Convergence of Genome-Wide Association and Candidate Gene Studies for Alcoholism

Emily Olfson and Laura Jean Bierut

Background: Genome-wide association (GWA) studies have led to a paradigm shift in how researchers study the genetics underlying disease. Many GWA studies are now publicly available and can be used to examine whether or not previously proposed candidate genes are supported by GWA data. This approach is particularly important for the field of alcoholism because the contribution of many candidate genes remains controversial.

Methods: Using the Human Genome Epidemiology (HuGE) Navigator, we selected candidate genes for alcoholism that have been frequently examined in scientific articles in the past decade. Specific candidate loci as well as all the reported single nucleotide polymorphisms (SNPs) in candidate genes were examined in the Study of Addiction: Genetics and Environment (SAGE), a GWA study comparing alcohol-dependent and nondependent subjects.

Results: Several commonly reported candidate loci, including rs1800497 in *DRD2*, rs698 in *ADH1C*, rs1799971 in *OPRM1*, and rs4680 in *COMT*, are not replicated in *SAGE* ($p > 0.05$). Among candidate loci available for analysis, only rs279858 in *GABRA2* ($p = 0.0052$, OR = 1.16) demonstrated a modest association. Examination of all SNPs reported in SAGE in over 50 candidate genes revealed no SNPs with large frequency differences between cases and controls, and the lowest p-value of any SNP was 0.0006.

Conclusions: We provide evidence that several extensively studied candidate loci do not have a strong contribution to risk of developing alcohol dependence in European and African ancestry populations. Owing to the lack of coverage, we were unable to rule out the contribution of other variants, and these genes and particular loci warrant further investigation. Our analysis demonstrates that publicly available GWA results can be used to better understand which if any of previously proposed candidate genes contribute to disease. Furthermore, we illustrate how examining the convergence of candidate gene and GWA studies can help elucidate the genetic architecture of alcoholism and more generally complex diseases.

Key Words: Alcohol Dependence, Candidate Genes, Genome-Wide Association Studies, Genetics.

G ENOME-WIDE ASSOCIATION (GWA) studies have revolutionized the search for common genetic variants that influence individual risk for complex diseases. Before this revolution, the discovery of genetic associations was dominated by candidate gene studies that used targeted gene approaches. Examination of these previous gene association studies demonstrates that most reported associations are not consistently replicated (Hirschhorn et al., 2002) and the strength of genetic associations in initial studies commonly erodes in subsequent research (Ioannidis et al., 2001). Despite this suggested irreproducibility, many candidate gene association studies continue to be published annually (Yu et al., 2008).

GWA studies rapidly evaluate millions of single nucleotide polymorphisms (SNPs) throughout the genome and therefore have the potential for identifying key variants in complex diseases. Since the publication of the first GWA study in 2005 (Klein et al., 2005), over 1,000 GWA studies have established genetic associations of more than 200 traits, many of which are complex diseases. SNP–trait associations from published GWA studies are readily available to investigators through "A Catalog of Genome-Wide Association Studies" by the National Human Genome Research Institute (www.genome.gov/gwastudies). More recently, several data sets from GWA studies have also become available to the scientific community through the database of Genotypes and Phenotypes (dbGaP) maintained by NCBI (Mailman et al., 2007). These online scientific databases provide opportunities for investigators to access GWA data.

Online databases can specifically be used to evaluate whether genes that were previously suggested in candidate gene studies are replicated in GWA studies. Research by Siontis and colleagues (2010) demonstrates that only a few of previously proposed candidate loci of common diseases reached genome-wide significance in GWA studies. The loci that did replicate, however, important genetic effects and

included variants implicated in Alzheimer's disease and statin-induced myopathy. Similarly, a recent analysis by Obeidat and colleagues (2011) examined genetic associations with lung function measures to evaluate the role of previously associated genes in a large GWA study and clarified the role of many controversial associations. This approach of comparing candidate gene and GWA studies is powerful because it highlights which findings have consistent scientific evidence and therefore merit being pursued in future studies. These findings prompted us to examine whether proposed candidate genes associated with alcohol dependence are supported by GWA data.

Genetic and environmental factors contribute to individual susceptibility to alcohol dependence. Twin studies estimate that heritable influences explain 47 to 64% of variance in risk for alcohol dependence (Heath et al., 1997; Knopik et al., 2004). Several past research efforts have focused on targeted gene approaches to shed light on the genes that underlie these heritable influences. This has led to the proposal of hundreds of candidate genes that contribute to the development of alcohol dependence (Yu et al., 2008). A few GWA studies have also explored genes potentially involved in alcohol dependence (Bierut et al., 2010; Edenberg et al., 2010; Farrer et al., 2009; Heath et al., 2011; Hodgkinson et al., 2010; Joslyn et al., 2010; Treutlein et al., 2009; Zlojutro et al., 2011). Despite extensive candidate gene studies and several GWA studies, little consensus exists over which genes contribute to the genetic basis of alcohol dependence.

The existence of many controversial candidate genes for alcoholism highlights the need for further research on whether or not these genes replicate in large data sets. Results from the Study of Addiction: Genetics and Environment (SAGE) have recently become available through dbGaP. SAGE compares DSM-IV alcohol-dependent individuals and nondependent, unrelated control subjects of European and African American descent. Using the SAGE data, we examined differences in SNP frequencies between cases and controls within previously reported candidate genes. These targeted candidate genes were selected using the Human Genome Epidemiology (HuGE) Navigator, a publicly searchable database established in 2001 of published genetic association and human genome epidemiological studies (Yu et al., 2008). The HuGE Navigator along with the SAGE results facilitated the systematic examination of candidate genes considered in many alcoholism studies over the last decade.

MATERIALS AND METHODS

Selection of Candidate Genes

The HuGE Navigator was developed using PubMed abstracts as the core data source and using data and text mining algorithms to develop a knowledge database (Yu et al., 2008). Each week since 2001, articles are systematically deposited in the database and represent a comprehensive list of recent articles. An automatic literature program screens PubMed for abstracts, and then, a genetic epidemiologist selects abstracts meeting inclusion criteria and indexes them. Phenopedia of the HuGE Navigator gives a disease-centered view

of genetic association studies by providing information about genes studied in relation to a queried phenotype (Yu et al., 2010). Phenopedia was queried in July 2011 for alcoholism, and 584 genes were retrieved.

We focused our study on genes that have been frequently characterized by candidate gene studies. In primary analysis, over 90% of the genes associated with alcoholism in the HuGE database have 5 or fewer publications (528 of 584 genes). The 56 candidate genes that have more than 5 publications vary substantially in the number of publications (6 to 103 publications). Figure 1A highlights that many genes have 1 or a few reported publications and there are some outliers that have been examined in many papers. This distribution may be explained in part by the fact that many of the genes with a low number of publications have been primarily identified in a GWA study and are not well characterized in targeted candidate gene studies. Figure 1B demonstrates that for almost 50% (176/386) of the genes with 1 publication, that publication is itself a GWA study. Based on these preliminary observations, we narrowed our investigation to genes with more than 5 publications to focus our analysis on well-studied genes.

As the X chromosome is not included in the publicly available SAGE results, the 2 candidate genes on the X chromosome *MAOA*
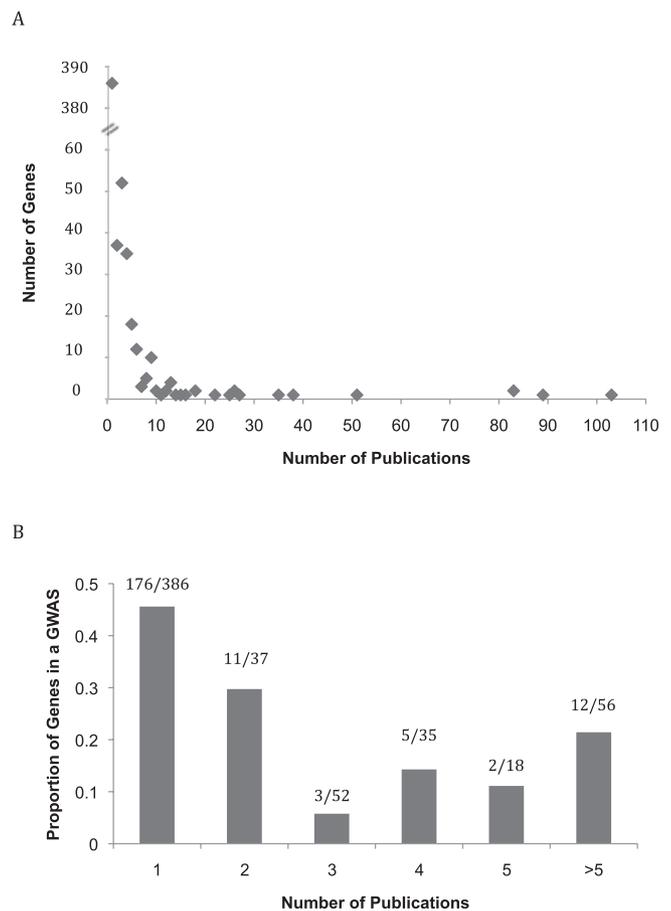
A

B

Fig. 1. Characteristics of genes associated with alcoholism in the Human Genome Epidemiology (HuGE) Navigator. (**A**) Distribution of number of publications on genes; and (**B**) proportion of genes observed in at least 1 GWA study stratified based on number of publications. Four genes were identified in 2 GWA studies (these genes had 2, 2, 10, and 51 publications). All other genes were found in none or 1 GWA study. A total of 8 GWA studies on alcoholism are listed in the HuGE database (Bierut et al., 2010; Edenberg et al., 2010; Farrer et al., 2009; Heath et al., 2011; Hodgkinson et al., 2010; Joslyn et al., 2010; Treutlein et al., 2009; Zlojutro et al., 2011).

and *HTR2C*, which have 26 and 8 publications, respectively, were excluded from the analysis. The 54 autosomal genes that had more than 5 publications were pursued using the SAGE data set. For the remainder of this paper, we will only refer to the 54 autosomal candidate genes.

SAGE Data

SAGE is a case–control study that analyzed genetic data on over 3,800 phenotyped subjects funded as part of the Gene Environment Association Studies (GENEVA) initiative supported by the National Human Genome Research Institute (Bierut et al., 2010). Alcohol-dependent cases and controls were selected from 3 large data sets: the Collaborative Study on the Genetics of Alcoholism (COGA), the Family Study of Cocaine Dependence (FSCD), and the Collaborative Genetic Study of Nicotine Dependence (COGEND). Cases are required to have a lifetime history of DSM-IV alcohol dependence. Controls are required to have been exposed to alcohol because alcohol use is necessary to develop dependence but not to have met lifetime diagnosis criteria for alcohol dependence or dependence for illicit drugs. A common assessment was performed for all cases and controls in the 3 studies that was based on the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz et al., 1994). The common methodology of interview administration, question format, and queried domains enabled phenotypic standardization across the 3 studies (Bierut et al., 2010). Characteristics of the cases and controls in the SAGE data set are listed in Table 1, and additional information is available at http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000092.v1.p1. The SAGE data set is publicly searchable through the Genome Brower under the Analysis tab on this website.

The power Calculator for Association (CATS) was used to determine what effect sizes the SAGE data set is able to detect (Skol et al., 2006). Using a sample size of 1,900 cases and 1,900 controls and an alpha level of 0.05, we calculated the power to detect associations at different allele frequencies and risk ratios.

Examination of SNPs in Candidate Genes

The HuGE database was used to survey articles on the 10 candidate genes that had the most publications (listed in Table 2). The

Table 1. Characteristics of Alcohol-Dependent Cases and Nondependent Controls in the SAGE Data Set

Characteristic	Cases $n = 1,897$	Controls $n = 1,932$	Total $n = 3,829$
Sex, n (%)			
Males	1.155 (60.9)	606 (31.4)*	1,761 (46.0)
Females	742 (39.1)	1,326 (68.6)	2,068 (54.0)
Age, years			
Mean ± SD	39.0 ± 9.3	39.3 ± 9.1	19.2 ± 9.2
Range	18.0–77.0	18.0–65.0	18.0–77.0
Self-reported race, n (%)			
European American	1,235 (65.1)	1,433 (74.2)*	2,668 (69.5)
African American	662 (34.9)	499 (25.8)	1,161 (30.3)
Self-reported Ethnicity, n (%)			
Hispanic	76 (4.0)	56 (2.8)	132 (3.4)
Alcohol dependence Diagnosis, n (%)	1,897 (100.0)	0 (0.0)*	1,897 (49.5)

Sex, age, race, ethnicity, and alcohol dependence characteristics of cases and controls in the Study of Addiction: Genetics and Environment (SAGE) dataset (Bierut et al., 2010).
*$p < 0.0001$ for difference between cases and controls.

most well-established loci based on expert opinion of the literature for each of the top 10 candidate genes were searched in the genome browser to test whether candidate loci that had been highly reported in candidate gene studies replicated in the SAGE data set. As allele A9 for *SLC6A3* is a VNTR, we examined the 2 SNPs rs27072 and rs27048 as proxies because they have been found to be associated with similar withdrawal symptoms and are roughly in the same region of the gene as the VNTR (Le Strat et al., 2008). As the originators of the SAGE data set, we were also able to compare the odds ratios and *p*-values within the original 3 data sets (COGA, FSCD, and COGEND) to verify whether there was any heterogeneity across the 3 contributing studies.

The 54 candidate genes with more than 5 publications were identified, and chromosomal regions containing the gene plus 10 kb both 5' and 3' of the gene were examined. These expanded regions were searched using the SAGE genome browser, and SNPs within these regions with $p < 0.05$ were recorded.

For each candidate gene, all SNPs with $p < 0.05$ were queried together in SNP Annotation and Proxy Search (SNAP) to assess linkage disequilibrium (Johnson et al., 2008). These searches were performed using the 1000 Genomes Pilot 1 SNP data set, an $r^2 \geq 0.8$, and a distance limit of 500. This analysis was performed with both the CEU and YPI population panels separately because of allele frequency differences between European American and African American subsets. All SNPs that had an $r^2 > 0.8$ and at least 1 other variant in a group were considered a cluster.

RESULTS

The SAGE data set contains half of the most commonly reported variants associated with the 10 most well-studied candidate genes (Table 2). Of the 5 candidate variants reported in SAGE, the only variant with a $p < 0.05$ is rs279858 in *GABRA2* ($p = 0.0052$, OR = 1.16). The commonly reported variants, rs1800497 in *DRD2,* rs698 in *ADH1C*, rs1799971 in *OPRM1*, and rs4680 in *COMT*, have $p > 0.05$. The minor allele for rs17999971 in *OPRM1* trends toward being protective (OR = 0.88) while the minor alleles of rs1800497 in *DRD2*, rs698 in *ADH1C*, and rs4680 in *COMT* trend toward being associated with alcohol dependence (OR = 1.11, 1000 genomes pilot 1 SNP 1.08, 1.02, respectively). The effects of these associations are in the expected direction based on previous candidate gene studies (Blum et al., 1990; Bond et al., 1998; Hendershot et al., 2011; Ponce et al., 2008; Thomasson et al., 1991; Tiihonen et al., 1999; Tolstrup et al., 2008, 2008; Zhang et al., 2006). In addition, these effects in SAGE were similar to the findings in the individual 3 studies that contributed to SAGE: COGA, FSCD, and COGEND. Across the 3 contributing studies, the odds ratios ranged from 1.07 to 1.13 for rs1800497 in *DRD2*, 1.06 to 1.11 for rs698 in *ADH1C*, 0.82 to 0.95 for rs1799971 in *OPRM1*, 1.09 to 1.17 for rs279858 in *GABRA2*, and 1.02 to 1.09 for rs4680 in *COMT* (data not shown).

Several commonly reported variants associated with alcoholism are not on the Illumina chip that was used to generate the SAGE data set and thus these variants could not be examined. These SNPs include rs671 in *ALDH2*, rs1229984 in *ADH1B*, rs4795541 in *SLC6A4*, and rs3813867 in *CYP2E1*. As the A9 allele in *SLC6A3* is a VNTR and therefore also not reported in SAGE, we examined 2 proxy SNPs

Table 2. Examination of Top 10 Candidate Loci for Alcoholism in SAGE

Candidate genes	Publications on alcoholism association	Commonly reported SNP	Common name of SNP	p-value in SAGE	Odds ratio in SAGE (CI)
ALDH2	103	rs671	ALDH2*2 (Harada et al., 1982)	–	–
ADH1B	89	rs1229984	ADH1B*2/ADH2*2 (Thomasson et al., 1991)	–	–
DRD2	83	rs1800497	TaqIA (Blum et al., 1990)	0.0900	1.1053 (0.9845–1.2408)
SLC6A4	83	rs4795541	5-HTTLPR/S allele (Sander et al., 1997)	–	–
ADH1C	51	rs698	ADH1C*2 (Thomasson et al., 1991)	0.1452	1.0819 (0.9732–1.2028)
OPRM1	38	rs1799971	Ala118Gly (Bond et al., 1998)	0.1372	0.8823 (0.7481–1.0407)
CYP2E1	35	rs3813867	CYPE1*c2 (Hayashi et al., 1991)	–	–
GABRA2	27	rs279858*	Edenberg et al. (2004)	0.0052	1.1572 (1.0445–1.2821)
COMT	26	rs4680	Val158Met (Tiihonen et al., 1999)	0.6328	1.0244 (0.9278–1.1311)
SLC6A3	25	**	A9 (VNTR) (Dobashi et al., 1997)		

In the 10 most frequently studied genes associated with alcoholism according to the Human Genome Epidemiology (HuGE) Navigator, the most well-studied variants were examined in Study of Addiction: Genetics and Environment (SAGE).

*One of over 20 single nucleotide polymorphisms (SNPs) significantly associated with alcohol dependence (Edenberg et al., 2004). This SNP was examined because it was the only one in an exon.

**Examined 2 SNPs, rs27072 and rs27048, as proxies (Le Strat et al., 2008).

(Le Strat et al., 2008). Neither of these 2 SNPs show a significant difference between the cases and controls ($p = 0.8646$ for rs27072 and $p = 0.3842$ for rs27048).

In every gene with more than 5 publications, few SNPs had impressive differences between cases and controls. Of the 2,175 SNPs reported in the 54 genes with more than 5 publications, approximately 5% have a $p < 0.05$ (116/2,175) and approximately 1% have a $p < 0.01$ (16/2,175) (Table 3). The lowest p-value of any variant was 0.0006 for rs925946, which is a SNP upstream of *BDNF*. Overall, this distribution of p-values is consistent with chance.

In a few genes, a large proportion of the SNPs have modest frequency differences between cases and controls. In 10 of the 54 genes examined, more than 10% of the SNPs have $p < 0.05$ and in 3 genes this portion exceeds 20%. Specifically, the proportion of SNPs in SAGE with $p < 0.05$ is 55% (16/29) in *GABRA2*, 34% (10/29) in *BDNF*, and 44% (4/9) in *HTR1A* (Table 3). These few genes with a clustering of p-values less than 0.05 may warrant further study. To test whether the large proportion of SNPs with small p-values in these genes could be explained by linkage disequilibrium, we performed SNAP analyses.

Many variants clustered as defined by $r^2 > 0.8$ within the genes, but the proportion of clusters containing SNPs with $p < 0.05$ remained quite similar with SNAP analyses in both CEU and YPI populations (data not shown). Of the variants with linkage disequilibrium data available in SNAP for the CEU population, 27 SNPs in *GABRA2* broke down into 10 clusters of which 5 clusters had SNPs with $p < 0.05$ (50%), 24 SNPs in *BDNF* broke down into 9 clusters of which 4 clusters had SNPs with $p < 0.05$ (44%), and 6 SNPs in *HTR1A* broke down into 3 clusters of which 1 cluster had SNPs with $p < 0.05$ (33%). Generally, fewer SNPs clustered in the YRI population than in the CEU population, but the proportion of clusters containing SNPs with $p < 0.05$ was comparable between the 2 populations. In the YRI population, 25 SNPs in *GABRA2* broke down into 16 clusters of which 12 clusters had SNPs with $p < 0.05$ (75%), 25 SNPs in

BDNF broke down into 19 clusters of which 10 clusters had SNPs with $p < 0.05$ (53%), and 9 SNPs in *HTR1A* broke down into 5 clusters of which 2 cluster had SNPs with $p < 0.05$ (40%).

Power calculations demonstrate that the SAGE data set has 90% power with an alpha level of 0.05 to detect a genetic variant with a minor allele frequency of 0.10 and an odds ratio of 1.25 or greater. The data set also has 90% power with an alpha level of 0.05 to detect a variant with a minor allele frequency of 0.40 and an odds ratio of 1.15 or greater.

DISCUSSION

Over the last decade, hundreds of candidate genes have been proposed for alcoholism. We used local and global approaches to specifically investigate variants within the most widely studied of previously proposed candidate genes. Our primary finding is that most of these candidate genes are not strongly supported by GWA data. This observation reduces the likelihood that these previously proposed genes individually have a strong effect on the genetic risk of alcohol dependence. The results mirror prior work that most candidate loci in common diseases are not strongly replicated in GWA studies except for a few biologically important variants (Obeidat et al., 2011; Siontis et al., 2010).

Analysis of well-characterized loci that were previously proposed in candidate gene studies in a large GWA study on alcoholism, SAGE, reveals unimpressive differences between cases and controls at most loci. The frequently studied variants associated with alcoholism in *DRD2*, *ADH1C*, *OPRM1*, and *COMT* demonstrate insignificant frequency differences in SAGE ($p > 0.05$, Table 2). Although several studies implicate a biological role of these variants in alcoholism (Blum et al., 1990; Bond et al., 1998; Hendershot et al., 2011; Ponce et al., 2008; Thomasson et al., 1991; Tiihonen et al., 1999; Tolstrup et al., 2008; Zhang et al., 2006), our results reveal that these variants are not strongly associated with alcoholism in European and African ancestry populations.

Table 3. SNPs in Frequently Studied Candidate Genes Associated with Alcoholism

Candidate genes	Publications on alcoholism	SNPs recorded in dbSNP	Total SNPs in SAGE	SNPs in SAGE			
				$p < 0.05$	$p < 0.01$	$p < 0.005$	$p < 0.001$
ALDH2	103	407	24	–	–	–	–
ADH1B	89	337	21	–	–	–	–
DRD2	83	826	41	7	1	–	–
SLC6A4	83	637	21	2	–	–	–
ADH1C	51	522	29	1	–	–	–
OPRM1	38	3,568	122	4	–	–	–
CYP2E1	35	210	49	–	–	–	–
GABRA2	27	1,692	29	16	5	1	–
COMT	26	752	55	–	–	–	–
SLC6A3	25	1,322	38	–	–	–	–
HTR2A	22	1,036	61	8	–	–	–
HTR1B	18	60	12	2	1	–	–
DRD4	18	184	9	–	–	–	–
BDNF	16	624	29	10	2	2	1
NPY	15	241	19	2	–	–	–
DRD3	14	693	31	3	–	–	–
APOE	13	106	12	1	–	–	–
MTHFR	13	324	49	–	–	–	–
GABRA6	13	215	18	–	–	–	–
TPH1	13	277	14	–	–	–	–
GRIN2B	12	5,233	245	11	4	3	–
CNR1	12	2,778	27	5	2	–	–
TPH2	11	1,415	54	1	–	–	–
ADH4	10	830	62	1	–	–	–
CHRM2	10	1,992	62	3	–	–	–
CRHR1	9	1,183	26	1	–	–	–
ANKK1	9	218	26	4	–	–	–
ALDH1A1	9	739	122	3	–	–	–
DRD1	9	82	19	–	–	–	–
GABRG2	9	991	30	–	–	–	–
GABRB2	9	2,699	80	1	–	–	–
HTR1A	9	45	9	4	–	–	–
GSTM1	9	123	3	–	–	–	–
OPRD1	9	585	20	1	–	–	–
OPRK1	9	226	37	–	–	–	–
GABRB3	8	23	104	9	1	–	–
GABRA1	8	583	20	–	–	–	–
DBH	8	562	47	1	–	–	–
ADH1A	8	293	19	2	–	–	–
ADH5	7	584	38	–	–	–	–
GAD1	7	692	31	1	–	–	–
HFE	7	188	28	1	–	–	–
GRIN1	6	513	18	–	–	–	–
GAD2	6	784	49	1	–	–	–
GABRB1	6	4,354	111	2	–	–	–
ADH7	6	288	32	–	–	–	–
ADRA2A	6	46	9	–	–	–	–
CHRNA5	6	254	16	–	–	–	–
POMC	6	102	10	–	–	–	–
SLC6A2	6	837	49	1	–	–	–
CCKBR	6	300	20	4	–	–	–
CCKAR	6	167	15	–	–	–	–
TNF	6	177	36	2	–	–	–
CCK	6	461	18	1	–	–	–
Total ($n = 54$)			2,175	116	16	6	1

All genes with more than 5 publications in Human Genome Epidemiology (HuGE) Navigator were examined in Study of Addiction: Genetics and Environment (SAGE). Genes were expanded by 10 kb on both sides before they were queried in the SAGE database. MAOA and HTR2C were excluded because they were on the X chromosome.

The only candidate that modestly replicated in SAGE, rs279858 in *GABRA2*, had a p-value of 0.0052 (OR = 1.572). This finding was anticipated because a previous GWA study on the SAGE data set demonstrated a similar association (Bierut et al., 2010). The replication of rs279858 in SAGE provides some support for future studies focused on the function of this variant and associated variants in *GABRA2* (Edenberg et al., 2004).

When examined globally, none of the well-studied candidate genes demonstrate impressive variant differences between cases and controls. More specifically, only 1 SNP reported in SAGE (rs925946 upstream of *BDNF*,

$p = 0.0006$) in the 54 candidate genes had a p-value < 0.0009, a corrected p-value for the number of genes ($0.05/54 = 0.0009$). Additionally, the overall number of variants with $p < 0.05$ and <0.01 is close to that predicted by chance considering the total number of SNPs examined in all proposed candidate genes. Although the individual p-values for variants in the examined candidate genes are modest, a few candidate genes have a large portion of SNPs with $p < 0.05$ (Table 3). The results support further research into whether *GABRA2*, which was the candidate gene with the largest proportion of SNPs with $p < 0.05$ (55%), contributes to risk of developing alcohol dependence. *BDNF* and *HTR1A* also had more than one-fifth of SNPs with $p < 0.05$, indicating that these genes merit further investigation to elucidate their potential contribution to alcohol dependence.

Lack of replication in SAGE does not exclude the possibility that some previously proposed candidate genes and specific loci are biologically important. Several of the most well-studied candidate loci for alcoholism were not available in SAGE, including rs671 in *ALDH2*, rs1229984 in *ADH1B*, rs4795541 in *SLC6A4*, and rs3813867 in *CYP2E1*. A recent study that specifically genotyped rs1229984 in SAGE reported that the minor allele has a significant protective effect on alcohol dependence ($p = 6.6 \times 10^{-10}$) (Bierut et al., 2012). Because rs1229984 is common in Asians but rare in European Americans, this variant in *ADH1B* was not genotyped in the original GWA study. This highlights that GWA studies may miss important variants because of lack of coverage of SNPs that are uncommon in European American populations. Additionally, GWA studies cannot assess all forms of inheritance that can be associated with candidate genes such as insertion/deletion mutations, copy number repeats, and epigenetic changes. Although SAGE is a valuable tool, it cannot exclude the possibility that aspects of genes contribute to genetic risk of alcohol dependence.

Even though the well-studied candidate variants in *DRD2*, *ADH1C*, *OPRM1*, and *COMT* were not significantly associated with alcohol dependence in SAGE, their odds ratios were in the expected direction based on previous candidate gene studies. More specifically, the odds ratio of 0.88 for rs1799971 in *OPRM1* supports previous studies that the minor allele variant is protective against alcohol dependence (Bond et al., 1998; Zhang et al., 2006) while the odds ratios >1 for rs1800497 in *DRD2*, rs698 in *ADH1C*, and rs4680 in *COMT* support previous studies that the minor allele of these variants are more common in alcohol-dependent individuals (Blum et al., 1990; Hendershot et al., 2011; Ponce et al., 2008; Thomasson et al., 1991; Tiihonen et al., 1999; Tolstrup et al., 2008). The fact that these odds ratios are in the expected direction but did not pass a threshold of 0.05 for significance may suggest that these variants have a small contribution to alcohol dependence and this study lacked the power to detect the association.

Our study design had several strengths. First, the literature search for candidate genes included all genetic associations irrespective of ethnicity and criteria for alcoholism. By including all genes with the most genetic association study publications, we comprehensively examined previously identified genes associated with alcoholism in a large GWA study on alcoholism. Second, the SAGE data set has the power to detect associations of small magnitude. SAGE included more than 3,800 subjects and had 90% power to detect a genetic variant with an odds ratio of 1.25 for a risk locus with 10% minor allele frequency. Third, our findings in SAGE regarding the well-characterized loci were found to be very similar to the results in the 3 independent data sets that contributed to SAGE, which indicates that there is no heterogeneity across these data sets. Fourth, our approach used data that are available to the scientific community and can be easily replicated in future studies of other phenotypes.

Despite these strengths, the selection of candidate loci and genes based on number of publications retrieved by the HuGE Navigator Phenopedia has some limitations. One limitation is that no data suggest that the potential significance of a given gene is directly proportional to the number of publications. Despite this, we felt that the number of publications is an indicator of research efforts devoted to a given gene. By selecting genes with the most publications, we sought to capture well-studied genes that had been the focus of the field in the past. A second potential limitation is that we did not exclude publications based on the same data sets. Because we used a low threshold of >5 publications in the initial analysis, however, we are confident that we did not exclude any genes that have been examined in many studies. Additionally, the most well-studied loci of the 10 genes with the most publications were selected based on expert opinion and were felt to be unambiguously widely studied even if the exact order may not be reflective of the number of data sets published on the genes.

Beyond limitations in our selection of candidate genes, the SAGE data set has limitations that restrict the interpretation of our results. First, some of the most well-studied variants were not covered in SAGE and therefore could not be assessed. Second, the X chromosome is not included in the publically available SAGE results so we were unable to investigate genes on the X chromosome. Specifically, 2 candidate genes on the X chromosome, *MAOA* and *HTR2C*, which had 26 and 8 publications, respectively, were not assessed. Third, SAGE is limited in its power to identify genotyped variants on the GWA chip that have small effect sizes. Despite the fact that the SAGE data set was relatively large when it was originally published, identifying common variants with small effect sizes (<1.1) remains challenging and we are unable to rule out the possibility of real but modest effects of these genes. Fourth, variants that are uncommon (1 to 5%) or rare ($<1\%$) in the study population may also not be detected in SAGE because of their individually small contribution to overall alcoholism. Fifth, the SAGE data set primarily consists of European Americans (69.5%), African Americans (30.3%), and a few Hispanics (3.4%) (Table 1) and association findings may be different in other

populations such as Asians. A recent evaluation of 108 GWA studies demonstrates varying consistency in genetic effect sizes across ancestral groups, highlighting that analysis of data from diverse ancestral groups may be more revealing for certain genes and variants than analysis of only European and African ancestry populations (Nitzani et al., in press). Some of the genes and variants examined in this analysis are more well studied and have a higher frequency in Asian ancestry than in European and African ancestry populations, such as the Asp40 allele of the candidate variant rs1799971 in *OPRM1* (Arias et al., 2006) and therefore may have a more impressive effect in studies that include Asian ancestry populations. Sixth, our analysis did not examine the effects of combinations of genes or the effect of different environmental factors. Analysis of multiple genes and populations enriched for specific environmental risk factors will likely explain a greater degree of the genetic risk of alcoholism. Despite these limitations, this analysis demonstrates that GWA studies are a powerful technique for verifying the importance of genes and particular variants that have been previously identified in the candidate gene era.

In summary, we provide evidence that for alcohol dependence, several extensively studied candidate loci and genes are not replicated in a large GWA study, indicating that these variants do not individually have a large contribution to risk of developing alcohol dependence in European and African ancestry populations. Our analysis was unable to rule out the possibility that some variants and genes are important for risk of alcoholism owing to the lack of coverage. Recent work demonstrates that at least one highly reported variant rs1229984 in *ADH1B* that is not reported in SAGE is significantly associated with alcoholism (Bierut et al., 2012), suggesting the possible importance of further research on highly supported variants that cannot be assessed in SAGE. Our approach may also have missed variants that have a real but small individual contribution to overall inheritance of alcoholism.

This analysis demonstrates that targeted candidate gene studies and GWA studies each provide important information and studying the convergence of these 2 experimental designs has the potential to advance understanding of the etiology of alcohol dependence and more generally complex diseases. While GWA studies provide important information about the genetic contribution of common variants to complex diseases across populations, hypothesis-driven candidate gene studies are also important to assess variants of lesser significance that may be missed because of the strict *p*-value thresholds required for the large number of comparisons in GWA studies. Incorporating knowledge from both GWA and candidate gene studies will help clarify the role of genetics in complex disease and guide future research.

Our study also shows how the HuGE Navigator and dbGaP databases can be used as tools by researchers to easily access and analyze information on candidate genes and GWA data. Beyond alcoholism, the HuGE Navigator provides an easy way for investigators to search over 2,000 diseases and 10,000 genes for summary information and primary articles about genetic associations and human genome epidemiology (Yu et al., 2008). The dbGaP database provides access to results of over 100 studies examining phenotype and genotype associations, including 40 GWAS studies on different diseases. As dbGaP currently contains a limited number of GWA studies, researchers examining phenotypes not available in dbGaP may benefit from directly contacting the authors of relevant GWA studies and meta-analysis. Because of this easy accessibility, researchers who intend to perform future candidate gene studies should reference the HuGE navigator to assess background information and use dbGaP and existing GWA data to test whether their gene of interest is replicated in GWA data. Candidate gene studies need replication to meet scientific standards. Simple dbGaP analyses may help to focus future research on genes that are supported by GWA data and therefore more likely to be biologically important for human disease.

REFERENCES

Arias A, Feinn R, Kranzler HR (2006) Association of an Asn40Asp (A118G) polymorphism in the mu-opioid receptor gene with substance dependence: a meta-analysis. Drug Alcohol Depend 83:262–268.

Bierut LJ, Agrawal A, Bucholz KK, Doheny KF, Laurie C, Pugh E, Fisher S, Fox L, Howells W, Bertelsen S, Hinrichs AL, Almasy L, Breslau N, Culverhouse RC, Dick DM, Edenberg HJ, Foroud T, Grucza RA, Hatsukami D, Hesselbrock V, Johnson EO, Kramer J, Krueger RF, Kuperman

S, Lynskey M, Mann K, Neuman RJ, Nothen MM, Nurnberger JI Jr, Porjesz B, Ridinger M, Saccone NL, Saccone SF, Schuckit MA, Tischfield JA, Wang JC, Rietschel M, Goate AM, Rice JP (2010) A genome-wide association study of alcohol dependence. Proc Natl Acad Sci USA 107:5082–5087.

Bierut LJ, Goate AM, Breslau N, Johnson EO, Bertelsen S, Fox L, Agrawal A, Bucholz KK, Grucza R, Hesselbrock V, Kramer J, Kuperman S, Nurnberger J, Porjesz B, Saccone NL, Schuckit M, Tischfield J, Wang JC, Foroud T, Rice JP, Edenberg HJ (2012) ADH1B is associated with alcohol dependence and alcohol consumption in populations of European and African ancestry. Mol Psychiatry 17:445–450.

Blum K, Noble EP, Sheridan PJ, Montgomery A, Ritchie T, Jagadeeswaran P, Nogami H, Briggs AH, Cohn JB (1990) Allelic association of human dopamine D2 receptor gene in alcoholism. JAMA 263:2055–2060.

Bond C, LaForge KS, Tian M, Melia D, Zhang S, Borg L, Gong J, Schluger J, Strong JA, Leal SM, Tischfield JA, Kreek MJ, Yu L (1998) Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: possible implications for opiate addiction. Proc Natl Acad Sci USA 95:9608–9613.

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA (1994) A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol 55:149–158.

Dobashi I, Inada T, Hadano K (1997) Alcoholism and gene polymorphisms related to central dopaminergic transmission in the Japanese population. Psychiatr Genet 7:87–91.

Edenberg HJ, Dick DM, Xuei X, Tian H, Almasy L, Bauer LO, Crowe RR, Goate A, Hesselbrock V, Jones K, Kwon J, Li TK, Nurnberger JI Jr, O'Connor SJ, Reich T, Rice J, Schuckit MA, Porjesz B, Foroud T, Begleiter H (2004) Variations in GABRA2, encoding the alpha 2 subunit of the GABA(A) receptor, are associated with alcohol dependence and with brain oscillations. Am J Hum Genet 74:705–714.

Edenberg HJ, Koller DL, Xuei X, Wetherill L, McClintick JN, Almasy L, Bierut LJ, Bucholz KK, Goate A, Aliev F, Dick D, Hesselbrock V, Hinrichs A, Kramer J, Kuperman S, Nurnberger JI Jr, Rice JP, Schuckit MA, Taylor R, Todd Webb B, Tischfield JA, Porjesz B, Foroud T (2010) Genome-wide association study of alcohol dependence implicates a region on chromosome 11. Alcohol Clin Exp Res 34:840–852.

Farrer LA, Kranzler HR, Yu Y, Weiss RD, Brady KT, Anton R, Cubells JF, Gelernter J (2009) Association of variants in MANEA with cocaine-related behaviors. Arch Gen Psychiatry 66:267–274.

Harada S, Agarwal DP, Goedde HW (1982) Mechanism of alcohol sensitivity and disulfiram-ethanol reaction. Subst Alcohol Actions Misuse 3:107–115.

Hayashi S, Watanabe J, Kawajiri K (1991) Genetic polymorphisms in the 5'-flanking region change transcriptional regulation of the human cytochrome P450IIE1 gene. J Biochem 110:559–565.

Heath AC, Bucholz KK, Madden PA, Dinwiddie SH, Slutske WS, Bierut LJ, Statham DJ, Dunne MP, Whitfield JB, Martin NG (1997) Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men. Psychol Med 27:1381–1396.

Heath AC, Whitfield JB, Martin NG, Pergadia ML, Goate AM, Lind PA, McEvoy BP, Schrage AJ, Grant JD, Chou YL, Zhu R, Henders AK, Medland SE, Gordon SD, Nelson EC, Agrawal A, Nyholt DR, Bucholz KK, Madden PA, Montgomery GW (2011) A quantitative-trait genome-wide association study of alcoholism risk in the community: findings and implications. Biol Psychiatry 70:513–518.

Hendershot CS, Lindgren KP, Liang T, Hutchison KE (2011) COMT and ALDH2 polymorphisms moderate associations of implicit drinking motives with alcohol use. Addict Biol 17:192–201.

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genet Med 4:45–61.

Hodgkinson CA, Enoch MA, Srivastava V, Cummins-Oman JS, Ferrier C, Iarikova P, Sankararaman S, Yamini G, Yuan Q, Zhou Z, Albaugh B, White KV, Shen PH, Goldman D (2010) Genome-wide association identifies candidate genes that influence the human electroencephalogram. Proc Natl Acad Sci USA 107:8695–8700.

Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001) Replication validity of genetic association studies. Nat Genet 29:306–309.

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24:2938–2939.

Joslyn G, Ravindranathan A, Brush G, Schuckit M, White RL (2010) Human variation in alcohol response is influenced by variation in neuronal signaling genes. Alcohol Clin Exp Res 34:800–812.

Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J (2005) Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389.

Knopik VS, Heath AC, Madden PA, Bucholz KK, Slutske WS, Nelson EC, Statham D, Whitfield JB, Martin NG (2004) Genetic effects on alcohol dependence risk: re-evaluating the importance of psychiatric and other heritable risk factors. Psychol Med 34:1519–1530.

Le Strat Y, Ramoz N, Pickering P, Burger V, Boni C, Aubin HJ, Ades J, Batel P, Gorwood P (2008) The 3' part of the dopamine transporter gene DAT1/SLC6A3 is associated with withdrawal seizures in patients with alcohol dependence. Alcohol Clin Exp Res 32:27–35.

Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Beck J, Kimelman M, Shevelev S, Preuss D, Yaschenko E, Graeff A, Ostell J, Sherry ST (2007) The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 39:1181–1186.

Nitzani EE, Liberopoulos G, Manolio TA, Ioannidis JP (in press) Consistency of genome-wide associations across major ancestral groups. Hum Genet ????:????–????, doi 10.1007/s00439-011-1124-4.

Obeidat M, Wain LV, Shrine N, Kalsheker N, Soler Artigas M, Repapi E, Burton PR, Johnson T, Ramasamy A, Zhao JH, Zhai G, Huffman JE, Vitart V, Albrecht E, Igl W, Hartikainen AL, Pouta A, Cadby G, Hui J, Palmer LJ, Hadley D, McArdle WL, Rudnicka AR, Barroso I, Loos RJ, Wareham NJ, Mangino M, Soranzo N, Spector TD, Glaser S, Homuth G, Volzke H, Deloukas P, Granell R, Henderson J, Grkovic I, Jankovic S, Zgaga L, Polasek O, Rudan I, Wright AF, Campbell H, Wild SH, Wilson JF, Heinrich J, Imboden M, Probst-Hensch NM, Gyllensten U, Johansson A, Zaboli G, Mustelin L, Rantanen T, Surakka I, Kaprio J, Jarvelin MR, Hayward C, Evans DM, Koch B, Musk AW, Elliott P, Strachan DP, Tobin MD, Sayers I, Hall IP (2011) A comprehensive evaluation of potential lung function associated genes in the SpiroMeta general population sample. PLoS ONE 6: e19382.

Ponce G, Hoenicka J, Jimenez-Arriero MA, Rodriguez-Jimenez R, Aragues M, Martin-Sune N, Huertas E, Palomo T (2008) DRD2 and ANKK1 genotype in alcohol-dependent patients with psychopathic traits: association and interaction study. Br J Psychiatry 193:121–125.

Sander T, Harms H, Lesch KP, Dufeu P, Kuhn S, Hoehe M, Rommelspacher H, Schmidt LG (1997) Association analysis of a regulatory variation of the serotonin transporter gene with severe alcohol dependence. Alcohol Clin Exp Res 21:1356–1359.

Siontis KC, Patsopoulos NA, Ioannidis JP (2010) Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. Eur J Hum Genet 18:832–837.

Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38:209–213.

Thomasson HR, Edenberg HJ, Crabb DW, Mai XL, Jerome RE, Li TK, Wang SP, Lin YT, Lu RB, Yin SJ (1991) Alcohol and aldehyde dehydrogenase genotypes and alcoholism in Chinese men. Am J Hum Genet 48:677–681.

Tiihonen J, Hallikainen T, Lachman H, Saito T, Volavka J, Kauhanen J, Salonen JT, Ryynanen OP, Koulu M, Karvonen MK, Pohjalainen T, Syvalahti E, Hietala J (1999) Association between the functional variant of the catechol-O-methyltransferase (COMT) gene and type 1 alcoholism. Mol Psychiatry 4:286–289.

Tolstrup JS, Nordestgaard BG, Rasmussen S, Tybjaerg-Hansen A, Gronbaek M (2008) Alcoholism and alcohol drinking habits predicted from alcohol dehydrogenase genes. Pharmacogenomics J 8:220–227.

Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, Fehr C, Scherbaum N, Steffens M, Ludwig KU, Frank J, Wichmann HE, Schreiber S, Dragano N, Sommer WH, Leonardi-Essmann F, Lourdusamy A, Gebicke-Haerter P, Wienker TF, Sullivan PF, Nothen MM, Kiefer F, Spanagel R, Mann K, Rietschel M (2009) Genome-wide association study of alcohol dependence. Arch Gen Psychiatry 66:773–784.

Yu W, Clyne M, Khoury MJ, Gwinn M (2010) Phenopedia and genopedia: disease-centered and gene-centered views of the evolving knowledge of human genetic associations. Bioinformatics 26:145–146.

Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ (2008) A navigator for human genome epidemiology. Nat Genet 40:124–125.

Zhang H, Luo X, Kranzler HR, Lappalainen J, Yang BZ, Krupitsky E, Zvartau E, Gelernter J (2006) Association between two mu-opioid receptor gene (OPRM1) haplotype blocks and drug or alcohol dependence. Hum Mol Genet 15:807–819.

Zlojutro M, Manz N, Rangaswamy M, Xuei X, Flury-Wetherill L, Koller D, Bierut LJ, Goate A, Hesselbrock V, Kuperman S, Nurnberger J Jr, Rice JP, Schuckit MA, Foroud T, Edenberg HJ, Porjesz B, Almasy L (2011) Genome-wide association study of theta band event-related oscillations identifies serotonin receptor gene HTR7 influencing risk of alcohol dependence. Am J Med Genet B Neuropsychiatr Genet 156:44–58.