted child, and "Offspring 2" identifies his or her sibling, who is the biological offspring of the family's parents.

The phenotypic covariance matrix ($\mathbf{\Sigma_k}$) for each family-type is estimated unconditionally and without constraints. This serves to account for the within-family covariance of observations, the structure of which depends upon family-type. The allelic association analysis is implemented by modeling the means of each family-type conditional on genotype and covariates. By designating genotype and covariates as "definition variables" in Mx, the expected observation for each element of family data vector *i* of family-type *k* is defined conditionally on these variables, as follows:

$$E(x_{i1}) = J_{1k} + b_1 S_{i1} + b_2 C_{i1} + b_3 A_{i1} + b_4 G_{i1}$$

$$E(x_{i2}) = J_{2k} + b_1 S_{i2} + b_2 C_{i2} + b_3 A_{i2} + b_4 G_{i2}$$

$$E(x_{iM}) = J_{Mk} \quad + \quad b_2 C_{iM} + b_3 A_{iM} + b_5 G_{iM}$$

$$E(x_{iF}) = J_{Fk} \quad + \quad b_2 C_{iF} + b_3 A_{iF} + b_5 G_{iF}$$

where:

$J_{jk} =$ Intercept for family-member type *j* in family-type *k*,

$b_1 =$ Effect of sex, S_{ij}, among offspring (since sex of parent is implicit in being a mother or father),

$b_2 =$ Effect of Caucasian ancestry, C_{ij} (dummy-coded),

$b_3 =$ Effect of Asian ancestry, A_{ij} (dummy-coded),

$b_4 =$ Effect of minor allele, G_{ij}, for offspring,

$b_5 =$ Effect of minor allele, G_{ij}, for parents.

As will be seen from the above, we are regressing phenotype onto genotype and covariates. Thus, four intercept terms are estimated for each family-type, whereas the regression slopes are estimated across the whole sample.

We implemented the model in Mx by way of a two-step process for each combination of SNP and phenotype. First, we used one Mx script to estimate the unconditional phenotypic covariance matrices for each family-type and write each to a text file. Then, we used a second Mx script that fit the means model described above, using the fixed, precomputed covariance matrices read from the text files. This approach enabled us to estimate the covariance matrices from all available phenotypic data, irrespective of whether or not an individual had genotypic data. Additionally, because any given SNP accounts for a very small portion of the phenotypic variance, the use of these precomputed covariance matrices can save a considerable amount of computation time, with negligible loss of power or increase in Type I error rate (Li et al., under review). Missing genotypes were only a problem for the means model fitted in the second step, since the means model is defined conditional on genotype. Consequently, we excluded individuals with missing genotype from the means model. We used a program written in the R programming language for data-management and automation of the analysis for all sixteen SNP-phenotype combinations.

The following *ad hoc* modifications to our model had to be made when applying it to our sample. First, the intercepts J_1 and J_2 were fixed to be equal in twin families, since the designation of the labels “Offspring 1” and “Offspring 2” to the twins is essentially arbitrary. Second, we had to adjust the model in light of the fact that we

collected tissue samples from only one parent in each family of type 4, which made it impossible for any family of type 4 to have complete observations for all four of its members. We did so by treating each family of this type as providing three observations: one from each offspring, and one from the parent who provided a tissue sample. Consequently, there was no distinction between mothers and fathers in family-type 4—they were both simply treated as the family's "parent," and in effect, the intercept for mothers and fathers in family type 4 were constrained equal. Finally, we treated all participants in family type 6 as though they were parents, since most of them were either step-parents, or parents from families where no other members provided tissue samples. Hence, the means model for family-type 6 used the minor-allele effect for the parents and did not include the offspring-sex effect.

For each of the sixteen SNP-phenotype combinations, we fitted five different models to the data, each representing a slight variation on the null hypothesis of no association with the phenotype in question. The first model is the "full model," described in the preceding, where the SNP effects for offspring (b_4) and for parents (b_5) are free parameters and are estimated separately. In submodel 1, b_4 and b_5 are constrained to be equal, and thus, a single SNP effect is estimated for parents and offspring. In submodel 2, both SNP effects are fixed to zero. We also fitted two additional models nested within the full model: one in which b_4 is fixed to zero, with b_5 remaining free to vary, and another in which b_5 is fixed to zero, with b_4 free to vary. The purpose of these two models was to detect heterogeneity of SNP effects among parents versus offspring, as a

guide to selecting follow-up analyses; we report results only from the nested series of three models.

We compared the fit of submodel 2 to that of the full model in a likelihood-ratio test, which in large samples is approximately distributed as chi-square, with degrees-of-freedom equal to the number of parameters fixed in the submodel relative to the full model. We draw our statistical inferences using both the p-values from this likelihood-ratio test, and the likelihood-based confidence intervals (see Neale & Miller, 1997) for the free parameters from each model, which both support the same conclusion.

Results

We present the results of the allelic association analysis in Table II. None of the SNP effects for any phenotype were statistically distinguishable from zero at the adjusted Type I error ($\alpha = 0.004$), either when estimated separately for parents and offspring, or when a single parameter was estimated from both. None of the likelihood-ratio tests provided significant evidence of nonzero SNP effects.

The three largest likelihood-ratio test statistics from submodel 2 were from the association analysis of NEM, with rs2350780 ($\chi^2 = 6.446$, $p = 0.040$), rs324640 ($\chi^2 = 5.009$, $p = 0.082$), and rs8191992 ($\chi^2 = 7.905$, $p = 0.019$). From the two additional models (results not shown in Table II) nested within the full model, in which the SNP effect was fixed to zero either for parents or offspring, we observed a pattern of nominally significant results for association of MPQ NEM with rs3250780, rs324640, and rs8191992, *only among parents*. We therefore conducted a post-hoc association

analysis of the SNP most suggestive of association, rs8191992, with the MPQ primary scale that most closely approximates Big-Five Agreeableness, Aggression. Even without further Type I error correction (i.e., $\alpha = 0.004$), the SNP effect was not significantly different from zero, either estimated separately for offspring ($b_4 = 0.152$ [99.6% CI: -0.856, 1.159]) and parents ($b_5 = -0.295$ [99.6% CI: -0.935, 0.345]), or as a "pooled" estimate (b = -0.170 [99.6% CI: -0.719, 0.379]). Likewise, the likelihood-ratio test from submodel 2 did not provide significant evidence against the null hypothesis of zero SNP effect in both parents and offspring (p = 0.369).

Discussion

Analysis of our data did not provide any clear evidence of CHRM2's association with personality traits or with substance-use pathology. We additionally hypothesized that personality traits may mediate CHRM2's influence on substance-use pathology, but the lack of evidence for a direct contribution of CHRM2 obviously precludes any kind of mediation analysis.

Conclusions from the present study are subject to several limitations, which should be explicitly acknowledged. First, we used self-reported ethnicity as a covariate. Ideally, and if such data were available, we would have inferred each participant's ancestry from a panel of ancestry-informative genetic markers. Second, even after log transformation, the distribution of the substance-use pathology variable deviated substantially from multivariate normality. Our results concerning this phenotype should therefore be interpreted cautiously, especially since the software did not reach its strictest

criterion for finding an optimal value of the fit function. An alternate fit function, or the use of bootstrap confidence intervals, might be preferable for making inferences regarding substance-use pathology. However, non-parametric bootstrapping with these data would be complicated by the need to take into account different patterns of missing data. We attempted bootstrapping with a built-in procedure in Mx, which makes "missingness-naïve" random draws from the dataset; replications of our association analysis from these bootstrap samples rarely converged on a solution.

The presence of large amounts of missing data is another major limitation of our study. Out of 8,123 participants who provided tissue samples, fewer than 4,100 had non-missing data for both genotype and phenotype for any analyzed SNP-phenotype combination. For missing phenotype data, the numerical maximum-likelihood procedure we used will "work around" the missing observations as described under Methods, and if phenotype data are missing at random, it will provide parameter estimates unbiased by the missing-data mechanism (Schafer & Graham, 2002; Rubin, 1976). On the other hand, when a genotype is missing for a given individual, the Mx software will exclude that individual's *entire* family-data vector from analysis. To prevent the loss of the whole family, when an individual was missing the genotype for the SNP under analysis, we set that individual's phenotype score also to missing. This is equivalent to deletion of incomplete individual-participant cases, which will leave parameter estimates unbiased only if genotype data are missing *completely* at random (Schaefer & Graham, 2002; Rubin, 1976). If genotype data are missing completely at random, the probability of missing a genotype does not depend upon the value of any observed variable. We

conducted an analysis to see if participants' amounts of missing SNP data predicted their phenotype scores. This analysis, conducted in Mx, was mostly identical to our allelic association analysis. The difference was that only one analysis was conducted for each phenotype, which analysis used as a predictor the number of missing SNPs an individual had (ranging zero through four), in place of minor-allele count. None of the three personality variables was nominally significantly predicted by SNP missingness, but substance-use pathology was. Interestingly, it was slightly *negatively* associated with SNP missingness: the point estimate for the regression coefficient was -0.018 (95% CI: -0.029, -0.008). This result indicates that it is untenable to assume that genotypes are missing completely at random, rendering our conclusions about CHRM2 and substance-use pathology very tentative indeed. It may have been preferable to impute the expected minor allele count for each missing genotype, conditional on the individual's ethnicity, other genotypes (if any known), and the genotypes of his or her biological relatives (if any known). Finally, there were only four SNPs employed in the analysis to begin with, and our sparse coverage of the CHRM2 gene limits the generalizability of our results.

SNP	Minor Allele	Major Allele	MAF, Caucasian founders	MAF, Asian founders	MAF, All available genotypes
rs10228878	T	C	0.30	0.02	0.29
rs2350780	G	A	0.36	0.53	0.37
rs324640	A	G	0.51	0.08	0.49
rs8191992	T	A	0.46	0.89	0.48

Table I. Allele frequencies for analyzed SNPs. MAF = minor allele frequency. MAF is calculated separately for Caucasian founders and Asian founders. MAF calculated for "all available genotypes" uses all non-missing genotypes in the sample, irrespective of ethnicity or founder status.

Phenotype	SNP	Full model df	Full model: SNP effect in offspring (99.6% CI)	Full model: SNP effect in parents (99.6% CI)	Submodel 1: Pooled SNP effect (99.6% CI)	Submodel 2: χ^2 (2 df) (p-value)
Substance-Use Pathology	rs10228878	3851	-0.016 (-0.109, 0.078)	0.032 (-0.054, 0.118)	0.010 (-0.056, 0.075)	1.441 (p = 0.486)
	rs2350780	3937	0.023 (-0.063, 0.108)	0.018 (-0.064, 0.100)	0.020 (-0.041, 0.081)	0.924 (p = 0.630)
	rs324640	3921	-0.009 (-0.097, 0.079)	-0.020 (-0.099, 0.059)	-0.015 (-0.076, 0.045)	0.598 (p = 0.741)
	rs8191992	3903	-0.008 (-0.095, 0.080)	0.004 (-0.076, 0.084)	-0.001 (-0.062, 0.060)	0.023 (p = 0.879)
MPQ CON	rs10228878	3783	0.244 (-0.738, 1.227)	-0.235 (-1.141, 0.671)	-0.016 (-0.698, 0.667)	1.129 (p = 0.569)
	rs2350780	3870	0.170 (-0.729, 1.068)	-0.177 (-1.034, 0.681)	-0.012 (-0.648, 0.624)	0.681 (p = 0.711)
	rs324640	3852	0.069 (-0.854, 0.992)	0.493 (-0.344, 1.133)	0.303 (-0.333, 0.939)	2.895 (p = 0.235)
	rs8191992	3835	0.034 (-0.884, 0.952)	-0.147 (-0.984, 0.691)	-0.065 (-0.701, 0.571)	0.273 (p = 0.873)
MPQ NEM	rs10228878	3783	0.287 (-0.729, 1.302)	-0.076 (-1.067, 0.915)	0.101 (-0.622, 0.824)	0.725 (p = 0.696)
	rs2350780	3870	-0.013 (-0.943, 0.917)	-0.828 (-1.766, 0.111)	-0.416 (-1.090, 0.257)	6.446 (p = 0.040)
	rs324640	3852	-0.223 (-1.178, 0.733)	0.671 (-0.247, 1.587)	0.243 (-0.432, 0.918)	5.009 (p = 0.082)
	rs8191992	3835	0.012 (-0.938, 0.961)	-0.894 (-1.812, 0.023)	-0.458 (-1.132, 0.216)	7.905 (p = 0.019)
MPQ PEM	rs10228878	3783	0.210 (-0.830, 1.250)	-0.267 (-1.346, 0.810)	-0.020 (-0.786, 0.746)	0.889 (p = 0.641)
	rs2350780	3870	0.692 (-0.256, 1.640)	0.284 (-0.732, 1.299)	0.503 (-0.207, 1.212)	4.913 (p = 0.086)
	rs324640	3852	-0.029 (-1.004, 0.946)	-0.145 (-1.139, 0.849)	-0.086 (-0.799, 0.627)	0.181 (p = 0.913)
	rs8191992	3835	-0.115 (-1.085, 0.855)	0.279 (-0.713, 1.272)	0.077 (-0.635, 0.790)	0.804 (p = 0.669)

Table II. Estimated SNP effects and likelihood-ratio test results.

References

Agrawal, A., Edenberg, H. J., Foroud, T., Bierut, L. J., Dunne, G., Hinrichs, A. L., et al. (2006). Association of GABRA2 with drug dependence in the Collaborative Study of the Genetics of Alcoholism sample. Behavior Genetics, *36*(5), 640-650.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics, 21*(2), 263-265.

Baxter, M. G., & Chiba, A. A. (1999). Cognitive functions of the basal forebrain. *Current Opinion in Neurobiology, 9,* 178-183.

Begleiter, H., Porjesz, B., Bihari, B., & Kissin, B. (1984). Event-related brain potentials in boys at risk for alcoholism. *Science, New Series, 225*(4669), 1493-1496.

Bouchard, T. J., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of Neurobiology, 54,* 4-45.

Calabresi, P., Centonze, D., Gubellini, P., Pisani, A., & Bernardi, G. (1998). Blockade of M2-like muscarinic receptors enhances long-term potentiation at corticostriatal synapses. *European Journal of Neuroscience, 10*, 3020-3023.

Church, A. T. (1994). Relating the Tellegen and Five-Factor models of personality structure. *Journal of Personality and Social Psychology, 67*(5), 898-909.

Conway, K. P., Swendsen, J. D., Rounsaville, B. J., & Merikangas, K. R. (2002). Personality, drug of choice, and comorbid psychopathology among substance abusers. *Drug and Alcohol Dependence, 65*, 225-234.

Costa, P. T., & McCrae, R. R. (1997). Stability and change in personality assessment: the revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment, 68*, 86-94.

Cragg, S. J. (2006). Meaningful silences: How dopamine listens to the ACh pause. *Trends in Neuroscience, 29*(3), 125-131.

DeYoung, C. G., & Gray, J. R. (2009). Personality neuroscience: Explaining individual differences in affect, behavior, and cognition. In Corr, P. J., & Matthews, G. (Eds.), *The Cambridge Handbook of Personality Psychology* (p. 323-346). New York: Cambridge University Press.

Dick, D. M., Agrawal, A., Wang, J. C., Hinrichs, A., Bertelsen, S., Bucholz, K. K., et al. (2007). Alcohol dependence with comorbid drug dependence: Genetic and phenotypic associations suggest a more severe form of the disorder with stronger genetic contribution to risk. *Addiction, 102*, 1131-1139.

Dick, D. M., Aliev, F., Wang, J. C., Grucza, R. A., Shuckit, M., Kuperman, S., et al. (2008). Using dimensional models of externalizing psychopathology to aid in gene identification. *Archives of General Psychiatry, 65*(3), 310-318.

Dick, D. M., Prescott, C., & McGue, M. (2009). The genetics of substance use and substance use disorders. In Yong-Kyu Kim (Ed.), *Handbook of Behavior Genetics* (pp. 433-453). New York: Springer Science+Business Media, LLC.

Doppelmayr, M., Klimesch, W., Schwaiger, J., Auinger, P., & Winkler, Th. (1998). Theta synchronization in the human EEG and episodic retrieval. *Neuroscience Letters, 257*, 41-44.

Elkins, I. J., King, S. M., McGue, M., & Iacono, W. G. (2006). Personality traits and the development of nicotine, alcohol, and illicit drug disorders: Prospective links from adolescence to young adulthood. *Journal of Abnormal Psychology, 115*(1), 26-39.

Everitt, B. J., & Robbins, T. W. (1997). Central cholinergic systems and cognition. *Annual Review of Psychology, 48*, 649-684.

Foroud, T., Edenberg, H. J., Goate, A., Rice, J., Flury, L., Koller, D. L., et al. (2000). Alcoholism susceptibility loci: Confirmation studies in a replicate sample and further mapping. *Alcoholism: Clinical and Experimental Research, 24*(7), 933-945.

Gauderman, W. J., & Morrison, J. M. (2006). QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies [software and manual]. Retrieved June 8, 2009 from http://hydra.usc.edu/gxe .

Hoebel, B. G., Avena, N. M., & Rada, P. (2007). Accumbens dopamine-acetylcholine balance in approach and avoidance. *Current Opinion in Pharmacology, 7*, 617-627.

Iacono, W. G., Carlson, S. R., Taylor, J., Elkins, I. J., & McGue, M. (1999). Behavioral disinhibition and the development of substance-use disorders: Findings from the Minnesota Twin Family Study. *Development and Psychopathology*, *11*, 869-900.

Iacono, W. G., & McGue, M. (2002). Minnesota Twin Family Study. *Twin Research, 5*(5), 482-487.

Jang, K. L., & Yamagata, S. (2009). Personality. In Yong-Kyu Kim (Ed.), *Handbook of Behavior Genetics* (pp. 223-237). New York: Springer Science+Business Media, LLC.

Jones, K. A., Porjesz, B., Almasy, L., Bierut, L., Goate, A., Wang, J. C., et al. (2004). Linkage and linkage disequilibrium of evoked EEG oscillations with CHRM2 receptor gene polymorphisms: Implications for human brain dynamics and cognition. *International Journal of Psychophysiology, 53*, 75-90.

Jones, K. A., Porjesz, B., Almasy, L., Bierut, L., Dick, D. M., Goate, A., et al. (2006). A cholinergic receptor gene affects event-related oscillations. *Behavior Genetics, 36*, 627-639.

Kirkpatrick, R. M., McGue, M., & Iacono, W. G. (2009). Shared-environmental contributions to high cognitive ability. *Behavior Genetics, 39*, 406-416.

Klimesch, W., Doppelmayr, M., Yonelinas, A., Kroll, N. E. A., Lazzara, M., Röhm, D., & Gruger, W. (2001). Theta synchronization during episodic retrieval: neural correlates of conscious awareness. *Cognitive Brain Research, 12*, 33-38.

Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior, and personality: Modeling the externalizing spectrum. *Journal of Abnormal Psychology, 111*(3), 411-424.

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology, 2*, 111-113.

Li, J., & Ji, L. (2005). Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity, 95*, 221-227.

Li, X., Basu, S., Miller, M.B., & McGue, M. (2010). A rapid generalized least squares model for genome-wide quantitative trait association analysis. Manuscript under review, University of Minnesota.

Luo, X., Kranzler, H. R., Zuo, L., Wang, S., Blumberg, H. P., & Gelernter, J. (2005). CHRM2 gene predisposes to alcohol dependence, drug dependence and affective disorders: Results from an extended case-control structured association study. *Human Molecular Genetics, 14*(16), 2421-2434.

Luo, X., Kranzler, H. R., Zuo, L., Zhang, H., Wang, S., & Gelernter, J. (2007). CHRM2 variation predisposes to personality traits of agreeableness and conscientiousness. *Human Molecular Genetics, 16*(13), 1557-1568.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70*(3), 151-159.

Markon, K. E., Krueger, R. F., & Watson, D. (2005). Delineating the structure of normal and abnormal personality: An integrative hierarchical approach. *Journal of Personality and Social Psychology, 88*(1), 139-157.

Markou, A. (2008). Neurobiology of nicotine dependence. *Philosophical Transactions of the Royal Society B, 363*, 3159-3168.

McGue, M., Keyes, M., Sharma, A., Elkins, I., Legrand, L., Johnson, W., et al. (2007). The environments of adopted and non-adopted youth: Evidence on range restriction from the Sibling Interaction and Behavior Study (SIBS). *Behavior Genetics, 37*, 449-462.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics, 27*(2), 113-120.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling.* VCU Box 900126, Richmond, VA 23298: Department of Psychiatry. 6th Edition.

Nyholt, D. R. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *American Journal of Human Genetics, 74*, 765-769.

Peralta, E. G., Winslow, J. W., Peterson, G. L., Smith, D. H., Ashkenazi, A., Ramachandran, J., et al. (1987). Primary structure and biochemical properties of an M2 muscarinic receptor. *Science, New Series 236*(4801), 600-605.

Rada, P., Pothos, E., Mark, G. P., & Hoebel, B. G. (1991). Microdialysis evidence that acetylcholine in the nucleus accumbens is involved in morphine withdrawal and its treatment with chloridine. Brain Research, *561*(2), 354-356. Abstract retrieved January 16, 2001, from PsycINFO database.

Rada, P., Jensen, K., Hoebel, B. G. (2001). Effects of nicotine and mecamylamine-induced withdrawal on extracellular dopamine and acetylcholine in the rat nucleus accumbens. *Psychopharmacology, 157*, 105-110.

Rada, P., Johnson, D. F., Lewis, M. J., & Hoebel, B. G. (2004). In alcohol-treated rats, naloxone decreases extracellular dopamine and increases acetylcholine in the nucleus accumbens: Evidence of opioid withdrawal. *Pharmacology, Biochemistry and Behavior, 79*, 599-605.

Rada, P., & Hoebel, B. G. (2005). Acetylcholine in the accumbens is decreased by diazepam and increased by benzodiazepine withdrawal: a possible mechanism for dependency. *European Journal of Pharmacology, 508*, 131-138.

Rahman, S., Lopez-Hernandez, G. Y., Corrigall, W. A., & Papke, R. L. (2008). Neuronal nicotinic receptors as brain targets for pharmacotherapy of drug addiction. *CNS & Neurological Disorders – Drug Targets, 7*(5), 422-441.

Reich, T., Edenberg, H. J., Goate, A., Williams, J. T., Rice, J. P., Van Eerdewegh, P., et al. (1998). Genome-wide search for genes affecting the risk for alcohol dependence. *American Journal of Medical Genetics (Neuropsychiatric Genetics), 81*, 207-215.

Rose, R. J., Broms, U., Korhonen, T., Dick, D. M., & Kaprio, J. Genetics of smoking behavior. In Yong-Kyu Kim (Ed.), *Handbook of Behavior Genetics* (pp. 411-432). New York: Springer Science+Business Media, LLC.

Rubin, D. B. (1976). Inference and missing data. (1976). *Biometrika, 63*(3), 581-592.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.

Sufuoglu, M., & Mooney, M. (2009). Cholinergic functioning in stimulant addiction: Implications for medications development. *CNS Drugs, 23*(11), 939-952.

Tellegen, A. (1982). Brief manual of the Multidimensional Personality Questionnaire. Unpublished manuscript. University of Minnesota.

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the Multidimensional Personality Questionnaire. In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *Handbook of Personality Theory and Testing: Vol. II. Personality Measurement and Assessment* (pp. 261-292). Thousand Oaks, CA: Sage.

Wang, J. C., Hinrichs, A. L., Stock, H., Budde, J., Allen, R., Bertelsen, S., et al. (2004). Evidence of common and specific genetic effects: Association of the muscarinic acetylcholine receptor M2 (CHRM2) gene with alcohol dependence and major depressive syndrome. *Human Molecular Genetics, 14*(17), 1903-1911.

Williams, M. J., & Adinoff, B. (2008). The role of acetylcholine in cocaine addiction. *Neuropsychopharmacology, 33*, 1779-1797.

Wolf, M. E. (2003). LTP may trigger addiction. *Molecular Interventions, 3*(5), 248-252.

Young, S. E., Rhee, S. H., Stallings, M. C., Corley, R. P., & Hewitt, J. K. (2006). Genetic and environmental vulnerabilities underlying adolescent substance use and problem use: General or specific? *Behavior Genetics, 36*(4), 603-615.

Young, S. E., Stallings, M. C., Corley, R. P., Krauter, K. S., & Hewitt, J. K. (2000). Genetic and environmental influences on behavioral disinhibition. *American Journal of Medical Genetics (Neuropsychiatric Genetics), 96,* 684-695.

Zhang, W., Yamada, M., Gomeza, J., Basile, A. S., & Wess, J. (2002). Multiple muscarinic acetylcholine receptor subtypes modulate striatal dopamine release, as studied with M1-M5 muscarinic receptor knock-out mice. *The Journal of Neuroscience, 22*(15), 6347-6352.

Zhuo, C., Fryer, A. D., & Jacoby, D. B. (2001). Structure of the human M2 muscarinic acetylcholine receptor gene and its promoter. *Gene, 271*, 97-92.

Zuckerman, M., Kuhlman, D. M., Joireman, J., Teta, P., & Kraft, M. (1993). A comparison of three structural models for personality: The Big Three, the Big Five, and the Alternative Five. *Journal of Personality and Social Psychology, 65*(4), 757-768.