The Genetics of Substance Dependence

Jen-Chyong Wang, Manav Kapoor, and Alison M. Goate

Department of Psychiatry, School of Medicine, Washington University in St. Louis, St. Louis, Missouri 63110; email: jcwang@psychiatry.wustl.edu, kapoorm@psychiatry.wustl.edu, goatea@psychiatry.wustl.edu

Keywords
addiction, alcoholism, smoking, nicotinic acetylcholine receptors, alcohol-metabolizing genes

Abstract
A large segment of the population suffers from addiction to alcohol, smoking, or illicit drugs. Not only do substance abuse and addiction pose a threat to health, but the consequences of addiction also impose a social and economic burden on families, communities, and nations. Genome-wide linkage and association studies have been used for addiction research with varying degrees of success. The most well-established genetic factors associated with alcohol dependence are in the genes encoding alcohol dehydrogenase (ADH), which oxidizes alcohol to acetaldehyde, and aldehyde dehydrogenase (ALDH2), which oxidizes acetaldehyde to acetate. Recently emerging genetic studies have linked variants in the genes encoding the α3, α5, and β4 nicotinic acetylcholine receptor subunits to smoking risk. However, the influence of these well-established genetic variants accounts for only a small portion of the heritability of alcohol and nicotine addiction, and it is likely that there are both common and rare risk variants yet to be identified. Newly developed DNA sequencing technologies could potentially advance the detection of rare variants with a larger impact on addiction risk.
INTRODUCTION

Substance abuse and addiction pose a worldwide threat to public health and have a devastating social and economic impact on individuals and their families. The World Health Organization (146) has estimated that there are 2 billion alcohol users, 1.3 billion tobacco users, and 185 million illicit drug users worldwide.

In the 2010 National Survey on Drug Use and Health conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA) (119), 51.8% of Americans aged 12 or older (131.3 million people) reported being current drinkers of alcohol, and 23.1% reported participating in binge drinking (defined as having five or more drinks on the same occasion on at least 1 day in the 30 days prior to the survey). The World Health Organization (145) has also estimated that approximately 20%–30% of esophageal cancer, liver cancer, cirrhosis of the liver, homicide, epilepsy, and motor vehicle accidents worldwide result from the harmful use of alcohol.

In the 2010 SAMHSA survey (119), approximately 58.3 million Americans aged 12 or older reported being current cigarette smokers, and a recent Surgeon General’s report (129) indicated that one-third of people who have tried smoking became daily smokers (defined as those who reported that they have smoked 100 or more cigarettes during their lifetime and currently smoke every day or some days). The SAMHSA survey (119) also showed that 59.6% of current smokers aged 12 or older smoked daily and that this proportion increased with age—going from 16.5% among those aged 12–17, to 27.8% among those aged 18–25, to 48.9% among those aged 26 or older. The detrimental effects of tobacco use or exposure to secondhand smoke include an increased risk of cancer, chronic lung disease, heart disease, and stroke. In the United States, cigarette smoking accounts for 30% of deaths from cancer and nearly 80% of deaths from chronic obstructive pulmonary disease (COPD) (22, 95), and it is also the primary causal factor for early cardiovascular disease and deaths (22). Globally, cigarette smoking kills 5.4 million people every year and accounts for 10% of adult deaths (146).

Other psychoactive substances, such as cannabis, cocaine, and opioids, also cause significant health and social problems for both the people who use them and their families. According to data from the United Nations Office on Drugs and Crime, 149–271 million people worldwide aged 15–64 used an illicit drug in 2009; of these, 15–39 million were classified as problem users (31). In the SAMHSA survey (119), an estimated 22.6 million Americans aged 12 or older reported having used an illicit drug in 2010 during the month prior to the survey interview; of these, 2.9 million abused or were dependent on both alcohol and illicit drugs, and 4.2 million abused or were dependent on illicit drugs but not alcohol. Major health consequences of illicit drug use include accidental and intended injury, drug-induced psychotic symptoms, and increased risk for heart, liver, and lung diseases (31). A World Health Organization report (146) indicated that in 2004, an estimated 0.7% of the global burden of disease resulted from cocaine and opioid use.

Drug addiction is a chronic psychiatric disorder characterized by the persistent, compulsive, and uncontrolled use of a drug despite harmful consequences. Scientific studies on addictive behaviors began in the 1930s and revealed that people with an addiction are not simply lacking in willpower; instead, they are unable to control their use of the drug (135). With advances in our understanding of the effects of alcohol, nicotine, and illicit drugs on brain physiology and behavior, it has become evident that addiction is a psychiatric disease attributable to biological and environmental factors (135). The development of addiction involves several steps: the initiation of substance use, the transition from experimental use to regular use, and the actual development of addiction. Environmental factors such as peer pressure, parental monitoring, and the accessibility of a substance play a major role in the initial decision to drink, smoke, or take illicit drugs. Beyond the initiation step, the transition
Figure 1
Interaction of genetic and environmental factors in the development of substance dependence. The initiation of substance use is influenced largely by environmental factors; the use of the addictive substance is affected largely by genetic factors.

from regular substance use to dependence differs from person to person and is largely under genetic control (Figure 1)(74, 133, 135).

GENETIC INFLUENCES ON THE RISK OF SUBSTANCE DEPENDENCE

Evidence for a genetic influence on substance dependence has been provided by many family, twin, and adoption studies. Family members of alcohol-dependent individuals have a higher probability of suffering from alcohol dependence (54). In a study of families severely affected by alcohol-abuse disorders, approximately 50% of brothers and 22%–25% of sisters of an alcohol-dependent proband were alcohol dependent (15). Similarly, siblings of marijuana-dependent, cocaine-dependent, or habitual-smoking probands were at increased risk (approximately 1.7-fold higher) of developing marijuana dependence, cocaine dependence, or habitual smoking compared with siblings of nondependent individuals (15).

Studies with large twin cohorts have shown that the risk of alcohol dependence in the co-twin of an affected monozygotic twin is significantly higher than the risk in the co-twin of an affected dizygotic twin pair, which is similar to that of full siblings of the affected individual (116). In adoption studies, children of alcoholics adopted by nonalcoholics who grow up in a nondrinking environment have a higher risk of becoming alcoholic than do children of nonalcoholics adopted by the same parents; children of alcoholics raised by their alcoholic father have a similar risk of developing alcohol dependence as their full brothers who were adopted by nonalcoholics (116). Overall, studies have shown that the heritability of alcohol-use disorders ranges from 40% to 60% (116, 131).

Cigarette smoking commonly co-occurs with alcohol abuse. A meta-analysis of twin studies showed that both genetic and environmental factors affect smoking and smoking-related behaviors (83). In women, the initiation of smoking is influenced largely by genetic factors; in men, the genetic impact is more significant on the persistence of smoking than on the initiation of smoking (83). The heritability of smoking initiation and nicotine dependence is estimated to be 50% and 59%, respectively (83).

Studies have also indicated the familial transmission of illicit-substance-use disorders, with heritability estimates ranging from 30% to 80% (2, 83, 127). A recent meta-analysis of twin studies on marijuana use clearly indicated that vulnerability to both initiation and persistent use was significantly affected by both genetic and environmental factors (132). Genetic factors accounted for 48% and 40% of the total variance in initiation of marijuana use in men and women, respectively, and for 51% and 59% of the total variance in marijuana abuse in men and women, respectively.

Epidemiological and clinical studies have shown that many people subsequently use multiple drugs after their initiation of one drug (105, 109). Twin studies have demonstrated the presence of shared environmental factors that contribute to substance use. The shared environmental influence has a significant effect on tobacco initiation, alcohol use, and any drug use; however, genetic factors have a higher impact than shared environmental influences on tobacco use, tobacco problem use, and marijuana initiation (109). Studies on the etiology of the comorbidity of multiple substances in
adolescents have suggested that genetic and environmental influences are common across substance classes (74, 109). In family studies involving adults, findings regarding general versus substance-specific familial risks have not been conclusive (15, 91, 128).

**IDENTIFICATION OF GENETIC RISK FACTORS FOR ALCOHOL DEPENDENCE**

**Genome-Wide Linkage Studies**

One of the earliest genome-wide approaches to identifying genetic risk factors for alcoholism employed linkage mapping in large extended families or in many sibling pairs affected by alcohol dependence. Studies using this method identified several chromosomal regions with LOD (logarithm of the odds, to the base 10) scores suggesting that they contain loci influencing risk for alcohol dependence. The Collaborative Study on the Genetics of Alcoholism (COGA) investigators performed linkage studies on a large sample from the general US population using multigenerational pedigrees densely affected by alcoholism (45, 108). In contrast, investigators at the National Institute on Alcohol Abuse and Alcoholism conducted a linkage study on a more homogeneous population from a southwestern Native American tribe (88). Both of these studies provided evidence that loci on human chromosome 4 increase the risk for alcohol dependence. However, the linkage peak detected in the COGA study is near the alcohol dehydrogenase (ADH) gene cluster, whereas the linkage signal observed in the Native American sample is near the GABRB1 gene. The separate locations of these linkage signals may reflect differences in the underlying etiology across distinct populations, a suggestion supported by the observation of genome-wide significant linkage of alcohol dependence with markers on human chromosome 10 in an African American sample (47). Alternatively, this difference could reflect the inability of this kind of linkage study to accurately pinpoint the location of the gene(s) underlying a signal, or could mean that one or more of these linkage regions is associated with a false-positive signal.

Recently, linkage analysis of a community-based sample of Australian adults detected a suggestive linkage peak on human chromosome 5p with a LOD score of 2.2 (58). A genome-wide scan performed with community samples recruited through the University of California, San Francisco (UCSF) Family Alcoholism Study identified several suggestive regions linked with DSM-IV alcohol dependence (on human chromosomes 1, 2, 8, 9, 18, and 22) (50). Unfortunately, there is little consensus among studies regarding the location of linkage signals for alcohol dependence. This may be due to underlying genetic heterogeneity in the risk for alcohol dependence, with many genetic loci contributing to risk. This is probably compounded by the fact that all of these studies were underpowered to detect genes of small effect size (110).

**Genome-Wide Association Studies**

Several genome-wide association studies (GWAS) examining the risk for alcohol dependence have been completed using a variety of designs, including case-control series of male alcoholics recruited from inpatient treatment facilities (125), individuals selected from densely affected families with alcohol dependence (35), a mixed case-control series drawn from treatment- and community-based samples (14), subjects ascertained from community-based sibships, and individuals selected for heavier alcohol use (61). GWAS using quantitative traits derived from alcohol-consumption and alcohol-dependence symptomatology have also been examined in controls from a population-based sample recruited for schizophrenia (73) and an Australian population of related individuals (61). One study identified two correlated intergenic single-nucleotide polymorphisms (SNPs) on human chromosome 2q35 that met genome-wide significance in the combined analysis of the GWAS and follow-up data sets.

in risk for alcohol-use disorders have focused on the overlap of the top genetic signals across studies has been limited.

Although the results to date have been somewhat disappointing, they underscore the prior observations from linkage studies and support the hypothesis that alcohol dependence is a genetically heterogeneous disorder influenced by many genes of small effect. The power to detect statistically significant association is also an important consideration. The sample sizes (n < 5,000) in GWAS of alcohol dependence to date are much smaller than those of successful GWAS of other diseases such as type 2 diabetes and breast cancer, which used >30,000 subjects (142).

**Candidate Gene Studies**

Genetically influenced metabolic factors have been implicated in the etiology of alcoholism in a number of ethnic groups. The conversion of alcohols to the corresponding aldehydes is catalyzed by ADHs. This is the rate-limiting step in the elimination of ethanol in humans and experimental animals (18). Seven ADH-encoding genes (ADH1A, ADH1B, ADH1C, ADH4, ADH5, ADH6, and ADH7) are located as a cluster on human chromosome 4q22-23 (33). The class 1 enzymes—encoded by ADH1A, ADH1B, and ADH1C (previously termed ADH3) in humans—have a high affinity for ethanol and contribute the most to its conversion to acetaldehyde, particularly during the elimination phase. This class of ADH enzymes includes the most important ADH isoforms for oxidizing ethanol in humans (33). ADH7 acts early in the time course of alcohol metabolism in stomach mucosa that is exposed to high concentrations of alcohol (42).

The majority of association studies investigating the role of alcohol-metabolizing genes in risk for alcohol-use disorders have focused on the well-characterized coding variants within ADH1B, ADH1C, and ALDH2 and on the phenotype of alcohol dependence. There are three different ADH1B alleles (33). The reference allele is ADH1B*1, which encodes the β1 subunit with an arginine at amino-acid positions 48 (Arg48) and 370 (Arg370). ADH1B*2, a common allele in Asians, encodes the β2 subunit with a histidine at position 48 (His48). The ADH1B*3 allele, which encodes the β3 subunit with a cysteine at position 370 (Cys370), is found primarily in people of African descent. Amino-acid substitutions at positions 48 (ADH1B*2) and 370 (ADH1B*3) result in 70–80-fold higher enzyme activity compared with that produced by the ADH1B*1 allele (33). The rapid conversion of ethanol to acetaldehyde causes facial flushing and aversive effects after alcohol consumption and is protective against alcohol dependence (Figure 2) (103). A meta-analysis of the ADH1B*2 allele in Han Chinese and Japanese showed that individuals who are homozygous for this variant (His48/His48) have a five-fold decrease in risk for alcohol dependence compared with individuals who are heterozygous for this variant (Arg48/His48) (144). In Europeans, the risk for developing alcohol dependence is twofold lower in His48/His48 carriers compared with Arg48/His48 carriers (144). Recently, a case-control study in populations of European and African ancestry demonstrated that the ADH1B*2 (His48) allele in these populations is associated with a lower maximum number of drinks in a 24-h period (p = 3 × 10^{-11}) and has a strong protective effect on DSM-IV alcohol dependence in both populations (odds ratio = 0.34, p = 6.6 × 10^{-10}) (16). The protective effect of ADH1B*2 was not detectable by a GWAS approach in studies involving populations of European or African descent because none of the variants on these genotyping chips showed high linkage disequilibrium with this rare variant. These studies demonstrate that the ADH1B*2 allele correlates with reduced alcohol consumption and risk for alcohol dependence in all populations, though the allele frequencies vary in people of different ethnicity. The ADH1B*3 allele also has a protective effect on risk for alcoholism in African American families and southwest California Native Americans (36, 136).
The reference allele of the ADH1C gene is ADH1C*1, with an arginine at position 272 (Arg272) and an isoleucine at position 350 (Ile350). The ADH1C*2 allele, with a glutamine at position 272 (Gln272) and a valine at position 350 (Val350), is common in Europeans and African Americans. The ADH1C*3 allele, with a threonine at position 352 (Thr352), is found in Native Americans (33). Studies have shown that ADH1C*2 also has protective effects on the risk for alcohol dependence in people of Asian and African descent (33, 96). However, some studies showed that the protective effect of the ADH1C*1 allele is not an independent effect owing to the linkage disequilibrium of this allele with the ADH1B*2 allele (24, 103).

Another well-known polymorphism is in the ALDH2 gene, which encodes the aldehyde dehydrogenase 2 family (mitochondrial). The ALDH2*2 allele, which substitutes lysine for glutamate at position 504 (Lys504), results in a nearly inactive protein subunit that is unable to metabolize acetaldehyde (150). This allele is relatively common in Asians but nearly absent in people of European or African descent (69, 102) and is strongly associated with a reduced risk for alcohol dependence (33). Polymorphisms in other ADH genes have also been associated with alcohol dependence (36, 77). Studies have shown that variation in the ADH genes contributes substantially to variation in alcohol metabolism and consequently affects the risk for alcohol dependence. Although the variants ADH1B Arg48His and ADH1C Arg272Gln/Ile350Val are known to have a major effect on enzyme activity in vitro, these variants account for only a very small amount of the genetic variance in vivo (18, 94). In vivo studies in Europeans demonstrated that variants in ADH7 are associated with the early stages of alcohol metabolism, with additional effects in ADH1A, ADH1B, and ADH4 (18). Postabsorptive alcohol metabolism is affected by variants in the ADH7-ADH1C-ADH1B gene cluster. Approximately 20% of the total genetic variance for alcohol metabolism was attributed to the combined effects of variants in the ADH gene region (18). Because patterns of linkage disequilibrium across this genomic region vary among different ethnic populations...
and the frequencies of functional variants differ from one population to another, the effects of functional variants may be population specific.

Gamma amino butyric acid (GABA) is the main inhibitory neurotransmitter in the central nervous system, and its transmission is considered to mediate the pharmacological effects of alcohol in the brain. The modulatory actions of GABA are mediated through two types of receptors: the ionotropic GABA_A receptor and the metabotropic GABA_B receptor (57, 137). A family study from the COGA group identified multiple SNPs in GABRA2 (which encodes the GABA_A receptor α2 subunit) associated with increased risk for alcohol dependence (34). Several subsequent studies using case-control samples replicated the association between GABRA2 and alcohol dependence, though the nature of the association and the specific variants associated with alcohol dependence differ in some samples (27, 40, 43, 80, 118). Furthermore, one study showed that GABRA2 alleles affect the SRE (self-rating of the effects of alcohol), suggesting that genetic variations in GABRA2 might play a role in the risk for alcohol-use disorders by moderating the SRE (111). Evidence from a functional MRI study suggested that a SNP in GABRA2 (rs279871) associated with alcohol dependence is also associated with the medial frontal response to alcohol cues (72).

Adjacent to GABRA2 is the GABRG1 gene, which encodes the GABA_A receptor γ1 subunit. Several studies have reported association of GABRG1 variants with the risk for alcohol dependence and drinking behaviors (39, 107). Haplotype analyses have suggested that markers in the GABRA2 gene associated with alcohol dependence are in linkage disequilibrium with markers in the GABRG1 gene in many populations, indicating that the association with GABRA2 may be driven by variants in GABRG1 (28, 71). Despite multiple studies implicating SNPs in GABRA2 and GABRG1 in the risk for alcohol-related behaviors, the specific functional alleles underlying these associations have yet to be identified.

In summary, candidate gene studies have successfully detected functional variants in alcohol metabolism genes such as ADH1B, ADH1C, and ALDH2 associated with a risk of developing alcohol dependence in populations of Asian descent. Although alleles associated with reduced risk in Asians are rare in populations of African and European descent, they also reduce risk for alcohol dependence in these populations. Several GWAS of alcoholism have produced no conclusive evidence for specific genetic risk factors. The heterogeneous nature of the ascertainment strategies and the phenotypic measures used across studies could potentially explain the lack of a replicated association. Furthermore, the sample sizes in current alcohol studies are small compared with those in GWAS of other psychiatric disorders, limiting the power to detect genetic risk factors.

IDENTIFICATION OF GENETIC RISK FACTORS FOR NICOTINE DEPENDENCE

Genome-Wide Linkage Studies

To identify susceptibility loci for nicotine dependence, more than 20 linkage analyses across the entire genome have been conducted using a family-based and/or sib-pair design (for a review, see 83). Although a number of genomic regions were identified as significant or suggestive for harboring susceptibility loci for nicotine dependence or smoking-related phenotypes, only four linkage regions have been replicated in four or more independent samples; these reside on human chromosomes 9q, 10q, 11p, and 17p (82). Recently a genome-wide linkage scan suggested that a region on human chromosome 2q31.1 confers risk for the development of nicotine dependence with a broad range of dependence symptoms rather than a specific aspect of the disorder (51).

Genome-Wide Association Studies

In contrast to the studies of alcohol dependence, GWAS of smoking behavior have
reported consistent and compelling genetic evidence for association. The first GWAS using a case-control sample reported evidence that variants within the nicotinic acetylcholine receptor (nAChR) subunit genes on human chromosomes 15q (CHRNA5–CHRNA3–CHRNA4) and 8p (CHRNA6–CHRNβ4) influence risk for nicotine dependence, as defined by scores on the Fagerström test for nicotine dependence (17). The chromosome 15 association has been replicated in subsequent GWAS either directly or indirectly using highly correlated SNPs ($r^2 > 0.8$), with cigarettes per day (CPD) as a quantitative variable to define heavy- and light-smoking individuals (13, 17, 122, 141). Genome-wide association meta-analyses for CPD further confirmed that variants in CHRNA5, CHRNA3, and CHRNA4 are associated with the risk of developing heavy smoking (87, 123, 124). In addition, the GWAS reported by Thorgeirsson et al. (123) showed that variation in the CHRNA6–CHRNβ3 gene cluster on human chromosome 8 is associated with CPD at a genome-wide significance level.

Genetic variation in nicotine metabolism also plays an important role in cigarette consumption (93, 115) and nicotine dependence (8). Conversion of nicotine to cotinine typically accounts for 70%–80% of nicotine metabolism, the majority of which is catalyzed by the cytochrome P450 2A6 (CYP2A6) enzyme (67). Recent GWAS meta-analyses using subjects of European descent identified SNPs in the region of CYP2A6 associated with CPD (123, 124).

**Candidate Gene Studies**

In parallel with these GWAS, several studies using a candidate gene approach have also reported the association of SNPs in the CHRNA5–CHRNA3–CHRNA4 gene cluster with nicotine dependence and smoking quantity (17, 113, 141). Furthermore, a fine mapping study (113) observed that the nonsynonymous SNP rs16969968 in exon 5 of CHRNA5 has consistent effects on the risk for nicotine dependence in both European (odds ratio = 1.40, 95% confidence interval = 1.23–1.59) and African (odds ratio = 2.04, 95% confidence interval = 1.15–3.62) populations, despite a large difference in allele frequency for the SNP. A second locus tagged by rs5787776 in the 3′ untranslated region of CHRNA3 that has low linkage disequilibrium with rs16969968 is associated with nicotine dependence in European Americans but not in African Americans. Another linkage disequilibrium bin tagged by an intronic SNP in CHRNA5, rs588765, confers a protective effect for nicotine dependence in populations of European descent (Figure 3) (113, 139). A comprehensive meta-analysis involving more than 32,000 subjects confirmed the three unique loci in this gene cluster that affect smoking quantity (112). In Asians, a locus tagged by rs5787776 overlapped with a locus tagged by rs588765, and variants in this distinctive linkage disequilibrium pattern were reported to influence smoking initiation, smoking cessation (84), and smoking quantity (84, 147).

There are at least two distinct biological mechanisms in the nAChR gene cluster on chromosome 15 that alter the risk for developing nicotine dependence. One mechanism involves the variant rs16969968 (D398N), which likely alters protein structure and receptor function. An in vitro functional analysis demonstrated that the maximal response to agonist per receptor was twofold higher for the $\alpha_4$β2α5D398 nAChR variant relative to the $\alpha_4$β2α5N398 nAChR variant (17). The second potential mechanism is altered mRNA expression of CHRNA5 (139, 140). Several variants located upstream of the coding region and within intronic regions of CHRNA5 (i.e., rs588765) are strongly associated with the variability in CHRNA5 mRNA expression observed in the human frontal cortex. Subjects homozygous for the minor allele of rs588765 showed a 2.9-fold increase in CHRNA5 mRNA expression compared with subjects homozygous for the major allele (139, 140). The rs588765 polymorphism and highly correlated variants are only weakly correlated with the D398N variant. The N398 variant, which greatly increases risk for nicotine dependence, occurs
Figure 3
Three distinct loci across the CHRNA5-CHRNA3-CHRNB4 gene cluster on human chromosome 15. Triangles represent single-nucleotide polymorphisms (SNPs). SNPs in the same color are highly correlated with each other but not with SNPs in other colors. SNP position is not drawn to scale.

primarily on the background of low mRNA expression of CHRNA5. The nonrisk variant D398 occurs on both high- and low-expression alleles. The risk for nicotine dependence is significantly lower when D398 occurs on a background of low CHRNA5 mRNA expression than when it occurs on a background of high CHRNA5 mRNA expression (139).

Studies examining genetic and environmental risks for nicotine dependence have shown that there is an interaction between environmental factors and the rs16969968 variant that has an effect on smoking. The genetic risk associated with rs16969968 was reduced in subjects with high levels of parent monitoring and increased in subjects with low levels of parent monitoring (23). Interaction between childhood adversity and rs16969968 is also associated with the risk for nicotine dependence in men (149): Among men who experienced childhood adversity, individuals who carry the AA risk genotype have the highest risk of developing nicotine dependence compared with individuals who carry the GA or GG genotype.

A study that sequenced all genes encoding nicotinic receptor subunits has demonstrated that the low-frequency coding variants R37H in CHRNA3 and T375I and T91I in CHRNB4 decrease the risk for nicotine dependence among regular smokers (55). It further showed that the minor allele of each polymorphism increases the cellular response to nicotine (β4T375I p = 0.01, β4T91I p = 0.02, α3R37H p = 0.003), but the largest effect on in vitro receptor activity was seen in the presence of both CHRNB4 T91I and CHRNA3 R37H (p = 2 × 10^{-6}), two SNPs in strong linkage disequilibrium in human populations (r^2 = 0.89, n = 2,035 European Americans; r^2 = 0.59, n = 710 African Americans).

Nicotine is the major substance in tobacco responsible for addiction among cigarette smokers (62). An in vivo study has shown that approximately 80% of nicotine consumed is metabolically inactivated to cotinine (12); approximately 90% of this conversion is mediated by CYP2A6 (92, 99). The next step of nicotine metabolism, which oxidizes cotinine to...
form *trans*-3′-hydroxycotinine, is entirely catalyzed by CYP2A6 (98). CYP2A6 is a highly polymorphic enzyme. Different CYP2A6 alleles have different functional consequences, and the frequency of CYP2A6 alleles varies among ethnic populations (97). A number of studies have reported association between reduced or absent CYP2A6 enzyme activity and lower risk of smoking, including decreased cigarette consumption, smoking intensity, and withdrawal symptoms; shorter smoking duration; and increased cessation. However, some studies have failed to detect any association between CYP2A6 variation and smoking status (64). A recent study using quantified measures of deuterated (D₂)-cotinine/(D₂-cotinine + D₂-nicotine) following oral administration in 189 European Americans demonstrated that CYP2A6*12 is a loss-of-function allele indistinguishable from CYP2A6*4 and CYP2A6*2 alleles, and that the CYP2A6*1B 5′ untranslated region conversion has a negligible impact on metabolism (19). After controlling for the CYP2A6 genotype, the authors found modest associations between increased metabolism and both gender and current smoking (19).

In summary, genetic studies of nicotine dependence have successfully identified risk factors using both GWAS and candidate gene approaches. The consistent phenotypic measure—CPD—is easily obtained in large cohort studies and has been successfully used in meta-analyses of the genetics of smoking. These studies have greatly increased the power to detect genetic risk factors for nicotine consumption. However, these associated genetic factors explain only a small percentage of the variance in nicotine consumption, indicating that further research to detect other genetic factors influencing smoking is warranted.

**IDENTIFICATION OF GENETIC RISK FACTORS FOR ILLICIT DRUG DEPENDENCE**

**Genome-Wide Linkage Studies**

Cannabis is the most widely used illicit drug. A genome-wide linkage analysis of cannabis dependence and related phenotypes in individuals from the UCSF Family Alcoholism Study identified genome-wide significant linkage (LOD score of 3 or higher) for cannabis craving and withdrawal symptoms for regions on human chromosomes 1, 3, 6, 7, and 9; no evidence for linkage with cannabis dependence reached genome-wide significance (38). Loci on human chromosomes 3 (3q21) and 9 (9q34), which are close to the regions linked to cannabis withdrawal in the UCSF study, were also suggested to influence cannabis-dependence symptoms in adolescents who participated in a Colorado Center on Antisocial Drug Dependence study (65). A Native American community study detected genome-wide significant linkage with the severe cannabis use/antisocial subtype on human chromosomes 16 (LOD score of 4.4) and 19 (LOD score of 6.4) (37).

For other illicit drugs, significant linkage peaks have been identified on human chromosomes 9 (a region approximately 40 cM upstream of the region linked with cannabis use) and 12 for cocaine dependence (48), on chromosome 17 at 103.5 cM for a heavy-opioid-use cluster-defined trait (49), and on 14q for DSM-IV opioid dependence (79). Genome-wide linkage analysis of heroin dependence in Han Chinese reported several linkage regions, but none reached genome-wide significance (52). An analysis in families severely affected by alcohol-use disorders reported significant linkage on human chromosome 2 (LOD score of 3.2) with illicit drug dependence (1).

**Genome-Wide Association Studies**

There have not been many GWAS of illicit-drug-use disorders. A study using 708 DSM-IV cannabis-dependent cases and 2,346 cannabis-exposed nondependent controls from the Study of Addiction: Genetics and Environment data set showed a suggestive association between cannabis dependence and variants in the ANKFRN1 gene on human chromosome 17 (4). In a sample of 325 methadone-stabilized, formerly severe heroin addicts and 250 control individuals, Nielsen et al. (100) used a pooled
GWAS approach to find variants associated with vulnerability to heroin addiction.

**Candidate Gene Studies**

Genes involved in dopamine neurotransmission are biologically plausible candidates for association with cocaine dependence because dopamine pathways play a major role in drug reward (70, 78). Genetic association analysis of dopamine receptors and transporter genes found both positive and negative associations (76). These discrepancies may be due to small sample size as well as the complex nature of the phenotype.

**OPRM1**, which codes for the G protein-coupled mu opioid receptor, is the primary site of action of most opioids. A nonsynonymous SNP in exon 1 of **OPRM1**, A118G, is the most commonly studied variant for opioid dependence, but its association is controversial. Several studies have reported a positive association between variants in **OPRM1** and opiate (including heroin) dependence (11, 20, 81, 120), whereas other studies did not detect an association (101, 138). A study using sensory neurons isolated from a humanized mouse model showed that the A118G missense variant of **OPRM1** modulates the morphine and fentanyl pharmacological profile (89). Morphine is approximately fivefold less potent and 26% less efficacious in neurons with the 118GG genotype than it is in neurons with the 118AA genotype. However, there is no difference in the potency and efficacy of the agonist fentanyl in neurons with different genotypes.

Two well-characterized cannabinoid receptors associated with the endocannabinoid signaling system, CB1 (CNR1) and CB2 (CNR2), have been reported to be associated with vulnerability to psychiatric disorders, including substance abuse (130). Studies using **CNR1**-knockout mice have reported that the mice display alterations in reward- and drug-seeking behaviors in response to psychostimulants, including alcohol (25, 106), nicotine (32, 44), cocaine, and amphetamine (90). The most-studied genetic variant in **CNR1** is the (AAT)ₙ trinucleotide short-tandem repeat, which was reported to be associated with intravenous administration of drugs of abuse (26). However, other studies have not confirmed this finding (10, 29, 85). Several other variants in **CNR1** have been reported to be associated with cannabis dependence (3), cannabis-dependence symptoms (66), cocaine dependence (152), and other substance dependences (63, 114, 151).

Interestingly, the rs16969968 nonsynonymous variant in the α₅ nAChR subunit is also associated with cocaine dependence, but the minor allele reduces the risk for cocaine dependence, which is the opposite of the effect reported for nicotine dependence (53).

**SEQUENCING APPROACHES TO IDENTIFY VARIANTS THAT COULD EXPLAIN THE MISSING HERITABILITY FOR SUBSTANCE DEPENDENCE**

One drawback of the GWAS method is its reliance on linkage disequilibrium. This means that this approach is good for identifying variants that are common in the general population (>1%) but misses rare variants with larger effects on risk that have low linkage disequilibrium with common variants detected with standard genotyping chips. As a way to uncover the missing heritability factors that influence the risk for psychiatric diseases, including addiction, next-generation DNA sequencing combined with the results of association and perhaps linkage studies holds the promise of identifying a larger set of susceptibility loci (9, 46). In contrast to GWAS, the sequencing of targeted genomic regions identified from GWAS or linkage analysis, or whole-exome or whole-genome sequencing, improves the ability to discover novel causative or highly penetrant mutations for human diseases.

Rare variants have been shown to be risk factors for some complex disorders, but their role in psychiatric disorders and especially addiction-related phenotypes is largely unexplored (75). A few studies have shown associations between rare variants in nicotinic
receptor genes and nicotine dependence (55, 143, 148).

Several variants in the alcohol metabolism genes (i.e., \textit{ADH1B}) and nicotine metabolism genes (i.e., \textit{CYP2A6}) have low frequencies (1\%–5\% minor allele frequency) and generally reduce risk for dependence, suggesting that human populations might be genetically predisposed to develop addiction, with rare variant alleles leading to reduced risk. This is supported in the nicotine literature, in which most people who smoke develop some symptoms of dependence, whereas only a small proportion smoke without developing any symptoms of dependence. It could be that as a result of some unknown evolutionary selection pressure, most people are predisposed to addiction when exposed to substances.

**IMPLICATIONS OF RECENT FINDINGS ON HEALTH BEYOND ADDICTION**

**Nicotinic Receptors and Lung Cancer and Chronic Obstructive Pulmonary Disease**

GWAS approaches have revealed that several SNPs within the \textit{CHRNA5-CHRNA3-CHRNB4} gene cluster are significantly associated with the risk of lung cancer and COPD (5, 68, 117, 122). The most strongly associated SNPs are the same as those that show association with nicotine dependence and CPD in other studies (13, 112, 122, 124). It is unclear whether the association of this locus with lung cancer is a direct biological effect on lung cancer susceptibility or is mediated through effects on increased risk of smoking. Although SNPs at this locus are only weakly associated with lung cancer risk in those who have never smoked, they are associated with risk for other smoking-associated cancers and diseases (86, 126). This implies that this locus predisposes individuals to increased tobacco consumption, leading to increased risk for cancer. However, some studies have suggested a direct link between \textit{CHRNA5-CHRNA3-CHRNB4} variants and lung cancer: The risk of lung cancer that can be attributed to the \textit{CHRNA5-CHRNA3-CHRNB4} variants is higher than can be explained by the variants’ effect on smoking quantity (122), and the genetic risk for lung cancer and COPD remains after the risk associated with smoking has been statistically accounted for using CPD and the duration of smoking (86). Other studies, however, have shown that the amount of nicotine absorbed by smokers is not fully accounted for by CPD owing to differences in how individuals smoke (56, 60).

The α5 nAChR subunit is expressed in lung tissue, and a 30-fold upregulation of \textit{CHRNA5} mRNA expression is seen in lung cancer tissue compared with normal lung tissue (41). In addition, tobacco smoke and nicotine can both mediate the stepwise overexpression of nAChR subtypes, which leads to increased Ca$^{2+}$ permeability in exposed cells (6). Thus, a switch in the nAChR composition (involving the α3 and α5 subunits, among others) could change receptor function, leading to pathologic effects in nicotine-exposed cells. SNPs in the \textit{CHRNA5-CHRNA3-CHRNB4} gene cluster could therefore contribute to an increased risk of nicotine dependence and to lung cancer independently and on two levels: (a) by increasing the number of cigarettes smoked and the likelihood of nicotine dependence, and (b) by inserting themselves into the pathophysiological cascade that leads to lung cancer (134).

**ADH and ALDH2 Genes and Esophageal Cancer**

Variants in \textit{ADH1B} and \textit{ALDH2} that influence alcohol consumption and alcohol dependence also play a role in the risk for upper aero-digestive tract (UADT) cancer. The protein encoded by the \textit{ADH1B}*2 allele, which is associated with reduced risk for alcohol dependence, has increased enzyme activity. The protein encoded by the \textit{ALDH2}*2 allele, a protective allele for alcohol dependence, has almost zero enzyme activity (33). Individuals who carry these two alleles have much higher levels of acetaldehyde (a carcinogen) compared with
noncarriers if they consume alcohol (Figure 2). Studies have shown that, owing to the accumulation of acetaldehyde in the blood, individuals who have the alcohol flushing response are at higher risk for esophageal cancer (21).

GWAS have also identified a significant association between esophageal squamous cell cancer and SNPs on human chromosomes 4q21-23 and 12q24, which include the functional variants rs1229984 in ADH1B and rs671 in ALDH2, respectively (30). Tanaka et al. (121) demonstrated that the interaction of ADH1B and/or ALDH2 risk alleles with smoking and alcohol consumption significantly increases the risk for the development of esophageal squamous cell carcinoma. Studies have also shown that the combination of alcohol consumption with the inactive heterozygous ALDH2 genotype (ALDH2∗1/∗2) and less-active homozygous ADH1B genotype (ADH1B∗1/∗1) increases the risk of UADT squamous cell carcinoma in central European (59) and Japanese (7) populations. The effect of ALDH2∗1/∗1 results from the high level of acetaldehyde; the effect of ADH1B∗1/∗1 is due to heavy drinking that leads to longer exposure of the UADT to salivary ethanol and acetaldehyde. These studies point to significant gene-environment interactions that potentially lead toward complex pathophysiological pathways for the development of such diseases.

CONCLUSION

Genomic approaches are beginning to provide clues to the underlying genetic etiology of addiction, and have demonstrated that exposure to these substances in combination with genetic vulnerability to addiction plays an important role in the risk for common cancers previously associated with substance use.

SUMMARY POINTS

1. Genetic studies have identified functional alleles in alcohol metabolism genes (ADH, which encodes alcohol dehydrogenase, and ALDH2, which encodes aldehyde dehydrogenase) that influence the risk for alcohol dependence. The interaction of genetic (polymorphisms of ADH and ALDH2 genes) and environmental (heavy drinking) factors is associated with risk for UADT cancers.

2. Recent GWAS have successfully identified variants in the α3, α5, and β4 subunits of nAChR associated with risk of nicotine dependence. However, GWAS on alcoholism have not provided conclusive evidence for specific genetic factors for this type of addiction. It is likely that CPD, a common phenotype used in nicotine-dependence GWAS, is a more consistent measurement than other quantitative measures of other substance-use disorders. Harmonized phenotypic measures provide a convenient method for combining small cohorts into a large sample with more than 10,000 subjects, increasing the power to detect statistically significant association.

3. Although the GWAS approach has been successfully used to investigate the genetic influence on smoking, the association of CHRNA5-CHRNA3-CHRNB4 with nicotine dependence explains only a small part of this addiction’s heritability. A significant fraction of the genetic variance remains unexplained despite the use of very large sample sizes. This missing heritability may be explained by rare variants with large effect. More robust DNA sequencing approaches can potentially identify this missing heritability.

4. The roles of alcohol metabolism genes in esophageal cancer and nicotinic receptors in lung cancer indicate significant gene-environment interaction in cancer vulnerability.
DISCLOSURE STATEMENT

A.M.G. and J.C.W. are listed as inventors on issued US patent 8,080,371, “Markers for Addiction,” covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction.

ACKNOWLEDGMENTS

This work is supported by the following projects: the Collaborative Study on the Genetics of Alcoholism, funded by NIH grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism and the National Institute on Drug Abuse (NIDA); the Collaborative Genetic Study of Nicotine Dependence, funded by NIH grant P01 CA89392 from NIDA and the National Cancer Institute; and the Genetics of Vulnerability to Nicotine Addiction, funded by NIH grant 5R01 DA12854 from NIDA.

LITERATURE CITED


   wide association study of alcoholism risk in the community: findings and implications. Biol. Psychiatry
   70:513–18
   141B:499–503
64. Ho MK, Tyndale RF. 2007. Overview of the pharmacogenomics of cigarette smoking. Pharmacogenomics
   7:81–98
   loci influencing adolescent cannabis dependence symptoms: evidence for linkage on chromosomes 3 and
   9. Drug Alcohol Depend. 89:34–41
   Rev. 57:79–115
   cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452:633–37
70. Hyman SE, Malenka RC, Nestler EJ. 2006. Neural mechanisms of addiction: the role of reward-related
   librium patterns of GABRA2 and GABRG1 genes at the GABA cluster locus on human chromosome 4.
   Genomics 91:61–69
   34:2169–78
   of symptoms of alcohol dependence in the molecular genetics of schizophrenia (MGS2) control sample.
74. Kendler KS, Schmitt E, Aggen SH, Prescott CA. 2008. Genetic and environmental influences on alcohol,
   caffeine, cannabis, and nicotine use from early adolescence to middle adulthood. Arch. Gen. Psychiatry
   65:674–82
   rare coding variants identify ABCA13 as a candidate gene in schizophrenia, bipolar disorder, and
   genetics of opiate and cocaine addictions and their treatments. Pharmacol. Rev. 57:1–26
   genes with alcohol dependence in the Irish Affected Sib Pair Study of alcohol dependence (IASPSAD)
   43
   linkage of opioid dependence to chromosome 1q4. Hum. Mol. Genet. 16:1327–34
   Res. 29:493–98


142. Wellcome Trust Case Control Consort. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–78


# Contents

Human Genetic Individuality  
*Maynard V. Olson*  ................................................................. 1

Characterization of Enhancer Function from Genome-Wide Analyses  
*Glenn A. Maston, Stephen G. Landt, Michael Snyder, and Michael R. Green*  ........... 29

Methods for Identifying Higher-Order Chromatin Structure  
*Samin A. Sajan and R. David Hawkins* ........................................ 59

Genomics and Genetics of Human and Primate Y Chromosomes  
*Jennifer F. Hughes and Steve Rozen* ......................................... 83

Evolution of the Egg: New Findings and Challenges  
*Katrina G. Claw and Willie J. Swanson* .................................... 109

Evolution of the Immune System in the Lower Vertebrates  
*Thomas Boehm, Norimasa Iwanami, and Isabell Hess* ..................... 127

The Human Microbiome: Our Second Genome  
*Elizabeth A. Grice and Julia A. Segre* ..................................... 151

Functional Genomic Studies: Insights into the Pathogenesis of Liver Cancer  
*Ze-Guang Han*  ........................................................................ 171

A Comparative Genomics Approach to Understanding Transmissible Cancer in Tasmanian Devils  
*Janine E. Deakin and Katherine Belov* .................................... 207

The Genetics of Sudden Cardiac Death  
*Dan E. Arking and Nona Sotoodehnia* ....................................... 223

The Genetics of Substance Dependence  
*Jen-Chyong Wang, Manav Kapoor, and Alison M. Goate* ............... 241

The Evolution of Human Genetic Studies of Cleft Lip and Cleft Palate  
*Mary L. Marazita* ................................................................. 263

Genomic Analysis of Fetal Nucleic Acids in Maternal Blood  
*Yuk Ming Dennis Lo and Rossa Wai Kwun Chiu* ......................... 285
Enzyme Replacement Therapy for Lysosomal Diseases: Lessons from 20 Years of Experience and Remaining Challenges
R.J. Desnick and E.H. Schuchman .................................................. 307

Population Identification Using Genetic Data
Daniel John Lawson and Daniel Falush ........................................ 337

Evolution-Centered Teaching of Biology
Karen Burke da Silva ................................................................. 363

Ethical Issues with Newborn Screening in the Genomics Era
Beth A. Tarini and Aaron J. Goldenberg ...................................... 381

Sampling Populations of Humans Across the World: ELSI Issues
Bartha Maria Knoppers, Ma’n H. Zawati, and Emily S. Kirby .......... 395

The Tension Between Data Sharing and the Protection of Privacy in Genomics Research
Jane Kaye .................................................................................. 415

Genetic Discrimination: International Perspectives
M. Otlowski, S. Taylor, and Y. Bombard ....................................... 433

Indexes

Cumulative Index of Contributing Authors, Volumes 4–13 .............. 455
Cumulative Index of Chapter Titles, Volumes 4–13 ......................... 459

Errata

An online log of corrections to Annual Review of Genomics and Human Genetics articles may be found at http://genom.annualreviews.org/errata.shtml